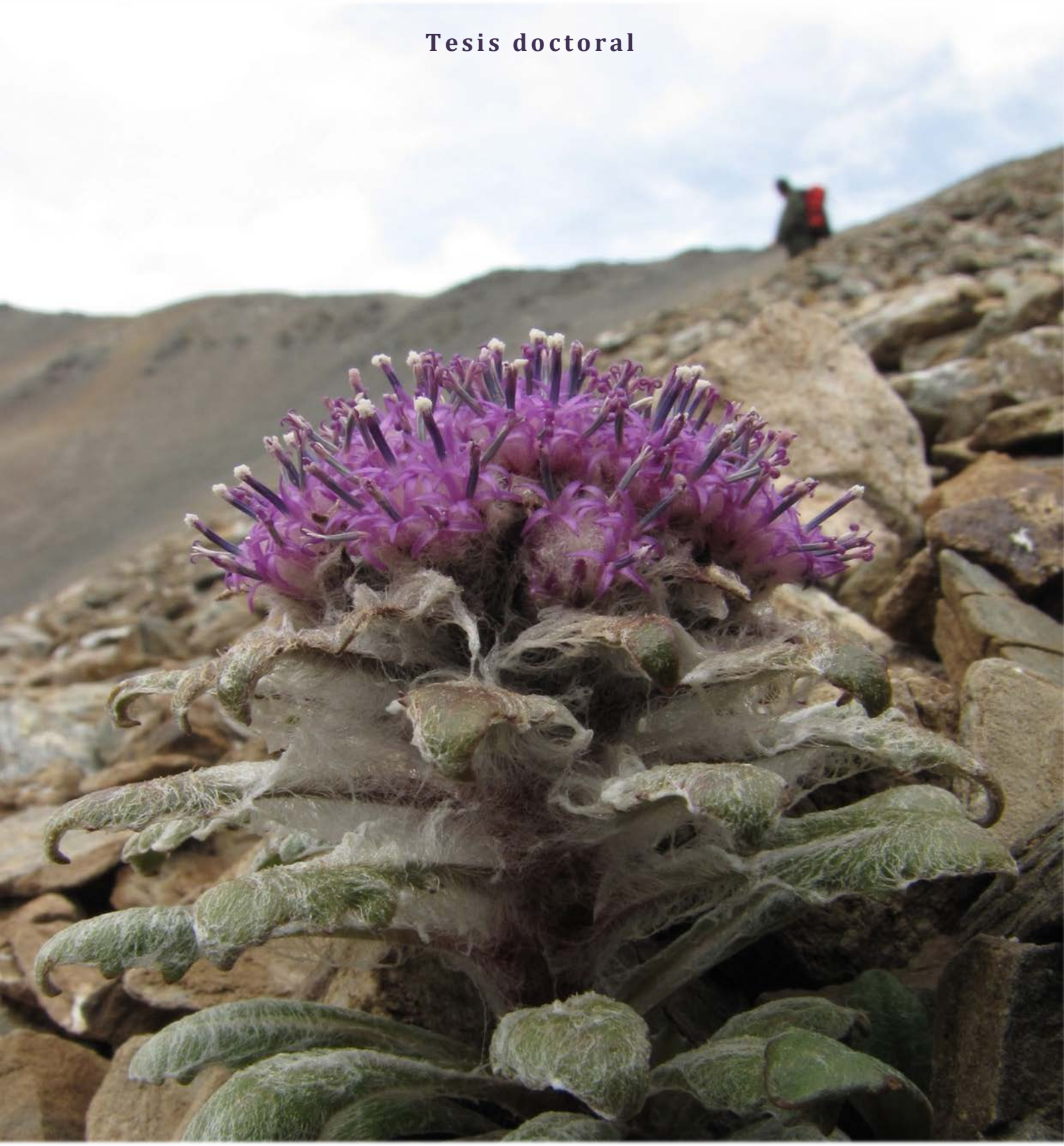
Universitat Autònoma de Barcelona

SECUENCIACIÓN DE NUEVA GENERACIÓN EN EL ESTUDIO DE LAS RADIACIONES ALPINAS: LOS GÉNEROS SAUSSUREA Y JURINEA

SONIA HERRANDO MORAIRA

SECUENCIACIÓN DE NUEVA GENERACIÓN EN EL ESTUDIO DE LAS RADIACIONES ALPINAS: LOS GÉNEROS *SAUSSUREA* Y *JURINEA*

Tesis doctoral



SONIA HERRANDO MORAIRA

2021

UNIVERSITAT AUTÒNOMA DE BARCELONA ▪ FACULTAT DE BIOCIÈNCIES
DEPARTAMENT DE BIOLOGIA ANIMAL, BIOLOGIA VEGETAL I ECOLOGIA
UNITAT DE BOTÀNICA

INSTITUT BOTÀNIC DE BARCELONA
CONSEJO SUPERIOR DE INVESTIGACIONES CIENTÍFICAS

Tesis doctoral

SECUENCIACIÓN DE NUEVA GENERACIÓN EN EL ESTUDIO DE LAS RADIACIONES ALPINAS:

LOS GÉNEROS *SAUSSUREA* Y *JURINEA*

Sonia Herrando Moraira

Barcelona, 2021
Programa de doctorado en Biología y Biotecnología Vegetal

UNIVERSITAT AUTÒNOMA DE BARCELONA ▪ FACULTAT DE BIOCIÈNCIES
DEPARTAMENT DE BIOLOGIA ANIMAL, BIOLOGIA VEGETAL I ECOLOGIA
UNITAT DE BOTÀNICA

INSTITUT BOTÀNIC DE BARCELONA
CONSEJO SUPERIOR DE INVESTIGACIONES CIENTÍFICAS

SECUENCIACIÓN DE NUEVA GENERACIÓN EN EL ESTUDIO DE LAS RADIACIONES ALPINAS:

LOS GÉNEROS *SAUSSUREA* Y *JURINEA*

Memoria presentada por

Sonia Herrando Moraira

para optar al título de Doctora por la Universitat Autònoma de Barcelona
bajo la supervisión de los directores y tutora de tesis

Dr. Alfonso Susanna de la Serna
(director)

Dra. Núria Garcia Jacas
(directora)

Dra. Mercè Galbany Casals
(directora y tutora)

Programa de doctorado en Biología y Biotecnología Vegetal

Fotografías de portada principal y portadas de capítulos:
Jordi López-Pujol, Sergi Massó, Alfonso Susanna

La presente tesis doctoral ha sido posible gracias a la financiación para la propia investigación y para la beca predoctoral FPI (formación de personal investigador) concedida por el Ministerio de Educación y Cultura durante el período 2017-2021 en el marco del proyecto “Radiaciones alpinas de tipo insular en Asia: los casos de *Saussurea* y *Jurinea* en el Himalaya y el Tian Shan” (Agencia Española de Investigación CGL2015-66703-P).

A David y Sam

Agradecimientos

Agradecimientos

El doctorado llega a su fin, parece increíble, pero es cierto. Aquí estoy, escribiendo las últimas líneas de mi tesis doctoral. El tren está a punto de llegar a su última parada. Ha sido un viaje largo, lleno de experiencias inolvidables; un aprendizaje continúo que me acompañará siempre.

Estoy enormemente agradecida por haber conseguido uno de esos *tickets* tan apreciados y escasos que me han permitido subir al tren y dedicarme a lo que me apasiona. Las personas que me han acompañado durante estos largos cinco años de trayecto han sido imprescindibles y un factor clave. Tengo mucho que agradecerles puesto que nada hubiera sido igual sin la presencia de todas ellas.

En primer lugar me gustaría mencionar a las dos personas que hicieron posible que en mi interior saltara la primera chispa que me permitió a sentir interés por la botánica. Sucedió durante el primero curso de mis estudios universitarios y me refiero a **Llorenç Sáez** y **Mercè Galbany**. La pasión de ambos por las plantas es admirable, y el entusiasmo que muestran por su trabajo contagia positivamente a los estudiantes. Yo fui una de las estudiantes afortunadas que se dejó llevar de la mano de ambos. Gracias a los dos por vuestra contribución y legado, por los conocimientos, experiencias, maneras de trabajar y entender la botánica… y la vida. Gracias por haberme otorgado el privilegio de ser vuestra discípula. Siempre os recordaré como mis maestros y mis referentes

Mercè Galbany me indicó dónde estaba la parada de inicio del tren, allí me dirigí y allí encontré a dos veteranos investigadores: **Alfonso Susanna** y **Núria García Jacas**. Los tres conformaron finalmente la comunidad de directores de esta tesis, los pilares fundamentales que me han ayudado a orientar el trabajo y superar las dificultades del viaje. Gracias a los tres por el apoyo personal y profesional que me habéis ofrecido. Os agradezo especialmente la confianza que en su momento depositasteis en mí y vuestra contribución a mi trayectoria profesional.

Gracias **Alfonso** por tu optimismo y vehemencia hasta al final. Todos sabemos que la motivación es un factor fundamental para elaborar una tesis doctoral y tú me has animado siempre. Has sido brújula y guía en todo momento. Fuiste además compañero de aventuras en China y tu sentido del humor ayudó a sobrellevar las circunstancias vividas en aquellas tierras, queridas y lejanas por todos nosotros. Me encantó conocer a tu familia en Madrid, y compartir contingo viajes y largas estancias en aeropuertos.

Gracias **Núria**, por tu serenidad y empatía. Siempre me has ayudado a solucionar todo tipo de problemas, y me has aportado esas dosis de calma que permiten relativizar las dificultades que durante la tesis parecen insalvables. Siempre recordaré el momento en que te dije: “Ser madre es duro, es una experiencia enorme”. Tú me miraste y asentiste, nada más, pero capté claramente tu comprensión y complicidad, me sentí apoyada y respaldada. Fue un momento muy emotivo.

Gracias **Mercè**, por tu gran inteligencia intelectual y emocional, y por tu función de cuidadora y guardiana. Te he sentido siempre como una *madre,* personal y profesionalmente. He contado con tu apoyo incondicional y, en los malos momentos, siempre me has reconfortado. Cómo no recordar la anécdota de la primera charla sobre NGS antes de empezar la tesis. Me derrumbé ante el reto que tenía que superar. Tus palabras de ánimo me dieron la fuerza necesaria para continuar el camino y han hecho posible que hoy este aquí escribiendo estas últimas líneas. ¡Qué afortunada me siento de haberte conocido y haber pasado juntas momentos tan especiales!

También me gustaría hacer una especial referencia a cuatro compañeros de viaje, que conocí en este tren, y que han sido para mí pilares fundamentales. Siempre me han brindado su ayuda tanto a nivel personal como profesional, y han contribuido de manera significativa a que la tesis saliera adelante. Me refiero a **Jordi López Pujol**, **Maria Luisa Gutiérrez**, **Roser Vilatersana** y **Neus Nualart**. Gracias por vuestra amistad y por tantos momentos divertidos que hemos compartido.

Gracias **Jordi**, por tu enorme nobleza y bondad. Eres una de las personas con más capacidad de trabajo, esfuerzo, tesón y entusiasmo que he conocido. ¡Tengo tanto que agradecerte que no sé por dónde empezar! Sin duda has sido mi mejor maestro y amigo durante todo el viaje. Me saltan lágrimas de felicidad al pensar en los buenos momentos que hemos vivido juntos. Es difícil encontrar un compañero con el que encajes al cien por cien en lo profesional y personal, y yo he tenido la suerte de encontrarlo. Gracias por tan buenos consejos que me has dado estos años, por enseñarme y motivarme para aprender técnicas y métodos nuevos; gracias por compartir conmigo tus experiencias y por ser luz cuando a mi alrededor sólo veía oscuridad. Sin duda, estos años habrían sido muy diferentes sin ti, sin tu compañía, apoyo y aprendizaje. Esta tesis es gran en parte fruto de la buena energía que me has transmitido; formas parte de ella.

Gracias **Maria Luisa**, bondadosa y formadora de equipos. Sin duda, capitana de grupo y apoyo fundamental para los que hemos estado a tu lado. Contigo el laboratorio era un sitio de trabajo intenso pero también de risas y confesiones. Nos lo hemos pasado tan bien... Qué suerte haber encontrado en este viaje a una amiga como tú, he disfrutado mucho de nuestra sintonía y conexión. Estoy segura de que seguirás transmitiendo buena energía y capacidad de unir a las personas allá donde vayas. Tienes un don muy valioso.

Gracias **Roser**, por tu alegría y valentía. Lo he pasado muy bien contigo, tanto en el Instituto como en el viaje a China. Eres una mujer aventurera, valiente y atrevida, capaz de enfrentarse a grandes retos y esforzarse hasta conseguirlos. Tu ejemplo me ha ayudado a ver los desafíos de la tesis como algo positivo y estimulante, en lugar de afrontarlos con temor y desconfianza. Gracias por tu gran apoyo. Me conoces tan bien que sabías cómo estaba mi ánimo únicamente por la expresión de mi cara. Tus enseñanzas y la paciencia que has tenido conmigo durante el máster han contribuido significativamente en la elaboración de esta tesis. Por ello puedes sentirla también parte tuya.

Gracias **Neus**, emprendedora y protectora. Nos conocimos mejor en un curso que hicimos en Madrid, con Jordi. Recuerdo trabajar contigo por la noche y acabar agotadas. Desde aquel viaje sentí una gran sintonía contigo. Siempre te he visto como una *hermana mayor* y has sido para mí un gran referente. He aprendido mucho de tu profesionalidad y de tu implicación en todos los proyectos o nuevos retos que surgían. Te agradezco tu protección y los ánimos que me dabas cuando lo necesitaba.

Gracias también a todo el equipo humano del *Institut Botànic de Barcelona*, por recibirme siempre con los brazos abiertos y hacer de esta institución *segundo hogar*. Gracias a las personas que forman parte del equipo permanente, y a las que han pasado por allí y han dejado su huella:

Aida Monge, **Airy Gras**, **Àngel Romo**, **Ángela Zarate**, **Bárbara Queralt**, **Bea Díaz**, **Carles Pérez**, **Carlos Gómez Bellver**, **Carlos Silva**, **Carmen Ceballos**, **Dani Vitales**, **David Pérez-Prieto**, **Diana Muñiz**, **Eduard Farràs**, **Fernando Castro**, **Jaume Pellicer**, **Jéssica Requena**, **Joan Pere Pascual**, **Jordi Antúnez**, **Josep Maria Montserrat**, **Laia Barres**, **Laura Gavioli**, **Manica Balant**, **Neus Ibáñez**, **Noemí Montes**, **Núria Gironell**, **Oriane Hidalgo**, **Pilar Alonso**, **Rafa Martín**, **Roi Rodríguez**, **Samuel Pyke**, **Sara López Vinyallonga**, **Sepand Riyahi**, **Sergi Massó**, **Sònia Garcia**, **Teresa Garnatje** y **Toni Forés**.

En especial, quiero agradecer a las mujeres (**Ángela**, **Bárbara**, **Carmen**, **Neus**, **Sònia**, **Teresa…**) por la ayuda que me han ofrecido durante toda la tesis, pero también por su apoyo durante el período de mi embarazo y maternidad. Os admiro mucho como mujeres y como madres. Admiro vuestra gran lucha y esfuerzo.

Nunca olvidaré el primer día que me incorporé al trabajo después de la baja por maternidad. Llegué al Instituto a lágrima viva por lo difícil que era separarme de la criatura. Pero allí estaban todas ellas, entre muchos otros y otras, para hacer que mi vuelta fuero mucho más llevadera.

También quisiera mencionar a **Sergi Massó**, un gran compañero que siempre me tendió su mano cuando más lo necesitaba. La chocolatina que me ofreció en aquella primera charla sobre NGS me hizo sonreír y me dio mucho ánimo. Fue un gesto sencillo pero precioso. También, una mención a **Laia Barres**. Gracias por tu ayuda en el laboratorio y por el apoyo emocional durante mi embarazo. Lo hiciste todo más fácil, con tu agradable compañía y gran profesionalidad. Las muestras estuvieron en muy buenas manos.

Gracias a mis compañeros de la *Universitat Autònoma de Barcelona*: **Cristina Roquet**, **Javi López Alvarado**, **Juancho Calleja**, **Llorenç Sáez** y **Pau Carnicero**. Cuando iba a trabajar con vosotros siempre he tenía la sensación de *volver a casa*, volver al lugar dónde me eduqué y crecí. **Cristina**, muchísimas gracias por tus maravillosos consejos y buenas ideas. Me han resultado muy útiles hasta el final de la tesis.

Gracias a los compañeros que conocí en la Universidad de Memphis (EEUU). Hicieron posible que las largas jornadas en el laboratorio fueran agradables. Muchas gracias **Jennifer Mandel**, **Carol Siniscalchi** y **Ram Thapa**. Y también a **Katy Jones** por aquel encuentro tan especial en Praga, donde pudimos conocernos y compartir nuestros proyectos.

Gracias también a los compañeros de proyecto. A muchos sólo os he conocido por la pantalla del ordenador pero esta tesis no hubiera concluido felizmente sin vuestra ayuda y colaboración: **Alexander Sennikov**, **Elizaveta Pyak**, **Hyoung-Tak Im**, **Iraj Mehregan**, **Jian-Quan Liu**, **Kazumi Fujikawa**, **Liansheng Xu**, **Seung-Chul Kim** y **You-Sheng Chen**.

Igualmente estoy inmensamente agradecida a toda mi familia y amigos por haber estado a mi lado a lo largo de este duro viaje. Gracias **Adrià Miralles** por tantas excursiones y charlas sobre el mundo de la ciencia. Gracias a mis padres (**Carmen** y **Jordi**) y a mi hermano (**Albert**) que siempre me han brindado su apoyo incondicional en esta montaña rusa que ha representado el doctorado. A mi madre, por ayudarme tantísimo en la maquetación de la tesis, a mi hermano, por su soporte en muchos análisis y en programación, y a mi padre por sus sabios consejos en los momentos más bajos.

Y al pilar fundamental durante todo este tiempo que lo sostiene todo, **David**, mi compañero de vida. Por los numerosos e incontables viajes en coche hasta el *Institut Botànic*, y sus largas esperas hasta que acabara *una última tarea*. Toda su esencia, apoyo y ayuda también están de manera implícita en esta tesis. No es fácil vivir con un doctorando/a. ¡Tienes el cielo ganado! Gracias por tu infinita paciencia. Han sido años muy intensos y sin tenerte a mi lado no habría llegado hasta aquí. Gran parte de esta tesis ha sido posible gracias a tu amor y apoyo.

Por último, a una persona, que con tan sólo 2 años de vida, me ha enseñado más que cualquier otra experiencia que haya vivido hasta ahora. Gracias **Sam**, tu llegada me ayudó a entender que la tesis era un reto relativamente fácil de superar, en comparación con lo que representa la maternidad. El nuevo reto, sin duda, era mucho mayor. Llegaste prácticamente a mitad de la tesis y lo revolucionaste todo. Tu llegada me enseñó a relativizar los problemas, a priorizar lo esencial, y a aprovechar al máximo y de manera intensa el tiempo necesario para concluir esta tesis. Pero sobre todo has hecho que todo adquiera mucho más sentido. Mil gracias, pequeño Sam.

Gracias sinceras a todas y a todos por haber contribuido a que esta tesis doctoral haya sido una experiencia inolvidable.

Índice

ГЕРБАРИЙ ФЛОРЫ СССР,
ИЗДАВАЕМЫЙ БОТАНИЧЕСКИМ ИНСТИТУТОМ им. В. Л. КОМАРОВА
АКАДЕМИИ НАУК СССР

6350. Saussurea oxyodontha Hult.

1930, Kungl. Sv. Vet.-Akad. Handl. 8, 2 : 207 (Fl. Kamtsch. 4); Лищиц, 1962, Фл. СССР **27** : 488; Ворошилов, 1966, Фл. сов. Дальн. Вост. : 433. — *Saussurea Ledebouri* g. *Kruhseana* Herd. 1868, Bull. Soc. Nat. Moscou 41,2 : 43. — *S. subsinuata* auct. non Ledeb.: Комаров, 1930, Фл. Камч. 3 : 181.

Охотско-чукотский гипоарктический вид, встречающийся в южной части Чукотки. Наиболее обычные местонахождения *S. oxyodontha* — луговые разнотравные сообщества по берегам ручьев, на террасах рек, у подножий и в нижних частях склонов. Вид отмечается также в зарослях пойменных кустарников и в пойменных лесах. Т. Плиева.

Магаданская обл., Билибинский р-н, Анюйское нагорье, гора Вулканная, луговина по берегу ручья на северном склоне. Собр. В. Петровский и Т. Плиева. — Prov. Magadan, distr. Bilibinsky, montes Anjuj, mons Vulkannaja, pratulum ad ripam rivuli in clivo septentrionali. Leg. V. Petrovsky et T. Plieva.
1982 VIII 7.

Índice

Introducción general

Contexto general

Actualmente se calcula que existen en la Tierra más de 308.000 plantas vasculares (Christenhusz & Byng, 2016). Por desgracia, algunas publicaciones recientes nos advierten de su elevada tasa de extinción, que sería de hasta 500 veces superior a lo que se esperaría que ocurriera sin la influencia antrópica (Humphreys et al., 2019). Por ello, urge destinar más recursos económicos y esfuerzos humanos a descubrir, estudiar y proteger la biodiversidad vegetal que habita en nuestro planeta. Muchos científicos del campo de la botánica recalcan la necesidad de continuar potenciando los estudios de ciencia básica en este campo (Crisci et al., 2019, 2020), ya que para proteger resulta esencial conocer el origen y la historia evolutiva de las especies.

Precisamente en ello se centra la sistemática, el área científica que trata de comprender cómo las especies pueden clasificarse sobre la base de su parentesco evolutivo y similitudes morfológicas (taxonomía) y qué patrones y procesos hay detrás de su historia evolutiva (diversificación o especiación; cf. Stuessy et al., 2014, Fig. 1).

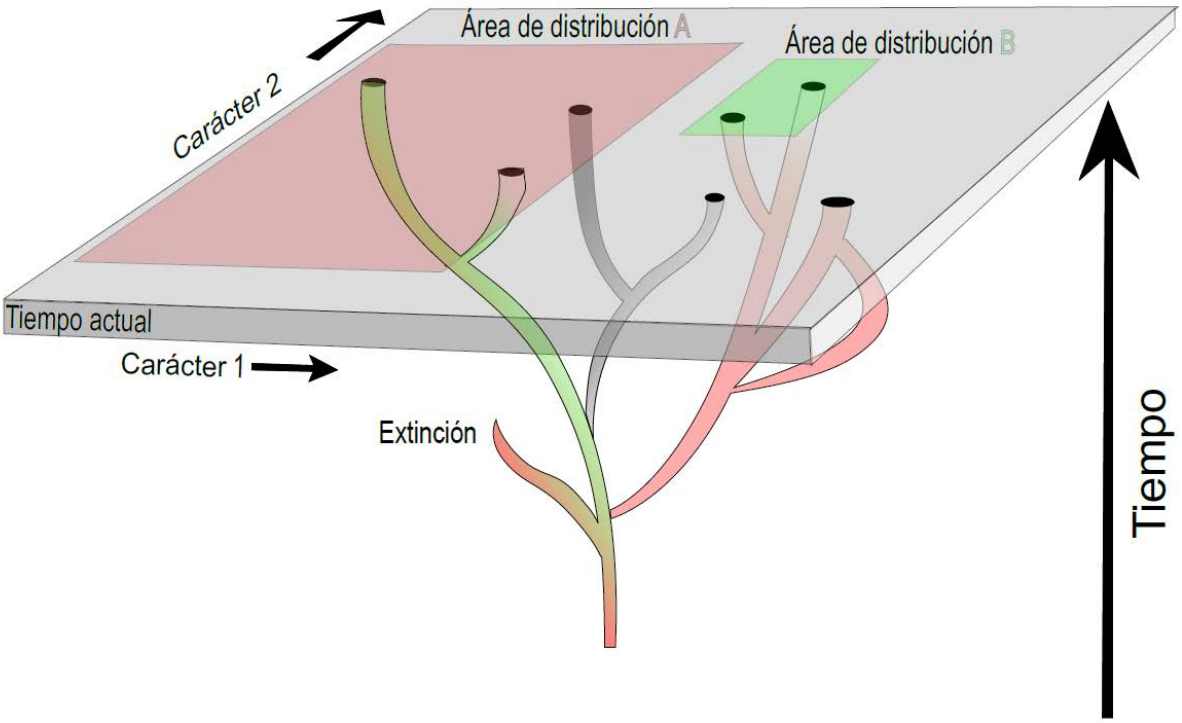


Fig. 1. Representación en tres dimensiones de la historia filogenética, distribución geográfica y variabilidad morfológica de un grupo a través del tiempo. Modificado de Shyamal (2007).

Dentro de esta disciplina, los investigadores siempre han mostrado un gran interés en aquellos grupos formados por numerosas especies o linajes, aparentemente originados en un corto período de tiempo. Este proceso es comúnmente denominado radiación evolutiva (Schluter, 2000). Las explosiones de diversidad pueden deberse a factores geológicos como levantamientos de montañas o ser el resultado de grandes cambios climáticos. Por ello, la diversidad en la Tierra no se distribuye de manera homogénea ni al azar, sino que aquellos enclaves que han proporcionado un mayor número de oportunidades

ecológicas son hoy en día los más diversos, los conocidos como *hotspots* de biodiversidad descritos inicialmente por Myers et al. (2000). En la Figura 2 se resaltan las zonas de mayor diversidad vegetal global estimados por Brummitt et al. (2021). El estudio de estas áreas y especies que habitan en ellas es clave para poder proteger estas zonas singulares del planeta. Muchas de estas regiones han sido propuestas como puntos de origen de diversificación de especies, donde persisten linajes antiguos, y a su vez son zonas muy activas de especiación *in situ* (conocidos como *museum* and *cradle centers*; Moreau & Bell, 2013). Además, también son consideradas zonas que han actuado, y se prevé que actuarán en el futuro, como refugios de biodiversidad, en las cuales las especies han sufrido en menor medida los efectos de grandes cambios paleoclimáticos pasados (Keppel et al., 2012; Harrison & Noss, 2017).

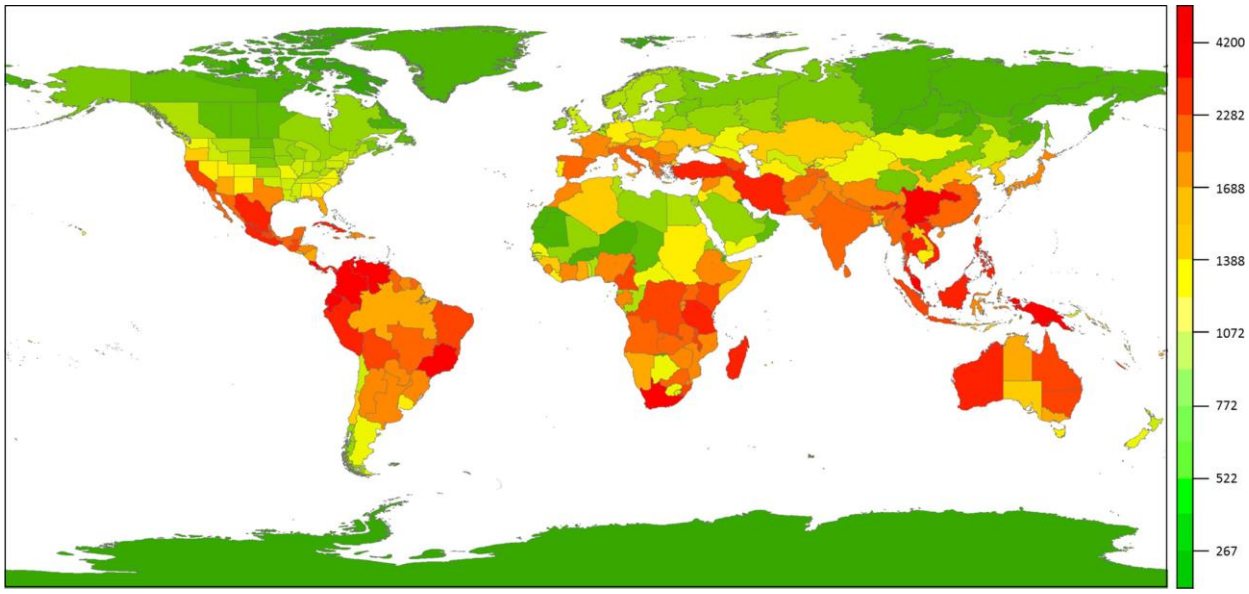


Fig. 2. Riqueza de especies reescalada utilizando la relación especie-área para cada región geopolítica. Las regiones coloreadas en verde muestran una baja riqueza de especies y las coloreadas en rojo una alta riqueza. Los valores de la escala representan el número de especies estimado. Mapa extraído de Brummitt et al. (2021).

Presentación de la temática de la tesis

La presente tesis doctoral se centra en el estudio sistemático y macroevolutivo de dos géneros de la familia de las Compuestas: *Saussurea* DC. y *Jurinea* Cass. (tribu *Cardueae*). Ambos géneros reúnen ciertas particularidades que los hacen un modelo ideal para abordar los grandes ejes de la sistemática: su gran riqueza de especies (ca. 700 especies; Susanna & Garcia-Jacas, 2007, 2009) y su distribución geográfica a lo largo del hemisferio norte: las regiones del Himalaya, la meseta del Qinghai-Tibet (QTP), y en particular las montañas Hengduan, para *Saussurea*, y la región iranoturania para *Jurinea*, son las zonas de mayor diversidad. Estas dos áreas principales de distribución son consideradas como uno de los mayores focos de biodiversidad del mundo, con tasas de endemicidad y riqueza de especies excepcionales (Boufford & Van Dyck, 1999; Myers et al., 2000; López-Pujol et al., 2006, 2011).

Sin embargo, hasta la fecha, estas áreas han sido poco exploradas por los biólogos evolutivos por causas de diversa índole que han dificultado el estudio en profundidad de los procesos de diversificación que han originado las especies que habitan allí. Por un lado, la dificultad en sí de recolección de material

biológico, ya que gran parte de la diversidad se encuentra en zonas montañosas de difícil acceso en las que además puede haber conflictos geopolíticos. Por otro lado, las limitaciones metodológicas de la secuenciación Sanger, prácticamente la única herramienta disponible que existía hace una década para la reconstrucción de filogenias moleculares. La diversificación explosiva de gran cantidad de especies en un corto período de tiempo deja una débil señal filogenética en el nivel de diferenciación interespecífico, por lo que a mayor cantidad de marcadores empleados mayor será la resolución de la historia evolutiva de un grupo y las relaciones entre sus especies. En el diseño de esta tesis doctoral se planteó abordar dos de los mayores retos actuales en la sistemática vegetal, que son la recolección de gran parte de las especies conocidas de un grupo de gran tamaño y su exploración filogenética con técnicas de secuenciación de alto rendimiento.

La revolución de la técnica *Hyb-Seq*

El uso de las técnicas conocidas como de secuenciación de alto rendimiento (*high throughput sequencing*, HTS) en organismos no modelo ha tenido un aumento exponencial estos últimos años. Ya se encuentran en marcha iniciativas para obtener genomas completos de diversos grupos como el "10KP" que tiene como objetivo secuenciar el genoma completo de 10.000 plantas en un periodo de 5 años (2017-2022; https://db.cngb.org/10kp/). Los campos de la biología evolutiva y la sistemática vegetal están experimentando una gran revolución tras la aplicación de estas nuevas técnicas, ayudando a resolver numerosas incógnitas que permanecían ocultas sobre la evolución de las plantas. En comparación con la secuenciación Sanger, las nuevas tecnologías y plataformas permiten la recuperación de genomas completos o miles de regiones específicas. Es precisamente el hecho de obtener gran cantidad de datos lo que ha supuesto uno de los mayores retos de la era genómica: cómo ensamblar y anotar los genomas (retos computacionales y analíticos) y dónde almacenarlos (retos de almacenamiento e infraestructura; Andermann et al., 2020).

En el caso de las especies vegetales, la complejidad se multiplica debido a las características de su genoma como sus enormes tamaños, o a factores como los efectos de la poliploidía, hibridación o reorganizaciones cromosómicas (Soltis et al., 2015; Pellicer et al., 2018). Para estudios sistemáticos con un enfoque filogenético, un factor también a tener presente es el amplio muestreo de especies que se debe utilizar. Además, hay que tener en cuenta que una de las fuentes principales de obtención de material vegetal son los herbarios. Estos testigos generalmente tienen el ADN muy fragmentado, e incluso contaminado, lo que ha dificultado su uso para secuenciar con Sanger (Särkinen et al., 2012) o para nuevas técnicas de alto rendimiento pero basadas en la secuenciación de largas moléculas enteras (p.ej. PacBio, NanoPore).

En este sentido, una de las técnicas de secuenciación masiva basada en fragmentos cortos (75–300 bp) con mayores ventajas en cuanto al coste, alcance taxonómico, complejidad computacional, bajos niveles de datos faltantes (lo que minimiza problemas de ortología), entre otras (ver Tabla 1), es la técnica llamada *Hyb-Seq* (Weitemier et al., 2014). Ésta combina el *genome skimming*, conocida también como *shotgun sequencing* que hace un barrido general de todo el genoma (Straub et al., 2012), con el *target enrichment*, *sequence/target/hybrid capture* o *bait hybridization* que tiene como objetivo la recuperación de un conjunto de genes diana o *targets* con una alta cobertura (Cronn et al., 2012; Lemmon et al., 2012; Dodsworth et al., 2019).

Tabla 1. Comparación de las técnicas de muestreo genómico más utilizadas en sistemática vegetal. Se detalla en escala de colores, de verde a naranja en función de lo favorable o no que sea la técnica para el factor mencionado. Información extraída de Lemmon & Lemmon (2013) y Dodsworth et al. (2019).*Para los casos de amplia aplicación taxonómica, la técnica puede utilizarse para comparaciones a diversos niveles, como familia, género, especies. Para los casos de reducida aplicación taxonómica, la técnica es útil para comparación de especies cercanas filogenéticamente o niveles por debajo de especie. **Depende; en caso de utilizar un panel o kit ya diseñado en una investigación previa su coste es reducido. Por el contrario, el diseño de un nuevo panel a partir de recursos genómicos o transcriptómicos previos o a secuenciar de nuevo encarece mucho el coste de esta fase.

Tipo	**Aleatorio**		**Estratificado**	**Mixto**
Técnica	**Genome skimming**	**GbS o RAD-Seq**	**RNA-Seq**	**Hyb-Seq**
Rango taxonómico aplicable*	Amplia	Reducida	Medio	Amplia
Acepta muestras degradadas	Sí	Medio	No	Sí
Precisa investigación genómica previa	No	No (pero útil)	No (pero útil)	Depende**
Recupera genes nucleares de baja copia	No (limitado)	No (SNPs)	Sí (miles)	Sí (variable)
Número de datos faltantes	Bajo	Alto	Medio	Medio
Coste por muestra	Medio	Bajo	Alto	Bajo
Coste en computación (análisis bioinformáticos)	Medio	Alto	Alto	Medio
Tiempo de recuperación de secuencias	Alto	Bajo	Medio	Bajo

Esta técnica tiene la particularidad de recuperar regiones nucleares conservadas de baja copia (*targets* o secuencias de genes enriquecidos) y a la vez permite recuperar gran parte del resto del genoma, como el ADN de los plastomas, ribosomas o mitocondrias. Los genes diana o *targets* se obtienen mediante un conjunto de sondas de ARN, diseñadas a partir de uno o varios transcriptomas de referencia, que se hibridan con la región de ADN complementaria de las muestras de estudio; posteriormente son amplificadas por PCR y finalmente secuenciadas. Existen paneles diseñados para recuperar información filogenética en múltiples niveles taxonómicos para un amplio rango de organismos, los llamados paneles universales, como los 353 loci para Angiospermas (Johnson et al., 2019). También se diseñan paneles específicos para grupos taxonómicos concretos como por ejemplo las familias *Cyperaceae* Juss. (Villaverde et al., 2020), *Arecaceae* Schultz Sch. (Harpe et al., 2019), o *Compositae* Giseke (Mandel et al., 2014) como se describe con detalle más adelante.

La técnica *Hyb-Seq* está resultando muy útil en los grupos en los que se está aplicando, llegando a resolver desde ramas profundas de los árboles filogenéticos (Buddenhagen et al., 2016; Léveillé-Bourret et al., 2018) hasta relaciones de especies cercanas (Villaverde et al., 2018; Larridon et al., 2020; Shee et al., 2020). Por ello se está convirtiendo en uno de los métodos más populares para obtener grandes cantidades de datos genéticos en estudios con un enfoque sistemático-filogenético, con un crecimiento exponencial durante los últimos 5 años (ver Fig. 3).

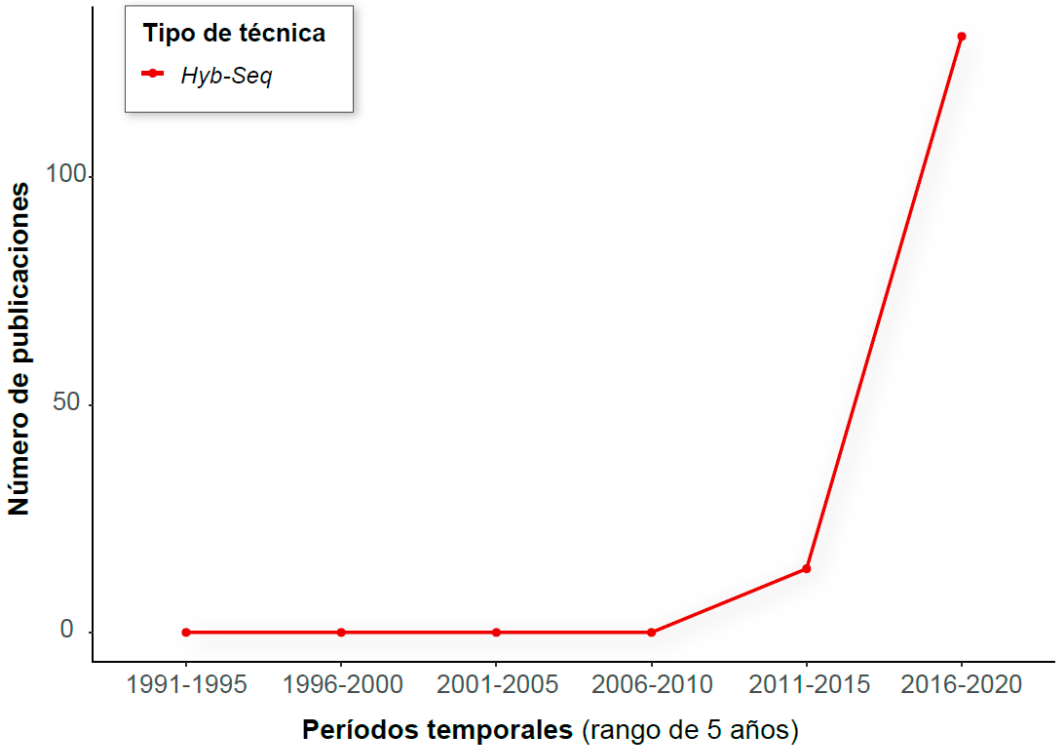


Fig. 3. Estudios publicados según WOS (Web os Science) que han utilizado la técnica *Hyb-Seq*. Datos extraídos de una búsqueda avanzada en WOS en diciembre de 2020 bajo la siguiente fórmula en la cual el período temporal varía: TS = (Hyb-Seq OR Hyb-seq OR target enrichment AND phylogenomics AND sequencing) AND PY = (1991−1995). TS son los temas y PY es el período de publicación en años.

El grupo de estudio

El siguiente apartado trata de contextualizar la posición taxonómica del grupo de estudio de la presente tesis doctoral. Se ofrece una descripción general en el nivel de familia (*Compositae* o *Asteraceae* Berchtold & J.Presl), tribu (*Cardueae* Cass.), y complejo de géneros (*Saussurea-Jurinea*). También se resumen cuáles han sido los últimos resultados obtenidos basados en los recientes avances en secuenciación de alto rendimiento y se abordan qué aspectos deberían desarrollarse para avanzar en el conocimiento del grupo de estudio.

La familia *Compositae*

Se estima que la familia de las Compuestas comprende el 10% de todas las Angiospermas, con un total de unas 25.000−35.000 especies repartidas entre 16 subfamilias, 50−51 tribus y 1600−1700 géneros (Bremer, 1994; Funk et al., 2009; Mandel et al., 2019; Susanna et al., 2020). En términos de diversidad solo son equiparables las familias *Orchidaceae* Juss. (20.000 especies; Peakall, 2007; Chase et al., 2015) y *Fabaceae* Lindley (ca. 19.500 species; Lewis et al., 2005). Su gran riqueza de especies también se traduce en una alta variabilidad morfológica y de formas vitales (Fig. 4), aunque principalmente está formada por hierbas perenes, subarbustos y arbustos; también se encuentran plantas suculentas, bejucos e incluso formas arbóreas. La familia tiene una distribución cosmopolita, con representantes en todos los continentes, incluso en la Antártida (Lewis-Smith et al., 2011).

Fig. 4. Ilustración de la variabilidad morfológica de las Compuestas. En orden de izquierda-derecha y arriba-abajo: *Werneria nubigena* Kunth, *Chuquiraga jussieui* J.F. Gmel., *Helianthus annuus* L., *Trixis californica* Kellogg, *Helenium flexuosum* Raf., *Arctotis stoechadifolia* P.J. Bergius, *Malacothrix glabrata* A. Gray, *Senecio auricula* Bourg. ex Coss., *Erigeron annuus* (L.) Pers., *Conoclinium greggii* Small, *Dendrosenecio kilimanjari* (Mildbr.) E.B. Knox, *Psathyrotes ramosissima* A. Gray. Las imágenes son de libre uso (CC BY-SA 2.0, CC-BY-NC-SA-3.0) y han sido extraídas de Panero & Crozier (2012; *The Tree of Life Web Project*).

Ecológicamente, se encuentran en cualquier tipo de hábitat terrestre y son especialmente abundantes en estepas, praderas, zonas semiáridas y áridas y regiones de clima temperado, montano o mediterráneo.

Los investigadores especialistas en la familia sugieren que su éxito evolutivo y ecológico radica en sus particulares estructuras florales (Bremer, 1994; Funk et al., 2005; Mandel et al., 2019). Todas las Compuestas comparten un mismo rasgo distintivo que es el capítulo, caracterizado por la organización de las flores en una estructura pseudoantial que se sustenta en un receptáculo, rodeado por brácteas involucrales, donde se insieren las flores. El capítulo de las Compuestas también presenta numerosas variaciones en el tamaño de la corola y color, grado de fusión de pétalos y la simetría (ver ejemplos en la Fig. 4). Su variación es atribuida a la adaptación al medio en el que viven, sus potenciales polinizadores o la transformación de estructuras de defensa contra la herbivoría. Otra característica de la familia es la presencia de vilano, que es un conjunto de cerdas o escamas pilosas simples o plumosas que rodean las flores y son posteriormente las encargadas de favorecer la dispersión de la semilla y evitar la herbivoría (Stuessy & Garver, 1996). Se ha sugerido que la aparición del vilano ha sido uno de los mayores contribuyentes a la alta diversificación de la familia y un factor evolutivo clave que habría facilitado la dispersión transoceánica e intercontinental, propiciando la colonización de lugares remotos (Panero & Crozier, 2016).

En el nivel taxonómico, su monofilia siempre ha sido clara y está apoyada tanto por datos morfológicos como moleculares (Panero & Funk, 2002, 2008; Panero & Crozier, 2016; Funk et al., 2005, 2009). Sin embargo, las relaciones filogenéticas entre subfamilias y tribus han sido, y permanecen en la actualidad, como uno de los puntos a resolver. Diversos factores han influido en la dificultad del estudio en el nivel intrafamiliar como son las rápidas radiaciones de especies, las hibridaciones antiguas o la poliploidía (Semple & Wantanabe, 2009; Huang et al., 2016). Las limitaciones de la secuenciación Sanger y la falta de muestreo de géneros clave son otros de los obstáculos que han impedido responder a cómo, cuándo y dónde se originaron o si su diversificación fue constante a lo largo del tiempo.

Con la aparición de las técnicas de secuenciación masiva, un grupo de investigadores liderado por la Dra. Jennifer Mandel desarrollaron un protocolo de laboratorio y bioinformático que permitía la extracción de cientos de loci nucleares ortólogos (llamados COS loci) que facilitaban la reconstrucción filogenética de grupos de Compuestas (Mandel et al., 2014). Este panel de marcadores para ser secuenciados con la técnica *Hyb-Seq* fue de los primeros en publicarse para grupos vegetales específicos, y el primero en estar disponible comercialmente para su uso en laboratorios moleculares (distribuido por MyBaits, Arbor Biosciences, USA). Los primeros resultados que se obtuvieron al aplicar este panel de marcadores y esta nueva técnica en un muestreo de especies reducido (de 15 a 23 especies) fueron una revolución para los botánicos especialistas en esta familia. Se obtuvieron altos apoyos estadísticos para nodos de Compuestas que no habían sido capaces de resolverse con tecnologías de secuenciación Sanger (Mandel et al., 2014, 2015).

Concretamente, se diseñaron sondas de captura de secuencia dirigidas a un conjunto de loci ortólogos, únicos o de baja copia homóloga (COS loci), identificados a través de etiquetas de secuencia expresada (*expressed sequence tags*; EST) previamente conocidos por su alto poder informativo en el nivel filogenético dentro de la familia (Chapman et al., 2007). Los transcriptomas utilizados como referencia para detectar los ESTs fueron: (1) *Carthamus tinctorius* L. (aprox. 20.000 ESTs); (2) *Helianthus annuus* L. (aprox. 70.000 ESTs) y (3) *Lactuca sativa* L. (aprox. 70.000 ESTs). El conjunto de sondas

incluyó un total de 9678 sondas dirigidas a un total de 1061 loci ortólogos. Además, la gran ventaja que presentaba el diseño del protocolo era la posibilidad de recuperar parcialmente el ADN plastidial (cloroplastos y mitocondrias) a partir de las lecturas de secuencias no pertenecientes a los COS loci.

Tras la publicación del método surgieron numerosas colaboraciones internacionales basadas en este protocolo entre el equipo de la Dra. Mandel y otros especialistas en sistemática de Compuestas para intentar esclarecer relaciones filogenéticas por resolver tanto en niveles taxonómicos elevados como familia o subfamilia (Mandel et al., 2014, 2015, 2017, 2019; Susanna et al., 2020), tribu y subtribus (Siniscalchi et al., 2019; Lichter-Marck et al., 2020; Watson et al., 2020), como bajos, género y especie (Jones et al., 2019; Thapa et al., 2020). Los trabajos presentados en la presente tesis doctoral se encuentran dentro de este grupo de estudios basados en la técnica *Hyb-Seq* y los COS loci.

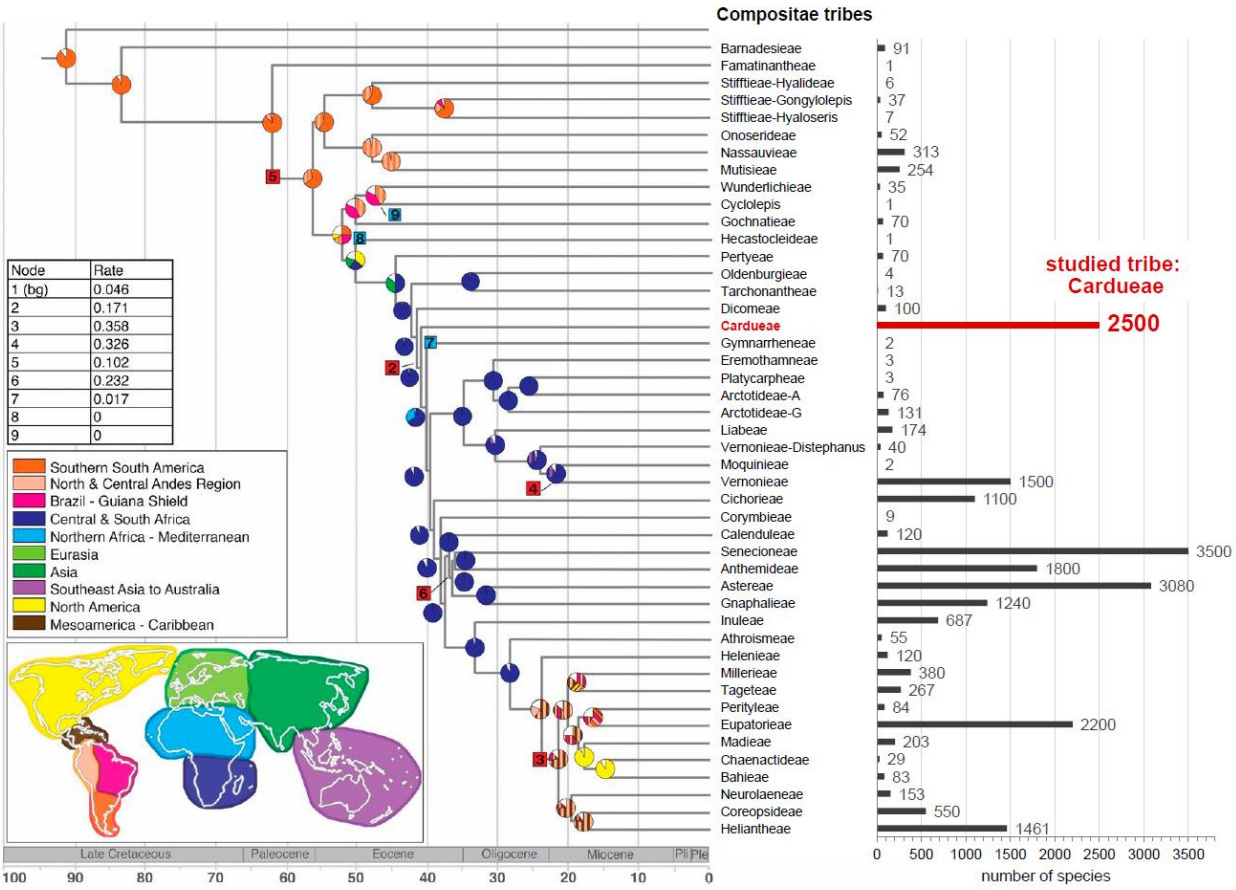


Fig. 5. Filogenia datada para las tribus y reconstrucción de áreas ancestrales ilustradas en los nodos internos. Los cambios en la tasa de diversificación se indican mediante los cuadros numerados (1–9) coloreados en rojo en caso de aumentos de la tasa o en azul en caso de disminución de la tasa. El histograma de la derecha muestra el número de especies por tribu, en el cual se resalta la tribu *Cardueae*, a la que pertenece el grupo de estudio. Figura adaptada de Mandel et al. (2019).

Basándose en datos genómicos extraídos de los COS loci y 256 táxones de la familia, Mandel et al. (2019) estimaron que tuvo su origen alrededor de 83 Ma (Cretácico tardío). La tasa de diversificación de las Compuestas se vio acelerada en el Eoceno, cuando numerosas radiaciones explosivas de especies tuvieron lugar (Fig. 5). Se considera que los desencadenantes de esta elevada especiación fueron el inicio del enfriamiento del clima y la colonización desde la región de origen (Sudamérica) hacia el continente

africano, donde la mayor parte de las tribus empezaron a divergir aprovechando nuevas oportunidades ecológicas. Posteriormente, tuvieron lugar múltiples dispersiones hacia América del Norte y luego hacia Asia (Fig. 5).

La implementación de los COS loci también ha supuesto nuevos avances en la clasificación de las Compuestas gracias a la obtención de nuevas reconstrucciones filogenéticas con mayor resolución. Respecto a la propuesta mayormente aceptada y utilizada hasta la fecha por Funk et al. (2009), se ha publicado recientemente una revisión con los cambios realizados basados en las evidencias moleculares. La nueva propuesta de Susanna et al. (2020) respecto a Funk et al (2009) sugiere el reconocimiento de 16 subfamilias en lugar de 12, y de 50 tribus en lugar de 43.

A pesar de los recientes avances, aún quedan numerosas incógnitas por resolver sobre la evolución de las Compuestas. Dado que el panel específico de los 1061 COS loci ha resultado altamente exitoso en la resolución de las relaciones filogenéticas del grupo, el gran reto al que se enfrentan los investigadores durante los próximos años es la recopilación de las especies y su secuenciación. Sin duda, se trata de un gran desafío, dada la altísima diversidad de la familia y su amplia distribución geográfica y ecológica.

La tribu *Cardueae*

La tribu *Cardueae*, compuesta por cerca de 2500 especies y 72 géneros (Susanna & Garcia-Jacas, 2009), es una de las tribus más diversas dentro la familia de las Compuestas (ver Fig. 5). Solo queda por debajo de las tribus *Senecioneae* (3500 especies) y *Astereae* (3080 especies; Funk et al., 2009; Mandel et al., 2019). Además, cabe destacar su elevado interés científico, ya que considerando las tribus más diversas (> 2000 especies), en la última década aparece como la primera o segunda tribu más citada (Fig. 6). Especialmente, se observa un crecimiento exponencial a partir de la década de los 2000 en adelante,

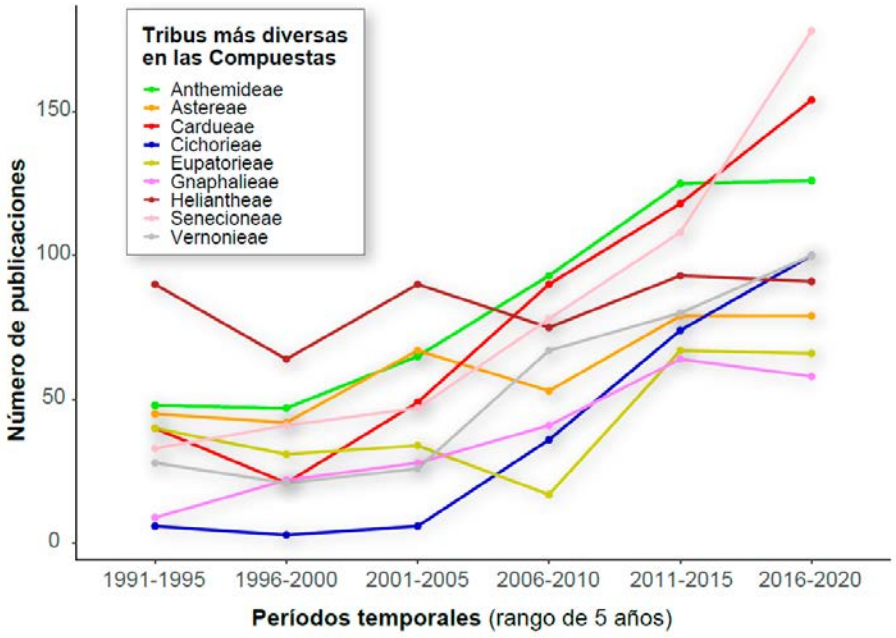


Fig. 6. Estudios publicados según WOS (Web os Science) sobre las tribus con mayor diversidad de especies de la familia *Compositae* o *Asteraceae*. Datos extraídos de una búsqueda avanzada en WOS en diciembre de 2020 bajo la siguiente fórmula en la cual el período temporal y la tribu varian: TS = (Cardueae*) AND PY = (1991–1995). TS son los temas y PY es el período de publicación en años.

coincidiendo con el surgimiento de las técnicas moleculares, que proporcionaron datos filogenéticos que confirmaban o rechazaban las hipótesis taxonómicas previas basadas en estudios morfológicos.

La tribu *Cardueae* comparte subfamilia, llamada *Carduoideae*, con tres tribus más, aunque éstas son mucho menos diversas y con una distribución restringida a África: *Dicomeae* (97 especies), *Tarchonantheae* (13 especies) y *Oldenburgieae* (4 especies; Funk et al., 2009; Ortiz et al., 2009). La subfamilia está distribuida por todos los continentes excepto la Antártida (Susanna & Garcia-Jacas, 2009), siendo el continente asiático el que alberga más del 75% de las especies (Fig. 7). Se considera una de las subfamilias con mayor éxito evolutivo. Estudios de diversificación recientes realizados por Panero & Crozier (2016) estimaron que el linaje de *Carduoideae* tenía la segunda tasa de diversificación más alta dentro de las Compuestas (0,32 especies/Ma). En las últimas filogenias de la familia, las *Carduoideae* no resultan ser un grupo monofilético (Mandel et al., 2019; Susanna et al., 2020). Sin embargo, la hipótesis más plausible sigue siendo que las *Cardueae* se originaron entre las tribus africanas *Dicomeae*, *Oldenburgieae* y *Tarchonantheae*.

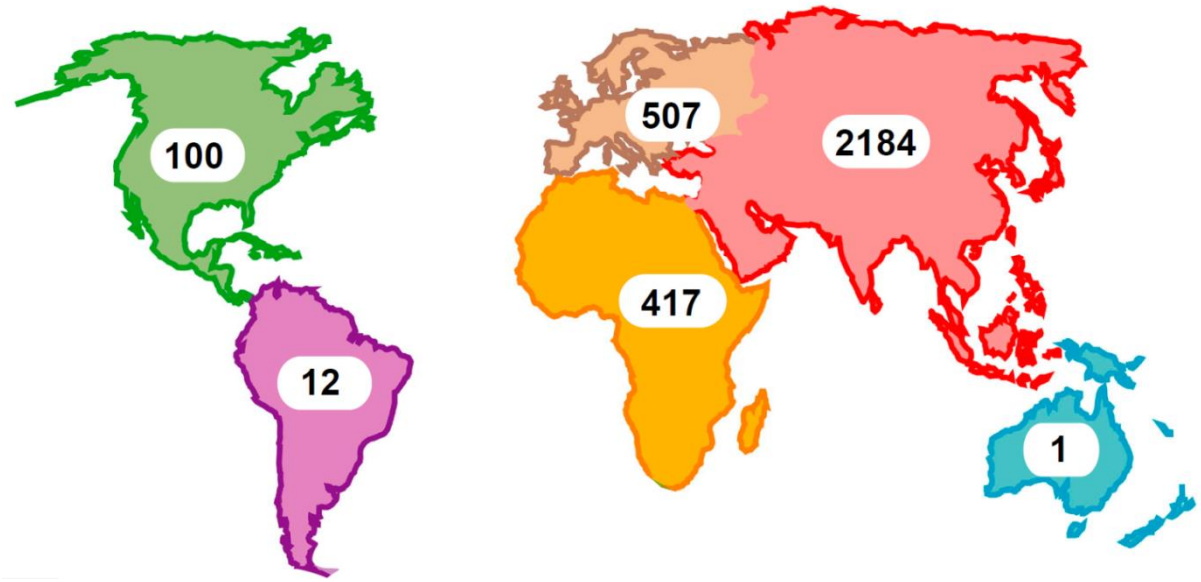


Fig. 7. Distribución global de la riqueza de especies de la subfamilia *Carduoideae* por continente. Modificado de Panero & Crozier (2016).

En cuanto a formas vitales, la tribu *Cardueae* se compone principalmente de hierbas perennes, bienales o monocárpicas, arbustos y subarbustos, a menudo espinosos; en menor grado, también se encuentran hierbas anuales e incluso formas arbóreas (Susanna & Garcia-Jacas, 2007). Morfológicamente, presentan una amplia variabilidad en cuanto a su capítulo (Fig. 8), pero todas ellas se caracterizan por una arquitectura básica del estilo, el cual tiene un anillo de pelos debajo del punto de división de los estigmas (Bremer, 1994; Susanna & Garcia-Jacas, 2009).

Fig. 8. Ilustración de la variabilidad morfológica del capítulo dentro de la tribu Cardueae. En orden de izquierda-derecha y arriba-abajo: *Echinops sphaerocephalus* L., *Carlina acaulis* L., *Carlina racemosa* L., *Cynara baetica* (Spreng.) Pau, *Atractylis cancellata* L., *Leuzea rhaponticoides* Graells, *Centaurea americana* Nutt., *Carthamus lanatus* L., *Cirsium vulgare* (Savi) Petr. Las imágenes son de libre uso (CC BY-SA 2.0, CC-BY-NC-SA-3.0) y han sido extraídas de Panero & Crozier (2012) y *The Tree of Life Web Project*.

En cuanto a su distribución, las *Cardueae* representan un elemento paisajístico clave de las regiones iranoturania y mediterránea (Quézel, 1978; Takhtajan, 1986). El mayor foco de diversidad se concentra en Asia Central (Susanna & Garcia-Jacas, 2007). Se hallan especies en una amplia variedad de hábitats (estepas, áreas semiáridas, desiertos, acantilados costeros, praderas alpinas, sabanas tropicales, etc.) y en un rango altitudinal muy amplio (desde nivel del mar hasta más de 5.000 m; Susanna & Garcia-Jacas, 2009).

Numerosas especies de *Cardueae* presentan un gran interés socio-económico, desafortunadamente, con connotaciones negativas debido a comportarse como “malas hierbas” extremadamente perjudiciales

(p.ej. *Carduus* L., *Cirsium* Mill. y *Picnomon* Adans.; Susanna & Garcia-Jacas 2009). Otras han ampliado su rango de distribución fuera de su área autóctona y causan numerosos problemas donde se establecen con un fuerte carácter invasor (*Centaurea stoebe* L. es invasora en EEUU y Canadá; Blair & Hufbauer, 2010). Por otro lado, también tienen aplicaciones beneficiosas en distintas ramas, como la alimentación (*Cynara* L. y *Carthamus* L.), la industria farmacéutica (*Carthamus tinctorius* L. como fuente de compuestos secundarios), la medicina popular (*Silybum marianum* Adans., *Cirsium ehrenbergii* Sch. Bip.) o su uso como ornamentales (*Centaurea cineraria* L.; Ellis, 1999). Algunas especies del género *Cousinia* Cass. son formadoras de paisaje y definen ecosistemas singulares a escala global (Djamali et al., 2012a).

En las últimas dos décadas ha habido importantes avances científicos en la exploración de la sistemática de *Cardueae*. En el aspecto taxonómico, la llegada de las técnicas de secuenciación molecular ayudó a definir la división subtribal mayormente aceptada hasta la fecha, extensamente descrita en Susanna & Garcia-Jacas (2007, 2009) basándose en los resultados publicados por Garcia-Jacas et al. (2002) y Susanna et al. (2006). Cinco grupos naturales resultaron bien definidos y apoyados filogenéticamente: *Cardopatiinae*, *Carduinae*, *Carlininae*, *Centaureinae* y *Echinopsinae*. Sin embargo, los autores ya resaltaron una de las limitaciones que tenía la clasificación presentada, la cual residía en las *Carduinae*, la subtribu más diversa (1700 especies, más del 70% de la diversidad de la tribu). Este grupo taxonómico englobaba todos aquellos géneros que no encajaban morfológicamente dentro de las otras subtribus, pero a su vez entre ellos eran altamente variables y de incierta relación filogenética dadas las limitaciones que presentaba la secuenciación Sanger.

En el aspecto biogeográfico, el descubrimiento de nuevos registros fósiles en *Compositae* y *Cardueae*, sumado al avance de los métodos para datar filogenias y reconstruir sus áreas ancestrales, supuso la publicación del primer trabajo donde se recopilaban las fechas de divergencia entre las subtribus y sus principales rutas de dispersión (Barres et al., 2013). Se infirió que la tribu podría haberse originado en el Eoceno medio en Asia occidental, una zona que también se estima como área ancestral para la mayoría de las subtribus. Los grandes eventos de diversificación de la tribu se asociaron a movimientos geológicos que tuvieron lugar entre el Oligoceno-Mioceno, como los ciclos continuos de conexión y división entre la microplaca de Anatolia y la cuenca Mediterránea occidental o el levantamiento de la cordillera del Himalaya. Posteriormente, durante el Plioceno-Pleistoceno se sugiere que el enfriamiento global propició la expansión de las *Cardueas* hacia, por ejemplo, el Nuevo Mundo y África.

A pesar de todos estos significativos avances, aún no se ha explorado el alcance que tendría la secuenciación de alto rendimiento y la aplicación de los COS loci especialmente para resolver la posición y división de la subtribu *Carduinae* y la actualización del marco de evolución espaciotemporal de la tribu en base a una filogenia bien resuelta.

El complejo *Saussurea-Jurinea*

A pesar de la riqueza de especies de la tribu *Cardueae*, ésta no está homogéneamente repartida entre los géneros que la componen (ver Fig. 9). Existen dos géneros hiperdiversos (> 500 especies) que son *Cousinia* y *Saussurea*. Luego, con entre 100−200 especies le siguen los géneros *Cirsium*, *Centaurea*, *Jurinea* y *Echinops*. Todos ellos ya representan alrededor del 80% de especies de la tribu. En cuanto a los antecedentes, han sido considerablemente explorados en el campo de la sistemática molecular, filogenética y biogeografía, en mayor o menor grado los siguientes: *Centaurea* L. (Garcia-Jacas et al., 2000, 2001, 2006; Font et al., 2002, 2009; Hilpold et al., 2014); *Cousinia* (Mehregan, 2008; López-

Vinyallonga et al., 2009, 2011; Mehregan & Kadereit, 2009; Minaeifar et al., 2016); *Echinops* (Garnatje et al., 2005; Sánchez-Jiménez et al., 2010, 2012; Montazerolghaem et al., 2017) y *Cirsium* (Kelch & Baldwin, 2003; Slotta et al., 2012; Ackerfield et al., 2020a, 2020b). Sin embargo, *Saussurea* y *Jurinea* siguen siendo los géneros menos explorados en el conjunto de las Cardueas a pesar de su alta diversidad, entre 400−500 y 200 especies, respectivamente (Lipschitz, 1979; Susanna & Garcia-Jacas, 2007, 2009; Shi & Raab-Straube, 2011; Chen, 2015; Raab-Straube, 2017).

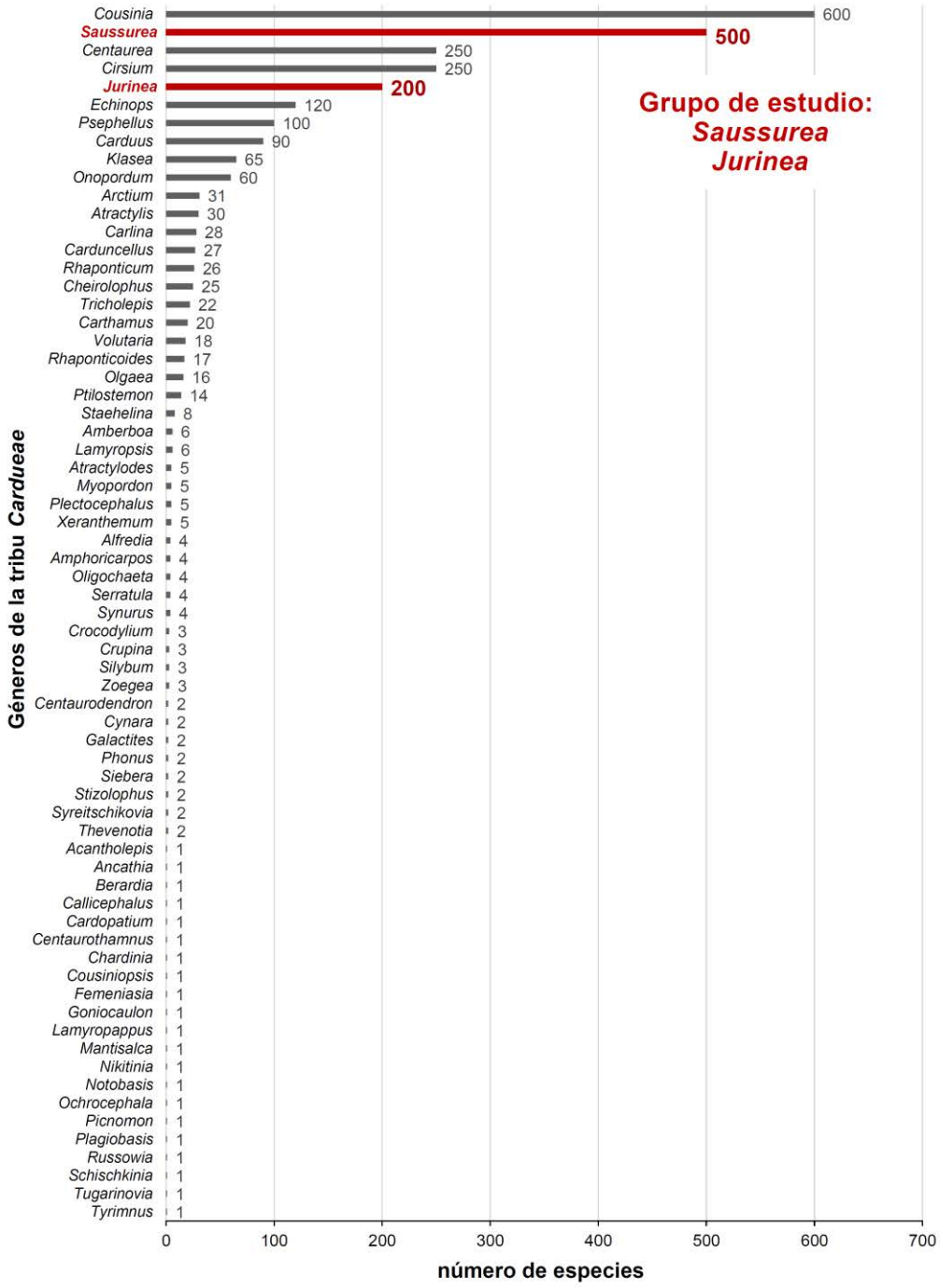


Fig. 9. Distribución del número de especies por género de la tribu *Cardueae* en orden ascendente. En rojo se resaltan los géneros sujetos a un estudio evolutivo en profundidad en esta tesis.

Morfológicamente, el complejo *Saussurea-Jurinea* es un grupo natural bien definido y comúnmente aceptado (Susanna et al., 2006; Susanna & Garcia-Jacas, 2007, 2009; Shi & Raab-Straube, 2011). Está compuesto por hierbas perenes o subarbustos (ver Figs. 10 y 11), raramente bienales o anuales. Sus hojas son por lo general glabras en la cara superior y blanco-plateadas en la cara inferior. Los capítulos son cilíndricos o globosos, a menudo con una disposición en tipo panícula, homógamos y con un vilano de páleas muy largas, vistosas y plumosas, agrupadas en una base formando un anillo (Susanna & Garcia-Jacas, 2007). Aunque ambos géneros exhiben una gran variabilidad morfológica, es especialmente reseñable la morfología que han adquirido algunas especies de *Saussurea* adaptadas a condiciones extremas de alta montaña (Shi & Raab-Straube, 2011; Chen, 2015), como las llamadas *snowball plants* del subgénero *Eriocoryne* con una densa cobertura de un indumento lanudo grueso o el tipo *greenhouse plants* del subgénero *Amphilaena*, en el cual la inflorescencia está oculta por brácteas semitransparentes de color blanco-amarillento o morado (ver ejemplos en la Fig. 10).

En lo tocante a la distribución general, se encuentran especies de *Saussurea* y *Jurinea* a lo largo de todo el hemisferio norte, siendo los hábitats de media montaña de Asia Central y Oriental los lugares con mayor diversidad. El género *Saussurea* ha especiado de manera notable en la región de la meseta del Qinghai-Tibet (QTP) y sus áreas adyacentes, la cordillera del Himalaya y las montañas Hengduan, que albergan cerca del 63% de endemismos en contraposición con el resto de Asia, y especialmente Europa (8 especies) y el continente Americano (6 especies; Lipschitz, 1976; Chen, 2015). En cuanto a las preferencias ecológicas, pueden encontrarse especies de *Saussurea* en una gran diversidad de hábitats: prados alpinos y sub-alpinos, estepas, bosques húmedos, zonas rupícolas o pedregales, etc (Fig. 12). Por otro lado, el mayor foco de especiación de *Jurinea* se sitúa en la región iranoturania. Especialmente, las montañas de Asia Central, Pamir-Alai y Tian Shan (Kazakhstan, Kyrgyzstan, Tajikistan, Turkmenistan, Uzbekistan) albergan cerca de 98 especies, y la zona sur de Asia occidental (Cáucaso, Irán, Turkmenistán y Afganistán) unas 50 especies (Cherneva & Tsukervanik, 1993; Rechinger & Wagenitz, 1979; Cherneva, 2008). Muchas especies de *Jurinea* son consideradas elementos esteparios (Szukala et al., 2019), aunque también hay especies alpinas e incluso adaptadas a sistemas dunares o acantilados costeros (Fig. 13).

Aunque la recopilación de los estudios realizados hasta la fecha en el grupo *Saussurea-Jurinea* pueda sugerir que ambos géneros han estado bien estudiados (ver Tabla 2), estos trabajos previos sobre sistemática muestran dos remarcables deficiencias o limitaciones a la hora de extraer conclusiones sólidas: (1) el muestreo incompleto, ya que ninguno de ellos es exhaustivo en el total de especies incluidas; ni por lo que respecta a área de distribución ni por la representación de los principales grupos; y (2) los marcadores moleculares utilizados para la reconstrucción filogenética, en todos los casos un bajo número de marcadores basados en la secuenciación Sanger (principalmente el espaciador interno transcrito o ITS), lo que ha resultado en relaciones o agrupaciones de clados no respaldadas estadísticamente.

Fig. 10. Ilustración de la variabilidad morfológica de *Saussurea*. En orden de izquierda-derecha y arriba-abajo: *Saussurea alpina* DC., *Saussurea pygmaea* Spreng., *Saussurea gossypiphora* D. Don, *Saussurea wellbyi* Hemsl., *Saussurea obvallata* Wall., *Saussurea gnaphalodes* Ostenf., *Saussurea stella* Maxim., *Saussurea medusa* Maxim., *Saussurea pygmaea* Spreng., *Saussurea costus* (Falc.) Lipsch., *Saussurea americana* Eaton, *Saussurea angustifolia* (L.) DC. Las imágenes son de libre uso (CC0, CC BY-SA 4.0, CC-BY-NC-SA-3.0) y han sido extraídas de Flora of China.

Fig. 11. Ilustración de la variabilidad morfológica de *Jurinea*. En orden de izquierda-derecha y arriba-abajo: *Jurinea roegneri* K. Koch, *Jurinea cyanoides* (L.) DC., *Jurinea longifolia* DC., *Jurinea moschus* (Hablitz) Bobrov, *Jurinea mongolica* Maxim., *Jurinea humilis* (Desf.) DC., *Jurinea ledebourii* Bunge, *Jurinea ruprechtii* Boiss. Las imágenes son de libre uso (CC-BY-NC-SA-3.0, 4.0) y han sido extraídas de GBIF Secretariat (2021a).

Fig. 12. Representación de los hábitats de *Saussurea*. Las imágenes son de libre uso (CC-BY-NC-SA-3.0, 4.0) y han sido extraídas de GBIF Secretariat (2021b).

Fig. 13. Hábitats de *Jurinea*. Las imágenes son de libre uso (CC-BY-NC-SA-3.0, 4.0) y han sido extraídas de GBIF Secretariat (2021a).

Una inspección general de los antecedentes por tipos de publicaciones de ambos géneros nos indica que hasta 2016 han predominado los artículos sobre descripción de nuevas especies o táxones, por detrás de las revisiones taxonómicas (ver Tabla 2). Este último hecho se explica por la larga y complicada historia taxonómica que tiene el complejo *Saussurea-Jurinea*, en el cual se han llegado a segregar hasta 15 géneros (ver Fig. 14). Hasta la fecha, no se ha podido llegar a un consenso sólido de cuantos géneros realmente alberga el complejo. Algunos de los factores que han dificultado su clasificación son: (1) la falta de muestreo en estudios filogenéticos, cuyos árboles resultantes en los estudios con mejor muestreo no superaban el 10% del total de especies de *Saussurea* (Raab-Straube, 2003; Wang et al., 2009) y 9% de *Jurinea* (Dogan et al., 2010a); (2) las adaptaciones morfológicas que presentan algunos táxones a condiciones extremas en zonas de montaña de gran elevación, los cuales han llevado a clasificaciones artificiales basadas en caracteres resultantes de la convergencia evolutiva (Wang et al., 2009); (3) clasificaciones basadas casi exclusivamente en el aquenio, propenso a presentar accesorios como surcos, picos, estrías o protuberancias (Häffner, 2000) que pueden haber sido interpretados erróneamente como caracteres taxonómicos verdaderamente informativos; y (4) la aceptación como agrupaciones válidas de grupos polifiléticos resultantes de análisis filogenéticos (p.ej. Raab-Straube, 2003; Kita et al., 2004; Wang et al., 2013). Por lo tanto, resulta urgente la exploración de todos estos géneros para clarificar los límites genéricos del complejo basados en nuevas filogenias bien resultas que podrían proporcionar las técnicas de secuenciación masiva.

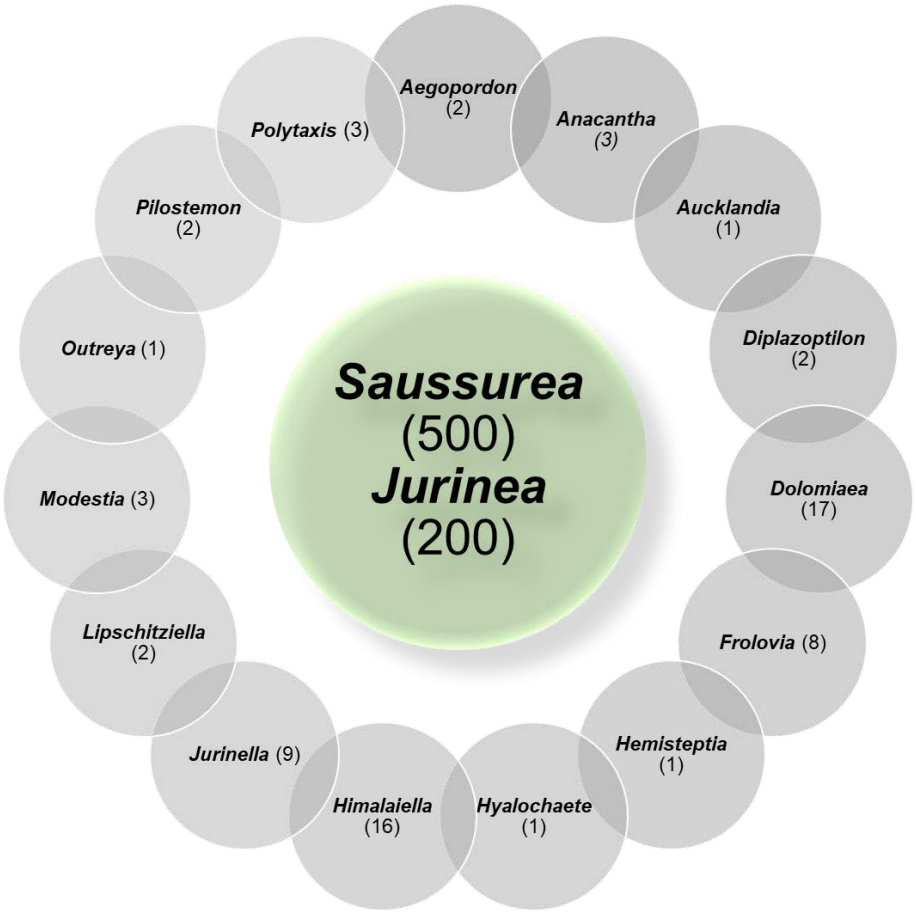


Fig. 14. Géneros satélite descritos y considerados en los tratamientos taxonómicos más recientes del complejo *Saussurea-Jurinea*. Entre paréntesis se especifican el número de especies descritas para cada género.

Uno de los indicadores que muestra el elevado interés que estos géneros están suscitando en los últimos años es el gran número de estudios que han sido publicados sobre ellos durante el transcurso de la presente tesis doctoral, 26 en el caso de *Saussurea* y 5 en *Jurinea*. Para el género *Saussurea* su aumento ha sido espectacular, probablemente debido al interés creciente del estudio de las radiaciones de especies, y en particular, en la región QTP (ver revisión por Wen et al., 2014). Desde 2016, la continua publicación de artículos de revisión taxonómica, así como la descripción de nuevas especies (11 en *Saussurea* y 2 en *Jurinea*), ponen de relieve que aún queda un largo recorrido para conocer más acerca de su historia evolutiva, así como la dificultad de establecer límites genéricos e infragenéricos en el grupo. También es destacable en *Saussurea* la aparición de trabajos basados en técnicas de secuenciación masiva, principalmente estudios centrados en la presentación de cloroplastos enteros para determinadas especies.

Tabla 2. Resumen de las publicaciones en el campo de la sistemática sobre los géneros *Saussurea* y *Jurinea*. Abreviatura: ref = referencias.

Tipo de estudio	*Saussurea*		*Jurinea*	
	Antes de la tesis (< 2016)	**Durante la tesis (2016-2021)**	**Antes de la tesis (< 2016)**	**Durante la tesis (2016-2021)**
Nº revisiones taxonómicas (ref)	6 (Narits et al., 2000; Fujikawa et al., 2007; Fujikawa, 2010; Ghimire et al., 2016; Raab-Straube, 2011; Yuan et al., 2015)	4 (Raab-Straube, 2017; Smirnov et al., 2018; Kasana et al., 2018, 2020)	3 (Conti, 1998; Sennikov & Lazkov, 2013; Altinordu & Crespo, 2016)	1 (Kasana et al., 2019)
Nº estudios de filogenia y/o filogeografía (ref)	6 (Raab-Straube, 2003; Kita et al., 2004; Wang & Liu, 2004; Wang et al., 2007, 2009; Gailite & Rungis, 2012)	5 (Xing & Ree, 2017; Chen et al., 2019; Xu et al., 2019; Zhang et al., 2019, en revisión)	3 (Dogan et al., 2007, 2009, 2010a)	1 (Szukala et al., 2019)
Nº artículos de descripción de nuevas especies (ref)	8 (Fujikawa & Ohba, 2002; Raab-Straube, 2009; Xu et al., 2013; Chen, 2014; Chen & Yuan, 2014, 2015; Wang et al., 2005, 2014)	9 (Kadota, 2017; Chen & Wang, 2018; Rana et al., 2018; Chen, 2020; Xu et al., 2020; Zhang et al., 2019; Chen & Xu, 2020; Pyak et al., 2020; Ri et al., 2020)	6 (Dogan et al., 2010b, 2010c; 2011, 2014; Mirtadzadini et al., 2011; Stevanović et al., 2010)	2 (Mirtadzadini & Joharchi, 2017; Aksoy et al., 2018)
Nº artículos de distribución y ecología (ref)	4 (Faju et al., 1999; Oh et al., 2002; Butola & Samant, 2010; Wen et al., 2014)	4 (Kou et al., 2017; Shurupova et al., 2017; Peng et al., 2019; Norris et al., 2020)	2 (Dogan & Duran, 2009; Balos & Akan, 2015)	1 (Ahmed et al., 2020)
Nº artículos de secuenciación de cloroplastos (ref)	0	7 (Cheon, et al., 2017; Xie et al., 2017; Yun et al., 2017; Peng et al., 2019; Wang et al., 2020a, 2020b; Wang et al., 2021)	0	0

La historia biogeográfica y evolutiva de *Jurinea* es prácticamente desconocida. Únicamente, Dogan et al. (2010a) reconstruyó un árbol filogenético a partir de marcadores ITS con 18 especies, pero sin una representación amplia de su área de distribución global. Para *Saussurea*, la filogenia que presentaron Wang et al. (2009), también basada en la región ITS, reveló que el género habría sufrido una radiación de tipo insular en ambientes alpinos de la región QTP hace unos 14–7 Ma, coincidiendo con los eventos de levantamientos geológicos de QTP.

Las radiaciones evolutivas

Conceptos generales

Una de las ramas de la sistemática vegetal que más está creciendo en los últimos años con la aplicación de datos filogenómicos es el estudio de las radiaciones evolutivas o historia de la diversificación de las especies (Fig. 15), especialmente en comparación con los estudios con un enfoque más taxonómico-nomenclatural. Comúnmente, se define como radiación el aumento significativo en la tasa de diversificación y/o disparidad comparado con las tasas de diversificación globales del grupo de estudio y que tiene consecuencias a escala macroevolutiva (Schluter, 2000; Donoghue & Sanderson, 2015; Nürk et al., 2019).

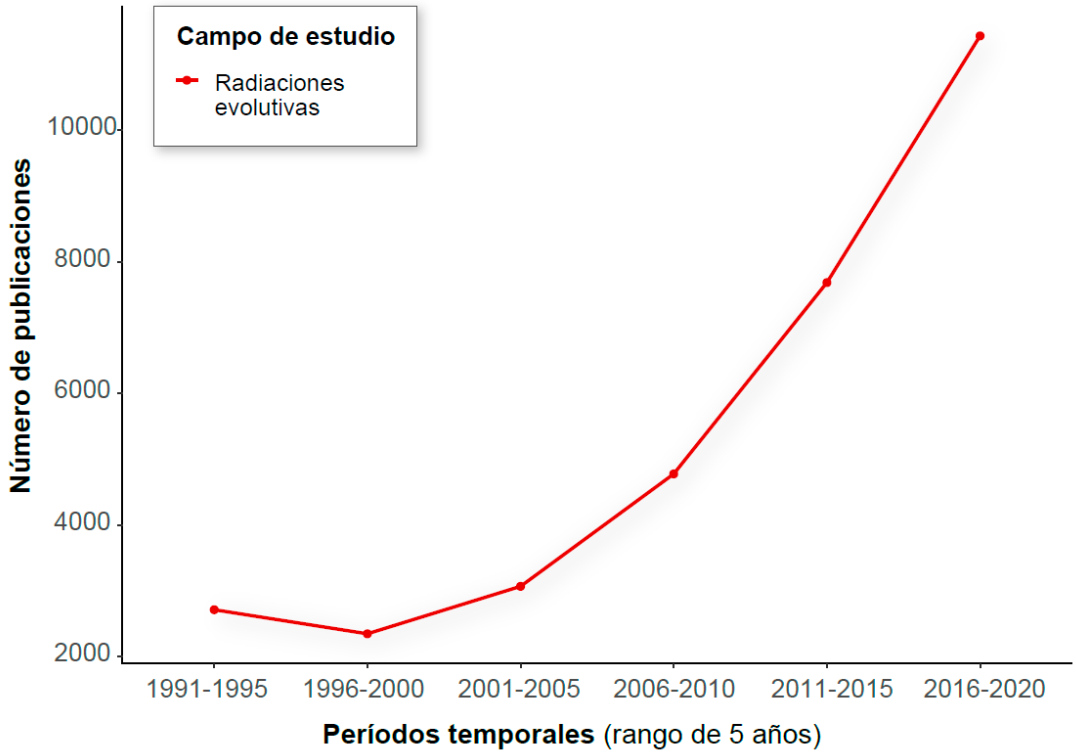


Fig 15. Estudios publicados según WOS (Web of Science) sobre radiaciones evolutivas. Datos extraídos de una búsqueda avanzada en WOS en diciembre de 2020 bajo la siguiente fórmula en la cual el período temporal varía: SU = Plant Science AND TS = (Evolutionary radiation OR Radiation OR Diversification OR Speciation AND phylogen* AND Sequencing) AND PY = (1991–1995). SU es el campo de estudio, TS son los temas y PY es el período de publicación en años.

Las radiaciones se clasifican comúnmente entre radiaciones adaptativas y no adaptativas. Por un lado, las adaptativas consisten la proliferación de especies que divergen por la adaptación a diversos recursos disponibles que proporciona el entorno, dando lugar a una amplia diversidad morfológica y ecológica (Schluter, 2000; Rundell & Price, 2009; Givnish, 2015). La aceleración en las tasas de diversificación se ve potenciada en este caso, y principalmente, por factores bióticos, dando lugar a especiaciones de tipo simpátrico (Simões et al., 2016). Por otro lado, las radiaciones no adaptativas, llamadas también geográficas o climáticas, consisten en la proliferación de especies únicamente por causas de aislamiento reproductivo por barreras geográficas, dando lugar a modos de especiación de tipo alopátrico (Czekanski-Moir & Rundell, 2019). También pueden darse radiaciones evolutivas simplemente por disminuciones de la tasa de extinción en vez de aumentos de la especiación *per se*, llamadas pseudorradiaciones (Simões et al., 2016).

Preguntas clave para el estudio de las radiaciones

Generalmente, los estudios sobre radiaciones evolutivas abordan tres grandes cuestiones: (1) la historia espacio-temporal de la evolución del grupo; (2) el tempo o ritmo que ha tenido la radiación; y (3) los impulsores o factores que han desencadenado la radiación. Según los resultados obtenidos, se suele atribuir uno o varios epítetos que describen qué tipo de radiación más probable ha experimentado el grupo de estudio (ver detalles en Tabla 3). Por ejemplo, aparte de los adjetivos más comunes como radiación adaptativa, geográfica, o ecológica, también se suelen describir los impulsores específicos de su diversidad como por ejemplo radiación alpina, si la radiación ha tenido lugar principalmente en regiones montañosas y una parte importante de las especies se ha adaptado a las condiciones del piso alpino.

En relación a los tres principales ejes, el primer punto se centra en la identificación de grupos monofiléticos, la estimación de tiempos de divergencia entre linajes, la reconstrucción de área ancestral y sus rutas de dispersión. El segundo punto tiene como objetivo la exploración de la dinámica de las tasas de diversificación, como localizar en qué puntos de la filogenia se producen cambios significativos en estas tasas, en qué momento en el tiempo, o si hay varios puntos de cambios de tasa. El último punto es sin duda uno de los aspectos que plantean más desafíos hasta la fecha (Bouchenak-Khelladi et al., 2015): encontrar o hipotetizar qué factores han promovido o impulsado la diversificación de las especies. Se intenta relacionar el marco de diversificación con eventos o factores que actuaron como "fondo" (presentes antes del inicio de una radiación), "desencadenantes" (contemporáneos a una radiación) o "moduladores" (que surgen después de una radiación; Donoghue & Sanderson 2015). Tradicionalmente, han sido reconocidos dos tipos de factores o condiciones que promueven la especiación (Simpson, 1953; Bouchenak-Khelladi et al., 2015): (1) los extrínsecos, como la elevación de montañas, oscilaciones del nivel del mar, aparición de islas oceánicas, cambios climáticos, fragmentación del paisaje, colonización de nuevos territorios, reordenamientos de ecosistemas, extinciones masivas, cambios en la presión de depredación; y (2) los rasgos intrínsecos, como adaptaciones morfológicas, por ejemplo, fruto de la coevolución con polinizadores, ventajas fisiológicas, cambios en la estrategia del ciclo vital, cambios en el modo de dispersión, poliploidía, hibridación, etc. Ambos tipos de factores, extrínsecos e intrínsecos, y varios dentro de cada categoría, pueden actuar a la vez e interactuar sinérgicamente (hipótesis de la confluencia; Donoghue & Sanderson, 2015), o dejar señales de mezclas complejas de oportunidades extrínsecas e innovaciones intrínsecas en las radiaciones de especies (Bouchenak-Khelladi et al., 2015).

Tabla 3. Categorización de las radiaciones evolutivas en plantas según los atributos comunes que se encuentran en la literatura sobre diversificación de especies. Aquí, usamos el término "radiación" bajo la circunscripción más general, equivalente a "divergencia de especies", es decir, un proceso de aumento de la diversidad en un linaje (Linder, 2008), en lugar de una asociación estricta con un aumento de la tasa de diversificación ("especiación explosiva") y/o diversificación de roles ecológicos (Givnish, 2015). Referencias: [1]Whitfield & Lockhart (2007); [2]Linder (2008); [3]Knope et al. (2012); [4]Givnish (2015).

Cuestiones clave	Factores	Atributos adscritos a las radiaciones	Breve descripción
(1) Historia espacio-temporal del grupo	Número de cambios en la tasa de diversificación	Única	Episodio único de cambio en la tasa de diversificación dentro de un grupo
		Múltiple	Varias subradiaciones (pulsos de diversificación acelerada o cambios) dentro del grupo
	Historias evolutivas	Independiente	Varias subradiaciones comienzan a divergir en diferentes rangos temporales
		Paralela	Varias subradiaciones comienzan a divergir al mismo tiempo
	Tiempo desde la divergencia	Antigua	Largo tiempo transcurrido desde la divergencia del grupo, correspondiente aproximadamente desde el Jurásico tardío al Paleógeno[1]
		Reciente	Corto tiempo transcurrido desde la divergencia del grupo, del Neógeno al Cuaternario[2]
	Escala geográfica (especies/área)	Amplia	Diversificación a macroescala en vastos territorios, áreas de distribución continentales o en todo el mundo (ejemplos [3])
		Estrecha	Diversificación a microescala con un rango geográfico restringido, p. ej. archipiélagos oceánicos[3]
(2) Tempo	Cambio en la tasa de diversificación	No explosiva	Divergencia de especies no asociada a un cambio en la tasa de diversificación
		Explosiva, rápida	Aumento significativo de la tasa de diversificación neta (especiación - extinción)[4] dentro de un grupo
(3) Impulsores	Factores de fondo, desencadenantes o moduladores	Condiciones extrínsecas	Elevación o deformaciones de montañas, oscilaciones del nivel del mar, aparición de islas oceánicas, cambios climáticos, fragmentación del paisaje, reordenamientos de ecosistemas, extinciones masivas, cambios en la presión de depredación, expansiones geográficas, colonizaciones o eventos de dispersión a larga distancia
		Rasgos intrínsecos	Innovaciones, sinnovaciones, adaptaciones morfológicas, ventajas fisiológicas, cambios en los rasgos del ciclo de vida, cambios en el modo de dispersión, poliploidía, hibridación

Retos actuales en el estudio de las radiaciones

Por lo general, se sigue un flujo de trabajo común en los estudios de radiación de especies, que comienza con la reconstrucción de la historia filogenética fechada para el grupo de interés. Este primer paso es crucial, ya que representa la base para dibujar hipótesis evolutivas sólidas y rastrear factores potenciales que dieron forma a la historia evolutiva del grupo. Tener esta buena base depende de poder superar varios factores que han resultado y siguen resultando hoy en día unos grandes desafíos: (1) un muestreo de especies casi completo, o incluso un muestreo denso en el nivel intraespecífico. Esto es debido a que un muestreo incompleto pude hacer variar sustancialmente las edades estimadas de la

filogenia (Linder et al., 2005), además de ocasionar resultados sesgados en las reconstrucciones de área ancestral si no se incluyen todos los posibles linajes hermanos (Bacon et al., 2016); (2) un apoyo robusto de los principales clados, es decir, filogenias estadísticamente respaldadas; (3) la disponibilidad de registros fósiles relacionados o un uso preciso de puntos de calibración secundarios previamente publicados; (4) una base de datos por especie con información sobre rasgos morfológicos, fisiológicos o cualquier carácter con potencial evolutivo; y (5) registros geográficos a escala fina para realizar análisis ambientales y basados en el espacio de nicho ocupado, tanto presente como pasado (referencias en Donoghue & Sanderson, 2015; Hughes et al., 2015).

En relación con el primer punto (1), ciertamente, una de las principales críticas que reciben los estudios de radiación es la baja cobertura de muestreo o representación del total de su diversidad (Heath et al., 2008). Es cuestionable cómo se pueden proponer conclusiones sobre la diversificación de un grupo si no se incluyen todas sus especies conocidas. En estos últimos años se están realizando grandes esfuerzos para aumentar los muestreos en los trabajos sobre radiaciones evolutivas, por ejemplo, la inclusión de un 60% de las 300 especies de *Saxifraga* L. (Ebersbach et al., 2017), un 60% de 200 especies en *Hypericum* L. (Nürk et al., 2018), 60% de 800 especies en *Erica* L. (Pirie et al., 2019), un 66% de 2000 especies en *Carex* L. (Martín-Bravo et al., 2019) o un 75% de 400 especies de la tribu *Antirrhineae* (Fernández-Mazuecos et al., 2019). Las colecciones depositadas en los herbarios están siendo una fuente indispensable de material que puede ser secuenciado, especialmente en el caso de grupos con muchas especies o para aquellos distribuidos en áreas remotas o políticamente inestables (Villaverde et al., 2018; Brewer et al., 2019).

Relacionado con el segundo punto (2), las técnicas moleculares de alto rendimiento están ofreciendo, en general, apoyos estadísticos de ramas más altos que con las técnicas tradicionales para grupos que previamente no podían resolverse y solo se conseguían obtener politomías no resueltas con grandes ensamblajes de especie (Parks et al., 2009). Con respecto al tercer punto (3), los nuevos descubrimientos de fósiles (p.ej. para Compuestas; Barreda et al., 2010) o los métodos de datación molecular revisados están mejorando las estimaciones de divergencia temporal (Clarke et al., 2011). Además, para fechar filogenias se están utilizando algoritmos mejorados basados en verosimilitud penalizada (Smith & O'Meara, 2012; Maurin, 2020), que reducen los tiempos de computación para fechar filogenias derivadas de la secuenciación masiva, los cuales parten directamente de la topología base de la filogenia (p.ej. Stubbs et al., 2018).

Para la generación de bases de datos de rasgos morfológicos y funcionales (punto 4), desafortunadamente no se observan grandes avances en su aplicación para el estudio de las radiaciones, más allá de los métodos desarrollados para verificar su influencia en la especiación o extinción de los grupos estudiados (modelos SSE; Ng & Smith, 2014; Rabosky & Goldberg, 2017; Harvey & Rabosky, 2018; Han et al., 2020). Probablemente este hecho sea debido a la disminución del interés científico de la disciplina más puramente taxonómica (Hopkins & Freckleton, 2002; Joppa et al., 2011). Esto puede afectar colateralmente a la investigación de las radiaciones, como por ejemplo: (a) obviando caracteres micro o macro morfológicos clave en la evolución de las especies; (b) no descubriendo especies crípticas; o (c) trabajando con un grupo de estudio para el que el tratamiento taxonómico consenso es insuficiente o inexistente (Hughes & Atchison, 2015).

El aumento de muchas fuentes disponibles para recopilar datos espaciales de biodiversidad (Jetz et al., 2012) está facilitando la incorporación de conjuntos de datos ambientales estructurados

geográficamente en un contexto filogeográfico (punto 5). Se están publicando resultados sorprendentes, como la radiación dependiente de la temperatura paleoclimática de las campanillas (*Campanulaceae*: *Lobelioideae*) neotropicales en los Andes, cuyas tasas de extinción fueron más bajas durante los intervalos pasados más fríos y la colonización de los bosques nubosos de altura media fue posible gracias a los linajes preadaptados al frío (Lagomarsino et al., 2016). Otro interesante ejemplo es la integración de tendencias de diversificación con datos de nicho ecológico y evolución fenotípica en la radiación global de plantas de *Saxifragales* realizado por Folk et al. (2019). Los autores destacaron un importante cambio de nicho en las especies coincidiendo con la proliferación de hábitats más fríos y secos a partir de los 5 Ma en adelante. Aunque recientemente están surgiendo resultados prometedores después de integrar datos ecológicos-paleoclimáticos con la historia filogeográfica de las especies, son aún muchos los grupos para los que se desconoce el papel que ha podido jugar la ecología y los eventos de cambio climático en la historia de su diversificación.

Regiones biogeográficas diversas y poco exploradas

Existen ciertos enclaves geográficos que resultan altamente atractivos para el estudio de las radiaciones evolutivas de plantas, como los sistemas alpinos o insulares oceánicos. El atractivo de su estudio reside en sus características intrínsecas como su tamaño limitado, aislamiento y gran diversidad de macro y micro hábitats debido a su acusada influencia de la tercera dimensión, la altitud (Seehausen, 2015). La colonización de estos enclaves por parte de linajes ancestrales pudo conducir a numerosas radiaciones de especies que se adaptaron a la gran multitud de nichos ecológicos distintos disponibles, como por ejemplo, en las Islas Galápagos las margaritas arborescentes de Darwin del género *Scalesia* Arn. (Fernández-Mazuecos et al., 2020), o la diversificación alpina de *Lupinus* L. en los Andes (Drummond, 2008; Hughes & Eastwood, 2006; Drummond et al., 2012).

En particular, los estudios de radiaciones en plantas se han centrado en estos últimos años en explorar la historia evolutiva de los grupos en determinados enclaves geográficos, como las zonas montañosas de los Andes y el Himalaya o la región del Cabo y la cuenca Mediterránea (Fig. 16). Por el contrario, dos áreas altamente diversas, pero aún poco investigadas, son las montañas Hengduan y la región iranoturania, que justamente son dos de los principales focos de diversificación de *Saussurea* y *Jurinea*, respectivamente. Dado que ambos géneros son elementos característicos de estas áreas y presentan una alta diversidad morfológica, funcional y ecológica, los dos representan casos de estudio ideales para investigar cómo llegaron a diversificarse en estas zonas y como se expandieron a sus áreas adyacentes llegando a tener representantes repartidos por prácticamente todo el hemisferio norte. La exploración de su historia biogeográfica puede servir para entender los procesos de especiación y los factores que han dado lugar a la diversidad que actualmente se encuentra en estas regiones. A continuación, se hace una breve descripción de las características y particularidades que presentan estas áreas y pueden influir en la historia evolutiva de su biota.

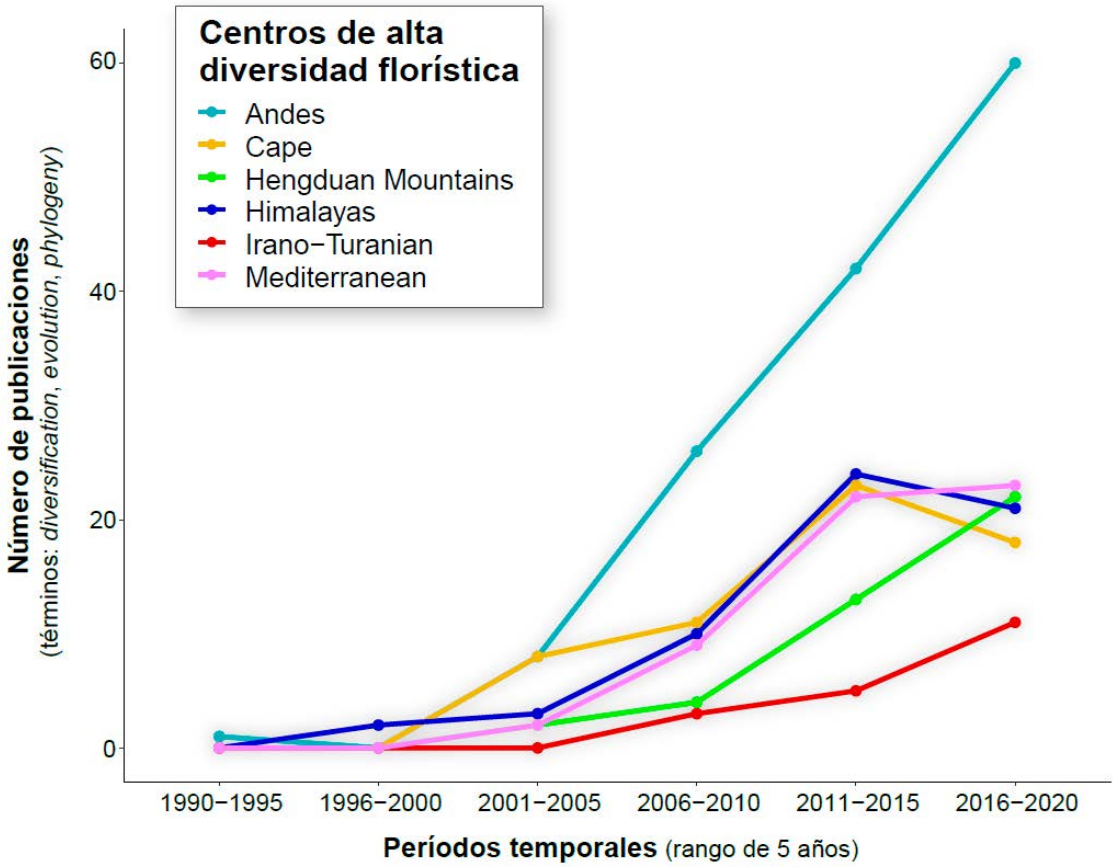


Fig. 16. Estudios publicados según WOS (Web os Science) sobre radiaciones evolutivas en algunos de las áreas más diversas en cuanto a especies vegetales. Datos extraídos de una búsqueda avanzada en WOS en diciembre de 2020 bajo la siguiente fórmula en la cual la región y el período temporal varían: SU = Plant Science AND TS = (Diversification OR Evolution OR Phylogeny AND Mediterranean) AND PY = (1991−1995). SU es el campo de estudio, TS son los temas y PY es el período de publicación en años.

La región iranoturania

La región iranoturania es el principal foco de radiación de especies de *Jurinea*, albergando más del 70% de su riqueza total. Esta región florística es también una de las más biodiversas y más extensas de toda Eurasia (Takhtajan, 1986), extendiéndose a lo largo de 16.000.000 km^2 (30% de Eurasia; Manafzadeh et al., 2017). Está ubicada en Asia central-occidental, en la zona tradicionalmente conocida como Oriente Medio. Limita con la región circumboreal al norte, con la mediterránea oriental en el oeste, la península Arábiga en el sur y Asia oriental en el este. La región iranoturania se distingue principalmente de sus áreas adyacentes (Mediterránea, Saharoarábiga y Euro-Siberiana) por el clima, caracterizado por una marcada continentalidad, estacionalidad de sus precipitaciones, temperaturas frías durante el invierno y veranos muy cálidos (Djamali et al., 2012b). La delimitación geográfica de la región ha generado una larga y controvertida historia de distintas divisiones. La reciente revisión realizada por Manafzadeh et al. (2017) propone la división de la región en dos áreas principales: iranoturania occidental (W-IT) e iranoturania oriental (E-IT; ver Fig. 17). La diversidad de plantas no está repartida equitativamente entre las dos áreas, ya que se estiman unas 27.000 especies para la región W-IT y unas 5000 para la E-IT (Manafzadeh et al., 2017).

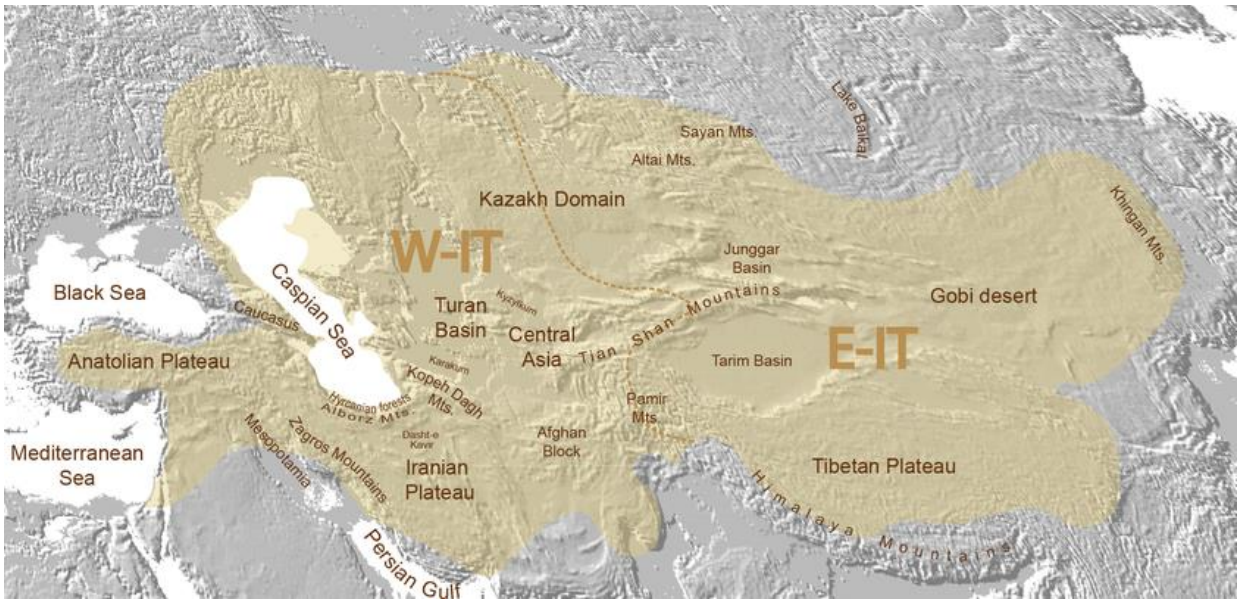


Fig. 17. Propuesta de los límites geográficos de la región florística iranoturania, basada en la delimitación de Takhtajan (1986) para las fronteras norte y este, y por Léonard (1988–1989) para las sur y oeste. Mapa extraído de Manafzadeh et al. (2017).

La característica principal de la región iranoturania es la gran representación de especies xerófitas adaptadas a ambientes áridos donde la precipitación y la disponibilidad de agua es escasa, por ejemplo elementos esteparios o plantas con metabolismo tipo C4 (Lauterbach et al., 2019). Son varias las teorías que se han postulado sobre cuándo se originaron estos elementos xerófitos característicos de la región. Las hipótesis más antiguas apuntaban a un origen en el Cretácico (145–66 Ma; Takhtajan, 1969) o durante la desecación del mar Tethys en el Neogeno (34 Ma; Zohary, 1973). Sin embargo, los estudios filogeográficos recientes sobre flora iranoturania están mostrando que el origen de estos elementos es mucho más reciente de los que se creía, a partir del Mioceno en delante (Manafzadeh et al., 2014; Wu et al., 2015; Lauterbach et al., 2019), y especialmente en el Plioceno-Pleistoceno (Moharrek et al., 2019; Mahmoudi Shamsabad et al., 2020). Este último marco temporal de origen tiene mayores coincidencias con eventos geológicos relevantes, como la segunda colisión de la placa Africana y Árabiga contra la Eurásiatica (13 Ma, Mioceno medio) que originó el levantamiento de la meseta iraní, o las posteriores deformaciones durante 15–5 Ma a lo largo de las montañas Zagros, Alborz, Kopeth Dagh y Cáucaso (Mouthereau, 2011). Además, también ocurrieron grandes cambios climáticos, como la aridificación de Asia central (Zhang et al., 2014) o el inicio de las oscilaciones térmicas del cuaternario (Zachos et al., 2001).

Otra de las características reseñables de la región iranoturania es su papel como “donante” de flora a las áreas adyacentes en períodos anteriores. En este sentido, ha sido especialmente documentada su contribución a la flora mediterránea actual (Jabbour & Renner, 2011; Manafzadeh et al., 2014; Peterson et al. 2019), la cual recibió a través de corredores terrestres, como el de la zona de Anatolia, linajes ancestrales con una pre-adaptación a climas áridos y semi-áridos que pudieron diversificarse con éxito en la cuenca Mediterránea.

Las Montañas Hengduan

Las Montañas Hengduan es el área que alberga una mayor concentración de especies de *Saussurea*, acogiendo cerca del 26% de su riqueza total (119 especies en Hengduan de las 400–500 totales; Zhang et al., 2009a). También son uno de los 34 *hotspots* mundiales de biodiversidad ("Mts. Of Southwest China"; Mittermeier et al., 2011), contando un total de 12.800 especies vegetales (Sun et al., 2017), de las cuales 3500 son endémicas (29,2%; López-Pujol et al., 2011) y es el principal *hotspot* de gimnospermas del mundo (20 son endémicas; Mutke & Barthlott, 2005). Su elevada diversidad se explica, en gran parte, por su compleja orografía, su amplia diversidad climática y de hábitats y su reciente formación geológica (es actualmente una zona con riesgo sísmico; Hubbard & Shaw, 2009). Además, son consideradas como uno de los principales refugios glaciares de China (Wen et al., 2016), lo cual pudo frenar la extinción masiva o migración de especies. Se estima que los glaciares del Cuaternario durante el último máximo glacial (LGM) no llegaron a cubrir grandes extensiones, especialmente en la región sur de las Hengduan (29ºN; Zhang et al., 2009b). Desafortunadamente, es una región botánicamente poco muestreada, lo que hace pensar que su exploración durante los próximos años puede incrementar sus valores de riqueza de especies (Yang et al., 2014).

Las Hengduan se encuentran en la parte más oriental del QTP (Fig. 18), con una extensión de 500.000 km^2 y una elevación media de 4.000 m. Están formadas por siete principales cadenas montañosas con una orientación Norte-Sur. El pico más elevado es el Gongga Shan (7556 m). Forman parte de la zona geológicamente más joven del QTP, pues se estima que su origen y deformaciones posteriores empezaron hace unos 15 Ma y siguieron durante el período Mioceno-Plioceno-Pleistoceno (Favre et al., 2015). Su particular topografía hace que existan unos gradientes altitudinales enormes, donde en tan solo 20 Km de distancia se pasa de un fondo de valle a 1500 m de altitud a un pico de 7000 m (ver ejemplo en Fig. 19). Desde un punto de vista biológico, este hecho proporciona una heterogeneidad de hábitats elevada en un espacio reducido. Por lo tanto, las Montañas Hengduan constituyen un escenario ideal para estudiar los procesos de diversificación en las regiones montañosas.

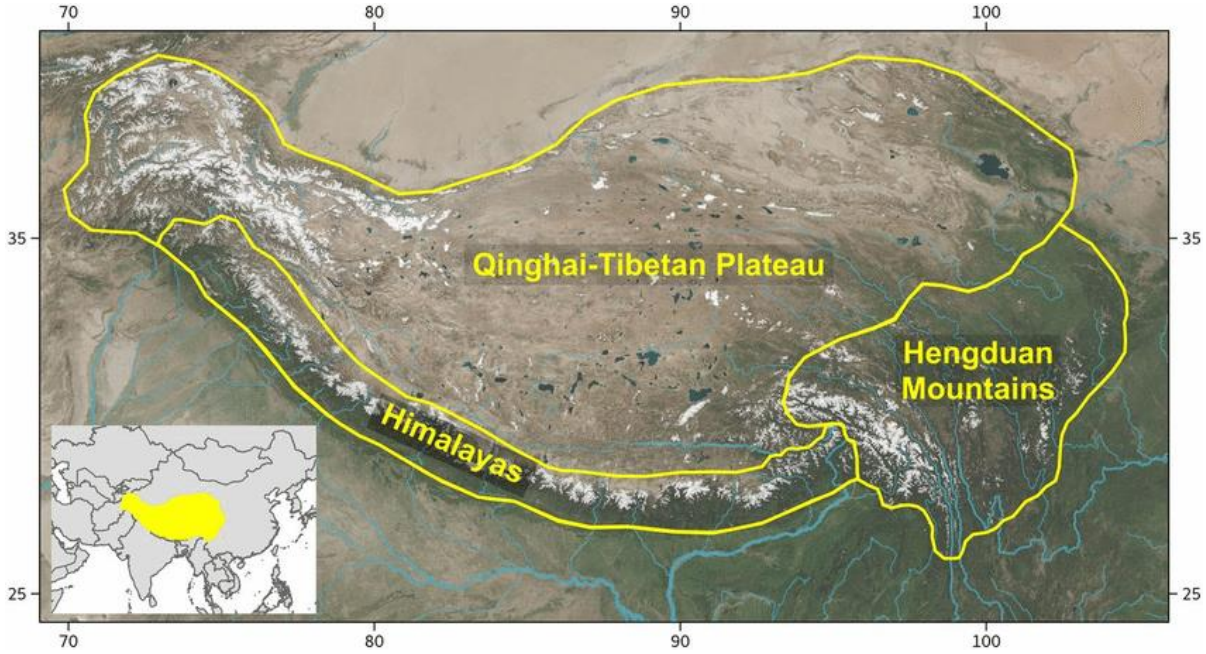


Fig. 18. Delimitación geográfica de las Montañas Hengduan respecto al Himalayas y la meseta del Qinghai-Tibet (QTP). Mapa extraído de Xing & Ree (2017).

Fig. 19. Ejemplo del gradiente altitudinal que existe en las montañas Hengduan (Jade Dragon Snow Mountain desde el sur). Imagen extraída de Neeltje (2017).

Referencias bibliográficas

Ackerfield, J.R., Keil, D.J., Hodgson, W.C., Simmons, M.P., Fehlberg, S.D., Funk, V.A. 2020a. Thistle be a mess: Untangling the taxonomy of *Cirsium* (Cardueae: Compositae) in North America. *Journal of Systematics and Evolution* 58: 881–912.

Ackerfield, J. et al. 2020b. A prickly puzzle: Generic delimitations in the *Carduus-Cirsium* group (Compositae: Cardueae: Carduinae). *Taxon* 69: 715–738.

Ahmed, M.J., Murtaza, G., Shaheen, H., Habib, T. 2020. Distribution pattern and associated flora of *Jurinea dolomiaea* in the western Himalayan highlands of Kashmir: An indicator endemic plant of alpine phytodiversity. *Ecological Indicators* 116: 106461.

Aksoy, N., Ataşlar, E., Cevher Keskin, B. 2018. Morphological and molecular data reveal a new species of *Jurinea* (Asteraceae) from Western Black Sea Region, Turkey. *Phytologia Balcanica* 24: 351–359.

Altinordu, F., Crespo, M.B. 2016. Notes on typification of three names in *Jurinea* (Asteraceae). *Phytotaxa* 284: 138–142.

Andermann, T. et al. 2020. A guide to carrying out a phylogenomic target sequence capture project. *Frontiers in Genetics* 10: 1407.

Bacon, C.D., Simmons, M.P., Archer, R.H., Zhao, L.-C., Andriantiana, J. 2016. Biogeography of the Malagasy Celastraceae: Multiple independent origins followed by widespread dispersal of genera from Madagascar. *Molecular Phylogenetics and Evolution* 94: 365–382.

Balos, M.M., Akan, H. 2015. *Jurinea mesopotamica* Hand.-Mazz. Asteraceae'nın Türkiye'deki varlığı ve taksonomisine katkılar. *Bağbahçe Bilim Dergisi* 2: 1–8.

Barreda, V., Palazzesi, L., Tellería, M.C., Katinas, L., Crisci, J. V. 2010. Fossil pollen indicates an explosive radiation of basal Asteracean lineages and allied families during Oligocene and Miocene times in the Southern Hemisphere. *Review of Palaeobotany and Palynology* 160: 102–111.

Barres, L., Sanmartín, I., Anderson, C. L., Susanna, A., Buerki, S., Galbany-Casals, M., Vilatersana, R. 2013. Reconstructing the evolution and biogeographic history of tribe Cardueae (Compositae). *American Journal of Botany* 100: 867–882.

Blair, A.C., Hufbauer, R.A. 2010. Hybridization and invasion: one of North America's most devastating invasive plants shows evidence for a history of interspecific hybridization. *Evolutionary Applications* 3: 40–51.

Bouchenak-Khelladi, Y., Onstein, R.E., Xing, Y., Schwery, O., Linder, H.P. 2015. On the complexity of triggering evolutionary radiations. *New Phytologist* 207: 313–326.

Boufford, D.E., Van Dyck, P.P. 1999. South-Central China. Pp. 338–351. En Mittermeier, R.A., Myers, N., Mittermeier, C.G. (eds.), *Hotspots: Earth's biologically richest and most endangered terrestrial ecoregions*. CEMEX, Mexico City.

Bremer, K. 1994. *Asteraceae: Cladistics & Classification*. Portland, Timber Press.

Brewer, G.E. et al. 2019. Factors affecting targeted sequencing of 353 nuclear genes from herbarium specimens spanning the diversity of angiosperms. *Frontiers in Plant Science* 10: 1102.

Brummitt, N., Araújo, A.C., Harris, T. 2021. Areas of plant diversity—What do we know? *Plants, People, Planet* 3: 33–44.

Buddenhagen, C. et al. 2016. Anchored phylogenomics of angiosperms I: assessing the robustness of phylogenetic estimates. *BioRxiv* 086298. https://doi.org/10.1101/086298

Butola, J.S., Samant, S.S. 2010. *Saussurea* species in Indian Himalayan Region: diversity, distribution and indigenous uses. *International Journal of Plant Biology* 1: e9.

Chapman, M.A., Chang, J., Weisman, D., Kesseli, R.V., Burke, J.M. 2007. Universal markers for comparative mapping and phylogenetic analysis in the Asteraceae (Compositae). *Theoretical and Applied Genetics* 115: 747–755.

Chase, M.W., Cameron, K.M., Freudenstein, J.V., Pridgeon, A.M., André, S. 2015. An updated classification of Orchidaceae. *Botanical Journal of the Linnean Society* 177: 151–174.

Chen, J., Wang, Y.J. 2018. New *Saussurea* (Asteraceae) species from Bogeda Mountain, eastern Tianshan, China, and inference of its evolutionary history and medical usage. *Plos ONE* 13: e0199416.

Chen, J., Zhao, Y.B., Wang, Y.J., Li, X.G. 2019. Identification of species and materia medica within *Saussurea* subg. *Amphilaena* based on DNA barcodes. *Peer J* 7: e6357.v

Chen, Y.S. 2014. Six new species of *Saussurea* (Asteraceae) from eastern Himalaya. *Phytotaxa* 177: 191–206.

Chen, Y.S. 2020. *Saussurea yilingii* (Asteraceae, Cardueae), a new species from Sichuan, China. *Phytotaxa* 452: 236–240.

Chen, Y.S., Xu, L.S. 2020. *Saussurea xinjiangensis* sp. nov. (Asteraceae, Cardueae), a new species from Xinjiang, China. *Nordic Journal of Botany*, 38: e02711.

Chen, Y.S., Yuan, Q. 2014. *Himalaiella lushaiensis* (Asteraceae), a new species from India. *Phytotaxa* 173: 293–298.

Chen, Y.S., Yuan, Q. 2015. Twenty-six new species of *Saussurea* (Asteraceae, Cardueae) from the Qinghai-Tibetan Plateau and adjacent regions. *Phytotaxa* 213: 159–211.

Chen, Y.S. 2015. *Flora of Pan-Himalaya*, vol. 48, Asteraceae II: *Saussurea*. Cambridge, Cambridge University Press.

Cheon, K.S., Kim, H.J., Han, J.S., Kim, K.A., Yoo, K.O. 2017. The complete chloroplast genome sequence of *Saussurea chabyoungsanica* (Asteraceae), an endemic to Korea. *Conservation Genetics Resources* 9: 51–53.

Cherneva, O.V. Tsukervanik, T.I. 1993. *Jurinea* Cass. Pp. 371–399. En: Vvedenskij, A.I. (ed.), *Opredelitel' rastenij Srednej Azii. Kriticheskij konspekt flory [Определитедь растений Средней Азии. Критический конспект флоры; Conspectus florae Asiae Mediae]*, vol. 10. Tashkent: Izdatel'stvo FAN Respubliki Uzbekistan. [en Ruso]

Cherneva, O.V. 2008. *Jurinea* Cass. Pp. 242–253. En: Takhtajan, A.L. (ed.), *Konspekt flory Kavkaza [Конспект флоры Кавказа; Caucasian flora conspectus]*. Saint-Petersburg & Moscow: KMK Scientific Press. [en Ruso]

Christenhusz, M.J., Byng, J.W. 2016. The number of known plants species in the world and its annual increase. *Phytotaxa* 261: 201–217.

Clarke, J.T., Warnock, R.C., Donoghue, P.C. 2011. Establishing a time-scale for plant evolution. *New Phytologist* 192: 266–301.

Conti, F. 1998. Contributions to the knowledge of *Jurinea mollis* sl (Compositae). 1. *Jurinea mollis* subsp. *moschata*, a synonym of *J. mollis* subsp. *mollis*. *Willdenowia* 28: 47–52.

Crisci, J.V., Apodaca, M.J., Katinas, L. 2019. El fin de la Botánica. *Revista del Museo de La Plata* 4: 41–50.

Crisci, J.V., Katinas, L., Apodaca, M.J., Hoch, P.C. 2020. The End of Botany. *Trends in Plant Science* 25: 1173–1176.

Cronn, R., Knaus, B.J., Liston, A., Maughan, P.J., Parks, M., Syring, J.V., Udall, J. 2012. Targeted enrichment strategies for next-generation plant biology. *American Journal of Botany* 99: 291–311.

Czekanski-Moir, J.E., Rundell, R.J. 2019. The ecology of nonecological speciation and nonadaptive radiations. *Trends in Ecology and Evolution* 34: 400–415.

Djamali, M. et al. 2012a. Ecological implications of *Cousinia* Cass. (Asteraceae) persistence through the last two glacial–interglacial cycles in the continental Middle East for the Irano-Turanian flora. *Review of Palaeobotany and Palynology* 172: 10–20.

Djamali, M., Brewer, S., Breckle, S.W., Jackson, S.T. 2012b. Climatic determinism in phytogeographic regionalization: a test from the Irano-Turanian region, SW and Central Asia. *Flora-Morphology, Distribution, Functional Ecology of Plants* 207: 237–249.

Dodsworth, S. et al. 2019. Hyb-Seq for flowering plant systematics. *Trends in Plant Science* 24: 887–891.

Dogan, B., Duran, A. 2009. Distribution and habitat features of the endemic *Jurinea* species for Turkey. *Plant, fungal and, habitat investigation and conservation. Proceedings of IV BBC Sofia* 2006: 220–223.

Dogan, B., Duran, A., Hakki, E. 2009. Numerical analyses of wild *Jurinea* spp. (Asteraceae) in Turkey. *Bangladesh Journal of Botany* 38: 47–53.

Dogan, B., Duran, A., Hakki, E.E. 2007. Phylogenetic analysis of *Jurinea* (Asteraceae) species from Turkey based on ISSR amplification. *Annales Botanici Fennici* 44: 353–358.

Dogan, B., Hakki, E.E., Duran, A. 2010a. A phylogenetic analysis of *Jurinea* (Compositae) species from Turkey based on ITS sequence data. *African Journal of Biotechnology* 9: 1741–1745.

Dogan, B., Duran, A., Hakki, E.E. 2010b. *Jurinea tortu*mensis sp. nov. (Asteraceae) from northeast Anatolia, Turkey. *Nordic Journal of Botany* 28: 479–483.

Dogan, B., Duran, A., Martin, E., Hakki, E. 2010c. *Jurinea turcica* (Asteraceae), a new species from North-west Anatolia, Turkey. *Biologia* 65: 28–32.

Dogan, B., Duran, A., Öztürk, M. 2011. *Jurinea cataonica* subsp. *mardinensis* subsp. nov. (Asteraceae) from southeast Anatolia, Turkey. *Nordic Journal of Botany* 29: 366–369.

Dogan, B., Kandemir, A., Osma, E., Duran, A. 2014. *Jurinea kemahensis* (Asteraceae), a new species from East Anatolia, Turkey. *Annales Botanici Fennici* 51: 75–79.

Donoghue, M. J., Sanderson, M.J. 2015. Confluence, synnovation, and depauperons in plant diversification. *New Phytologist* 207: 260–274.

Drummond, C.S. 2008. Diversification of *Lupinus* (Leguminosae) in the western New World: derived evolution of perennial life history and colonization of montane habitats. *Molecular Phylogenetics and Evolution* 48: 408–421.

Drummond, C.S., Eastwood, R.J., Miotto, S.T., Hughes, C.E. 2012. Multiple continental radiations and correlates of diversification in *Lupinus* (Leguminosae): testing for key innovation with incomplete taxon sampling. *Systematic Biology* 61: 443–460.

Ebersbach, J., Schnitzler, J., Favre, A., Muellner-Riehl, A.N. 2017. Evolutionary radiations in the species-rich mountain genus *Saxifraga* L. *BMC Evolutionary Biology* 17: 1–13.

Ellis, B.W. 1999. *Taylor's guide to annuals – How to select and grow more than 400 annuals, biennials, and tender perennials*. Haughton Mifflin Company, New York.

Faju, C., Yinggen, Y., Dexiu, Z., Yaolin, G., Zhongchen, G. 1999. Advances in studies of species, habitats distribution and chemical composition of snow lotuses (*Saussurea*) in China. *Chinese Bulletin of Botany* 16: 561–566.

Favre, A., Päckert, M., Pauls, S.U., Jähnig, S.C., Uhl, D., Michalak, I., Muellner-Riehl, A.N. 2015. The role of the uplift of the Qinghai-Tibetan Plateau for the evolution of Tibetan biotas. *Biological Reviews* 90: 236–253.

Fernández-Mazuecos, M. et al. 2019. Macroevolutionary dynamics of nectar spurs, a key evolutionary innovation. *New Phytologist* 222: 1123–1138.

Fernández-Mazuecos, M. et al. 2020. The radiation of Darwin's giant daisies in the Galápagos Islands. *Current Biology* 30: 4989–4998.

Flora of China Vol. 20-21: *Saussurea*. En eFloras.org, Missouri Botanical Garden. [consultado en Enero 2021]. http://www.efloras.org/florataxon.aspx?flora_id=2&taxon_id=129338

Folk, R.A. et al. 2019. Rates of niche and phenotype evolution lag behind diversification in a temperate radiation. *Proceedings of the National Academy of Sciences* 116: 10874–10882.

Font, M., Garcia-Jacas, N., Vilatersana, R., Roquet, C., Susanna, A. 2009. Evolution and biogeography of *Centaurea* section *Acrocentron* inferred from nuclear and plastid DNA sequence analyses. *Annals of Botany* 103: 985–997.

Font, M., Garnatje, T., Garcia-Jacas, N., Susanna, A. 2002. Delineation and phylogeny of *Centaurea* sect. *Acrocentron* based on DNA sequences: a restoration of the genus *Crocodylium* and indirect evidence of introgression. *Plant Systematics and Evolution* 234: 15–26.

Fujikawa, K. 2010. Taxonomic study of the genus *Saussurea* section *Eriocoryne* (Asteraceae) in the Himalayas. *Makinoa* 8: 25–80.

Fujikawa, K., Ohba, H. 2002. Two new species of *Saussurea* subgenus *Eriocoryne* (Asteraceae) from the Nepal Himalaya. *Edinburgh Journal of Botany* 59: 283–289.

Fujikawa, K., Kita, Y., Ohba, H. 2007. Taxonomic status of *Saussurea yakla* (Asteraceae) from the Himalayas. *Journal of Japanese Botany* 82: 130–136.

Funk, V.A. et al. 2005. Everywhere but Antarctica: Using a supertree to understand the diversity and distribution of the Compositae. *Biologiske Skrifter* 55: 343–374.

Funk, V.A., Susanna, A., Stuessy, T.F., and Robinson, H. 2009. Classification of Compositae. Pp. 171-189. En: Funk, V.A., Susanna, A., Stuessy, T.F. and R.J. Bayer (eds.), *Systematics, evolution, and biogeography of Compositae*. IAPT: Vienna.

Gailite, A., Rungis, D. 2012. An initial investigation of the taxonomic status of *Saussurea esthonica* Baer ex Rupr. utilising DNA markers and sequencing. *Plant Systematics and Evolution* 298: 913–919.

Garcia-Jacas, N., Garnatje, T., Susanna, A., Vilatersana, R. 2002. Tribal and subtribal delimitation and phylogeny of the Cardueae (Asteraceae): a combined nuclear and chloroplast DNA analysis. *Molecular Phylogenetics and Evolution* 22: 51–64.

Garcia-Jacas, N., Susanna, A., Garnatje, T., Vilatersana, R. 2001. Generic delimitation and phylogeny of the subtribe Centaureinae (Asteraceae): a combined nuclear and chloroplast DNA analysis. *Annals of Botany* 87: 503–515.

Garcia-Jacas, N., Susanna, A., Mozaffarian, V., Ilarslan, R. 2000. The natural delimitation of *Centaurea* (Asteraceae: Cardueae): ITS sequence analysis of the *Centaurea jacea* group. *Plant Systematics and Evolution* 223: 185–199.

Garcia-Jacas, N., Uysal, T., Romashchenko, K., Suárez-Santiago, V.N., Ertuğrul, K., Susanna, A. 2006. *Centaurea* revisited: a molecular survey of the *Jacea* group. *Annals of Botany* 98: 741–753.

Garnatje, T., Susanna, A., Garcia-Jacas, N., Vilatersana, R., Vallès, J. 2005. A first approach to the molecular phylogeny of the genus *Echinops* (Asteraceae): Sectional delimitation and relationships with the genus Acantholepis. *Folia Geobotanica* 40: 407–419.

GBIF Secretariat. 2021a. GBIF Backbone Taxonomy. *Jurinea* Cass. Checklist dataset. [consultado en Enero 2021]. https://doi.org/10.15468/39omei

GBIF Secretariat. 2021b. GBIF Backbone Taxonomy. *Saussurea* DC. Checklist dataset. [consultado en Enero 2021]. https://doi.org/10.15468/39omei

Ghimire, B., Jeong, M.J., Lee, K.M., Heo, K., Lee, C.H., Suh, G.U. 2016. Achene morphology of *Saussurea* species (Asteraceae, Cardueae) in Korea and its systematic implications. *Botanical Journal of the Linnean Society* 181: 692–710.

Givnish, T.J. 2015. Adaptive radiation versus 'radiation'and 'explosive diversification': why conceptual distinctions are fundamental to understanding evolution. *New Phytologist* 207: 297–303.

Häffner, E. 2000. On the phylogeny of the subtribe Carduinae (tribe Cardueae, Compositae). *Englera* 21: 3–208.

Han, T.S., Zheng, Q.J., Onstein, R.E., Rojas-Andrés, B.M., Hauenschild, F., Muellner-Riehl, A.N., Xing, Y.W. 2020. Polyploidy promotes species diversification of *Allium* through ecological shifts. *New Phytologist* 225: 571–583.

Harpe, M., Hess, J., Loiseau, O., Salamin, N., Lexer, C., Paris, M. 2019. A dedicated target capture approach reveals variable genetic markers across micro-and macro-evolutionary time scales in palms. *Molecular Ecology Resources* 19: 221–234.

Harrison, S., Noss, R. 2017. Endemism hotspots are linked to stable climatic refugia. *Annals of Botany* 119: 207–214.

Harvey, M.G., Rabosky, D.L. 2018. Continuous traits and speciation rates: Alternatives to state-dependent diversification models. *Methods in Ecology and Evolution* 9: 984–993.

Heath, T.A., Hedtke, S.M., Hillis, D.M. 2008. Taxon sampling and the accuracy of phylogenetic analyses. *Journal of Systematics and Evolution* 46: 239–257.

Hilpold, A. et al. 2014. Phylogeny of the *Centaurea* group (*Centaurea*, Compositae)–geography is a better predictor than morphology. *Molecular Phylogenetics and Evolution* 77: 195–215.

Hopkins, G.W., Freckleton, R.P. 2002. Declines in the numbers of amateur and professional taxonomists: implications for conservation. *Animal Conservation* 5: 245–249.

Huang, C.H., Zhang, C., Liu, M., Hu, Y., Gao, T., Qi, J., Ma, H. 2016. Multiple polyploidization events across Asteraceae with two nested events in the early history revealed by nuclear phylogenomics. *Molecular Biology and Evolution* 33: 2820–2835.

Hubbard, J., Shaw, J.H. 2009. Uplift of the Longmen Shan and Tibetan plateau, and the 2008 Wenchuan (M= 7.9) earthquake. *Nature* 458: 194–197.
Hughes, C.E., Atchison, G.W. 2015. The ubiquity of alpine plant radiations: from the Andes to the Hengduan Mountains. *New Phytologist* 207: 275–282.
Hughes, C.E., Eastwood, R. 2006. Island radiation on a continental scale: exceptional rates of plant diversification after uplift of the Andes. *Proceedings of the National Academy of Sciences* 103: 10334–10339.
Hughes, C.E., Nyffeler, R., Linder, H.P. 2015. Evolutionary plant radiations: where, when, why and how? *New Phytologist* 207: 249–253.
Humphreys, A.M., Govaerts, R., Ficinski, S.Z., Lughadha, E.N., Vorontsova, M.S. 2019. Global dataset shows geography and life form predict modern plant extinction and rediscovery. *Nature Ecology and Evolution* 3: 1043–1047.
Jabbour, F., Renner, S.S. 2011. *Consolida* and *Aconitella* are an annual clade of *Delphinium* (Ranunculaceae) that diversified in the Mediterranean basin and the Irano-Turanian region. *Taxon* 60: 1029–1040.
Jetz, W., McPherson, J.M. Guralnick, R.P. 2012. Integrating biodiversity distribution knowledge: toward a global map of life. *Trends in Ecology and Evolution* 27: 151–159.
Johnson, M.G. et al. 2019. A universal probe set for targeted sequencing of 353 nuclear genes from any flowering plant designed using k-medoids clustering. *Systematic Biology* 68: 594–606.
Jones, K. E. et al. 2019. An empirical assessment of a single family-wide hybrid capture locus set at multiple evolutionary timescales in Asteraceae. *Applications in Plant Sciences* 7: e11295.
Joppa, L.N., Roberts, D.L., Pimm, S.L. 2011. The population ecology and social behaviour of taxonomists. *Trends in Ecology and Evolution* 26: 551–553.
Kadota, Y. 2017. Systematic studies of Asian *Saussurea* (Asteraceae) VIII. Three new species from Honshu, Japan. *The Journal of Japanese Botany* 92: 69–81.
Kasana, S., Dwivedi, M.D., Uniyal, P.L. 2018. Taxonomic status of *Saussurea costus* (Falc.) Lipsch. (Asteraceae: Cardueae): a critically endangered species from Himalaya, India. *Pleione* 12: 81–84.
Kasana, S., Dwivedi, M.D., Uniyal, P.L., Pandey, A.K. 2020. An updated circumscription of *Saussurea* (Cardueae, Asteraceae) and allied genera based on morphological and molecular data. *Phytotaxa* 450: 173–187.
Kasana, S., Uniyal, P.L., Pandey, A.K. 2019. Lectotypification of *Lipschitziella ceratocarpa* (Asteraceae: Cardueae): an endemic species from Indian Himalaya. *Kew Bulletin* 74: 1–2.
Kelch, D.G., Baldwin, B.G. 2003. Phylogeny and ecological radiation of New World thistles (*Cirsium*, Cardueae–Compositae) based on ITS and ETS rDNA sequence data. *Molecular Ecology* 12: 141–151.
Keppel, G. et al. 2012. Refugia: identifying and understanding safe havens for biodiversity under climate change. *Global Ecology and Biogeography* 21: 393–404.
Kita, Y., Fujikawa, K., Ito, M., Ohba, H., Kato, M. 2004. Molecular phylogenetic analyses and systematics of the genus *Saussurea* and related genera (Asteraceae, Cardueae). *Taxon* 53: 679–690.
Knope, M.L., Morden, C.W., Funk, V.A., Fukami, T. 2012. Area and the rapid radiation of Hawaiian Bidens (Asteraceae). *Journal of Biogeography* 39: 1206–1216.
Kou, J., Wang, Y. 2017. Research advances in allocation of resources and elevation of *Saussurea* in eastern margin of Tibetan Plateau. *Journal of Traditional Chinese Veterinary Medicine* 32: 1433–1438.
Lagomarsino, L.P., Condamine, F.L., Antonelli, A., Mulch, A., Davis, C.C. 2016. The abiotic and biotic drivers of rapid diversification in Andean bellflowers (Campanulaceae). *New Phytologist* 210: 1430–1442.
Larridon, I. et al. 2020. Tackling rapid radiations with targeted sequencing. *Frontiers in Plant Science* 10: 1655.
Lauterbach, M., Veranso-Libalah, M.C., Sukhorukov, A.P., Kadereit, G. 2019. Biogeography of the xerophytic genus *Anabasis* L. (Chenopodiaceae). *Ecology and Evolution* 9: 3539–3552.
Lemmon, A.R., Emme, S.A., Lemmon, E.M. 2012. Anchored hybrid enrichment for massively high-throughput phylogenomics. *Systematic Biology* 61: 727–744.
Lemmon, E.M., Lemmon, A.R. 2013. High-throughput genomic data in systematics and phylogenetics. *Annual Review of Ecology, Evolution, and Systematics* 44 : 99–121.
Léonard, J. 1988–1989. Contribution à l'étude de la flore et de la végétation des déserts d'Iran: Etude des aires de distribution les phytochories, les chorotypes, Fascicules 8–9. Jardin Botanique National de Belgique, Meise.
Léveillé-Bourret, É., Starr, J.R., Ford, B.A., Moriarty Lemmon, E., Lemmon, A.R. 2018. Resolving rapid radiations within angiosperm families using anchored phylogenomics. *Systematic Biology* 67: 94–112.
Lewis, G., Schrire, B., MacKinder, B., Lock, M. 2005. *Legumes of the world*. Royal Botanical Gardens, Kew, UK.
Lewis-Smith, R.I., Richardson, M., 2011. Fuegian plants in Antarctica: Natural or anthropogenically assisted immigrants? *Biological Invasions* 13: 1–5.
Lichter-Marck, I.H. et al. 2020. Phylogenomics of *Perityleae* (Compositae) provides new insights into morphological and chromosomal evolution of the rock daisies. *Journal of Systematics and Evolution* 58: 853–880.
Linder H.P. 2008. Plant species radiations: where, when, why? *Philosophical Transactions of the Royal Society B: Biological Sciences* 363: 3097–3105.
Linder, H.P., Hardy, C.R., Rutschmann, F. 2005. Taxon sampling effects in molecular clock dating: an example from the African Restionaceae. *Molecular Phylogenetics and Evolution* 35: 569–582.
Linder, H.P., Hardy, C.R. Rutschmann, F. 2005. Taxon sampling effects in molecular clock dating: an example from the African Restionaceae. *Molecular Phylogenetics and Evolution* 35: 569–582.
Lipschitz, S. 1976. *Saussurea* DC. Pp. 216–217. En: Tutin, T.G., Heywood, V.H., Burges, N.A., Moore, D.M., Valantine, D.H., Walters, S.M., Webb, D.A. (eds.), *Flora Europaea*, vol. 4. Cambridge University Press, Cambridge.
Lipschitz, S.J. 1979. Rod *Saussurea* DC. (Asteraceae) [Genus *Saussurea* DC. (Asteraceae)]. Leningrad, Nauka, Leningradskoje Otdelenie. [en Ruso]
López-Pujol, J., Zhang, F.M., Ge, S. 2006. Plant biodiversity in China: richly varied, endangered, and in need of conservation. *Biodiversity and Conservation* 15: 3983–4026.

López-Pujol, J., Zhang, F.-M., Sun, H.-Q., Ying, T.-S., Ge, S. 2011. Centres of plant endemism in China: places for survival or for speciation? *Journal of Biogeography* 38: 1267–1280.

López-Vinyallonga, S., Mehregan, I., Garcia-Jacas, N., Tscherneva, O., Susanna, A., Kadereit, J.W. 2009. Phylogeny and evolution of the *Arctium*-*Cousinia* complex (Compositae, Cardueae-Carduinae). *Taxon* 58: 153–171.

López-Vinyallonga, S., Romaschenko, K., Susanna, A., Garcia-Jacas, N. 2011. Systematics of the Arctioid group: Disentangling *Arctium* and *Cousinia* (Cardueae, Carduinae). *Taxon* 60: 539–554.

Mahmoudi Shamsabad, M., Moharrek, F., Assadi, M., Nieto Feliner, G. 2020. Biogeographic history and diversification patterns in the Irano-Turanian genus *Acanthophyllum* sl (Caryophyllaceae). *Plant Biosystems-An International Journal Dealing with all Aspects of Plant Biology* 155. https://doi.org/10.1080/11263504.2020.1756974

Manafzadeh, S., Salvo, G., Conti, E. 2014. A tale of migrations from east to west: The Irano-Turanian floristic region as a source of Mediterranean xerophytes. *Journal of Biogeography* 41: 366–379.

Manafzadeh, S., Staedler, Y.M., Conti, E. 2017) Visions of the past and dreams of the future in the Orient: the Irano-Turanian region from classical botany to evolutionary studies. *Biological Reviews* 92: 1365–1388.

Mandel, J.R. et al. 2017. The Compositae Tree of Life in the age of phylogenomics. *Journal of Systematics and Evolution* 55: 405–410.

Mandel, J.R., Dikow, R.B., Funk, V.A. 2015. Using phylogenomics to resolve mega-families: An example from Compositae. *Journal of Systematics and Evolution* 53: 391–402.

Mandel, J.R. et al. 2014. A target enrichment method for gathering phylogenetic information from hundreds of loci: An example from the Compositae. *Applications in Plant Science* 2: 1300085.

Mandel, J.R. et al. 2019. A fully resolved backbone phylogeny reveals numerous dispersals and explosive diversifications throughout the history of Asteraceae. *Proceedings of the National Academy of Sciences of the United States of America* 116: 14083–14088.

Martín-Bravo, S. et al. 2019. A tale of worldwide success: Behind the scenes of *Carex* (Cyperaceae) biogeography and diversification. *Journal of Systematics and Evolution* 57: 695–718.

Maurin, K.J. 2020. An empirical guide for producing a dated phylogeny with treePL in a maximum likelihood framework. arXiv preprint arXiv:2008.07054.

Mehregan, I. 2008. Systematics, phylogeny and biogeography of *Cousinia* (Asteraceae). Unpublished PhD thesis, Johannes Gutenberg Universität, Mainz.

Mehregan, I., Kadereit, J.W. 2009. The role of hybridization in the evolution of *Cousinia* s. str. (Asteraceae, Cardueae). *Willdenowia* 39: 35–47.

Minaeifar, A.A., Sheidai, M., Attar, F., Noormohammadi, Z., Ghasemzadeh-Baraki, S. 2016. Biosystematic study in the genus *Cousinia* Cass. (Asteraceae), section *Cousinia*. *Biochemical Systematics and Ecology* 69: 252–260.

Mirtadzadini, M., Joharchi, M.R. 2017. *Jurinea khorassanica* sp. nov. *(J.* sect. *Stechmannia*, Asteraceae) from Iran. *Nordic Journal of Botany* 35: 569–572.

Mirtadzadini, M., Reza Rahiminejad, M., Vitek, E. 2011. *Jurinea giviensis* sp. nov. (Compositae) from Iran. *Nordic Journal of Botany* 29: 159–162.

Mittermeier, R.A., Turner, W.R., Larsen, F.W., Brooks, T.M., Gascon, C. 2011. *Global biodiversity conservation: the critical role of hotspots*. En: Zachos, F., Habel, J.C. Biodiversity hotspots. Springer, Berlin, Heidelberg.

Moharrek, F., Sanmartín, I., Kazempour-Osaloo, S., Nieto Feliner, G. 2019. Morphological innovations and vast extensions of mountain habitats triggered rapid diversification within the species-rich Irano-Turanian genus *Acantholimon* (Plumbaginaceae). *Frontiers in Genetics* 9: 698.

Montazerolghaem, S., Susanna, A., Calleja, J.A., Mozaffarian, V., Rahiminejad, M.R. 2017. Molecular systematics and phylogeography of the genus *Echinops* (Compositae, Cardueae-Echinopsinae): focus on the Iranian centre of diversification. *Phytotaxa* 297: 101–138.

Moreau C.S., Bell C.D. 2013. Testing the museum versus cradle tropical biological diversity hypothesis: phylogeny, diversification, and ancestral biogeographic range evolution of the ants. *Evolution* 67: 2240–2257.

Mouthereau, F. 2011. Timing of uplift in the Zagros belt/ Iranian plateau and accommodation of late Cenozoic Arabia–Eurasia convergence. *Geological Magazine* 148: 726–738.

Mutke, J., Barthlott, W. 2005. Patterns of vascular plant diversity at continental to global scales. *Biologiske Skrifter* 55: 521–531.

Myers, N., Mittermeier, R.A., Mittermeier, C.G., Da Fonseca, G.A., Kent, J. 2000. Biodiversity hotspots for conservation priorities. *Nature* 403: 853–858.

Narits, A., Leht, M., Paal, J. 2000. Taxonomic status of *Saussurea alpina* subsp. *esthonica* (Asteraceae): phenetical analysis. *Annales Botanici Fennici* 37: 197–206.

Neeltje. 2017. Hutao Xia Yulong Xue Shan. Wikimedia Commons. [consultado en Enero 2021]. https://commons.wikimedia.org/wiki/File:HutaoXia_YulongXueShan.jpg

Ng, J., Smith, S.D. 2014. How traits shape trees: new approaches for detecting character state-dependent lineage diversification. *Journal of Evolutionary Biology* 27: 2035–2045.

Norris, J. et al. 2020. *Synthesis of the local knowledge and ecology of vansemberuu (Saussurea dorogostaiskii Palib.) for conservation, Darkhad Valley*. Khuvsgul, Mongolia.

Nurk, N.M., Michling, F., Linder, H.P. 2018. Are the radiations of temperate lineages in tropical alpine ecosystems pre-adapted? *Global Ecology and Biogeography* 27: 334–345.

Nürk, N.M. et al. 2019. Diversification in evolutionary arenas—Assessment and synthesis. *Ecology and Evolution* 10: 6163–6182.

Oh, Y.J., Paik, W.K., Lee, W.C. 2002. Habitat characteristics of *Saussurea chabyoungsanica*. *The Korean Journal of Ecology* 25: 145–152.

Ortiz, S. et al. 2009. The basal grade of the Compositae: Mutisieae (sensu Cabrera) and Carduoideae. Pp. 193–21. En: Funk, V.A., Susanna, A., Stuessy, T.F., Bayer, R.J. (Eds.), *Systematics, Evolution, and Biogeography of Compositae*. International Association for Plant Taxonomy. Vienna, Austria.

Panero, J.L., Crozier, B.S. 2012. Asteraceae. Sunflowers, daisies. Version 27 January 2012. http://tolweb.org/Asteraceae/20780/2012.01.27 en The Tree of Life Web Project. [consultado en Enero 2021] http://tolweb.org/

Panero, J.L., Crozier, B.S. 2016. Macroevolutionary dynamics in the early diversification of Asteraceae. *Molecular Phylogenetics and Evolution* 99: 116–132.

Panero, J.L. Funk, V.A. 2002. Toward a phylogenetic subfamilial classification for the Compositae (Asteraceae). *Proceedings of the Biological Society of Washington* 115: 909–922.
Panero, J.L., Funk, V.A. 2008. The value of sampling anomalous taxa in phylogenetic studies: Major clades of the Asteraceae revealed. *Molecular Phylogenetics and Evolution* 47: 757–782.
Parks, M., Cronn, R., Liston, A. 2009. Increasing phylogenetic resolution at low taxonomic levels using massively parallel sequencing of chloroplast genomes. *BMC Biology* 7: 1–17.
Peakall, R. 2007. Speciation in the Orchidaceae: confronting the challenges. *Molecular Ecology* 16: 2834–2837.
Pellicer, J. et al. 2018. Genome size diversity and its impact on the evolution of land plants. *Genes* 9: 88.
Peng, D., Sun, L., Pritchard, H.W., Yang, J., Sun, H., Li, Z. 2019. Species distribution modelling and seed germination of four threatened snow lotus (*Saussurea*), and their implication for conservation. *Global Ecology and Conservation* 17: e00565.
Peterson, A., Harpke, D., Peterson, J., Harpke, A., Peruzzi, L. 2019. A pre-Miocene Irano-Turanian cradle: Origin and diversification of the species-rich monocot genus *Gagea* (Liliaceae). *Ecology and Evolution* 9: 5870–5890.
Pirie, M.D., et al. 2019. Leaps and bounds: geographical and ecological distance constrained the colonisation of the Afrotemperate by *Erica*. *BMC Evolutionary Biology* 19: 1–12.
Pyak, E.A., Pyak, A.I., Madyka, V.V. 2020. *Saussurea odorata* (Asteraceae), a new species from western Mongolia. *Phytotaxa* 470: 235–242.
Quézel, P. 1978. Analysis of the fl ora of Mediterranean and Saharan Africa. *Annals of the Missouri Botanical Garden* 65: 479–534.
Raab-Straube, E. von 2017. Taxonomic revision of *Saussurea* subgenus *Amphilaena* (Compositae, Cardueae). *Englera* 34. Botanic Garden and Botanical Museum Berlin, Freie Universität Berlin, Berlin.
Raab-Straube, von E. (2003). Phylogenetic relationships in *Saussurea* (Compositae, Cardueae) sensu lato, inferred from morphological, ITS and trnL-trnF sequence data, with a synopsis of *Himalaiella* gen. nov., *Lipschitziella* and *Frolovia*. *Willdenowia* 33: 379–402.
Raab-Straube, von E. 2009. *Saussurea luae* (Compositae, Cardueae), a new species of Snow Lotus from China. *Willdenowia* 39: 101–106.
Raab-Straube, von E. 2011. The genus *Saussurea* (Compositae, Cardueae) in China: taxonomic and nomenclatural notes. *Willdenowia* 41: 83–95.
Rabosky, D.L., Goldberg, E.E. 2017. FiSSE: A simple nonparametric test for the effects of a binary character on lineage diversification rates. *Evolution* 71: 1432–1442.
Rana, H.K., Sun, H., Paudel, A., Ghimire, S.K. 2018. *Saussurea ramchaudharyi* (Asteraceae), a new species from Nepal. *Phytotaxa* 340: 271–276.
Rechinger, K.H., Wagenitz, G. 1979. *Jurinea*. Pp. 180–209. En: Rechinger, K.H. (ed.), *Flora Iranica*, vol. 139a. Graz, Akademische Druck- u. Verlagsanstalt.
Ri, X., Zhao, L.Q., Qing, H. 2020. *Saussurea yiwuensis* (Asteraceae), a new species from Xinjiang, China. *Annales Botanici Fennici* 57: 159–161.
Rundell, R.J., Price, T.D. 2009. Adaptive radiation, nonadaptive radiation, ecological speciation and non ecological speciation. *Trends in Ecology and Evolution* 24: 394–399.
Sánchez-Jiménez, I., Hidalgo, O., Canela, M. Á., Siljak-Yakovlev, S., Šolić, M. E., Vallès, J., Garnatje, T. 2012. Genome size and chromosome number in *Echinops* (Asteraceae, Cardueae) in the Aegean and Balkan regions: technical aspects of nuclear DNA amount assessment and genome evolution in a phylogenetic frame. *Plant Systematics and Evolution* 298: 1085–1099.
Sánchez-Jiménez, I., Lazkov, G.A., Hidalgo, O., Garnatje, T. 2010. Molecular systematics of *Echinops* L. (Asteraceae, Cynareae): A phylogeny based on ITS and trnL-trnF sequences with emphasis on sectional delimitation. *Taxon* 59: 698–708.
Särkinen, T. et al. 2012. How to open the treasure chest? Optimising DNA extraction from herbarium specimens. *PLoS ONE*: e43808.
Schluter, D. 2000. *The Ecology of Adaptive Radiation*. Oxford University Press, UK.
Seehausen, O. 2015. Process and pattern in cichlid radiations–inferences for understanding unusually high rates of evolutionary diversification. *New Phytologist* 207: 304–312.
Semple, J.C., Watanabe, K. 2009. A review of chromosome numbers in Asteraceae with hypotheses on chromosomal base number evolution. Pp. 61–72. En: *Systematics, Evolution, and Biogeography of Compositae*, V.A. Funk, A. Susanna, T.F. Stussey, R.J. Bayer (Eds.) International Association for Plant Taxonomy, Vienna.
Sennikov, A.N., Lazkov, G.A. 2013. Taxonomic corrections and new records in vascular plants of Kyrgyzstan. *Memoranda Societatis pro Fauna et Flora Fennica* 87: 41–64.
Shee, Z.Q., Frodin, D.G., Cámara-Leret, R., Pokorny, L. 2020. Reconstructing the complex evolutionary history of the Papuasian *Schefflera* radiation through herbariomics. *Frontiers in Plant Science* 11: 258.
Shi, Z., Raab-Straube, E. von. 2011. *Saussurea* group. Pp. 42–149. En: Wu, Z.Y., Raven, P.H., Hong, D.Y. (eds.), *Flora of China*, vol. 20–21 (Asteraceae). Science Press, Beijing; Missouri Botanical Garden Press, St Louis.
Shurupova, M.N., Zverev, A.A., Gureyeva, I.I. 2017. Ecological ranges and types of rarity in the Kuznetsk Alatau of some *Saussurea* DC. species. *Contemporary Problems of Ecology* 10: 28–37.
Shyamal, L. 2007. Phylogenetics. Wikimedia Commons. [consultado en Enero 2021]. https://commons.wikimedia.org/wiki/File:Phylogenetics.svg
Simões, M. et al. 2016. The evolving theory of evolutionary radiations. *Trends in Ecology and Evolution* 31: 27–34.
Simpson, G.G. 1953. *The Major Features of Evolution*. Columbia University Press, New York, UK.
Siniscalchi, C.M., Loeuille, B., Funk, V.A., Mandel, J.R., Pirani, J.R. 2019. Phylogenomics yields new insight into relationships within *Vernonieae* (Asteraceae). *Frontiers in Plant Science* 10: 1224.
Slotta, T.A.B., Horvath, D.P., Foley, M.E. 2012. Phylogeny of *Cirsium* spp. in North America: host specificity does not follow phylogeny. *Plants* 1: 61–73.
Smirnov, S.V., Kechaykin, A.A., Sinitsyna, T.A., Shmakov, A.I. 2018. Synopsis of the genus *Saussurea* DC. (Asteraceae) of Eurasia. *Ukrainian Journal of Ecology* 8: 207–285.
Smith, S.A., O'Meara, B.C. 2012. treePL: divergence time estimation using penalized likelihood for large phylogenies. *Bioinformatics* 28: 2689–2690.
Soltis, P.S., Marchant, D.B., Van de Peer, Y., Soltis, D.E. 2015. Polyploidy and genome evolution in plants. *Current Opinion in Genetics and Development* 35: 119–125.

Stevanović, V., Matevski, V., Tan, K. 2010. *Jurinea micevskii* (Asteraceae), a new species from the Republic of Macedonia. *Phytologia Balcanica* 16: 249–254.
Straub, S.C.K., Parks, M., Weitemier, K., Fishbein, M., Cronn, R.C., Liston, A. 2012. Navigating the tip of the genomic iceberg: Next-generation sequencing for plant systematics. *American Journal of Botany* 99: 349–364.
Stubbs, R.L., Folk, R.A., Xiang, C.L., Soltis, D.E., Cellinese, N. 2018. Pseudo-parallel patterns of disjunctions in an Arctic-alpine plant lineage. *Molecular Phylogenetics and Evolution* 123: 88–100.
Stuessy, T.F, Garver, D. 1996. The defensive role of pappus in heads of Compositae. En: Caligari, P.D.S., Hind D.J.N. (eds.), *Compositae: Biology & Utilization*. Proceedings of the International Compositae Conference, Royal Botanic Gardens, Kew, UK.
Stuessy, T.F., Crawford, D.J., Soltis, D.E., Soltis, P.S. 2014. *Plant systematics: the origin, interpretation, and ordering of plant biodiversity*. Koeltz Scientific Books, Königstein.
Sun, H., Zhang, J., Deng, T., Boufford, D.E. 2017. Origins and evolution of plant diversity in the Hengduan Mountains, China. *Plant Diversity* 39: 161.
Susanna, A., Garcia-Jacas, N. 2007. Tribe *Cardueae* Cass. (1819). Pp. 123–146. En: Kadereit, J.W., Jeffrey, C. (eds.), *The families and genera of vascular plants*, vol. 8. Springer, Berlin & Heidelberg.
Susanna, A., Garcia-Jacas, N. 2009. *Cardueae* (Carduoideae). Pp. 293–313. En: Funk, V.A., Susanna, A., Stuessy, T.F. Bayer, R.J. (eds.), *Systematics, evolution, and biogeography of Compositae*. International Association for Plant Taxonomy, Vienna.
Susanna, A. et al. 2020. The classification of the Compositae: A tribute to Vicki Ann Funk (1947–2019). *Taxon* 69: 807–814.
Susanna, A., Garcia-Jacas, N., Hidalgo, O., Vilatersana, R., Garnatje, T. 2006. The *Cardueae* (Compositae) revisited: Insights from ITS, trnL-trnF, and matK nuclear and chloroplast DNA analysis. *Annals of Missouri Botanical Garden* 93: 150–171.
Szukala, A. et al. 2019. Phylogeny of the Eurasian genus *Jurinea* (Asteraceae: Cardueae): Support for a monophyletic genus concept and a first hypothesis on overall species relationships. *Taxon* 68: 112–131.
Takhtajan, A. 1969. *Flowering plants: origin and dispersal*. Smithsonian Institution Press, Washington, DC.
Takhtajan, A. 1986. *Floristic regions of the world*. University of California Press, Berkeley, California.
Thapa, R., Bayer, R.J., Mandel, J.R. 2020. Phylogenomics resolves the relationships within *Antennaria* (Asteraceae, Gnaphalieae) and yields new insights into its morphological character evolution and biogeography. *Systematic Botany* 45: 387–402.
Villaverde, T. et al. 2018. Bridging the micro-and macroevolutionary levels in phylogenomics: Hyb-Seq solves relationships from populations to species and above. *New Phytologist* 220: 636–650.
Villaverde, T. et al. 2020. A new classification of *Carex* (Cyperaceae) subgenera supported by a HybSeq backbone phylogenetic tree. *Botanical Journal of the Linnean Society* 194: 141–163.
Wang, J., He, R., Zhang, H., Hu, Y., Wang, J., Wang, L., Li, Y. 2021. Complete chloroplast genome of *Saussurea inversa* (Asteraceae) and phylogenetic analysis. *Mitochondrial DNA Part B* 6: 8–9.
Wang, R., Liu, J., Liu, S., Guan, S., Jiao, P. 2020a. Characterization of the complete chloroplast genome of *Saussurea involucrata* (Compositae), an endangered species endemic to China. *Mitochondrial DNA Part B* 5 : 511–512.
Wang, J. et al. 2020b. The complete chloroplast genome of *Saussurea medusa* (Asteraceae), a rare subnival plant in Qinghai-Tibetan Plateau. *Mitochondrial DNA Part B* 5: 3581–3582.
Wang, Y.F., Li, Q.J., Du, G.Z., Lian, Y.S. 2014. *Saussurea pseudograminea* sp. nov. (Asteraceae) from the Qinghai–Tibetan plateau, China. *Nordic Journal of Botany* 32: 185–189.
Wang, Y.J., Liu, J.Q. 2004. Phylogenetic analyses of *Saussurea* sect. *Pseudoeriocoryne* (Asteraceae: Cardueae) based on chloroplast DNA trnL–F sequences. *Biochemical Systematics and Ecology* 32: 1009–1023.
Wang, Y.J., Liu, J.Q., Miehe, G. 2007. Phylogenetic origins of the Himalayan endemic *Dolomiaea*, *Diplazoptilon* and *Xanthopappus* (Asteraceae: Cardueae) based on three DNA regions. *Annals of Botany* 99: 311–322.
Wang, Y.J., Pan, J.T., Liu, S.W., Liu, J.Q. 2005. A new species of *Saussurea* (Asteraceae) from Tibet and its systematic position based on ITS sequence analysis. *Botanical Journal of the Linnean Society* 147: 349–356.
Wang, Y.J., Susanna, A., Von Raab-Straube, E., Milne, R., Liu, J.Q. 2009. Island-like radiation of *Saussurea* (Asteraceae: Cardueae) triggered by uplifts of the Qinghai–Tibetan Plateau. *Biological Journal of the Linnean Society* 97: 893–903.
Watson, L.E., Siniscalchi, C.M., Mandel, J.R. 2020. Phylogenomics of the hyperdiverse daisy tribes: Anthemideae, Astereae, Calenduleae, Gnaphalieae, and Senecioneae. *Journal of Systematics and Evolution* 58: 841–852.
Weitemier, K., Straub, S.C., Cronn, R.C., Fishbein, M., Schmickl, R., McDonnell, A., Liston, A. 2014. Hyb-Seq: Combining target enrichment and genome skimming for plant phylogenomics. *Applications in Plant Sciences* 2: 1400042.
Wen, J., Zhang, J., Nie, Z.L., Zhong, Y., Sun, H. 2014. Evolutionary diversifications of plants on the Qinghai-Tibetan Plateau. *Frontiers in Genetics* 5: 4.
Wen, Z., Yang, Q., Quan, Q., Xia, L., Ge, D., Lv, X. 2016. Multiscale partitioning of small mammal β-diversity provides novel insights into the Quaternary faunal history of Qinghai–Tibetan Plateau and Hengduan Mountains. *Journal of Biogeography* 43: 1412–1424.
Whitfield, J.B., Lockhart, P.J. 2007. Deciphering ancient rapid radiations. *Trends in Ecology and Evolution* 22: 258–265.
Wu, S.D., Lin, L., Li, H.L., Yu, S.X., Zhang, L.J., Wang, W. 2015. Evolution of Asian interior arid-zone biota: evidence from the diversification of Asian *Zygophyllum* (Zygophyllaceae). *PloS ONE* 10: e0138697.
Xie, Q. et al. 2017. The complete chloroplast genome of Tianshan snow lotus (*Saussurea involucrata*), a famous traditional Chinese medicinal plant of the family Asteraceae. *Mitochondrial DNA Part A* 28: 294–295.
Xing, Y., Ree, R.H. 2017. Uplift-driven diversification in the Hengduan Mountains, a temperate biodiversity hotspot. *Proceedings of the National Academy of Sciences of the United States of America* 114: E3444–E3451.
Xu, B.Q., Hao, G., Xia, N.H. 2013. *Saussurea wenchengiae* (Asteraceae), a new species from Qinghai, China. *Annales Botanici Fennici* 50: 83–86.
Xu, L.S., Herrando-Moraira, S., Susanna, A., Galbany-Casals, M., Chen, Y.S. 2019. Phylogeny, origin and dispersal of *Saussurea* (Asteraceae) based on chloroplast genome data. *Molecular Phylogenetics and Evolution* 141: 106613.
Xu, L.S., Yi, S.Y., Chen, Y.S. 2020. *Saussurea sagittifolia* (Asteraceae, Cardueae), a new species from the Bashan Mountains region of China. *Phytotaxa* 472: 295–298.

Yang, W., Ma, K., Kreft, H. 2014. Environmental and socio-economic factors shaping the geography of floristic collections in China. *Global Ecology and Biogeography* 23: 1284–1292.

Yuan, Q., Bi, Y.C., Chen, Y.S. 2015. *Diplazoptilon* (Asteraceae) is merged with *Saussurea* based on evidence from morphology and molecular systematics. *Phytotaxa* 236: 53–61.

Yun, S.A., Gil, H.Y., Kim, S.C. 2017. The complete chloroplast genome sequence of *Saussurea polylepis* (Asteraceae), a vulnerable endemic species of Korea. *Mitochondrial DNA Part B* 2: 650–651.

Zachos, J., Pagani, M., Sloan, L., Thomas, E., Billups, K. 2001. Trends, rhythms, and aberrations in global climate 65 Ma to present. *Science* 292: 686–693.

Zhang, D.C., Boufford, D.E., Ree, R.H., Sun, H. 2009b. The 29 N latitudinal line: an important division in the Hengduan Mountains, a biodiversity hotspot in southwest China. *Nordic Journal of Botany* 27: 405–412.

Zhang, D.C., Zhang, Y.H., Boufford, D.E., Sun, H. 2009a. Elevational patterns of species richness and endemism for some important taxa in the Hengduan Mountains, southwestern China. *Biodiversity and Conservation* 18: 699–716.

Zhang, X. et al. 2019. Plastome phylogenomics of *Saussurea* (Asteraceae: Cardueae). *BMC Plant Biology* 19: 1–10.

Zhang, X. et al. (en revisión). Insights into the drivers of radiating diversification in biodiversity hotspots using *Saussurea* (Asteraceae) as a case. https://www.biorxiv.org/content/10.1101/2021.03.15.435394v1

Zhang, Y. et al. 2014. Cenozoic record of aeolian sediment accumulation and aridification from Lanzhou, China, driven by Tibetan Plateau uplift and global climate. *Global and Planetary Change* 120: 1–15.

Zhang, Y., Tang, R., Huang, X., Sun, W., Ma, X., Sun, H. 2019. *Saussurea balangshanensis* sp. nov. (Asteraceae), from the Hengduan Mountains region, SW China. *Nordic Journal of Botany* 37: e02078.

Zohary, M. 1973. *Geobotanical foundations of the Middle East*. Gustav Fischer, Stuttgart.

Objetivos


cm 1 2 3 4 5 6 7 8
0.5mm
mm

Los objetivos del presente proyecto de tesis son los siguientes:

- **Obtener una muestra lo más completa posible del complejo *Jurinea* y *Saussurea***, especialmente centrada, respectivamente, en los dos núcleos de radiación, la región iranoturania y las montañas Hengduan, incluyendo también a ser posible otros centros menores de especiación;

- **Generar filogenias bien resueltas de ambos géneros** y esclarecer su posición dentro de la **tribu** y **sus límites genéricos**, mediante análisis de concatenación y coalescencia a partir de datos de secuenciación masiva obtenidos por *Hyb-Seq*;

- Sobre la base de las nuevas filogenias y los **análisis** de **inferencia biogeográfica**, **datación, diversificación y nicho ecológico**, analizar las radiaciones de los dos géneros para responder a las siguientes preguntas:

  - ¿**Cuándo** se separaron los dos géneros de sus grupos hermanos y empezaron a diversificarse?

  - ¿**Qué grupos monofiléticos** existen y cuándo y dónde empezaron a diversificarse?

  - ¿**Cuál es el área geográfica ancestral** más plausible de *Saussurea* y *Jurinea* y cuáles habrían sido sus rutas de dispersión?

  - ¿Constituyen las especies de *Saussurea* y *Jurinea* clados con **tasas de diversificación** constantes a lo largo del tiempo o ha habido algún cambio significativo en esta tasa a lo largo de su evolución?

  - En caso de detectarse un cambio significativo en la tasa de diversificación, ¿puede este cambio coincidir en el tiempo y relacionarse con algún **evento geológico** o **climático**?

  - ¿**Qué factores** (intrínsecos y/o extrínsecos) habrían sido los impulsores de estas radiaciones?

  - ¿Cómo han podido influir las **oscilaciones paleoclimáticas** en la diversificación de ambos géneros? ¿Son las regiones climáticamente estables las más ricas en especies?

Estructura de la tesis doctoral

La presente tesis doctoral está estructurada en un total de seis CAPÍTULOS con temáticas independientes, pero relacionadas entre sí y ordenados cronológicamente. Los CAPÍTULOS están diseñados con el formato de artículo científico, tres de ellos ya publicados: en la revista *Molecular Phylogenetics and Evolution* (Herrando-Moraira et al., 2018, **CAPÍTULO 1**; Herrando-Moraira et al., 2019, **CAPÍTULO 2**) y *Taxon* (Herrando-Moraira et al., 2020, **CAPÍTULO 3**). Los **CAPÍTULOS 4** y **5** han sido enviados para su revisión a las revistas *Scientific Data* y *Systematic Biology*, respectivamente. El **CAPÍTULO 6** se encuentra en fase de preparación para la revista *National Science Review*.

- El **CAPÍTULO 1** es un estudio con un enfoque básicamente metodológico y exploratorio de la técnica, los análisis y su aplicabilidad/resultado en el grupo de estudio. Utilizando datos *Hyb-Seq*, con el panel de 1061 COS loci, trata de esclarecer las relaciones taxonómicas del complejo *Saussurea-Jurinea* con su grupo hermano *Arctium-Cousinia* (muestreo de 76 especies), unas relaciones que eran confusas en filogenias previas con secuencias Sanger. También son comparados varios métodos (extracción de secuencias, técnicas de filtrado e inferencia filogenética) para determinar si nuestras elecciones metodológicas afectan significativamente a los resultados filogenéticos obtenidos.

- El **CAPÍTULO 2** aborda la clasificación subtribal dentro de la tribu *Cardueae*. Con un muestreo representativo de todos los grupos de la tribu y sus grupos hermanos más cercanos (86 especies) aplicando el panel de los COS loci y extrayendo también las secuencias del ADN del cloroplasto, se propone un nuevo marco filogenético y taxonómico con 12 subtribus. Del total de subtribus reconocidas, seis son nuevas y una de ellas es la formada por el complejo *Saussurea-Jurinea*, ahora llamada subtribu *Saussureinae*. La nueva filogenia también se ha datado. La metodología y resultados obtenidos de la datación sirven como fundamentos en los análisis de tiempos de divergencia en capítulos posteriores (5 y 6).

- El **CAPÍTULO 3** trata de esclarecer los límites genéricos dentro de la subtribu *Saussureinae* basándonos en un nuevo marco filogenético realizado con el panel de los COS loci y, por primera vez, con representantes de los 17 géneros descritos (167 especies). La nueva delimitación queda reducida a tres géneros y sirve para definir en qué rama del árbol de la subtribu *Saussureinae* empiezan las diversificaciones de los dos géneros principales, *Saussurea* y *Jurinea*.

- El **CAPÍTULO 4** es un artículo en formato *Data paper* donde es presentado a escala global un índice de estabilidad climática mediante mapas de alta resolución (2,5 arc-min). Se exponen dos rangos temporales distintos: por un lado, el conjunto pasado-presente, que abarca desde el Plioceno (3,3 Ma) hasta la actualidad; y por otro lado, presente-futuro, que llega hasta el año 2100 bajo diferentes escenarios de cambio climático. El índice presentado puede ser útil para numerosos campos científicos como la filogeografía, las ciencias de la tierra, la sociología, etc.

- Los **CAPÍTULOS 5 y 6** tienen un enfoque muy similar, ambos son estudios sobre las radiaciones evolutivas de *Jurinea* (con 187 especies incluidas) y *Saussurea* (con 324). Usando COS loci, se reconstruye su historia filogenética en el espacio y el tiempo, y se formulan y verifican hipótesis sobre cómo, cuándo y qué factores han promovido la diversificación de las especies.

Chapter 1

Exploring data processing strategies in NGS target enrichment to disentangle radiations in the tribe *Cardueae* (Compositae)

Sonia Herrando-Moraira[a*] and The Cardueae Radiations Group
In alphabetical order:

Juan Antonio Calleja[b]
Pau Carnicero[b]
Kazumi Fujikawa[c]
Mercè Galbany-Casals[b]
Núria Garcia-Jacas[a]
Hyoung-Tak Im[d]
Seung-Chul Kim[e]
Jian-Quan Liu[f]
Javier López-Alvarado[b]
Jordi López-Pujol[a]
Jennifer R. Mandel[g]
Sergi Massó[a]
Iraj Mehregan[h]
Noemí Montes-Moreno[a]
Elizaveta Pyak[i]
Cristina Roquet[j]
Llorenç Sáez[b]
Alexander Sennikov[k]
Alfonso Susanna[a]
Roser Vilatersana[a]

a Botanic Institute of Barcelona (IBB, CSIC-ICUB), Pg. del Migdia, s. n., 08038 Barcelona, Spain
b Systematics and Evolution of Vascular Plants (UAB) – Associated Unit to CSIC. Departament de Biologia Animal, Biologia Vegetal i Ecologia, Facultat de Biociències, Universitat Autònoma de Barcelona, ES-08193 Bellaterra, Spain
c Kochi Prefectural Makino Botanical Garden, 4200-6, Godaisan, Kochi 781-8125, Japan
d Department of Biology, Chonnam National University, Gwangju 500 -757, Republic of Korea
e Department of Biological Sciences, Sungkyunkwan University, Gyeonggi-do 440-746, Republic of Korea
f Key Laboratory for Bio-Resources and Eco-Environment, College of Life Sciences, Sichuan University, Chengdu, China
g Department of Biological Sciences, University of Memphis, Memphis, Tennessee 38152, USA
h Department of Biology, Science and Research Branch, Islamic Azad University, Tehran, Iran
i Department of Botany, Inst. of Biology, Tomsk State University, RU-634050 Tomsk, Russia
j Univ. Grenoble Alpes, Univ. Savoie Mont Blanc, CNRS, LECA (Laboratoire d'Ecologie Alpine), FR-38000 Grenoble, France
k Botanical Museum, Finnish Museum of Natural History, PO Box 7, FI-00014 University of Helsinki, Finland; and Herbarium, Komarov Botanical Institute of Russian Academy of Sciences, Prof. Popov str. 2, 197376 St. Petersburg, Russia

*Corresponding author at: Botanic Institute of Barcelona (IBB, CSIC-ICUB), Pg. del Migdia, s. n., ES-08038 Barcelona, Spain. E-mail address: sonia.herrando@gmail.com (S. Herrando-Moraira).

Journal, volume, pages (year): Molecular Phylogenetics and Evolution, 128, 69–87 (2018).
DOI: https://doi.org/10.1016/j.ympev.2018.07.012

Impact factor year 2018: 3.992 (Clarivate Analytics, 2019)
Ranking ISI Journal Citation Reports ©:
- 2018: 87/298 Biochemistry & Molecular Biology
- 2018: 12/50 Evolutionary Biology
- 2018: 42/173 Genetics Heredity

Received: 14 April 2018; Received in revised form: 13 July 2018; Accepted: 14 July 2018

Abstract

Target enrichment is a cost-effective sequencing technique that holds promise for elucidating evolutionary relationships in fast-evolving lineages. However, potential biases and impact of bioinformatic sequence treatments in phylogenetic inference have not been thoroughly explored yet. Here, we investigate this issue with an ultimate goal to shed light into a highly diversified group of Compositae (Asteraceae) constituted by four main genera: *Arctium*, *Cousinia*, *Saussurea,* and *Jurinea*. Specifically, we compared sequence data extraction methods implemented in two easy-to-use workflows, PHYLUCE and HybPiper, and assessed the impact of two filtering practices intended to reduce phylogenetic noise. In addition, we compared two phylogenetic inference methods: 1) the concatenation approach, in which all loci were concatenated in a supermatrix; and 2) the coalescence approach, in which gene trees were produced independently and then used to construct a species tree under coalescence assumptions. Here we confirm the usefulness of the set of 1061 COS targets (a nuclear conserved orthology loci set developed for the Compositae) across a variety of taxonomic levels. Intergeneric relationships were completely resolved: there are two sister groups, *Arctium-Cousinia* and *Saussurea-Jurinea*, which are in agreement with a morphological hypothesis. Intrageneric relationships among species of *Arctium*, *Cousinia,* and *Saussurea* are also well defined. Conversely, conflicting species relationships remain for *Jurinea*. Methodological choices significantly affected phylogenies in terms of topology, branch length, and support. Across all analyses, the phylogeny obtained using HybPiper and the strictest scheme of removing fast-evolving sites was estimated as the optimal. Regarding methodological choices, we conclude that: 1) trees obtained under the coalescence approach are topologically more congruent between them than those inferred using the concatenation approach; 2) refining treatments only improved support values under the concatenation approach; and 3) branch support values are maximized when fast-evolving sites are removed in the concatenation approach, and when a higher number of loci is analyzed in the coalescence approach.

Keywords
Asteraceae
COS targets
HybPiper
NGS filtering strategies
Phylogenetic noise
PHYLUCE

Index

1. Introduction

1.1. Target enrichment strategies

The advent of the "target/hybrid enrichment" or "sequence capture" method has emerged in the last years as one of the most useful techniques in the field of phylogenomics and evolutionary studies (Cronn et al., 2012; Grover et al., 2012; Mamanova et al., 2009). This approach has provided significant advances, shedding light on previously unresolved evolutionary lineages analyzed using Sanger sequencing (Nicholls et al., 2015). This next generation sequencing (NGS) tool allows the recovery of hundreds to thousands of genetic markers from specific regions of the genome, even from degraded and ancient samples (Cronn et al., 2012). Remarkable advantages of this technique are: its reasonable sequencing cost, its power to resolve relationships at different taxonomic levels, and its reduced bioinformatic complexity compared to whole genome sequencing (Lemmon and Lemmon, 2013). The target DNA regions are enriched using probes or "baits". These can be specifically designed for the group of study via a known genome or transcriptome of a closely related species (e.g. Folk et al., 2015; García et al., 2017; Schmickl et al., 2016; Syring et al., 2016), or universally conserved loci (e.g., anchored hybrid enrichment, AHE) as for vertebrates (Lemmon et al., 2012) or angiosperms (Buddenhagen et al., 2016).

Concerning the Compositae or Asteraceae (both terms used to refer to the sunflower family; hereafter Compositae), Mandel et al. (2014) recently developed a target enrichment method, which uses the Hyb-Seq (sequence capture) approach (Weitemier et al., 2014), comprising a probe set of 9678 baits targeting a total of 1061 conserved orthology loci (hereafter COS) in this family. These COS loci were identified from thousands of expressed sequence tags (EST) across three available genomes of the family (see Mandel et al., 2014). This method has already proven useful at varied taxonomical scales, from deep Compositae nodes to shallower ones (Mandel et al., 2014, 2015, 2017; Siniscalchi et al., in prep.). In addition, the method allows the recovery of plastome data captured from off-target sequenced reads (Mandel et al., 2015). Nevertheless, the analytical power of this approach to resolve species relationships of recently and rapidly radiated genera in the family remains untested. In addition, the above cited previous works using the Compositae COS targets (Mandel et al., 2014, 2015, 2017) were performed following only one bioinformatics workflow for target sequences extraction, called PHYLUCE (Faircloth, 2015).

The last point seems crucial, since it has not been thoroughly investigated yet whether different bioinformatics extraction approaches yield congruent phylogenetic results, and whether these methodological choices could lead to bias in phylogenetic reconstruction. In recent years, a great number of easy-to-use workflows and automated pipelines are emerging to be used as target extraction procedures. The pipeline PHYLUCE (Faircloth, 2015) was initially designed for ultra-conserved elements (UCEs, Faircloth et al., 2012) and applied to a wide range of animal groups: birds (Hosner et al., 2015; Moyle et al., 2016), skinks (Bryson et al., 2017), ants (Ješovnik et al., 2017) and fishes (Burress et al., 2017; Longo et al., 2017). A bioinformatic approach for AHE was proposed in Prum et al. (2015) and used in several plant studies (Buddenhagen et al., 2016; Fragoso-Martínez et al., 2017; Mitchell et al., 2017; Wanke et al., 2017). Another method, HybPiper (Johnson et al., 2016) was designed specifically for Hyb-Seq data, implementing the ability to target exons and introns separately. The HybPiper workflow also offers the option to identify and separate paralogous copies. HybPiper has already been successfully applied to analyse data from captured target loci in plants (e.g. Crowl et al., 2017; Landis et al., 2017; Chau et al., 2018; Gernandt et al., 2018; Kates et al., 2018; Medina et al., 2018; Stubbs et al., 2018; Vatanparast et al., 2018). Other new and promising tools are aTRAM (Allen et al., 2015, 2017), HybPhyloMarker (Fér and Schmickly, 2018), and SECAPR (Andermann et al., 2018). From these published pipelines, we selected for this study two of the most commonly used approaches, PHYLUCE and HybPiper, to explore the technical differences between them and asses the consequences in inferred phylogenies of choosing one or another.

1.2. Parsing phylogenetic signal from noise in NGS studies

Despite the large amount of DNA sequence characters generated with NGS, the true gene genealogy can be obscured by various kinds of "phylogenetic noise" (Straub et al., 2014; Townsend et al., 2012). Potential sources of noise in nucleotide sequences include unusually fast-evolving sites, rich-indel regions, and ambiguous sequence calls, which may lead to substitution saturation, i.e. convergence in nucleotide states (homoplasy) that contradicts the real phylogenetic signal, and bias the ancestry character-state reconstructions (Rokas and Carroll, 2006). Additional noise may accumulate in all study phases due to sequencing errors, inaccurate assembly, or incorrect orthology assignment. Another possible source of error that should be taken into account with NGS data is the incorrect allele phasing in polyploid systems (Eriksson et al., 2018), in which phylogenetic trees can be often reconstructed from consensus sequences or chimeric consensus sequences rather than real allele

sequences (Kates et al., 2018). Consequences of ignoring possible phylogenetic noise are well documented (Kostka et al., 2008; Straub et al., 2014; Townsend et al., 2012), and may lead to long-branch attraction artifacts, topological differences among alternative reconstructions, or high support values for erroneous relationships (Dornburg et al., 2014; Jeffroy et al., 2006; Salichos and Rokas, 2013).

Part of this phylogenetic noise can be reduced with standard practices such as cleaning raw reads by filtering based on quality scores and alignment trimming (i.e. removal of ambiguously aligned and indel-rich positions). However, final trimmed alignments commonly used to perform phylogenetic inferences may still contain considerable levels of noise. Currently, standard procedures to deal with this issue are not well established, and we still lack a widely applicable refining metric to minimize the negative effects of phylogenetic noise and maximize the likelihood of an accurate phylogenetic reconstruction. Many recent studies based on target enrichment incorporate diverse filtering strategies at different components of data matrices, such as species, positions, or even entire sets of loci (see Table 1). Among all these practices, the most commonly used is the exclusion of loci recovered for a low number of species, which aims to reduce the effects of missing data and systematic bias on tree inference (see Hosner et al., 2015 for further details on potential impacts of missing data).

1.3. Resolving radiations and the case of the groups *Arctium-Cousinia* and *Jurinea-Saussurea* (tribe Cardueae)

Explosive diversification events (referred here as radiations) represent events in which many species or lineages evolved from a common ancestor in a short time period (Wen et al., 2013, 2014), caused by geographic isolation, dispersal barriers, sexual selection, or in some cases by ecological divergence or acquisition of novel key traits (Givnish, 2015). These events may leave few genomic traces, yielding few nucleotide differences among species derived from a common radiation, and thus hindering the reconstruction of phylogenetic relationships among them. As a consequence, unresolved phylogenies with short internal branches or large polytomies have been often recovered with traditional Sanger sequenced markers in recently diverged genera, hampering the in-depth study of radiations. With the emergence of NGS techniques, research focused on plant radiations are significantly increasing (*Heuchera* L., Folk et al., 2015; *Inga* Mill., Nicholls et al., 2015; Cariceae-Dulichieae-Scirpeae clade in Cyperaceae Juss., Léveillé-Bourret et al. 2016; order Zingiberales Griseb., Sass et al., 2016; *Salvia* L. subgenus *Calosphace* (Benth.) Epling, Fragoso-Martínez et al., 2017; *Protea* L., Mitchell et al., 2017; *Aristolochia* L., Wanke et al., 2017; "*Adenocalymma-Neojobertia*" clade from Bignoniaceae Juss., Fonseca and Lohmann, 2018; Iochrominae clade from Solanaceae Juss., Gates et al., 2018; *Pinus* subsection *Australes* Loudon Gernandt et al., 2018). Most of these studies obtained well resolved phylogenies, but they sampled only a small proportion of their study group. However, such first step of method testing is essential before performing studies with more complete species sampling, a type of research that will probably rise in coming years.

The tribe Cardueae (Compositae) is one of the most species-rich of the family with more than 2500 species, which accounts for one tenth of Compositae (Susanna and Garcia-Jacas, 2007, 2009). Three of the complexes described within the Cardueae rank among the largest radiations in the family: the *Arctium-Cousinia* group, with 600 species; the *Saussurea-Jurinea* group, involving ca. 550 species; and the *Carduus-Cirsium* group, with 350 species (Susanna and Garcia-Jacas, 2007). *Saussurea* DC. and *Jurinea* Cass. are especially interesting because they constitute two paradigmatic cases of mountain radiations. Previous molecular phylogenies of these genera resulted in large and undefined polytomies (*Saussurea*, Kita et al., 2004; Wang et al., 2009), as is usually the case with radiations. Another difficulty associated with the study of the radiations of *Saussurea* and *Jurinea* is the high number of satellite genera (up to 16) described within the complex (Susanna and Garcia-Jacas, 2009), considered at some point either *Saussurea* or *Jurinea*. A complete phylogenetic reconstruction of the whole group has never been performed and the taxonomic validity of the described genera remains unexplored with molecular data. In addition, species of both *Saussurea* and Jurinea always appeared entangled with the genera *Arctium* L. and *Cousinia* Cass. (Barres et al., 2013; Garcia-Jacas et al., 2002; Susanna et al., 2006). Thus, generic delimitation among these four genera is also unclear. Therefore, it is essential to obtain a well resolved phylogeny of these groups as a first step towards the improvement of the knowledge on the evolutionary processes that led to such diversified lineages.

Accordingly, we gathered for this study a representative sample of the four genera *Arctium*, *Cousinia*, *Saussurea*, and *Jurinea* together with several species of the tribe Cardueae and used the COS target enrichment approach with three main aims: 1) to evaluate the potential of COS loci for resolving relationships at inter- and intrageneric level of recently radiated genera in tribe Cardueae; 2) to elucidate the relationships among the genera Arctium, Cousinia, Saussurea, and Jurinea; 3) to test the differences between two extraction methods of target enriched data (PHYLUCE and HybPiper); and 4) to evaluate the effects of different filtering strategies on phylogenetic reconstruction and determine whether a widely applicable approach exists as a refining metric.

2. Materials and methods

2.1. Sampling strategy

In order to evaluate the usefulness of COS target enrichment methodology to resolve generic radiations in the Compositae, we included several representatives of the four genera of interest: 11 species of *Arctium*, 22 species of *Cousinia*, 19 species of *Saussurea*, 24 species of *Jurinea*, and four species described under different genera within the *Saussurea-Jurinea* complex depending on the taxonomical treatment (see section 4.1 for details). On the basis of previous phylogenetic studies of the tribe Cardueae (Barres et al., 2013; Garcia-Jacas et al., 2002; Susanna and Garcia-Jacas, 2007, 2009; Susanna et al., 2006), the following taxa were also added: *Alfredia acantholepis* Kar. & Kir., *Carduus pycnocephalus* L., *Cirsium sairamense* O. Fedtsch. & B. Fedtsch., *Olgaea petriprimi* B. A. Sharipova, and *Cynara*

Table 1. Strategies of filtering target enrichment sequencing data before phylogenetic analyses.

Criteria of filtering	Software that can be used	Examples of studies applying the filtering method
FILTERING SPECIES		
Exclusion of quickly evolving or unstable species	PHYLUCE[1] HybPhyloMarker[2]	Salichos and Rokas (2013) Ješovnik et al. (2017) Streicher et al. (2018) Gates et al. (2018)
FILTERING POSITIONS		
Exclusion of sites with high substitution rates	PhyDesign[3] OV[4] TIGER[5]	Dornburg et al. (2017) Fragoso-Martínez et al. (2017) Wanke et al. (2017) Streicher et al. (2018)
Exclusion of sites containing gaps	trimAL[6]	Salichos and Rokas (2013)
Inclusion of sites with high substitution rates	TIGER[5]	Streicher et al. (2018)
Inclusion of positions with high read coverage	Custom scripts and ape[7]	Grover et al. (2015)
FILTERING LOCI		
Exclusion of loci recovered in a low number of taxa	HybPhyloMarker[2] PHYLUCE[8]	Borowiec et al. (2015) Hosner et al. (2015) Streicher et al. (2016) Ješovnik et al. (2017) Longo et al. (2017) Mitchell et al. (2017) Streicher et al. (2018) Gernandt et al. (2018)
Exclusion of loci detected as potential paralogs	HybPiper[9]	Crowl et al. (2017) Chau et al. (2018) Gernandt et al. (2018) Vatanparast et al. (2018)
Exclusion of loci of short length	Geneious[10]	Gernandt et al. (2018)
Exclusion of highly variable loci	Geneious[10]	Gernandt et al. (2018)
Exclusion of loci with high number of missing data	Geneious[10]	Gernandt et al. (2018)
Exclusion of poorly aligned loci	Not specified	Salichos and Rokas (2013)
Exclusion of loci with low long-branch score from long-branched species	HybPhyloMarker[2] TreSpEx[11] R script[12]	Borowiec et al. (2015)
Inclusion of loci with strong phylogenetic signal (based on gene-trees with high mean bootstrap values)	HybPhyloMarker[2] TreSpEx[11] R script[12] Newick utilities[13]	Salichos and Rokas (2013) Bossert et al. (2017) Branstetter et al. (2017) Ješovnik et al. (2017) Ward and Branstetter (2017)
Inclusion of the most informative loci (high informative characters or parsimony informative sites)	HybPhyloMarker[2] PhyDesign[3] AMAS[14] Phyloch[15]	Hosner et al. (2015) Léveillé-Bourret et al. (2016) Meiklejohn et al. (2016) Longo et al. (2017)
Inclusion of slowly evolving loci (based on the smallest mean branch length)	HybPhyloMarker[2] TreSpEx[11] R script[12]	Salichos and Rokas (2013) Borowiec et al. (2015)
Inclusion of less saturated loci	R script[12]	Borowiec et al. (2015)
Inclusion of the most informative loci scored by some of the previous metrics	HybPhyloMarker[2] HybPiper[9] Geneious[10] TreSpEx[11] R script[12]	Borowiec et al. (2015) Gernandt et al. (2018)

[1]PHYLUCE (Faircloth, 2015) script "PHYLUCE_align_extract_taxa_from_alignments.py".
[2]HybPhyloMarker pipeline package (Fér and Schmickly, 2018).
[3]PhyDesign online application (López-Giráldez and Townsend, 2011; http://phydesign.townsend.yale.edu/).
[4]OV (observed variability) algorithm (Goremykin et al. 2010).
[5]TIGER software (Cummins and McInerey, 2011).
[6]trimAL program (Capella-Gutiérrez et al., 2009).
[7]R package "ape" (Paradis et al., 2004) and custom scripts (Grover et al., 2015) available at https://github.com/Wendellab/phylogenetics.
[8]PHYLUCE (Faircloth, 2015) script "get_only_loci_with_min_taxa.py".
[9]HybPiper pipeline (Johnson et al., 2016), script "paralog_investigator.py".
[10]Geneious software (Kearse et al., 2012).
[11]TreSpEx pipeline package (Struck, 2014).
[12]R script "gene_stats.R" (Borowiec et al. 2015).
[13]Newick utilities package (Junier and Zdobnov, 2010), function "nw_ed".
[14]AMAS software (Borowiec, 2016).
[15]R package "phyloch" (Heibl, 2008), function pis.

cardunculus L. For the last species, we directly incorporated the raw reads from Mandel et al. (2017) into our bioinformatics workflow. The information of location and voucher specimens of the 85 sampled species is summarized in Appendix, Table S1.

2.2. DNA extraction, library preparation, target enrichment, and sequencing

Dried leaf tissue was weighed to obtain a total amount of 200 mg per sample, which was later homogenized using Mixer Mill MM 301 (Retsch®, Haan, Germany). Genomic DNA was extracted using the DNeasy plant mini kit (Qiagen, Valencia, CA, USA) following manufacturer's specifications. The quantity of each extraction was checked with Qubit™ 3.0 Fluorometer (Thermo Scientific, Waltham, MA, USA). In order to obtain an average fragment size of 400–500 bp, approximately 1 µg in 70 µl per sample was sheared using a Q800R2 Sonicator® machine (QSonica, Newtown, CT, USA). The sonication step was conducted with the following parameters: 3 min (with 10 s pulse on, and 10 s pulse off), and the amplitude set at 20%. To ensure that genomic DNA was sheared at approximately the selected fragment size, all samples were checked and evaluated on a 1.2% (w/v) agarose gel. After shearing, we prepared the barcoded sequencing libraries using the NEBNext Ultra II DNA Library Prep kit for Illumina (New England Biolabs, Ipswich, MA, USA), following the standard protocol provided by the manufacturer. We added 25 µl of AMPure XP beads (Beckman Coulter, La Brea, CA, USA) for the first step of size selection, and 10 µl for the second step. The PCR amplification was performed using 15 cycles and each library was barcoded employing a unique index primer using NEBNext Multiplex Oligos for Illumina.j Library quantities were checked using the Qubit Fluorometer and then pooled in groups of four samples, aiming for a quantity of 500 ng per group. Pools were evaporated in a speed vacuum centrifuge, and then were resuspended in 7 µl of dH20. For sequence capture, we used MyBaits COS 1Kv1 (MYcroarray, Ann Arbor, MI, USA; http://www.mycroarray.com/mybaits/mybaits-UCEs.html). We followed the specifications in the manufacturer's protocol with slight modifications, such as the time and temperature to allow baits to hybridize to their targets (40 h at 65ºC). A post-capture PCR reaction of 16 cycles was performed using KAPA® HiFi HotStart ReadyMix (Kapa Biosystems, USA) and ~~"re~~amp" primers described in Meyer and Kircher (2010). To avoid adapter dimers problems, we added a supplementary cleanup magnetic bead-based step after the post-capture PCR reaction as specified in the NEBNext manual. Finally, target-enriched library pools were sent for sequencing to the DNA Sequencing Core CGRC/ICBR of the University of Florida or to Macrogen Co. (Seoul, South Korea) on one lane on a HiSeq 3000 sequencing platform (Illumina, USA) using 100 bp paired-end reads.

2.3. Raw data processing

A first quality control of raw sequence reads demultiplexed by sequencing cores was conducted in FastQC v.0.10.1 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Raw FASTQ data were then cleaned using ILUMIPROCESSOR (Faircloth, 2013), a wrapper program which incorporates Trimmomatic v.0.36 (Bolger et al., 2014) to remove Illumina adapters and to trim low-quality nucleotides of the reads. A trimming step was conducted with a sliding-window set to 5:20, cutting the start or the end of a read when the average of five terminal positions falls below 20 of the quality Phred+33 score. Cleaned reads finally retained were those with a minimum length of 36 bp and with both corresponding forward and reverse pair.

2.4. Extraction of target enrichment data: PHYLUCE and HybPiper pipelines

Two different orthology-detection methods were followed to extract and identify the sequence data that matched the 1061 target COS loci: the PHYLUCE pipeline package v.1.5 (Faircloth, 2015) and the HybPiper pipeline v.1.1 (Johnson et al., 2016). The main difference between both procedures is that the PHYLUCE pipeline begins with a de novo assembly of reads into contigs followed by a mapping step, aligning contigs back to the reference sequences. HybPiper first maps the reads against each target separately, and then assembles de novo the mapped reads into contigs, which are later mapped to targets (Fig. 1).

For the PHYLUCE method (Fig. 1), the trimmed reads were de novo assembled into contigs using the software SPAdes v.3.9.0 (Bankevich et al., 2012) testing several k-mer lengths: 21, 33, 55 and 77. Then, we mapped resultant contigs to the COS target sequences using LASTZ (Harris, 2007) with the python script ~~"a~~ssembly_match_contigs_to_probes.py". This program ensures that matches are 80% identical in 80% of the total length, and also removes potential paralogs. These potential paralogs are identified as assembled contigs that match multiple loci, or different contigs that match the same COS locus. After COS identification, the ~~"g~~et_match_counts.py" script was used to generate a relational list of contig names, generated by the assembler, with the names of each COS target across taxa indicated in a ~~"ta~~xon-set" file. This relational database was used for the script ~~"g~~et_fastas_from_match_counts.py" to generate a monolithic FASTA-formatted file containing all loci recovered for all taxa specified. Separate files for each locus were obtained running ~~"a~~ssembly_explode_get_fastas_file.py". For the final step of dataset creation, we used ~~"a~~lign_remove_locus_name_from_nexus_lines.py" to remove locus name from the FASTA header line to only retain the taxon name as is required for downstream analyses. The majority of raw extracted sequences were longer than the target length because reads were first assembled into contigs and then contigs were mapped to the reference targets. Therefore, extracted sequences could encompass part of non-coding regions outside the targets, which are derived from exonic regions.

For the HybPiper method, we used three sets of input files: the cleaned pair-end reads, a text-formatted list with the species names, and the target file that contains one or several orthologous sequences for each locus (see Mandel et al., 2014). We executed the entire pipeline with the script ~~"r~~eads_first.py" which, in a first phase, maps the reads to each target locus using the BWA mapper (Li and Durbin, 2009), selecting the best target sequence as a reference according to a mapping score. Secondly, reads mapped for each locus were de novo assembled into contigs with the best k-mer automatically detected by SPAdes assembler. In a third phase, ~~"e~~xonerate.py" was used to extract a unique longest contig

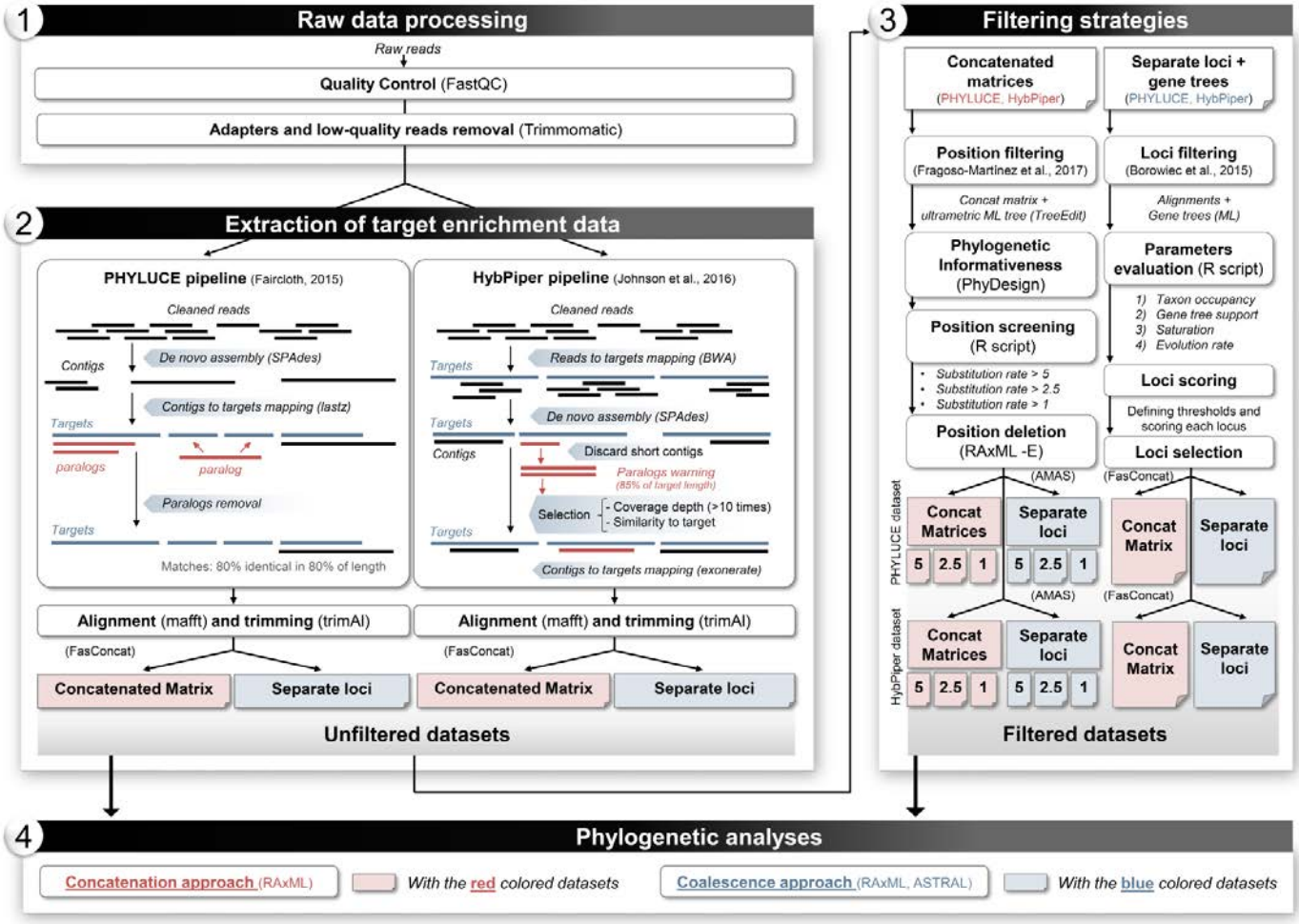


Fig. 1. Workflow representation of bioinformatic and phylogenetic analyses. The process followed consists of two alternative methods of target sequence extraction, PHYLUCE (Faircloth, 2015) and HybPiper (Johnson et al., 2016), and two approaches of sequence data refining applied to each: filtering by positions (Fragoso-Martínez et al., 2017) and filtering by loci (Borowiec et al., 2015). Squares in red and blue represent all the datasets analyzed (see Table 3 for details), in red showing analyses performed under the concatenation approach and in blue under the coalescence approach. The main programs used for the analyses are shown in brackets.

that aligned to the reference sequence. If multiple equally long contigs coexisted for the same locus (potential paralogs), the contig with greater coverage depth (10 times more) or the one with greater similarity to the target was retained (for details see Johnson et al., 2016). As a rule, we extracted exons because our target set comes from EST (expressed sequence tags), but some contigs may contain an extension of flanking non-coding regions. In these cases, contigs are usually longer than the target sequence. Finally, to retrieve sequences recovered for each species in a multi-fasta file for each gene, the ~~"r~~etrieve_sequences.py" script was executed.

In order to show the differences in recovery efficiency between the PHYLUCE and HybPiper pipelines, ~~"g~~et_seq_lengths.py" from HybPiper package was applied with slight modifications to the individual unaligned loci.

2.5. Alignment, alignment trimming, loci concatenation, and summary statistics

For both PHYLUCE and HybPiper methods, the multi-fasta files generated were aligned, for each locus separately, using the auto setting of MAFFT v.7.266 (Katoh and Standley, 2013). The resulting alignments were trimmed with trimAl v.14 (applying the automated1 flag) with the aim of removing positions ambiguously aligned (Capella-Gutiérrez et al., 2009). For subsequent phylogenetic inference based on supermatrix analysis (concatenation approach), gene alignments were concatenated with FASconCAT-G v.1.02 (Kück and Longo, 2014), which also provides the necessary information of gene partitions for subsequent steps. Finally, summary statistics of concatenated matrices were computed with AMAS (Borowiec, 2016).

2.6. Phylogenetic analyses without filtering step

The phylogenetic reconstruction analyses were conducted twice: first, under the concatenation approach using a supermatrix for tree estimation (hereafter concatenation approach), and second, under coalescence assumptions, in which species tree is estimated based on individual gene trees resulting from phylogenetic analyses of each locus separately (hereafter coalescence approach).

Concerning the concatenation approach, we ran Maximum Likelihood (ML) analyses with the software RAxML v.8.2.9 (Stamatakis, 2014) implemented on XSEDE in the CIPRES

Science Gateway v.3.1 (Miller et al., 2010). Specifically, we ran a simultaneous rapid bootstrapping and best ML tree search (Stamatakis et al., 2008), with 10 randomized maximum parsimony starting trees and a bootstrap resampling of 500 replicates to assess branch support values. We considered that only branches with bootstrap (BS) support values > 70% were statistically supported (Hillis and Bull, 1993). In the RAxML analysis, each locus was treated as a unit partition, and the GTRGAMMA evolution model was applied as recommended in Stamatakis (2006). Resulting trees were visualized in FigTree v.1.4.3 (Rambaut, 2016).

Regarding the coalescence approach, we first searched for individual gene trees with RAxML applying the same search options specified above but running 200-bootstrap replicates. Species tree inference under the coalescence approach was then performed using ASTRAL (Mirarab et al., 2014), which estimates the species tree that maximizes the number of quartets from a given input of unrooted gene trees under the assumption that all of them are correct. Branch support values were inferred through local posterior probabilities (LPP; Sayyari and Mirarab, 2016) calculated in ASTRAL-III v.5.5.3 (Zhang et al., 2018). The use of LPP as branch support metric has been proved to be more precise than multi-locus bootstrapping, especially when the error in estimating gene trees is low (Sayyari and Mirarab, 2016). Values of LPP > 0.95 were considered as strong branch support with very high precision, although lower values (LPP = 0.7–0.9) also give high precision (Sayyari and Mirarab, 2016).

2.7. Phylogenetic informativeness and position filtering

As a filtering step recently recommended by Fragoso-Martínez et al. (2017), we evaluated the effect of eliminating the "phantom" spike positions (ambiguous, indel-rich positions, or positions with high substitution rates) that can add phylogenetic noise and bias phylogenetic reconstructions. To identify these fast-evolving sites in our alignments, a first Phylogenetic Informativeness (PI) analysis and net PI profiles were performed in the web application PhyDesign (López-Giráldez and Townsend, 2011), specifically calculating the substitution rates per site with the implemented program HyPhy (Kosakovsky Pond et al., 2005). For the rate calculations, we used two inputs: 1) the partitioned concatenated matrices, from both the PHYLUCE and HybPiper methods; and 2) their respective ML trees, that were transformed to ultrametric with TreeEdit v.1.0a10 (Rambaut, 2002), applying the non-parametric rate smoothing algorithm (Sanderson, 1997) and scaled to a total height of 1.

To detect which positions exceeded the substitution rate (SR) values higher than three arbitrary pre-defined cut-off thresholds (5, 2.5, and 1), we imported the rate files per locus from PhyDesign to the R script "nmc3.R" developed by Fragoso-Martínez et al. (2017) and ran it in R v.3.1.2 (R Core Team, 2014) three times for the PHYLUCE and HybPiper datasets. The resulting spreadsheet of each analysis contained specific positions to remove from each locus (spreadsheet available in Appendix A). The final filtered matrices were generated in RAxML with the command –E, using the lists of positions to be removed, original matrices, and partition information. Next, ML and PI analyses were performed with the six filtered matrices. We used the AMAS software to separate each locus and re-ran RAxML per gene to later perform the coalescence analysis with ASTRAL-III.

2.8. Selection of the most informative loci

In order to reduce phylogenetic noise, another filtering strategy based on the selection of the most informative loci according to several parameters was implemented as suggested by Borowiec et al. (2015). We used the script "gene_stats.R" (available in Borowiec et al., 2015) for the loci selection procedure. As input, we used the individual ML gene trees and the trimmed alignments corresponding to both the PHYLUCE and the HybPiper unfiltered datasets. Then, the loci were scored for each of the following parameters: 1) the number of species recovered (accounting for taxon occupancy and missing data); 2) the average BS support value of the ML gene tree obtained (depicting information content); 3) the R^2 of mutational saturation regression curves (Philippe and Forterre, 1999), obtained from the inferred substitution values based on ML gene tree branch lengths against the number of observed differences in sequences for a given pair of species (representing saturation); and 4) average branch length of the ML gene tree calculated from the division of total tree length by total tree nodes (characterizing the rate of molecular evolution).

For each parameter, we scored each locus with 0 or 1 point, depending on whether it exceeded the arbitrary thresholds defined here (0 for parameter value below the threshold and 1 above the threshold). The thresholds selected were the following for each parameter. For taxon occupancy, the loci recovered for at least 50% of taxa (43 species) in the PHYLUCE dataset, and 95% of taxa (81 species) in the HybPiper dataset, were scored with 1 point. For the average BS value, the loci that yielded a tree with at least 60% mean BS in the PHYLUCE dataset, and 40% in the HybPiper dataset, were scored with 1 point. For saturation, the 25% of loci with the highest R2 of saturation curves were scored with 1 point. Finally, for evolution rates, the 25% of loci with lowest average branch length were scored with 1 point. Accordingly, a binary matrix with 0 and 1 points for each locus and each parameter was obtained. Finally, the selection of the best informative loci was performed in two steps: first, we calculated the points obtained for each locus, which ranged from 0 to 4, considering the four parameters together. And second, we selected those loci that obtained at least 2 points. Note that the four parameters were equally weighted, without any additional ponderation step, and that threshold values can be modified by the user depending on the parameter scores or the characteristics of the dataset analyzed.

The spreadsheets with parameters and scores are provided in Appendix A. After applying this locus filtering strategy, new datasets that only contained the selected loci were created accordingly, one comprising the supermatrix that was analyzed under the concatenation approach, and the other with each locus in a separate file, analyzed under the coalescence approach (both approaches described in g).

2.9. Topological comparisons

Differences in topology among all trees generated (unfiltered and filtered matrices, and in each case under the concatenation and coalescence approaches) were estimated with the Robinson-

Foulds distance (RF; Robinson and Foulds, 1981). First, we computed pairwise RF distances using PAUP v.4.0a (Swofford, 2003) and adjusted RF (RFadj) manually, which was estimated from RFajd = RF/(2n–6) being n the number of tree nodes (Mitchell et al., 2017; Steel and Penny, 1993), and ranging from 0 (same topology) to 1 (completely discordant topology). Secondly, the RF distances were exported as a tree distance matrix in R to compare all trees in the same tree space using the multidimensional scaling approach (Hillis et al., 2005) implemented in the R function "cmdscale" from the R package "stats".

3. Results

3.1. Target capture sequencing and efficiency

The average of raw pair-end reads was 4,263,196 per species. The outgroup Carduus pycnocephalus was the species sequenced here with the lowest number of reads (741,845), whereas Saussurea davurica had the highest number of reads (11,202,023).

From the 1061 targeted loci, we recovered a total of 675 loci (63.6%) with the PHYLUCE method and a total of 1055 loci (99.4%) with the HybPiper method (Table 2). Per species, the mean of on-target loci was 341 with the PHYLUCE method (the lowest number of loci recovered for a species = 208, the highest number of loci recovered for a species = 424), and 991 (510–1018) with the HybPiper method. In addition to this remarkable difference in the percentage of captured targets, our results show that the recovered loci per species are not equally distributed across the matrix in the PHYLUCE method (Fig. 2A–B). We pruned 48 loci recovered with PHYLUCE and four loci recovered with HybPiper because they were captured only in one or two species. Consequently, the final set of loci comprised 627 loci (59.1%) with the PHYLUCE method and 1051 loci (99.1%) with the HybPiper method. Only 9.2% of the loci selected with PHYLUCE were captured in 90% or more of the species sampled. In contrast, the taxa recovery was greater with HybPiper, with 89.6% of the selected loci captured in 90% or more of the sampled species. Despite these differences in missing data, the mean alignment length per locus was higher with the PHYLUCE method (823 bp; 139–3134 bp) than with the HybPiper method (317 bp; 63–1475 bp). Regarding the length of the captured loci in relation to the length of the respective reference target, we found that, in the case of PHYLUCE, 77.18% of the loci recovered were longer than the corresponding target (Fig. 2C); whereas with HybPiper, only 15.2% of the loci recovered were longer (Fig. 2D). The final aligned and trimmed concatenated matrices were composed by 515,875 bp with the PHYLUCE method and 333,614 bp with the HybPiper method, with 48% and 36% of variable sites, respectively (Table 2).

3.2. Phylogeny estimation

The capture probes designed for Compositae targeting 1061 COS loci have been useful to elucidate relationships among *Arctium*, *Cousinia*, *Saussurea*, and *Jurinea*, their generic delimitation, and also many of the relationships among closely related species. All the inferred phylogenies across the total 10 evaluated approximations (Fig. 1 and Table 3) support the monophyly of the four main genera (Supplementary Figs. S1–S4). The relationships as sister groups between *Arctium-Cousinia* and *Saussurea-Jurinea* were fully resolved with maximum support values also in all datasets analyzed (BS = 100 and LPP = 1). At lower taxonomical levels, shallow relationships were generally reconstructed with high support values for *Arctium*, *Cousinia*, and *Saussurea*, with only slight differences between analyses. In contrast, species relationships within Jurinea were not clearly outlined, presenting low-moderate support values (Supplementary Figs. S1–S4).

Across all the analyses under the concatenation approach, the best-resolved tree was that obtained with the HybPiper method

Table 2. Extraction performance of the 1061 COS targets (Mandel et al., 2014) and creation of the unfiltered datasets with the two methods compared, PHYLUCE (Faircloth, 2015) and HybPiper (Johnson et al., 2016). The evaluated parameters from 3 to 10 were calculated based on the dataset specified in the parameter 2. (i.e. from the total recovered loci, removing those loci captured only in one or two species). Abbreviations used: bp = base pairs; max = maximum; min = minimum; Nº = number; sd = standard deviation.

Evaluated parameters	PHYLUCE unfiltered dataset	HybPiper unfiltered dataset
1. Total Nº of recovered loci (%)	675 (63.3)	1055 (99.4)
2. Total Nº recovered loci removing those captured only in one or two species (%)	627 (59.1)	1051 (99.1)
3. Nº of captured loci in ≥ 90% of the species (%)	58 (9.2)	942 (89.6)
4. Average of recovered loci per species (sd; min–max)	340 (37; 208–410)	989 (68.2; 510–1061)
5. Average of species recovered per loci (sd; min–max)	46 (24.3; 4–84)	80 (12; 4–85)
6. Mean alignment length per locus in bp (sd; min–max)	823 (450; 139–3134)	317 (185; 63–1475)
7. Nº of loci longer than respective target length (%)	521 (77.2)	161 (15.2)
8. Length of the final concatenated matrix in bp	515,875	333,614
9. Proportion of missing data in the final concatenated matrix (%)	57.1	9.5
10. Proportion of variable sites in the final concatenated matrix (%)	48	36

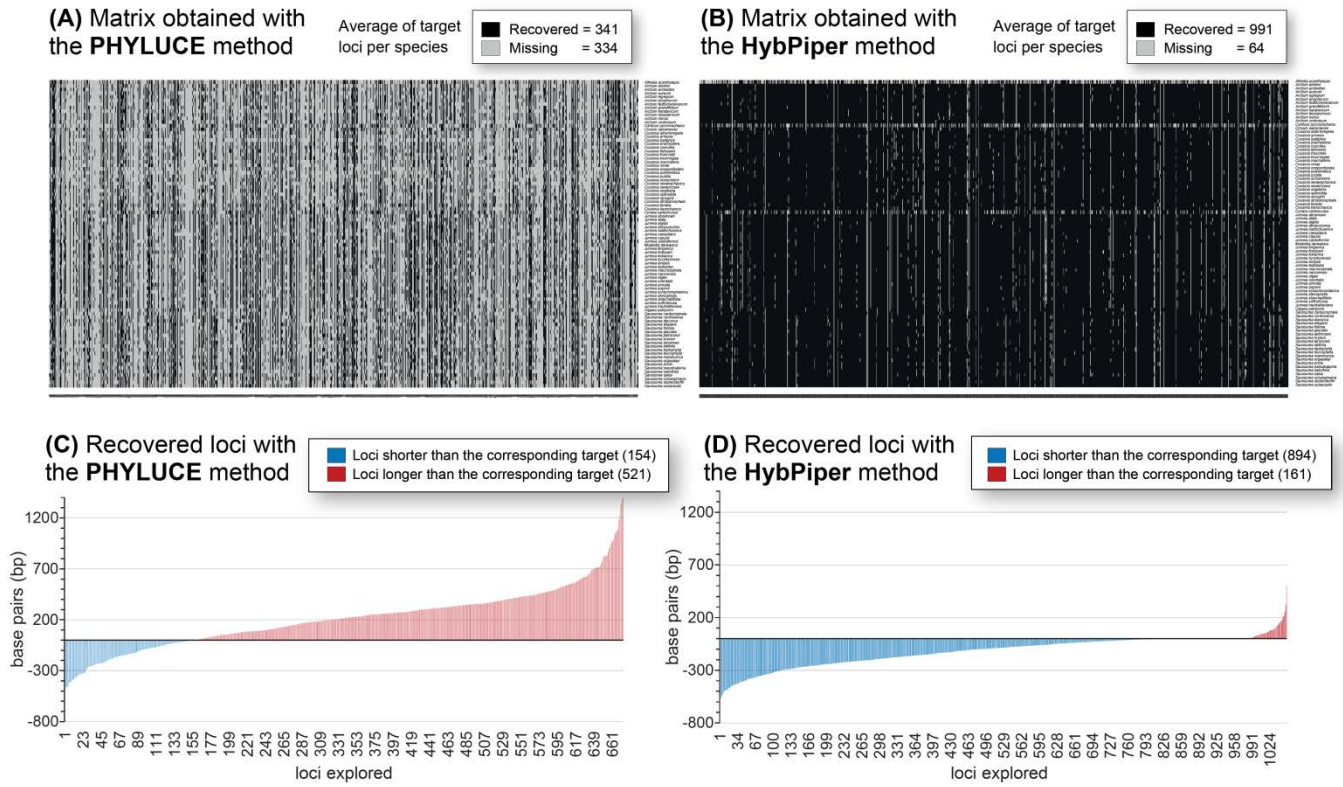


Fig. 2. Recovery efficiency of the 1061 COS loci using two target extraction methods: (A) PHYLUCE (Faircloth, 2015), and (B) HybPiper (Johnson et al., 2016). Columns represent each target locus and rows represent the 85 sampled species. The cells of the heat map in black represent loci on-target, and missing loci are showed in grey. Differences in length in base pairs (bp) between the reference target and the captured sequence (not aligned and trimmed) are represented for the PHYLUCE dataset (C) and for the HybPiper dataset (D). Blue bars represent loci shorter than the corresponding target and red bars represent loci that exceed the corresponding target in length. When the target length is equal to the captured locus length, the value of y-axis is zero.

and removing positions with SR > 1, with an average of 93.4% of BS support value and only five branches with BS < 70% (Fig. 3A and Table 3). The individual gene trees showed low average BS support values for both target extraction methods, although the values obtained with the PHYLUCE method (BS = 58.5) were considerably higher than the ones obtained with HybPiper (BS = 32.2), probably due to the longer loci and a few number of recovered species per locus (Table 2). A positive correlation was detected when the length of locus alignments and average BS support values of gene trees were compared for the PHYLUCE unfiltered dataset (Pearson's r = 0.54, p < 0.0001) and the HybPiper unfiltered dataset (Pearson's r = 0.60, p < 0.0001) methods. We also found another correlation, this time negative, between the number of taxa recovered and the average BS support value per locus obtained with the PHYLUCE unfiltered dataset (Pearson's r = -0.616, p < 0.0001) and the HybPiper unfiltered dataset (Pearson's r = -0.161, p < 0.0001). This indicated that gene trees obtained with those loci with a lower number of taxa recovered tended to be more supported. The lack of support in individual gene trees or the incongruence among them was reflected in short internal branches in coalescence units in the trees inferred under the coalescence approach (Fig. 3 and Supplementary Figs. S1–S4). Also, support values of coalescence trees were lower than those of the trees inferred with the concatenation approach (Table 3).

3.3. Phylogenetic informativeness and position filtering results

The amount of positions selected as ~~"f~~fast-evolving" sites (and thus removed from the alignments) varied considerably depending on the target extraction method. For all thresholds tested, a greater number of fast-evolving sites were removed from the PHYLUCE dataset, which were distributed in a greater number of loci than in the HybPiper dataset (see Table 3 for more details). For example, for a given filtering scenario (SR > 1) the positions trimmed were 9244 in the PHYLUCE dataset and 1885 in the HybPiper dataset. When these values were corrected for the matrix length, the number of positions filtered in the PHYLUCE dataset remained higher than in the HybPiper dataset, representing 1.8% of the PHYLUCE dataset and 0.6% of the HybPiper dataset.

The net phylogenetic informativeness (PI) mean value was markedly higher for the unfiltered PHYLUCE dataset (193.55) than for the unfiltered HybPiper dataset (26.20). The maximum PI value, which is related to the presence of ~~"p~~phantom spikes", was also higher for the unfiltered PHYLUCE dataset (7832.29) than for the HybPiper dataset (370.51). Overall, the highest PI values were coincident with the divergence of the four genera and their respective subsequent lineages, approximately at the time of 0.2–0.7 (PHYLUCE; Fig. 4A) and 0.3–0.8 (HybPiper; Fig. 4C). At earlier timing range (0–0.2), coincident with the main diversification of Jurinea and Saussurea, the PI profiles showed

Table 3. Characteristics of the datasets obtained with the two target extraction methods (PHYLUCE and HybPiper) under different filtering strategies: positions and loci. All concatenated matrices are included in the Supplementary Material. Abbreviations used: bp = base pairs; BS = bootstrap; coalesc = coalescence approach; concat = concatenation approach; LPP = local posterior probability; Nº = number; SR = substitution rate.

Dataset name	Nº of loci	Alignment length (bp)	Missing data (%)	Variable sites in bp (%)	Support mean (BS / LPP)	Nº of unsupported nodes (concat / coalesc)	Description of data filtering
PHYLUCE_627	627	515,875	57.1	247,320 (48)	87.2 / 0.88	17 / 28	Unfiltered
PHYLUCE_675_5	627	514,460	57.0	245,905 (48)	87.0 / 0.89	16 / 28	Filtering positions with SR > 5: removing 1415 (0.3%) characters from 93 (14.8%) loci
PHYLUCE_675_2.5	627	513,490	57.0	244,944 (48)	86.0 / 0.88	16 / 28	Filtering positions with SR > 2.5: removing 2385 (0.5%) characters from 131 (20%) loci
PHYLUCE_675_1	627	506,631	56.7	238,076 (47)	94.7 / 0.88	7 / 26	Filtering positions with SR > 1: removing 9244 (1.8%) characters from 252 (40.2%) loci
PHYLUCE_304	304	234,118	52.6	103,947 (44)	90.0 /0.85	9 / 33	48% of the original loci scoring best in taxon occupancy, mean bootstrap, saturation and evolution rate
HybPiper_1051	1051	333,614	9.5	118,542 (36)	91.2 / 0.87	11 / 28	Unfiltered
HybPiper_1051_5	1051	333,576	9.5	118,504 (36)	91.6 / 0.88	10 / 28	Filtering positions with SR > 5: removing 38 (0.01%) characters from 9 (0.9%) loci
HybPiper_1051_2.5	1051	333,556	9.5	118,484 (36)	91.4 / 0.87	12 / 28	Filtering positions with SR > 2.5: removing 58 (0.02%) characters from 18 (1.7%) loci
HybPiper_1051_1	1051	331,729	9.4	116,657 (35)	93.4 / 0.88	5 / 29	Filtering positions with SR > 1: removing 1885 (0.6%) characters from 213 (20.3%) loci
HybPiper_570	570	200,632	7.5	70,774 (35)	92.6 / 0.86	8 / 31	54% of the original loci scoring best in taxon occupancy, mean bootstrap, saturation and evolution rate

several peaks of loci visualized as ~~"p~~hantom spikes" that represent fast-evolving sites.

The removal of positions with substitution rates higher than 5 or 2.5 from unfiltered alignments did not improve the BS support values of trees (Supplementary Figs. S5 and S6). In contrast, with the strictest filtering scheme (SR > 1) of fast evolving sites removal, the number of resolved nodes in the concatenation approach notably increased (Fig. 4), and the curves were completely softened in the case of the HybPiper dataset (Fig. 4D). However, for the PHYLUCE dataset, some peaks close to zero, i.e. towards the present and the shallowest clades, appeared in all three filtering schemes (Fig. 4B). This could indicate that a stricter threshold would probably be needed to remove very fast-evolving positions and produce more refined PI profiles in the case of the PHYLUCE dataset.

3.4. Selection of the most informative loci

In the loci filtering strategy (using measures of taxon sampling, information content, saturation, and rate of evolution), we finally retained 304 loci (48% of the loci initially recovered; 234,118 bp) in the PHYLUCE dataset, and 570 loci (54% of the loci initially recovered; 200,632 bp) in the HybPiper dataset, which accounted for the highest phylogenetic signal (Table 3).

In the concatenation approach, the selection of the most informative loci resulted in significantly higher BS support values, decreasing the number of unsupported nodes from 17 to 9 in PHYLUCE and from 11 to 8 in HybPiper. In the coalescence approach, the selection of the best loci was not effective in terms of improving the LPP support values or the number of unsupported nodes, both values of these two metrics were even lower than the ones obtained with the corresponding unfiltered alignments (Figs. 5 and 6, and Table 3).

3.5. Comparison of tree topologies

Varied tree topologies were recovered in the global tree space among the approaches of concatenation/coalescence and the non-filtering/filtering strategies (Fig. 7). Discordant topologies between the concatenation and the coalescence approaches are well illustrated in the distant position that they occupy in the bidimensional tree space along both first and second dimensions (Fig. 7). On average, the RFadj between all the trees under concatenation vs. all the trees under coalescence was relatively high, with a 0.44 value (0.33–0.59; Supplementary Table S3). When topologies were compared among all those obtained under the concatenation approach and among all those obtained under the coalescence approach, we found that the topologies obtained with the coalescence approach were more similar between them (RFadj = 0.26) than the topologies obtained with the

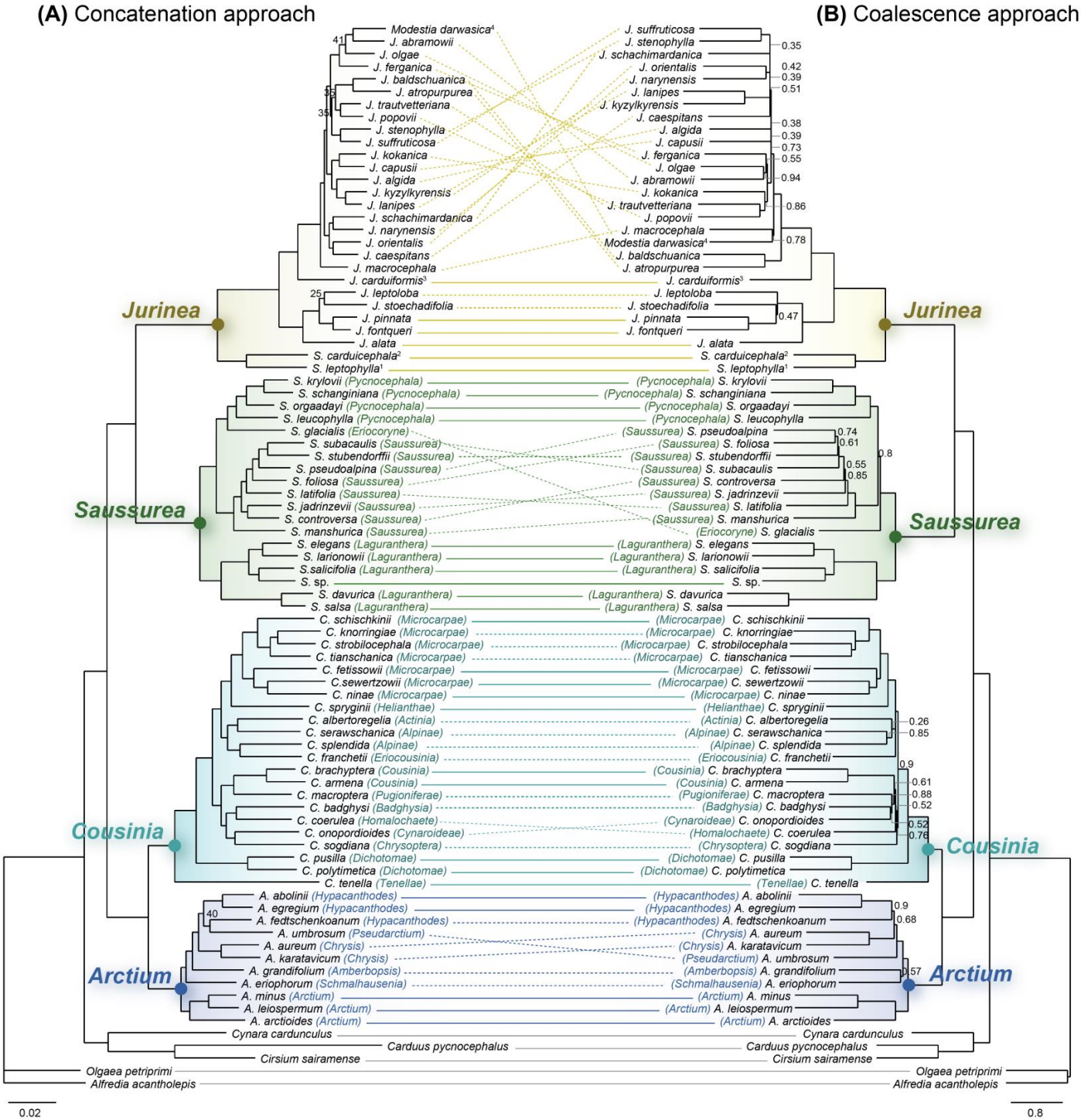


Fig. 3. Phylogenetic trees drawn opposite to each other (tanglegram) inferred from: (A) the concatenation approach from the HybPiper dataset, filtering positions with substitution rates (SR) > 1, and (B) the coalescence approach from the unfiltered HybPiper dataset. Only bootstrap (BS) support values < 70 and local posterior probabilities (LPP) < 0.95 are shown over branches. Continuous lines that link the species represent congruent positions between both trees and dashed lines represent incongruent positions. The section where each species belongs is specified in brackets, except for *Jurinea* (see text for details). Taxonomic treatments followed are López-Vinyallonga et al. (2011) for *Arctium*, Tscherneva (1997) for *Cousinia*, and Lipschitz (1979) for *Saussurea*. The species with a superscript were originally described under a different genus within the *Saussurea-Jurinea* complex: [1]*Saussurea leptophylla* = *Jurinea ancistrophylla*; [2]*Saussurea carduicephala* = *Jurinea ceratocarpa* = *Lipschitziella*; [3]*Jurinea carduiformis* = *Outreya carduiformis*; [4]*Modestia darwasica* = *Cirsium darwasicum*.

concatenation approach (RFadj = 0.31).

In relation to the impact of filtering, in general the softest filtering strategies of fast-evolving sites removal (SR > 5 and 2.5) did not significantly alter the tree topologies with respect to those obtained with the corresponding unfiltered alignments, both in the PHYLUCE and the HybPiper methods, under concatenation (RFadj = 0–0.01) or under coalescence (RFadj = 0–0.15). In contrast, for the strictest threshold scheme (SR > 1), topologies were more variable between unfiltered and filtered (SR > 1) datasets (Fig. 7). This effect was especially evident for the trees

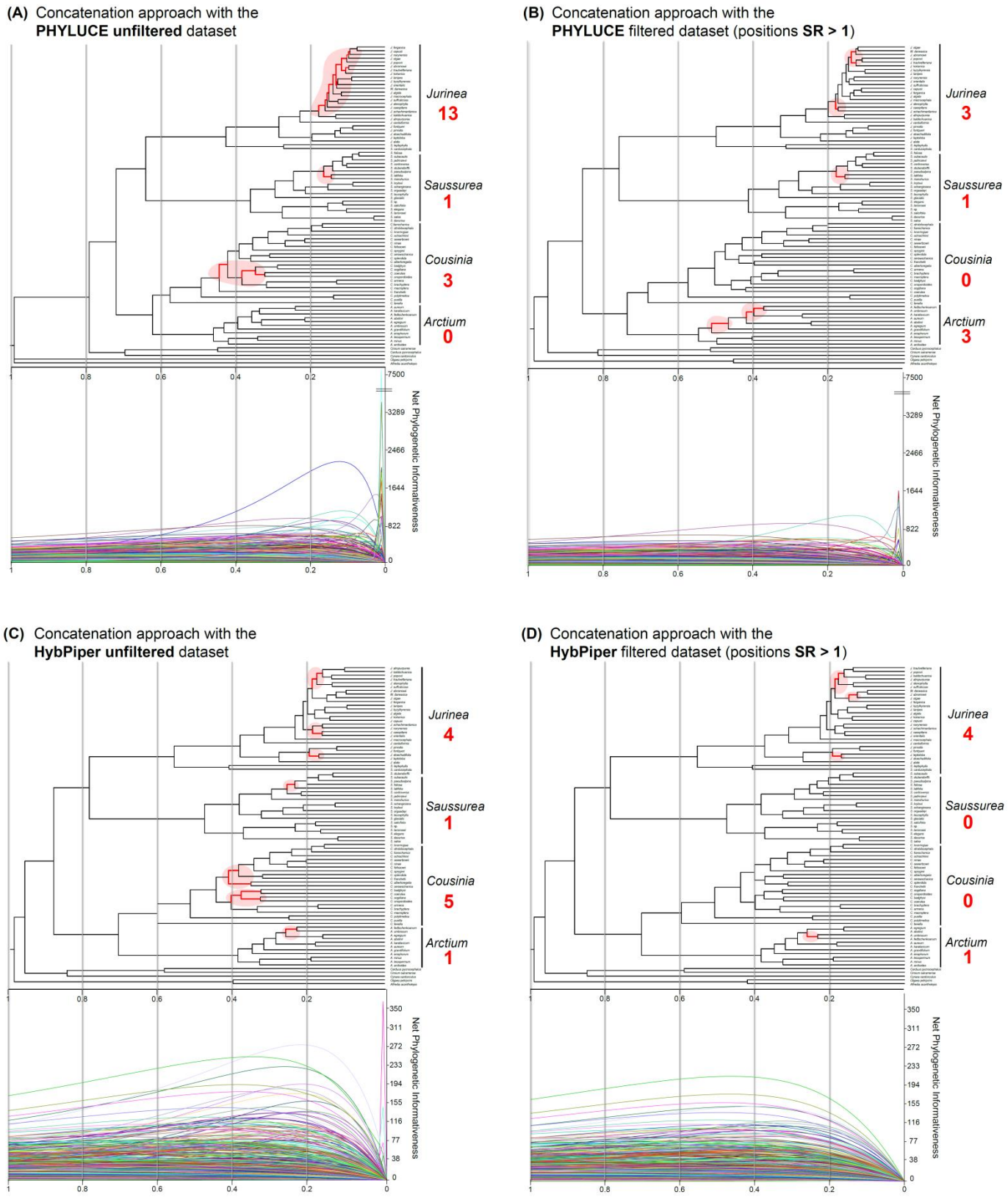


Fig. 4. Phylogenetic informativeness (PI) analyses showing the ultrametric trees scaled to an arbitrary scale of 1 (at the root) to 0 (at the tips) obtained with maximum likelihood analyses using the concatenation approach, and net phylogenetic informativeness profiles displaying curves for each locus in different colors. In this figure, we show the PI analyses performed to the PHYLUCE dataset, in particular (A) to the unfiltered alignments and (B) to the same alignments after filtering positions with substitution rates (SR) > 1, and to the HybPiper dataset, in particular (C) to the unfiltered alignments, and (D) to the same alignments after filtering positions with SR > 1. For a threshold of SR > 5 and SR > 2.5 see Supplementary Figs. S5 and S6. Branches with low bootstrap (BS) support values (BS < 70) are marked and highlighted in red. The number of unsupported nodes is specified for each genus at the right of the tree to see the differences between unfiltered datasets and the best positions filtering scheme (SR > 1) datasets.

inferred under concatenation, when the unfiltered PHYLUCE dataset was compared to the filtered (SR > 1) scheme (RFadj = 0.26). The selection of the most informative loci was the filtering strategy that resulted in the most discordant topologies when compared to the unfiltered datasets (Fig. 7 and Supplementary Table S3). In particular, the tree based on the best loci selected from the PHYLUCE dataset and under concatenation resulted in highly incongruent topologies compared to all the rest (mean of RFadj = 0.54), i.e. this dataset yielded the trees more distantly related to the other tree topologies inferred.

4. Discussion

4.1. COS loci resolve previously obscure generic relationships in Cardueae

The COS probe targets tested for deep nodes in Compositae (Mandel et al., 2014, 2015, 2017) are also useful to resolve close relationships at intergeneric levels. This evidence adds to previous studies (Mandel et al., 2014, 2015, 2017) and confirms the wide taxonomic range of COS loci applicability for phylogenomic and evolutionary studies on the largest family of flowering plants (Stebbins, 1970). For the first time, we were able to recover almost the entire set of target loci (99%, 1051 from 1061) using the novel pipeline HybPiper. Conversely, the pipeline PHYLUCE (the one used in previous studies using the COS loci set) captured only 627 loci (59%), a similar amount to those obtained in other studies for shallow species range (694 in Siniscalchi et al., in prep.) and higher taxonomical levels (763 and 795 in Mandel et al., 2014, 2015, respectively).

Here, we confidently resolved the historically obscure relationships among *Arctium, Cousinia*, *Saussurea*, and *Jurinea*. All phylogenies inferred in this study supported the sister relationships between *Arctium-Cousinia* and *Saussurea-Jurinea*, forming two separate complexes, a result that is congruent with the morphological hypothesis proposed by Susanna and Garcia-Jacas (2007, 2009). None of the preceding phylogenies built on Sanger sequencing data had been able to resolve the evolutionary relationships between these four genera with statistical support (Barres et al., 2013; Garcia-Jacas et al., 2002; Kita et al., 2004; Raab-Straube, 2003; Wang et al., 2007, 2013). In some cases, the genera were correctly nested but without support (López-Vinyallonga et al., 2009; Susanna et al., 2003, 2006; Susanna and Garcia-Jacas, 2009; Wang et al., 2009). Our study illustrates that controversial plant complexes with cryptic backbone relationships can be resolved with NGS target enriched data. Indeed, this NGS approach represents one of the most promising methodologies to date in the field of systematics and evolutionary biology (Buddenhagen et al., 2016), allowing the disentangling of both deep and shallow relationships of complex plant groups (e.g. Nicholls et al., 2015).

Certainly, the generic delimitation obtained here represents the first step toward increasing our knowledge of the evolution of highly diversified genera of the tribe Cardueae. The infrageneric relationships of the *Arctium-Cousinia* complex have been extensively explored with Sanger sequencing (see López-Vinyallonga et al., 2009, 2011; Susanna et al., 2003), but a complete phylogenetic assessment of *Saussurea-Jurinea* including all of the 16 small satellite genera described is still missing. Despite our reduced sampling, we have been able to clarify four possible cases of problematic classifications in the *Saussurea-Jurinea* complex. The first case concerns *Saussurea leptophy*lla Hemsl., which is here sampled for the first time in a phylogenetic tree. This species had been considered either as belonging to *Saussurea* (Lipschitz, 1979) *or Jurinea* (as *Jurinea ancistrophylla* Boiss., cf. Boissier, 1888). Phylogenies inferred in the present study indicate that the species should be placed in *Jurinea* (Fig. 3). Second, the satellite genus *Lipschitziella* R. Kam. (included here under Ju*rinea*), was described to accommodate *Saussurea carduicephala* and *Jurinea ceratocarpa* (Raab-Straube, 2003). Our results show that *Lipschitziella* groups with *Jurinea* (Fig. 3), matching previous phylogenies (Kita et al., 2004; Raab-Straube, 2003; Susanna et al., 2006; Wang et al. 2009). Third, we confirm that the monotypic genus *Outreya* Jaub. & Spach [included here as *Jurinea carduiformis* (Jaub. & Spach) Boiss., according to Garcia-Jacas et al., 2002] belongs to the *Jurinea* clade, as it had been shown previously (Garcia-Jacas et al., 2002; Susanna et al., 2006; Wang et al., 2013). Thus, its distinction as a separate genus is not *supported with the present data (Fig. 3). The last case concerns Modestia darwasica* (C. Winkl.) Kharadze & Tamamsch., which has been treated as a different genus within the complex. However, we found that this species was clearly nested in the Jurinea clade, as previously reported by Susanna et al. (2006). Despite these results, a more completely sampled phylogeny of the *Saussurea-Jurinea* complex should be conducted to confirm generic boundaries within the complex.

4.2. COS loci resolve species relationships within the radiated genera *Arctium, Cousinia* and *Saussurea*

Our study shows that COS loci are able to resolve the relationships among species at shallow taxonomic levels for the genera *Arctium*, *Cousinia*, and *Saussurea*. Previous studies on these genera based on chloroplast and nuclear conventional markers (e.g. for *Arctium-Cousinia* in López-Vinyallonga et al., 2009; for S*aussurea* in Wang et al., 2009) retrieved large polytomies, which hindered the phylogenetic assessment of subgeneric classifications. With the target enrichment technique, we have been able to recover dichotomous relationships highly supported in most clades, especially under the concatenation approach.

In general, species from the same section grouped together, which reflects congruence between molecular and morphological assemblies. It should be noted that the topology obtained with the coalescence approach matched the morphological sections in a higher number of cases than the tree inferred with the concatenation approach (Fig. 3). For example, the three species of *Arctium* sect. *Hypacanthodes* clustered together in the coalescence tree, whereas this section was paraphyletic in the concatenation based one. This was also the case for *Cousinia* and the two taxa of sect. *Alpinae* (Fig. 3). This fact highlights the usefulness of exploring both concatenation and coalescence approaches in phylogenetic reconstructions as currently recommended for phylogenomic data.

The comparison of *Arctium* and *Cousinia* species relationships obtained here (Fig. 3) with previously published

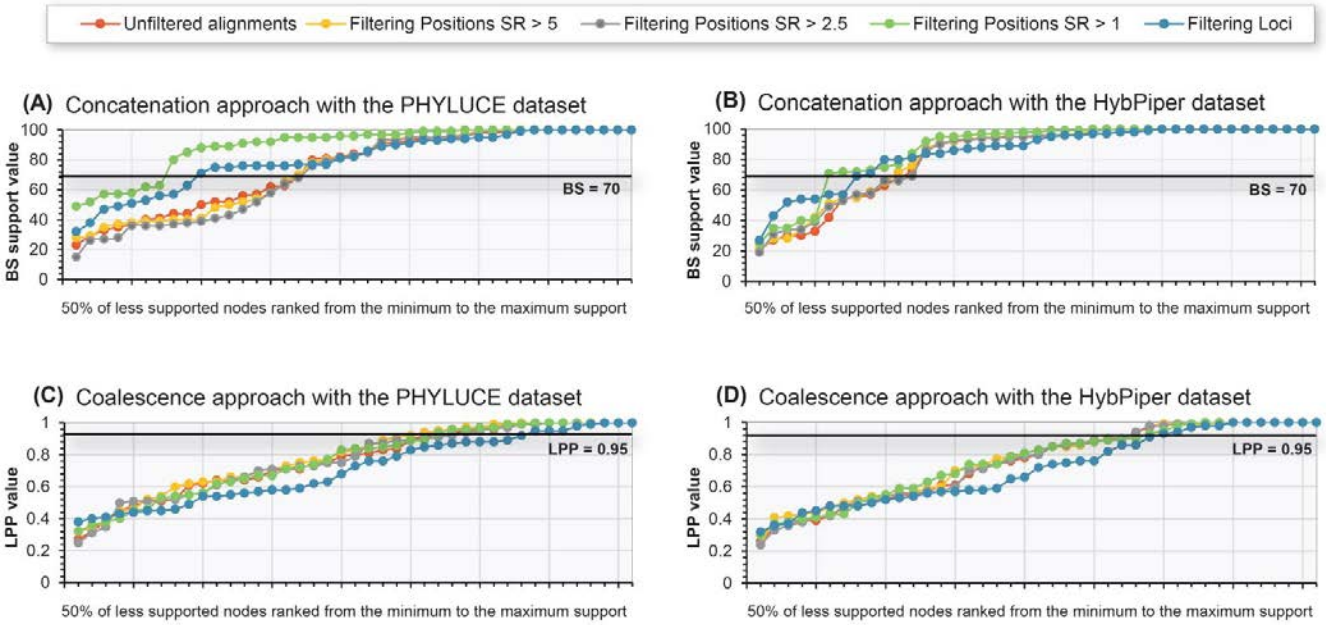


Fig. 5. Variation in support values across 50% of the less supported nodes, ranked from the minimum support to the maximum support obtained, according to different filtering treatments: unfiltered alignments, filtered alignments removing positions with substitution rates (SR) > 5, 2.5 and 1, and loci filtered alignments selecting the most informative loci. Support values were extracted from trees obtained with the concatenation approach, using (A) the dataset obtained with the PHYLUCE extracting method and (B) the dataset obtained with the HybPiper extracting method, and from species trees obtained with the coalescence approach, using (C) the dataset obtained with the PHYLUCE extracting method and (D) the dataset obtained with the HybPiper extracting method. Abbreviations used: BS = bootstrap; LPP = local posterior probability.

ones (López-Vinyallonga et al., 2009, 2011; Mehregan and Assadi, 2016; Mehregan and Kadereit, 2009; Susanna et al., 2003) shows that they are congruent except for a few cases. The phylogenies here presented provide the following new findings: 1) *Arctium grandifolium* (sect. *Amberbopsis*) and *A. eriophorum* (sect. *Schmalhausenia*) are not nested within sect. *Arctium* as previously recovered with ITS and *rpS4-trnT-trnL* markers; 2) *Cousinia tenella* (sect. *Tenellae*) groups with other *Cousinia* species, in contrast with the unusual grouping at the base of the whole *Arctium-Cousinia* complex retrieved in previous papers; and 3) after the divergence of *C. tenella*, the clade composed by *C. pusilla* and *C. polytimet*ica (both from sect. *Dichotomae*) is sister to the rest of *Cousinia*. The last two points are very interesting since it is observed that the annual species of *Cousinia* (*C. tenella*, *C. pusilla*, and *C. polytimetica*) are in separate lineages from all the other species (usually monocarpic and often biennial), which are grouped together in a different and much more diversified clade. These results suggest that a life strategy shift from annual to perennial would have allowed *Cousinia* to expand into new habitats, triggering higher diversification rates in a similar way to that reported for *Lupinus* L. in montane habitats (Drummond, 2008). In the case of *Cousinia*, we observed that when the monocarpic clade begins to diverge, individual gene trees became fairly incongruent and the resultant coalescence species tree, at this part, was poorly-moderately supported (Fig. 3B). This pattern of incongruent gene trees could be caused by an ancient hybridization event or polyploidization (Folk et al., 2018), but given that these processes are very rare in *Cousinia* (*Mehregan and Kadereit, 2009; Watanabe, 2002*), incomplete lineage sorting (ILS) may be more likely. However, this hypothesis needs further confirmation given that our taxon sampling is limited.

Concerning *Saussurea*, the high support values found for all the clades analyzed holds promise for future resolution of this radiation with a higher taxa sampling. Previous phylogenies, also with a reduced taxa sampling, retrieved poor-moderate resolution at the species level (Kita et al., 2004; Raab-Straube, 2003; Wang et al., 2009). Species relationships within sect. *Saussurea* were different in the trees obtained with the concatenation and the coalescence approaches (Fig. 3), causing a topological incongruence (see 4.3. for possible methodological tools to explore causes of incongruence). These differences could derive from fast and island-like radiation events in the major diversity center of *Saussurea* located in China (Wang et al., 2009; Wen et al., 2014), where more than 150 endemic species of the sect. S*aussurea* are found (Shi et al., 2011).

4.3. Conflicting species relationships within *Jurinea*

Compared to the other genera, the branch support values of interspecific relationships within *Jurinea* was surprisingly low. Specifically, relationships and topologies recovered were highly variable among the different phylogenomic approaches (concatenation/coalescence, among target extraction methods, and among posterior filtering treatments; Supplementary Figs. S1–S4). Whereas the optimal phylogenetic tree inferred with the concatenation approach resulted in moderate-high supported branches (only 16.7% of the internodes were unsupported; Fig. 3A), almost all branches of the coalescence tree were unsupported (93.8%; Fig. 3B), revealing high incongruence among gene trees. In addition, branch lengths of the *Jurinea* group were shorter than the branches of the other genera. Overall,

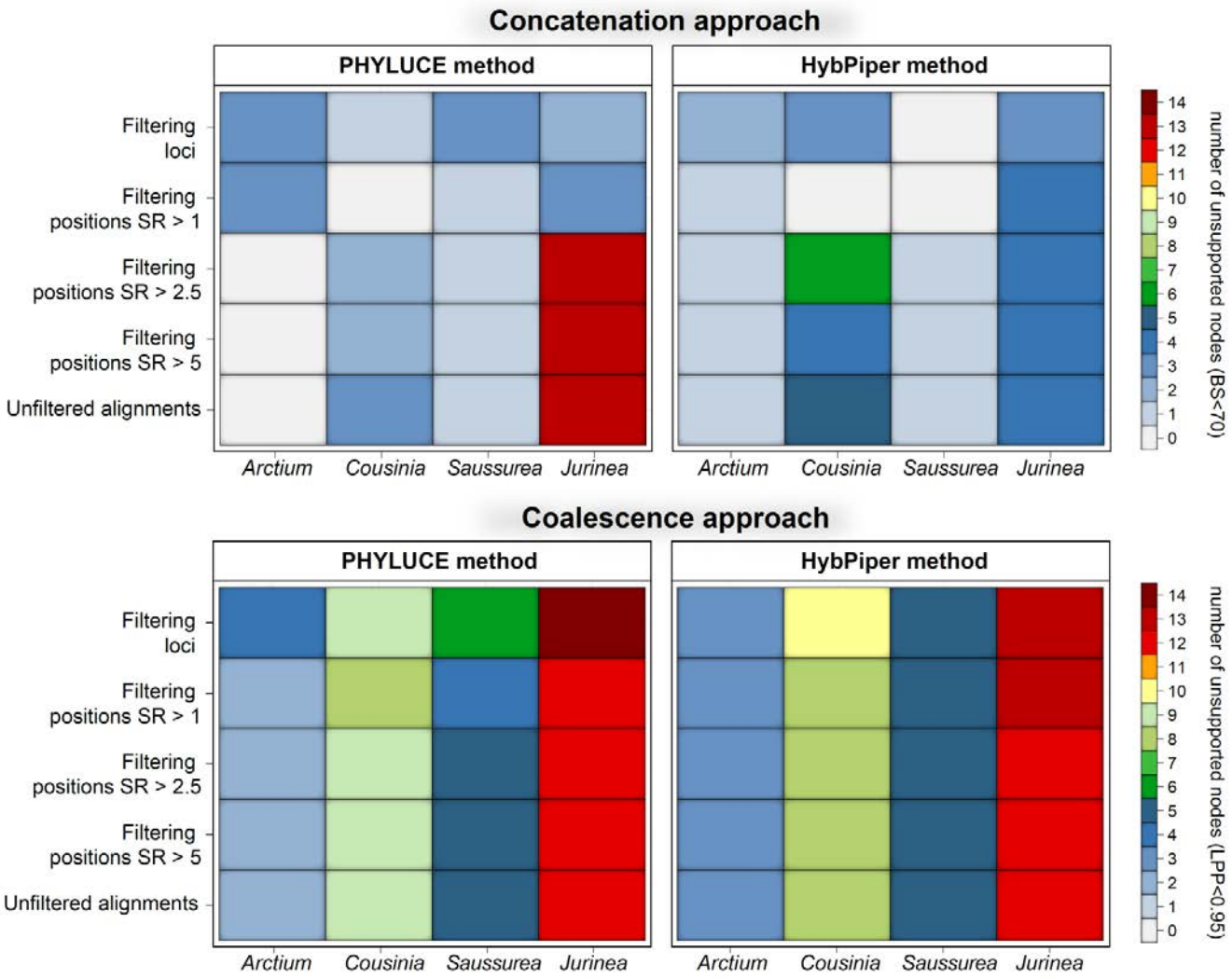


Fig. 6. Number of unsupported nodes represented in a color scale for all the executed analyses with the concatenation approach, considering nodes with bootstrap (BS) support values < 70, and with the coalescence approach, considering nodes with local posterior probabilities (LPP) < 0.95. For the two approaches, both types of target extraction methods are considered: PHYLUCE and HybPiper. Columns represent the four genera examined: *Arctium*, *Cousinia*, *Saussurea* and *Jurinea*, and rows the different analyzed datasets (see Fig. 1): unfiltered alignments, filtered alignments removing positions with substitution rates (SR) higher than 5, 2.5 and 1, and loci filtered alignments removing some loci following the criteria of taxon occupancy, support content, saturation, and evolution rate.

this topological structure could indicate: 1) ILS (persistence of ancestral polymorphisms of genes after species splitting), which could be common in cases of rapid radiations (Oliver, 2013; Rokas and Caroll, 2006; Whitfield and Lockhart, 2007); or 2) introgression phenomena or hybrid speciation, in which gene tree histories are discordant due to events of genetic admixture with other lineages (Folk et al., 2018). The limited taxon sampling of the present and previous studies on *Jurinea* (14–18 species with ISSR or ITS in Dogan et al., 2007, 2010; Salmerón-Sánchez et al., 2015) does not allow to discriminate between these two hypotheses. The high gene tree discordance found here for this group could indicate one of most common ILS effects, which is the occurrence of the inferred species tree in the ~~–~~anomaly zone" (Degnan and Rosenberg, 2006; Linkem et al., 2016). This term was described to refer to a tree space area where the most likely gene tree topologies do not reflect the true species tree topology. In these cases, phylogenetic inference methods fail to reconstruct the true species tree, especially when a concatenation approach is used (Mendes and Hahn, 2018). In future investigations, the relative influence of ILS and hybridization could be tested through multiple approaches recently proposed for Hyb-Seq data (see García et al., 2017; Kamneva et al., 2017; Mitchell et al., 2017; Simmons et al., 2016). The evolutionary role of polyploidy could also be explored as suggested by Crowl et al. (2017), Eriksson et al. (2018), or Grover et al. (2015). Although COS loci have been designed from low-copy nuclear genes, several possible paralog copies have been detected (Mandel et al., 2015) as we found here (see section 4.4.), and as had been reported for AHE data (Buddenhagen et al., 2016). However, polyploidy seems to be as rare in Jurinea as it is in *Cousinia* (Watanabe, 2002).

Several strategies may be followed to shed light into the evolutionary history of rapidly diversified genera (e.g. *Helianthus* L., Stephens et al., 2015) in which gene tree discordance prevails, and is even magnified, with phylogenomic data. Certainly, the first step to improve branch support values is to obtain a complete sampling of species, which is essential for reconstructing well resolved phylogenies (Lecointre et al., 1993; Philippe et al.,

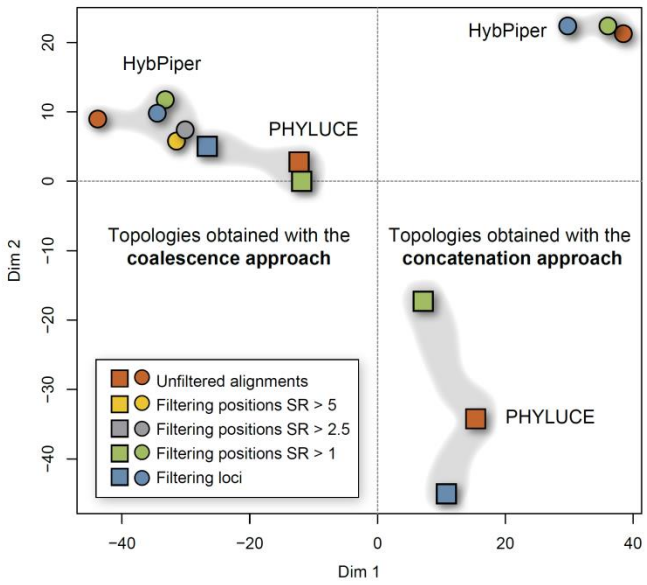


Fig. 7. Tree space from a multidimensional scaling of Robinson-Foulds (RF) pairwise distance comparisons among all topologies of trees inferred. Trees obtained with the PHYLUCE extracting method are represented in squares and trees obtained with the HybPiper extracting method are represented in circles. Trees that resulted in equal topology to unfiltered alignments are displayed in orange, i.e. as the same color of unfiltered alignments. The same topologies or almost equivalent with RFadj = 0 or RFadj = 0.01 are, for the coalescence approach, between the PHYLUCE unfiltered alignments and filtered alignments removing positions with substitution rates (SR) > 5 and 2.5 and, for the concatenation approach, for the two pipelines PHYLUCE and HybPiper, between unfiltered alignments and filtered alignments removing positions SR > 5 and 2.5.

2011). We included here a very small representation of *Jurinea* (26 out of the 200 described species; Susanna and Garcia-Jacas, 2007), so a broader representation is crucial to extract solid conclusions about its evolution. In agreement, we found that the position of species that are unique representatives of a section were the most variable cases in different phylogenetic analyses (Fig. 3). Another possible improvement in relation to sampling is the addition of several individuals per species (Kubatko and Degnan, 2007; Maddison and Knowles, 2006; McCormack et al., 2009), particularly for the coalescence approach when considerable levels of gene tree heterogeneity exist in the clade of interest.

Another option would be to increase gene alignment length by concatenating compatible loci through methods like naive or statistical binning (Bayzid and Warnow, 2013; Bayzid et al., 2015; Mirarab et al., 2014). In this way, the possible effect to incorporate gene trees derived from short alignments with a weak phylogenetic signal, which could lead to a poorly resolved species tree under the coalescence approach, would be minimized. However, disparate results have been found when applying binning procedures, recovering well resolved coalescence trees in some cases (Blaimer et al., 2016) and poorly resolved phylogenies in others (Streicher et al., 2018). Another alternative in study cases with low ILS effect, which seems not appropriate for *Jurinea*, would be to recover a higher number of variable positions such as those located in introns or flanking regions of conserved probes set. Indeed, several studies have showed that variability in the alignments increases with increasing distance from the center of UCE anchored loci (Bossert et al., 2017; Faircloth et al., 2012; Gilbert et al., 2015; Van Dam et al., 2017). In plants, this strategy has proved useful to resolve species divergence in the genus Heuchera (Folk et al., 2015).

Finally, other variants of high throughput sequencing, like restriction-site associated sequencing (RAD-seq; Baird et al., 2008), could help to clarify evolutionary relationships in rapidly diversified lineages, as it has been successfully achieved for other radiations (e.g. Darwell et al., 2016; Tripp et al., 2017; Wagner et al., 2013). Nevertheless, important drawbacks should be considered for this method: the short length of the loci captured (< 300 bp; Andrews et al., 2016), an increase of uncertain homology in relation to time since species divergence (Wagner et al., 2013), the difficulties to link data from different studies, and the problems already detected to resolve short internal branches (Leaché et al., 2015).

4.4. Evaluating differences between target extraction methods: PHYLUCE and HybPiper pipelines

This study represents the first evaluation of the impact in phylogenies of two target extraction methods implemented in the automated pipelines PHYLUCE (Faircloth, 2015) and HybPiper (Johnson et al., 2016). One of the notable differences observed between the two approaches is the length of the final matrix recovered: the PHYLUCE matrix was 35.3% longer than the HybPiper one (see Table 2 for details). At first sight, this result is quite surprising given that the total number of loci found was lower with PHYLUCE (627) than with HybPiper (1051). However, this is likely due to the fact that with PHYLUCE the reads are assembled into contigs before being mapped to the targets, which results in contigs that are longer than the targets themselves (Fig. 2C). In contrast, with HybPiper the reads are assembled after being mapped to the targets and thus the resulting contigs cannot be much longer than the targets, and actually tend to be shorter (Fig. 2D). This is also reflected in length differences of the individual locus alignments (on average, 823 bp per locus with PHYLUCE and 317 bp per locus with HybPiper).

Despite the fact that longer alignments are desirable for gene tree reconstructions, regions outside the core of the COS targets (identified from EST; Mandel et al., 2014) might include non-coding regions and, consequently, a high number of positions could have abnormally high substitution rates. Thus, in this non-coding regions saturation and multiple hits effects may tend to be high and, accordingly, the positional homology would be questionable. To this point, the phylogenetic informativeness analysis detected great amounts of ~~"p~~hantom spikes" in the PHYLUCE matrix, and even in the strictest scheme of fast-evolving positions removal (SR > 1, 9244 bp removed; Table 3) the curves of locus profiles were not smoothed sufficiently (Fig. 4B). However, this result could be due not only to the recovery of highly variable regions outside the conservative core of the EST regions, it could also be related to the lack of a target reference sequence to map the non-target sequences, thus resulting in poorly aligned regions with considerable homoplasy problems. Overall, the conservative core of the EST regions (i.e. the COS targets) showed enough variation to infer robust phylogenetic relationships, as shown by the HybPiper dataset. Therefore,

highly variable sites located outside the target length could be decreasing phylogenetic signal-to-noise ratio instead of adding valuable phylogenetic information. However, it should be tested for recently diversified lineages with low ILS with other bioinformatics methods if the inclusion of the COS flanking regions with great amounts of variation would provide valuable information to resolve entangled phylogenetic histories.

Another notable difference between the two methods is the different treatment of sequence variants (potentially paralogous copies or alleles; see section 2.4. for methodological details). Briefly, in PHYLUCE only the single copy loci are retained; in contrast, HybPiper retrieves multiple-copy loci, but only one of the copies (potential paralogs) is retained in the dataset, based on the criteria described before (see section 2.4). In our sequence dataset, between 0 and 167 (144 on average) loci were flagged with paralog warnings in the HybPiper method, from a total of 1051 target loci. Such multiple copies could originate from different sources: real paralog coexistence, recent polyploidy, contamination, sequencing errors, or allelic variants. Altogether, the species analyzed do not seem to be strongly affected by potential paralogs. However, flagged loci with paralog warnings detected with HybPiper should be further evaluated or removed from downstream analyses in a conservative framework given that small-scale duplications (segmental, tandem, and retro-duplications) have been shown to occur commonly in plant genomes (Hudson et al., 2011; Rensing, 2014).

In sum, how reads are assembled into contigs is probably the factor that contributes most to differences in the number of targets recovered between both analysis packages, rather than paralog treatment. This is evident from the fact that, with HybPiper, an average of 144 potential paralogs was detected, a number that is much lower than the difference in the number of loci retrieved by PHYLUCE (675) and HybPiper (1055). In other extraction pipelines like aTRAM (Allen et al., 2015, 2017) or the recently published HybPyloMarker (Fér and Schmickly, 2018), the predominant procedure, and probably the best strategy to recover the target loci, is to perform assembly after the reads are mapped to the targets (see Table 2 in Fér and Schmickly, 2018).

Concerning their influence on phylogenetic results, we found that both reference-based extraction methods were successful in the resolution of backbone relationships among the evaluated genera. The high amounts of missing data per loci retrieved with PHYLUCE (only 9.2% of genes were recovered in 90% or more of the species) did not affect the branch support values of intergeneric relationships. This is in agreement with Hosner et al. (2015), who reported that missing positions in alignments could be more problematic than entire missing sequences of a given locus. At shallow taxonomic levels, both packages were also able to detect gene tree discordances in the same proportion, independently of the data analysis pipeline used (Fig. 6). However, topologies built under the concatenation approach and under coalescence with the PHYLUCE dataset were more different between them than the ones obtained under the two approaches with the HybPiper dataset (Fig. 7). Nonetheless, in the concatenation approach analyses, considerable differences between the two extraction methods were found at species relationships level. The PHYLUCE method failed to estimate with confidence species relationships in Jurinea, resulting in an entangled topology with fairly low branch support values compared to the results found with HybPiper (Supplementary Figs. S1A, S2A). A possible explanation could be that the high number of missing loci for some species in the concatenated dataset hindered ancestry state reconstructions in resampled data matrices when bootstrap replicates were calculated under the concatenation approach.

As observed here and as García et al. (2017) reported with other extracting methods, the use of different procedures of target extraction can lead to different estimates of topology and branch lengths of tree reconstructions. Thus, it is evident that the choice of a given bioinformatic workflow can have a critical impact on the results obtained. In summary, the PHYLUCE method seems to present more limitations and introduces more phylogenetic noise than the HybPiper method. However, in taxonomical groups with low-moderate degrees of ILS, hybridization and polyploidy, PHYLUCE is more conservative in terms of avoiding potential paralog copies, more efficient in computational time demanded, memory used, and number of files produced, compared to HybPiper.

4.5. The coalescence approach yields higher topological robustness of phylogenetic trees

High throughput sequencing has provided extensive genome-scale datasets and has been useful to resolve many prior uncertain branches of the tree of life. However, incongruence between nuclear, mitochondrial or chloroplast based phylogenies, and conflicting gene tree histories persist across phylogenetic reconstructions (Jeffroy et al., 2006). This incongruence could be masked when gene sequences recovered are concatenated as a single supergene unit (supermatrix or concatenation approach). However, this analytical practice is currently under discussion in phylogenomics since it tends to produce maximum bootstrap support values and completely resolved phylogenies even when biological factors (like ILS, hybridization, horizontal gene transfer, recombination, and gene duplication/loss), random biases, or systematic errors (compositional heterogeneity, long-branch attraction, gene-tree discordances, and missing sequence data) are present in the input data (Kubatko and Degnan, 2007; Liu et al., 2015b; Salichos and Rokas, 2013). In our study, we obtained higher support values and almost fully resolved phylogenies applying the concatenation approach (Supplementary Figs. S1 and S2), but the resulting trees showed considerable conflicting topologies among the different extraction and filtering procedures (Fig. 7). These results support the claim of previous researchers (Kubatko and Degnan, 2007; Salichos and Rokas 2013), who suggested avoiding the use of traditional bootstrap values as a metric to quantify tree certainty in the concatenation approach.

Alternatively, analyzing sequence data under the coalescence assumptions may aid in avoiding reconstruction artifacts, detect possible gene incongruences, and better integrate different gene histories (see review in Liu et al., 2015b). Here, it has been confirmed that our study group presents high gene-tree heterogeneity, which is reflected in the weakly supported internal branches of the coalescence tree (Fig. 3B). Causes of incongruence may be derived from several factors. One is the relatively short length of our gene alignments (average of 823 bp in PHYLUCE and 317 bp in HybPiper), which could result in insufficient phylogenetic signal yielding poorly resolved gene trees (average of bootstrap 58.7 in PHYLUCE and 32.2 in HybPiper). Indeed, we found positive correlations between loci

lengths and mean BS support values in gene trees. In light of this observation, future studies should consider using a limited number of naive bins or a statistical binning approach (Mirarab et al., 2014) in order to improve gene trees reconciliation. It has also been proposed that high levels of missing data (missing locus per species) could lead to low support and accuracy of coalescence trees (Gatesy and Springer, 2014). However, our phylogenetic analyses were resilient to the effects of this type of missing data, since no remarkable differences were observed between tree topologies obtained with the PHYLUCE and the HybPiper methods (Fig. 5) despite their significantly distinct proportion of missing data (Fig. 2A and 2B). Such resilience was also shown in the simulation study by Hovmöller et al. (2013). It is well documented that the coalescence approach can consistently yield trees closer to the correct species tree as the number of loci increases (Liu et al., 2015a). Concordantly, we observed that phylogenies estimated with a reduced loci dataset showed lower branch support values in our coalescence approach (see section 4.6. for details).

Despite the incongruence detected across coalescence trees (Supplementary Figs. S3 and S4) and their lower support values with respect to concatenation trees (Fig. 5 and Table 3), we detected that coalescence tree topologies obtained with alternative extraction and refining methods were more congruent or similar among them than those obtained under the same conditions using the concatenation approach (Fig. 7). This pattern is in agreement with results reported by other researchers (Buddenhagen et al., 2016; Edwards et al., 2016; Mitchell et al., 2017), which highlighted the topological robustness of coalescence methods.

4.6. Impact of filtering target-enriched data

Recent target-enriched studies have added an additional step of sequence refining to minimize the impact of phylogenetic noise (Table 1). We explored the effectiveness of two types of dataset filtration: on the one hand removing positions with unusually high substitution rates (fast-evolving sites; Fragoso-Martínez et al., 2017), and on the other selecting and analysing only the most informative loci under different criteria (Borowiec et al., 2015).

First, it should be noted that all coalescence analyses were unaffected by the application of any filtering scheme, indicating that gene-tree discordances cannot be attributed to phylogenetic noise derived from fast-evolving sites or the addition of uninformative loci (see section 4.5 for possible sources of gene trees incongruence). In contrast, in the case of the concatenation approach, both strategies of filtering initial matrices before phylogenetic inference were in general effective (Table 3). This result is in agreement with similar findings reported by Xi et al. (2014), which showed that coalescence approaches were more robust in the presence of positions with high substitution rates compared to concatenation approaches.

The first strategy of position filtering proved to be more useful when the strictest threshold was applied (SR > 1), improving the bootstrap support values (Fig. 5) and increasing the number of supported nodes (Fig. 6) for the two different target extraction pipelines. Previous works (Goremykin et al., 2010; Parks et al., 2012; Straub et al., 2014; Xi et al., 2014) and the first studies applying this filtering workflow (Fragoso-Martínez et al., 2017; Wanke et al., 2017) already suggested its benefits to reduce phylogenetic noise and saturation. In particular, the noise in our study was especially mitigated in Jurinea, in which the filtering strategies employed here resulted in resolving initially unsupported nodes, for instance varying from 13 to 3 in the PHYLUCE method (Figs. 4A and 4B). However, as previously highlighted, removing too many positions may lead to an inappropriate exclusion of phylogenetically informative characters and consequently to the loss of robustly supported clades (Drew et al., 2014; Streicher et al., 2018). This occurred in Arctium, for which node resolution decreased in greatest refining scenarios (filtering by positions SR > 1 in PHYLUCE; Fig. 6). For this reason, it would be desirable to test several thresholds of filtering positions and see which one fits better the entire tree or the particular clade of interest. Additionally, less restrictive cut-offs for position filtering can result in an increase of unresolved nodes, as we observed in HybPiper dataset for Cousinia and scheme SR > 2.5 (green square in Fig. 6).

Currently, one of the main questions in phylogenomics is how many loci are needed to produce robust phylogenies. The answer is complex, and there is an increasing number of studies evaluating the effects of prioritizing the quality (information-rich loci or loci recovered in high number of taxa) or the quantity (as many loci as possible) (e.g. Borowiec et al., 2015; Hosner et al., 2015; Misof et al., 2013; Salichos and Rokas, 2013; Streicher et al., 2016). Here we observed that, in the concatenation approach, the use of a lower number of loci but those with the highest phylogenetic signal increased the resolution of entangled clades (Fig. 6), a trend observed in other works (Borowiec et al., 2015; Salichos and Rokas, 2013). However, in the coalescence approach, the retention of only the most informative loci (approximately half of them) resulted in low LPP support values and low phylogenetic resolution (Figs. 5 and 6 and Table 3). Moreover, incongruence between gene trees persisted and former unsupported branches in coalescence trees remained unsupported after locus filtering, in agreement with Longo et al. (2017). Therefore, our outcomes suggest that in coalescence approaches it seems preferable to keep all loci, rather than keeping only the most informative ones, as outlined by Liu et al. (2015a, b) and Streicher et al. (2016). Nonetheless, the strategy of eliminating relatively uninformative gene trees was successful in Hosner et al. (2015).

In sum, filtering by positions (in our case at threshold SR > 1) was the best refining strategy given the notable increase of tree resolution and the minimum topological differences in respect to the topologies recovered with unfiltered sequences (Fig. 7 and Supplementary Table 3). However, generalizing for future investigations, an optimal comprehensive filtration metric may not exist, given the different impacts of each filtering strategy depending on the clade of interest. The described methodologies of performing a heat map (Buddenhagen et al., 2016) and calculating internode certainty (Salichos and Rokas, 2013) could help to detect the most highly confident reconstructed clades and the more sensitive groups to particular data treatments. Additionally, trees inferred under concatenation and coalescence approaches benefit differently from sampling, filtering and post-processing strategies. In our case, it would be preferable to give priority to loci quality (removing fast-evolving positions or using the most informative ones) in the concatenation approach, and to maximize the number of loci in the coalescence approach.

5. Acknowledgments

Authors thank Carolina Siniscalchi, Maria Luisa Gutiérrez and Fernando Castro for providing technical support during the laboratory process. We also thank the herbaria that provided material for the study: BC, DUSH, E, ERE, FRU, GDA, LE, MJG, TK, and W. We would like to thank Stefan Wanke and two anonymous reviewers for their comments and suggestions that improved the manuscript. Financial support from the Spanish –Ministerio de Ciencia e Innovación" (Project CGL2015-66703-P and Ph.D. grant to Sonia Herrando-Moraira) and the Catalan government (–Ajuts a grups consolidats" 2014/SGR/514 and 2017-SGR1116) is gratefully acknowledged.

6. References

Allen, J.M., Boyd, B., Nguyen, N.-P., Vachaspati, P., Warnow, T., Huang, D.I., Grady, P. G., Bell, K.C., Cronk, Q.C., Mugisha, L., Pittendrigh, B.R., Leonardi, M.S., Reed, D.L. & Johnson, K.P. 2017. Phylogenomics from whole genome sequences using aTRAM. *Systematic Biology* 66: 786–798. https://doi.org/10.1093/sysbio/syw105

Allen, J.M., Huang, D.I., Cronk, Q.C. & Johnson, K.P. 2015. ATRAM - automated target restricted assembly method: a fast method for assembling loci across divergent taxa from next-generation sequencing data. *BMC Bioinformatics* 16: 98. https://doi.org/10.1186/s12859-015-0515-2

Andermann, T., Cano, A., Zizka, A., Bacon, C. & Antonelli, A. 2018. SECAPR-A bioinformatics pipeline for the rapid and user-friendly processing of Illumina sequences, from raw reads to alignments. *PeerJ* Preprints 6: e26477v3. https://doi.org/10.7287/peerj.preprints.26477v3

Andrews, K.R., Good, J.M., Miller, M.R., Luikart, G. & Hohenlohe, P.A. 2016. Harnessing the power of RADseq for ecological and evolutionary genomics. *Nature Reviews Genetics* 17: 81–92. https://doi.org/10.1038/nrg.2015.28

Baird, N.A., Etter, P.D., Atwood, T.S., Currey, M.C., Shiver, A.L., Lewis, Z.A., Selker, E.U., Cresko, W.A. & Johnson E.A. 2008. Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS ONE* 3: e3376. https://doi.org/10.1371/journal.pone.0003376

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin V.M., Nikolenko, S.I., Pham, S., Prjibelski A.D., Pyshkin, A.V., Sirotkin A.V., Vyahhi, N., Tesler, G., Alekseyev, M.A. & Pyshkin, A.V. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology* 19: 455–477. https://doi.org/10.1089/cmb.2012.0021

Barres, L., Sanmartín, I., Anderson, C.L., Susanna, A., Buerki, S., Galbany-Casals, M. & Vilatersana, R. 2013. Reconstructing the evolution and biogeographic history of tribe Cardueae (Compositae). *American Journal of Botany* 100: 867–882. https://doi.org/10.3732/ajb.1200058

Bayzid, M.S. & Warnow, T. 2013. Naive binning improves phylogenomic analyses. *Bioinformatics* 29: 2277–2284. https://doi.org/10.1093/bioinformatics/btt394

Bayzid, M.S., Mirarab, S., Boussau, B. & Warnow, T. 2015. Weighted statistical binning: enabling statistically consistent genome-scale phylogenetic analyses. *PLoS ONE* 10: e0129183. https://doi.org/10.1371/journal.pone.0129183

Blaimer, B.B., LaPolla, J.S., Branstetter, M.G., Lloyd, M.W. & Brady, S.G. 2016. Phylogenomics, biogeography and diversification of obligate mealybug-tending ants in the genus *Acropyga*. *Molecular Phylogenetics and Evolution* 102: 20–29. https://doi.org/10.1016/j.ympev.2016.05.030

Boissier, P.E. 1888. *Flora Orientalis sive enumeratio plantarum in Oriente a Graecia et Aegypto ad Indiae fines hucusque observatarum*. H. Georg, Basilea.

Bolger, A.M., Lohse, M. & Usadel, B. 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* 30: 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Borowiec, M.L. 2016. AMAS: a fast tool for alignment manipulation and computing of summary statistics. *PeerJ* 4: e1660. https://doi.org/10.7717/peerj.1660

Borowiec, M.L., Lee, E.K., Chiu, J.C. & Plachetzki, D.C. 2015. Extracting phylogenetic signal and accounting for bias in whole-genome data sets supports the Ctenophora as sister to remaining Metazoa. *BMC Genomics* 16: 987. https://doi.org/10.1186/s12864-015-2146-4

Bossert, S., Murray, E.A., Blaimer, B.B. & Danforth, B.N. 2017. The impact of GC bias on phylogenetic accuracy using targeted enrichment phylogenomic data. *Molecular Phylogenetics and Evolution* 111: 149–157. https://doi.org/10.1016/j.ympev.2017.03.022

Branstetter, M.G., Ješovnik, A., Sosa-Calvo, J., Lloyd, M.W., Faircloth, B.C., Brady, S.G. & Schultz, T.R. 2017. Dry habitats were crucibles of domestication in the evolution of agriculture in ants. *Proceedings of the Royal Society B* 284: 20170095. https://doi.org/10.1098/rspb.2017.0095

Bryson, R.W., Linkem, C.W., Pavón-Vázquez, C.J., Nieto-Montes de Oca, A., Klicka, J. & McCormack, J.E. 2017. A phylogenomic perspective on the biogeography of skinks in the *Plestiodon brevirostris* group inferred from target enrichment of ultraconserved elements. *Journal of Biogeography* 44: 2033–2044. https://doi.org/10.1111/jbi.12989

Buddenhagen, C., Lemmon, A.R., Lemmon E.M., Bruhl, J., Cappa, J., Clement, W.L., Donoghue, M., Edwards, E.J., Hipp, A.L., Kortyna, M., Mitchell, N., Moore, A., Prychid, C.J., Segovia-Salcedo, M.C., Simmons, M.P., Soltis, P.S., Wanke, S. & Mast, A. 2016. Anchored phylogenomics of angiosperms I: Assessing the robustness of phylogenetic estimates. https://doi.org/10.1101/086298

Burress, E.D., Alda, F., Duarte, A., Loureiro, M., Armbruster, J.W. & Chakrabarty, P. 2017. Phylogenomics of pike cichlids (Cichlidae: *Crenicichla*): the rapid ecological speciation of an incipient species flock. *Journal of Evolutionary Biology* 31: 14–30. https://doi.org/10.1111/jeb.13196

Capella-Gutiérrez, S., Silla-Martínez, J.M. & Gabaldón, T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25: 1972–1973. https://doi.org/10.1093/bioinformatics/btp348

Chau, J.H., Rahfeldt, W.A. & Olmstead, R.G. 2018. Comparison of taxon-specific versus general locus sets for targeted sequence capture in plant phylogenomics. *Applications in Plant Sciences* 6: e1032. https://doi.org/10.1002/aps3.1032

Cronn, R., Knaus, B.J., Liston, A., Maughan, P.J., Parks, M., Syring, J.V. & Udall, J. 2012. Targeted enrichment strategies for next-generation plant biology. *American Journal of*

Botany 99: 291–311. https://doi.org/10.3732/ajb.1100356

Crowl, A.A., Myers, C. & Cellinese, N. 2017. Embracing discordance: Phylogenomic analyses provide evidence for allopolyploidy leading to cryptic diversity in a Mediterranean *Campanula* (Campanulaceae) clade. *Evolution* 71: 913–922. https://doi.org/10.1111/evo.13203

Cummins, C.A. & McInerney, J.O. 2011. A method for inferring the rate of evolution of homologous characters that can potentially improve phylogenetic inference, resolve deep divergence and correct systematic biases. *Systematic Biology* 60: 833–844. https://doi.org/10.1093/sysbio/syr064

Darwell, C.T., Rivers, D.M. & Althoff, D.M. 2016. RAD-seq phylogenomics recovers a well-resolved phylogeny of a rapid radiation of mutualistic and antagonistic yucca moths. *Systematic Entomology* 41: 672–682. https://doi.org/10.1111/syen.12185

Degnan J.H. & Rosenberg N.A. 2006. Discordance of species trees with their most likely gene trees. *PLoS Genetics* 2: 762–768. https://doi.org/10.1371/journal.pgen.0020068

Dogan, B., Duran, A. & Hakki, E.E. 2007. Phylogenetic analysis of *Jurinea* (Asteraceae) species from Turkey based on ISSR amplification. *Annales Botanici Fennici* 44: 353–358.

Dogan, B., Hakki, E.E. & Duran, A. 2010. A phylogenetic analysis of *Jurinea* (Compositae) species from Turkey based on ITS sequence data. *African Journal of Biotechnology* 9: 1741–1745. https://doi.org/10.5897/AJB10.1525

Dornburg, A., Townsend, J.P., Brooks, W., Spriggs, E., Eytan, R.I., Moore, J.A., Wainwright, P.C., Lemmon, A., Lemmon, E.M. & Near, T.J. 2017. New insights on the sister lineage of percomorph fishes with an anchored hybrid enrichment dataset. *Molecular Phylogenetics and Evolution* 110: 27–38. https://doi.org/10.1016/j.ympev.2017.02.017

Dornburg, A., Townsend, J.P., Friedman, M. & Near, T.J. 2014. Phylogenetic informativeness reconciles ray-finned fish molecular divergence times. *BMC Evolutionary Biology* 14: 169. https://doi.org/10.1186/s12862-014-0169-0

Drew, B.T., Ruhfel, B.R., Smith, S.A., Moore, M.J., Briggs, B.G., Gitzendanner, M.A., Soltis, P. & Soltis, D.E. 2014. Another look at the root of the angiosperms reveals a familiar tale. *Systematic Biology* 63: 368–382. https://doi.org/10.1093/sysbio/syt108

Drummond, C.S. 2008. Diversification of *Lupinus* (Leguminosae) in the western New World: derived evolution of perennial life history and colonization of montane habitats. *Molecular Phylogenetics and Evolution* 48: 408–421. https://doi.org/10.1016/j.ympev.2008.03.009

Edwards, S.V., Xi, Z., Janke, A., Faircloth, B.C., McCormack, J.E., Glenn, T.C., Zhong, B., Wu, S., Lemmon, E.M. & Lemmon, A.R. 2016. Implementing and testing the multispecies coalescent model: a valuable paradigm for phylogenomics. *Molecular Phylogenetics and Evolution* 94: 447–462. https://doi.org/10.1016/j.ympev.2015.10.027

Eriksson, J.S., de Sousa, F., Bertrand, Y.J., Antonelli, A., Oxelman, B. & Pfeil, B.E. 2018. Allele phasing is critical to revealing a shared allopolyploid origin of *Medicago arborea* and *M. strasseri* (Fabaceae). *BMC Evolutionary Biology* 18: 9. https://doi.org/10.1186/s12862-018-1127-z

Faircloth, B.C. 2013. Illumiprocessor: A trimmomatic wrapper for parallel adapter and quality trimming. Available at https://github.com/faircloth-lab/illumiprocessor. http://dx.doi.org/10.6079/J9ILL

Faircloth, B.C. 2015. PHYLUCE is a software package for the analysis of conserved genomic loci. *Bioinformatics* 32: 786–788. https://doi.org/10.1093/bioinformatics/btv646

Faircloth, B.C., McCormack, J.E., Crawford, N.G., Harvey, M.G., Brumfield, R.T. & Glenn T.C. 2012. Ultraconserved elements anchor thousands of genetic markers spanning multiple evolutionary timescales. *Systematic Biology* 61: 717–726. https://doi.org/10.1093/sysbio/sys004

Fér, T. & Schmickl, R.E. 2018. HybPhyloMaker: Target enrichment data analysis from raw reads to species trees. *Evolutionary Bioinformatics* 14: 1–9. https://doi.org/10.1177/1176934317742613

Folk, R.A., Mandel, J.R. & Freudenstein, J.V. 2015. A protocol for targeted enrichment of intron-containing sequence markers for recent radiations: a phylogenomic example from *Heuchera* (Saxifragaceae). *Applications in Plant Sciences* 3: 1500039. https://doi.org/10.3732/apps.1500039

Folk, R.A., Soltis, P.S., Soltis, D.E. & Guralnick, R. 2018, in press. New prospects in the detection and comparative analysis of hybridization in the tree of life. *American Journal of Botany*. https://doi.org/10.1002/ajb2.1018

Fonseca, L.H.M. & Lohmann, L.G. 2018. Combining high-throughput sequencing and targeted loci data to infer the phylogeny of the ~~"~~*Adenocalymma-Neojobertia*" clade (Bignonieae, Bignoniaceae). *Molecular Phylogenetics and Evolution* 123: 1–15. https://doi.org/10.1016/j.ympev.2018.01.023

Fragoso-Martínez, I., Salazar, G.A., Martínez-Gordillo, M., Magallón, S., Sánchez-Reyes, L., Lemmon, E.M., Lemmon, A.R., Sazatornil, F. & Mendoza, C.G. 2017. A pilot study applying the plant Anchored Hybrid Enrichment method to New World sages (*Salvia* subgenus *Calosphace*; Lamiaceae). *Molecular Phylogenetics and Evolution* 117: 124–134. https://doi.org/10.1016/j.ympev.2017.02.006

García, N., Folk, R.A., Meerow, A.W., Chamala, S., Gitzendanner, M.A., de Oliveira, R.S., Soltis, D.E. & Soltis, P.S. 2017. Deep reticulation and incomplete lineage sorting obscure the diploid phylogeny of rain-lilies and allies (Amaryllidaceae tribe Hippeastreae). *Molecular Phylogenetics and Evolution* 111: 231–247. https://doi.org/10.1016/j.ympev.2017.04.003

Garcia-Jacas, N., Garnatje, T., Susanna, A. & Vilatersana, R. 2002. Tribal and subtribal delimitation and phylogeny of the Cardueae (Asteraceae): A combined nuclear and chloroplast DNA analysis. *Molecular Phylogenetics and Evolution* 22: 51–64. https://doi.org/10.1006/mpev.2001.1038

Gates, D.J., Pilson, D. & Smith, S.D. 2018. Filtering of target sequence capture individuals facilitates species tree construction in the plant subtribe Iochrominae (Solanaceae). *Molecular Phylogenetics and Evolution* 123: 26–34. https://doi.org/10.1016/j.ympev.2018.02.002

Gatesy, J. & Springer, M.S. 2014. Phylogenetic analysis at deep timescales: unreliable gene trees, bypassed hidden support, and the coalescence/concatalescence conundrum. *Molecular Phylogenetics and Evolution* 80: 231–266. https://doi.org/10.1016/j.ympev.2014.08.013

Gernandt, D.S., Aguirre Dugua, X., Vázquez-Lobo, A., Willyard, A., Moreno Letelier, A., Pérez de la Rosa, J.A., Piñero, D. & Liston, A. 2018. Multi-locus phylogenetics, lineage sorting, and reticulation in *Pinus* subsection *Australes*. *American Journal of Botany* 105: 1–15.

https://doi.org/10.1002/ajb2.1052

Gilbert, P.S., Chang, J., Pan, C., Sobel, E.M., Sinsheimer, J.S., Faircloth, B.C. & Alfaro, M.E. 2015. Genome-wide ultraconserved elements exhibit higher phylogenetic informativeness than traditional gene markers in percomorph fishes. *Molecular Phylogenetics and Evolution* 92: 140–146. https://doi.org/10.1016/j.ympev.2015.05.027

Givnish, T.J. 2015. Adaptive radiation versus ‘radiation’ and ‘explosive diversification’: why conceptual distinctions are fundamental to understanding evolution. *New Phytologist* 207: 297–303. https://doi.org/10.1111/nph.13482

Goremykin, V.V., Nikiforova, S.V. & Bininda-Emonds, O.R. 2010. Automated removal of noisy data in phylogenomic analyses. *Journal of Molecular Evolution*: 71: 319–331. https://doi.org/10.1007/s00239-010-9398-z

Grover, C.E., Gallagher, J.P., Jareczek, J.J., Page, J.T., Udall, J.A., Gore, M.A. & Wendel, J.F. 2015. Re-evaluating the phylogeny of allopolyploid *Gossypium* L. *Molecular Phylogenetics and Evolution* 92: 45–52. https://doi.org/10.1016/j.ympev.2015.05.023

Grover, C.E., Salmon, A. & Wendel, J.F. 2012. Targeted sequence capture as a powerful tool for evolutionary analysis 1. *American Journal of Botany* 99: 312–319. https://doi.org/10.3732/ajb.1100323

Harris, R.S. 2007. *Improved pairwise alignment of genomic DNA*. PhD Thesis, Pennsylvania State University, Pennsylvania.

Heibl, C. 2008. PHYLOCH: R language tree plotting tools and interfaces to diverse phylogenetic software packages. http://www.christophheibl.de/Rpackages.html

Hillis, D.M. & Bull, J.J. 1993. An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. *Systematic Biology* 42: 182–192. https://doi.org/10.1093/sysbio/42.2.182

Hillis, D.M., Heath, T.A. & John, K.S. 2005. Analysis and visualization of tree space. *Systematic Biology* 54: 471–482. https://doi.org/10.1080/10635150590946961

Hosner, P.A., Faircloth, B.C., Glenn, T.C., Braun, E.L. & Kimball, R.T. 2015. Avoiding missing data biases in phylogenomic inference: an empirical study in the landfowl (Aves: Galliformes). *Molecular Biology and Evolution* 33: 1110–1125. https://doi.org/10.1093/molbev/msv347

Hovmöller, R., Knowles, L.L. & Kubatko, L.S. 2013. Effects of missing data on species tree estimation under the coalescent. *Molecular Phylogenetics and Evolution* 69: 1057–1062. https://doi.org/10.1016/j.ympev.2013.06.004

Hudson, C.M., Puckett, E.E., Bekaert, M., Pires, J.C. & Conant, G.C. 2011. Selection for higher gene copy number after different types of plant gene duplications. *Genome Biology and Evolution* 3: 1369–1380. https://doi.org/10.1093/gbe/evr115

Jeffroy, O., Brinkmann, H., Delsuc, F. & Philippe, H. 2006. Phylogenomics: the beginning of incongruence? *Trends in Genetics* 22: 225–231. https://doi.org/10.1016/j.tig.2006.02.003

Ješovnik, A., Sosa-Calvo, J., Lloyd, M.W., Branstetter, M.G., Fernández, F. & Schultz, T.R. 2017. Phylogenomic species delimitation and host-symbiont coevolution in the fungus-farming ant genus *Sericomyrmex* Mayr (Hymenoptera: Formicidae): ultraconserved elements (UCEs) resolve a recent radiation. *Systematic Entomology* 42: 523–542. https://doi.org/10.1111/syen.12228

Johnson, M.G., Gardner, E.M., Liu, Y., Medina, R., Goffinet, B., Shaw, A.J., Zerega, N.J.C. & Wickett, N.J. 2016. HybPiper: Extracting coding sequence and introns for phylogenetics from high-throughput sequencing reads using target enrichment. *Applications in Plant Sciences* 4: 1600016. https://doi.org/10.3732/apps.1600016

Junier, T. & Zdobnov, E.M. 2010. The Newick utilities: high-throughput phylogenetic tree processing in the UNIX shell. *Bioinformatics* 26: 1669–1670. https://doi.org/10.1093/bioinformatics/btq243

Kamneva, O.K., Syring, J., Liston, A. & Rosenberg, N.A. 2017. Evaluating allopolyploid origins in strawberries (*Fragaria*) using haplotypes generated from target capture sequencing. *BMC Evolutionary Biology* 17: 180. https://doi.org/10.1186/s12862-017-1019-7

Kates, H.R., Johnson, M.G., Gardner, E.M., Zerega, N.J. & Wickett, N.J. 2018. Allele phasing has minimal impact on phylogenetic reconstruction from targeted nuclear gene sequences in a case study of *Artocarpus*. *American Journal of Botany* 105: 404–416. https://doi.org/10.1002/ajb2.1068

Katoh, K. & Standley, D.M. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution* 30: 772–780. https://doi.org/10.1093/molbev/mst010

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P. & Drummond, A. 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28: 1647–1649. https://doi.org/10.1093/bioinformatics/bts199.

Kita, Y., Fujikawa, K., Ito, M., Ohba, H. & Kato, M. 2004. Molecular phylogenetics analyses and systematics of the genus *Saussurea* and related genera (Asteraceae, Cardueae). *Taxon* 53: 679–690. https://doi.org/10.2307/4135443

Kosakovsky Pond, S.L.K., Frost, S.D.W. & Muse, S.V. 2005. HyPhy: hypothesis testing using phylogenies. *Bioinformatics* 21: 676–679. https://doi.org/10.1093/bioinformatics/bti079Kostka, M., Uzlikova, M., Cepicka, I. & Flegr, J. 2008. SlowFaster, a user-friendly program for slow-fast analysis and its application on phylogeny of *Blastocystis*. *BMC Bioinformatics* 9: 341. https://doi.org/10.1186/1471-2105-9-341

Kubatko, L.S. & Degnan, J.H. 2007. Inconsistency of phylogenetic estimates from concatenated data under coalescence. *Systematic Biology* 56: 17–24. https://doi.org/10.1080/10635150601146041

Kück, P. & Longo, G.C. 2014. FASconCAT-G: extensive functions for multiple sequence alignment preparations concerning phylogenetic studies. *Frontiers in Zoology* 11: 81. https://doi.org/10.1186/s12983-014-0081-x

Landis, J.B., Soltis, D.E. & Soltis, P.S. 2017. Comparative transcriptomic analysis of the evolution and development of flower size in *Saltugilia* (Polemoniaceae). *BMC Genomics* 18: 475. https://doi.org/10.1186/s12864-017-3868-2

Leaché, A.D., Chavez, A.S., Jones, L.N., Grummer, J.A., Gottscho, A.D. & Linkem, C.W. 2015. Phylogenomics of phrynosomatid lizards: conflicting signals from sequence capture versus restriction site associated DNA sequencing. *Genome Biology and Evolution* 7: 706–719.

https://doi.org/10.1093/gbe/evv026

Lecointre, G., Philippe, H., Vân Lê, H.L. & Le Guyader, H. 1993. Species sampling has a major impact on phylogenetic inference. *Molecular Phylogenetics and Evolution* 2: 205–224. https://doi.org/10.1006/mpev.1993.1021

Lemmon, A.R., Emme, S.A. & Lemmon, E.M. 2012. Anchored hybrid enrichment for massively high-throughput phylogenomics. *Systematic Biology* 61: 727–744. https://doi.org/10.1093/sysbio/sys049

Lemmon, E.M. & Lemmon, A.R. 2013. High-throughput genomic data in systematics and phylogenetics. *Annual Review of Ecology, Evolution and Systematics* 44: 99–121. https://doi.org/10.1146/annurev-ecolsys-110512-135822

Léveillé-Bourret, E., Starr, J.R., Ford, B.A., Lemmon, E.M. & Lemmon, A.R. 2016. Resolving rapid radiations within angiosperm families using anchored phylogenomics. *Systematic Biology* 67: 94–112. https://doi.org/10.1093/sysbio/syx050

Li, H. & Durbin, R. 2009. Fast and accurate short read alignment with burrowswheeler transform. *Bioinformatics* 25: 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Linkem, C.W., Minin, V.N. & Leaché, A.D. 2016. Detecting the anomaly zone in species trees and evidence for a misleading signal in higher-level skink phylogeny (Squamata: Scincidae). *Systematic Biology* 65: 465–477. https://doi.org/10.1093/sysbio/syw001

Lipschitz, S.J. 1979. *Genus Saussurea DC (Asteraceae)* [in Russian]. Nauka, Leningrad.

Liu, L., Wu, S. & Yu, L. 2015a. Coalescent methods for estimating species trees from phylogenomic data. *Journal of Systematics and Evolution* 53: 380–390. https://doi.org/10.1111/jse.12160

Liu, L., Xi, Z., Wu, S., Davis, C.C. & Edwards, S.V. 2015b. Estimating phylogenetic trees from genome-scale data. *Annals of the New York Academy of Sciences* 1360: 36–53. https://doi.org/10.1111/nyas.12747

Longo, S.J., Faircloth, B.C., Meyer, A., Westneat, M.W., Alfaro, M.E. & Wainwright, P.C. 2017. Phylogenomic analysis of a rapid radiation of misfit fishes (Syngnathiformes) using ultraconserved elements. *Molecular Phylogenetics and Evolution* 113: 33–48. https://doi.org/10.1016/j.ympev.2017.05.002

López-Giráldez, F. & Townsend, J.P. 2011. PhyDesign: an online application for profiling phylogenetic informativeness. *BMC Evolutionary Biology* 11: 152. https://doi.org/10.1186/1471-2148-11-152

López-Vinyallonga, S., Mehregan, I., Garcia-Jacas, N., Tscherneva, O., Susanna, A. & Kadereit, J.W. 2009. Phylogeny and evolution of the *Arctium-Cousinia* complex (Compositae, Cardueae-Carduinae). *Taxon* 58: 153–171.

López-Vinyallonga, S., Romaschenko, K., Susanna, A. & Garcia-Jacas, N. 2011. Systematics of the arctioid group: disentangling *Arctium* and *Cousinia* (Cardueae, Carduinae). *Taxon* 60: 539–554.

Maddison, W.P. & Knowles, L.L. 2006. Inferring phylogeny despite incomplete lineage sorting. *Systematic Biology* 55: 21–30. https://doi.org/10.1080/10635150500354928

Mamanova, L., Coffey, A.J., Scott, C.E., Kozarewa, I., Turner, E.H., Kumar, A., Howard, E., Shendure, J. & Turner, D.J. 2009. Target-enrichment strategies for next-generation sequencing. *Nature Methods* 7: 111–118. https://doi.org/10.1038/nmeth.1419

Mandel, J.R., Barker, M.S., Bayer, R.J., Dikow, R.B., Gao, T., Jones, K.E., Keely, S., Kilian, N., Ma, H., Siniscalchi, C.M., Susanna, A., Thapa, R., Watson, L. & Funk, V.A. 2017. The Compositae Tree of Life in the age of phylogenomics. *Journal of Systematics and Evolution* 55: 405–403. https://doi.org/10.1111/jse.12265

Mandel, J.R., Dikow, R.B. & Funk, V.A. 2015. Using phylogenomics to resolve mega-families: An example from Compositae. *Journal of Systematics and Evolution* 53: 391–402. https://doi.org/10.1111/jse.12167

Mandel, J.R., Dikow, R.B., Funk, V.A., Masalia, R.R., Staton, S.E., Kozik, A., Michelmore, R.W., Rieseberg, L.H. & Burke, J.M. 2014. A target enrichment method for gathering phylogenetic information from hundreds of loci: An example from the Compositae. *Applications in Plant Sciences* 2: 1300085. https://doi.org/10.3732/apps.1300085

McCormack, J.E., Huang, H. & Knowles, L.L. 2009. Maximum likelihood estimates of species trees: how accuracy of phylogenetic inference depends upon the divergence history and sampling design. *Systematic Biology* 58: 501–508. https://doi.org/10.1093/sysbio/syp045

Medina, R., Johnson, M., Liu, Y., Wilding, N., Hedderson, T.A., Wickett, N. & Goffinet, B. 2018. Evolutionary dynamism in bryophytes: Phylogenomic inferences confirm rapid radiation in the moss family Funariaceae. *Molecular Phylogenetics and Evolution* 120: 240–247. https://doi.org/10.1016/j.ympev.2017.12.002

Mehregan, I. & Assadi, M. 2016. A synopsis of *Cousinia* sect. *Pseudactinia* (Cardueae, Asteraceae) including a new species from NE Iran. *Phytotaxa* 257: 271–279. https://doi.org/10.11646/phytotaxa.257.3.5

Mehregan, I. & Kadereit, J.W. 2009. The role of hybridization in the evolution of *Cousinia* s.str. (Asteraceae, Cardueae). *Willdenowia* 39: 35–47. https://doi.org/10.3372/wi.39.39102

Meiklejohn, K.A., Faircloth, B.C., Glenn, T.C., Kimball, R.T. & Braun, E.L. 2016. Analysis of a rapid evolutionary radiation using ultraconserved elements: evidence for a bias in some multispecies coalescent methods. *Systematic Biology* 65: 612–627. https://doi.org/10.1093/sysbio/syw014

Mendes, F.K. & Hahn, M.W. 2018. Why concatenation fails near the anomaly zone. *Systematic Biology* 67: 158–169. https://doi.org/10.1093/sysbio/syx063

Meyer, M. & Kircher, M. 2010. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harbor Protocols* 2010: pdb.prot5448. https://doi.org/10.1101/pdb.prot5448

Miller, M.A., Pfeiffer, W. & Schwartz, T. 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In: *Proceedings of the Gateway Computing Environments Workshop (GCE), 14 Nov. 2010*. New Orleans: 1–8. https://doi.org/10.1109/GCE.2010.5676129

Mirarab, S., Bayzid, M.S., Boussau, B. & Warnow, T. 2014. Statistical binning enables an accurate coalescent-based estimation of the avian tree. *Science* 346: 1250463. https://doi.org/10.1126/science.1250463

Misof, B., Meyer, B., von Reumont, B.M., Kück, P., Misof, K. & Meusemann, K. 2013. Selecting informative subsets of sparse supermatrices increases the chance to find correct trees. *BMC Bioinformatics* 14: 348. https://doi.org/10.1186/1471-2105-14-348

Mitchell, N., Lewis, P.O., Lemmon, E.M., Lemmon, A.R. & Holsinger, K.E. 2017. Anchored phylogenomics improves the resolution of evolutionary relationships in the rapid radiation of *Protea* L. *American Journal of Botany* 104: 102–115. https://doi.org/10.3732/ajb.1600227

Moyle, R.G., Oliveros, C.H., Andersen, M.J., Hosner, P.A., Benz, B.W., Manthey, J.D., Travers, S.L., Brown, R.M & Faircloth, B.C. 2016. Tectonic collision and uplift of Wallacea triggered the global songbird radiation. *Nature communications* 7: 12709. https://doi.org/10.1038/ncomms12709

Nicholls, J.A., Pennington, R.T., Koenen, E.J., Hughes, C.E., Hearn, J., Bunnefeld, L., Dexter, K.G., Stone, G.N. & Kidner, C.A. 2015. Using targeted enrichment of nuclear genes to increase phylogenetic resolution in the neotropical rain forest genus *Inga* (Leguminosae: Mimosoideae). *Frontiers in Plant Science* 6: 710. https://doi.org/10.3389/fpls.2015.00710

Oliver, J.C. 2013. Microevolutionary processes generate phylogenomic discordance at ancient divergences. *Evolution* 67: 1823-1830. https://doi.org/10.1111/evo.12047

Paradis, E., Claude, J. & Strimmer, K. 2004. APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* 20: 289–290. https://doi.org/10.1093/bioinformatics/btg412

Parks, M., Cronn, R. & Liston, A. 2012. Separating the wheat from the chaff: mitigating the effects of noise in a plastome phylogenomic data set from *Pinus* L. (Pinaceae). *BMC Evolutionary Biology* 12: 100. https://doi.org/10.1186/1471-2148-12-100

Philippe, H. & Forterre, P. 1999. The rooting of the universal tree of life is not reliable. *Journal of Molecular Evolution* 49: 509–523. https://doi.org/10.1007/PL00006573

Philippe, H., Brinkmann, H., Lavrov, D.V., Littlewood, D.T.J., Manuel, M., Wörheide, G. & Baurain, D. 2011. Resolving difficult phylogenetic questions: why more sequences are not enough. *PLoS Biology* 9: e1000602. https://doi.org/10.1371/journal.pbio.1000602

Prum, R.O., Berv, J. S., Dornburg, A., Field, D.J., Townsend, J.P., Lemmon, E.M. & Lemmon, A.R. 2015. A comprehensive phylogeny of birds (Aves) using targeted next-generation DNA sequencing. *Nature* 526: 569. https://doi.org/10.1038/nature15697

R Core Team. 2014. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna.

Raab-Straube, E.V. 2003. Phylogenetic relationships in *Saussurea* (Compositae, Cardueae) sensu lato, inferred from morphological, ITS and trn L-trn F sequence data, with a synopsis of *Himalaiella* gen. nov., *Lipschitziella* and *Frolovia*. *Willdenowia* 33: 379–402. https://doi.org/10.3372/wi.33.33214

Rambaut, A. 2002. *TreeEdit, Version 1.0a10*. Evolutionary Biology Group, University of Oxford, Oxford. Available at http://evolve.zoo.ox.ac.uk

Rambaut, A. 2016. *FigTree ver. 1.4.3*. Department of Zoology, University of Oxford, Oxford. Available at http://tree.bio.ed.ac.uk/software/figtree/

Rensing, S.A. 2014. Gene duplication as a driver of plant morphogenetic evolution. *Current opinion in Plant Biology* 17: 43–48. https://doi.org/10.1016/j.pbi.2013.11.002

Robinson, D.F. & Foulds, L.R. 1981. Comparison of phylogenetic trees. *Mathematical Biosciences* 53: 131–147. https://doi.org/10.1016/0025-5564(81)90043-2

Rokas, A. & Carroll, S.B. 2006. Bushes in the tree of life. *PLoS Biology* 4: e352. https://doi.org/10.1371/journal.pbio.0040352

Salichos, L. & Rokas, A. 2013. Inferring ancient divergences requires genes with strong phylogenetic signals. *Nature* 497: 327–331. https://doi.org/10.1038/nature12130

Salmerón-Sánchez, E., Pérez-García, F.J., Medina-Cazorla, J.M., Martínez-Nieto, M.I., Martínez-Hernández, F., Garrido-Becerra, J.A., Mendoza-Fernández, A.J., Merlo Calvente, M.E. & Poveda, J.M. 2015. Genetic analysis based on plastidial and ribosomal sequences of the endemic bi-edaphic taxon *Jurinea pinnata* (Lag.) DC. (Compositae) in the Guadix-Baza Basin. *Plant Biosystems* 149: 922–932. https://doi.org/10.1080/11263504.2014.983203

Sanderson, M.J. 1997. A nonparametric approach of rate constancy to estimating divergence times in the absence of rate constancy. *Molecular Biology and Evolution* 14: 1218–1231.

Sass, C., Iles, W.J., Barrett, C.F., Smith, S.Y. & Specht, C.D. 2016. Revisiting the Zingiberales: using multiplexed exon capture to resolve ancient and recent phylogenetic splits in a charismatic plant lineage. *PeerJ* 4: e1584. https://doi.org/10.7717/peerj.1584

Sayyari, E. & Mirarab, S. 2016. Fast coalescent-based computation of local branch support from quartet frequencies. *Molecular Biology and Evolution* 33: 1654–1668. https://doi.org/10.1093/molbev/msw079

Schmickl, R., Liston, A., Zeisek, V., Oberlander, K., Weitemier, K., Straub, S.C., Cronn, R.C., Dreyer, L.L. & Suda, J. 2016. Phylogenetic marker development for target enrichment from transcriptome and genome skim data: the pipeline and its application in southern African *Oxalis* (Oxalidaceae). *Molecular Ecology Resources* 16: 1124–1135. https://doi.org/10.1111/1755-0998.12487

Shi, Z., Raab-Straube, E.V., Greuter, W. & Martins, L. 2011. *Cardueae*. In: Wu, Z.Y., Raven, P.H. & Hong, D.Y. (eds.), *Flora of China*, vol. 20–21 (Asteraceae). Science Press and Missouri Botanical Garden Press, Beijing and St. Louis: 42–194.

Simmons, M.P., Sloan, D.B. & Gatesy, J. 2016. The effects of subsampling gene trees on coalescent methods applied to ancient divergences. *Molecular Phylogenetics and Evolution* 97: 76–89. https://doi.org/10.1016/j.ympev.2015.12.013

Stamatakis, A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22: 2688–2690. https://doi.org/10.1093/bioinformatics/btl446

Stamatakis, A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30: 1312–1313. https://doi.org/10.1093/bioinformatics/btu033

Stamatakis, A., Hoover, P. & Rougemont, J. 2008. A rapid bootstrap algorithm for the RAxML web servers. *Systematic Biology* 57: 758–771. https://doi.org/10.1080/10635150802429642

Stebbins, G.L. 1970. Adaptive radiation of reproductive characteristics in angiosperms, I: pollination mechanisms. *Annual Review of Ecology and Systematics* 1: 307–326.

Steel, M.A. & Penny, D. 1993. Distributions of tree comparison metrics—some new results. *Systematic Biology* 42: 126–141.

https://doi.org/10.2307/2992536

Stephens, J.D., Rogers, W.L., Mason, C.M., Donovan, L.A. & Malmberg, R.L. 2015. Species tree estimation of diploid *Helianthus* (Asteraceae) using target enrichment. *American Journal of Botany* 102: 910–920. https://doi.org/10.3732/ajb.1500031

Straub, S.C., Moore, M.J., Soltis, P.S., Soltis, D.E., Liston, A. & Livshultz, T. 2014. Phylogenetic signal detection from an ancient rapid radiation: Effects of noise reduction, long-branch attraction, and model selection in crown clade Apocynaceae. *Molecular Phylogenetics and Evolution* 80: 169–185. https://doi.org/10.1016/j.ympev.2014.07.020

Streicher, J.W., Miller, E.C., Guerrero, P.C., Correa, C., Ortiz, J.C., Crawford, A.J., Pie, M.R. & Wiens, J.J. 2018. Evaluating methods for phylogenomic analyses, and a new phylogeny for a major frog clade (Hyloidea) based on 2214 loci. *Molecular Phylogenetics and Evolution* 119: 128–143. https://doi.org/10.1016/j.ympev.2017.10.013

Streicher, J.W., Schulte, J.A. & Wiens, J.J. 2016. How should genes and taxa be sampled for phylogenomic analyses with missing data? An empirical study in iguanian lizards. *Systematic Biology* 65: 128–145. https://doi.org/10.1093/sysbio/syv058

Struck, T.H. 2014. TreSpEx--Detection of Misleading Signal in Phylogenetic Reconstructions Based on Tree Information. *Evolutionary Bioinformatics* 10: 51–67. https://doi.org/10.4137/EBO.S14239

Stubbs, R.L., Folk, R.A., Xiang, C.L., Soltis, D.E. & Cellinese, N. 2018. Pseudo-parallel patterns of disjunctions in an Arctic-alpine plant lineage. *Molecular Phylogenetics and Evolution* 123: 88–100. https://doi.org/10.1016/j.ympev.2018.02.016

Susanna, A. & Garcia-Jacas, N. 2007. Tribe Cardueae. In: Kadereit, J.W. & Jeffrey, C. (eds.), *Flowering Plants. Eudicots. Asterales*, vol. 8. In: Kubitzki, K. (ed.), *The families and genera of vascular plants*. Springer Verlag, Berlin: 123–146.

Susanna, A. & Garcia-Jacas, N. 2009. Cardueae (Carduoideae). In: Funk, V.A., Susanna, A., Stuessy, T.F. & Bayer, R.J. (eds.), *Systematics, evolution, and biogeography of Compositae*. IAPT, Vienna: 293–313.

Susanna, A., Garcia-Jacas, N., Hidalgo, O., Vilatersana, R. & Garnatje, T. 2006. The Cardueae (Compositae) revisited: Insights from ITS, *trn*L-*trn*F, and *mat*K nuclear and chloroplast DNA analysis. *Annals of the Missouri Botanical Garden* 93: 150–171. https://doi.org/10.3417/0026-6493(2006)93[150:TCCRIF]2.0.CO;2

Susanna, A., Garcia-Jacas, N., Vilatersana, R. & Garnatje, T. 2003. Generic boundaries and evolution of characters in the *Arctium* group: a nuclear and chloroplast DNA analysis. *Collectanea Botanica* 26: 101–118. https://doi.org/10.3989/collectbot.2003.v26.17

Swofford, D.L. 2003. PAUP*: phylogenetic analysis using parsimony, version 4.0 b10. Sinauer, Sunderland.

Syring, J.V., Tennessen, J.A., Jennings, T.N., Wegrzyn, J., Scelfo-Dalbey, C. & Cronn, R. 2016. Targeted capture sequencing in whitebark pine reveals range-wide demographic and adaptive patterns despite challenges of a large, repetitive genome. *Frontiers in Plant Science* 7: 484. https://doi.org/10.3389/fpls.2016.00484

Townsend, J.P., López-Giráldez, F. & Friedman, R. 2008. The phylogenetic informativeness of nucleotide and amino acid sequences for reconstructing the vertebrate tree. *Journal of Molecular Evolution* 67: 437–447. https://doi.org/10.1007/s00239-008-9142-0

Townsend, J.P., Su, Z. & Tekle, Y.I. 2012. Phylogenetic signal and noise: predicting the power of a data set to resolve phylogeny. *Systematic Biology* 61: 835–849. https://doi.org/10.1093/sysbio/sys036

Tripp, E.A., Tsai, Y.H.E., Zhuang, Y. & Dexter, K.G. 2017. RADseq dataset with 90% missing data fully resolves recent radiation of *Petalidium* (Acanthaceae) in the ultra-arid deserts of Namibia. *Ecology and Evolution* 7: 7920–7936. https://doi.org/10.1002/ece3.3274

Tscherneva, O.V. 1997. Genus 1578. *Cousinia* Cass. [translated from Russian]. In: Schischkin, B.K. & Bobrov, E.G. (eds.), *Flora SSSR*. Bishen Singh Mahendra Pal Singh and Koeltz Scientific Books, Dehra Dun: 135–442.

Van Dam, M.H., Lam, A.W., Sagata, K., Gewa, B., Laufa, R., Balke, M., Faircloth, B.C. & Riedel, A. 2017. Ultraconserved elements (UCEs) resolve the phylogeny of Australasian smurf-weevils. *PLoS ONE* 12: e0188044. https://doi.org/10.1371/journal.pone.0188044

Vatanparast, M., Powell, A., Doyle, J.J. & Egan, A.N. 2018. Targeting legume loci: A comparison of three methods for target enrichment bait design in Leguminosae phylogenomics. *Applications in Plant Sciences* 6: e1036. https://doi.org/10.1002/aps3.1036

Wagner, C.E., Keller, I., Wittwer, S., Selz, O.M., Mwaiko, S., Greuter, L., Sivasundar, A. & Seehausen, O. 2013. Genome-wide RAD sequence data provide unprecedented resolution of species boundaries and relationships in the Lake Victoria cichlid adaptive radiation. *Molecular Ecology* 22: 787–798. https://doi.org/10.1111/mec.12023

Wang, Y. J., Raab-Straube, E.V., Susanna, A. & Liu, J.Q. 2013. *Shangwua* (Compositae), a new genus from the Qinghai-Tibetan Plateau and Himalayas. *Taxon* 62: 984–996. https://doi.org/10.12705/625.19

Wang, Y.J., Liu, J.Q. & Miehe, G. 2007. Phylogenetic origins of the Himalayan endemic *Dolomiaea*, *Diplazoptilon* and *Xanthopappus* (Asteraceae: Cardueae) based on three DNA regions. *Annals of Botany* 99: 311–322. https://doi.org/10.1093/aob/mcm284

Wang, Y.J., Susanna, A., Raab-Straube, E.V., Milne, R. & Liu, J.Q. 2009. Island-like radiation of *Saussurea* (Asteraceae: Cardueae) triggered by uplifts of the Qinghai–Tibetan Plateau. *Biological Journal of the Linnean Society* 97: 893–903. https://doi.org/10.1111/j.1095-8312.2009.01225.x

Wanke, S., Mendoza, C.G., Müller, S., Guillén, A.P., Neinhuis, C., Lemmon, A.R., Lemmon, E.M. & Samain, M.S. 2017. Recalcitrant deep and shallow nodes in *Aristolochia* (Aristolochiaceae) illuminated using anchored hybrid enrichment. *Molecular Phylogenetics and Evolution* 117: 111–123. https://doi.org/10.1016/j.ympev.2017.05.014

Ward, P.S. & Branstetter, M.G. 2017. The acacia ants revisited: convergent evolution and biogeographic context in an iconic ant/plant mutualism. *Proceedings of the Royal Society B* 284: 20162569. https://doi.org/10.1098/rspb.2016.2569

Watanabe, K. 2002. *Index to chromosome numbers in Asteraceae*. Available at http://www.lib.kobe-u.ac.jp/infolib/meta_pub/G0000003asteraceae_e

Weitemier, K., Straub, S.C., Cronn, R.C., Fishbein, M., Schmickl, R., McDonnell, A. & Liston, A. 2014. Hyb-Seq:

Combining target enrichment and genome skimming for plant phylogenomics. *Applications in Plant Sciences* 2: 1400042. https://doi.org/10.3732/apps.1400042

Wen, J., Ree, R.H., Ickert-Bond, S.M., Nie, Z. & Funk, V. 2013. Biogeography: Where do we go from here? *Taxon* 62: 912–927. https://doi.org/10.12705/625.15

Wen, J., Zhang, J.-Q., Nie, Z.-L., Zhong Y. & Sun, H. 2014. Evolutionary diversifications of plants on the Qinghai-Tibetan Plateau. *Frontiers in Genetics* 5: 4. https://doi.org/10.3389/fgene.2014.00004

Whitfield, J.B. & Lockhart, P.J. 2007. Deciphering ancient rapid radiations. *Trends in Ecology and Evolution* 22: 258–265. https://doi.org/10.1016/j.tree.2007.01.012

Xi, Z., Liu, L., Rest, J.S. & Davis, C.C. 2014. Coalescent versus concatenation methods and the placement of *Amborella* as sister to water lilies. *Systematic Biology* 63: 919–932. https://doi.org/10.1093/sysbio/syu055

Zhang, C., Rabiee, M., Sayyari, E. & Mirarab, S. 2018. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC bioinformatics* 19: 153. https://doi.org/10.1186/s12859-018-2129-y

7. Appendix A

All alignments and tree files for each dataset are deposited in Mendeley Data repository (http://dx.doi.org/10.17632/hgpn6g27c6.1).

8. Supplementary material

Table S1. Species sampled, authority and collection information.

Species	Authority	Location and voucher
Alfredia acantholepis	Kar. & Kir.	Kazakhstan: Almatinskaya oblast, Alatau mt. above Almaty, *Susanna 2092 et al.* (BC)
Arctium abolinii	(Kult. ex Tscherneva) S. López , Romasch., Susanna & N. Garcia	Kyrgyzstan: SW, Jalal Abad Oblast, Kara Saj Tal, Aksy Rayan, *Lazkov s.n.* (LE)
Arctium arctioides	(Schrenk.) Kuntze	Kazakhstan: Dzhezkazganskaya reg., Turgajskaya lowland, 49 km to SW from Dzhezkazgana, *Kamelin 6434* (LE)
Arctium aureum	(C. Winkl.) Kuntze	Tajikistan: Schtut, road to Penjikent, *Susanna 2514 et al.* (BC)
Arctium egregium	(Juz.) S. López , Romasch., Susanna & N. Garcia	Uzbekistan: Angren valley, rise to Kamchik pass, *Kamelin 420* (LE)
Arctium eriophorum	(Regel & Schmalh.) Kuntze	Kazakhstan: Almatinskaya oblast, Alatau mt., above Almaty, *Susanna 2088 et al.* (BC)
Arctium fedtschenkoanum	(Bornm.) S. López , Romasch., Susanna & N. Garcia	Tajikistan: *Romaschenko 632 & Susanna* (BC)
Arctium grandifolium	(Kult.) S. López , Romasch., Susanna & N. Garcia	Kazakhstan: Zambylskaya oblast, Talaski Alatau, 6 km W from Il Tai, *Susanna 2181 et al.* (BC)
Arctium karatavicum	(Regel & Schmalh.) Kuntze	Tadjikistan: *Romaschenko 607 & Susanna* (BC)
Arctium leiospermum	Juz. & Ye. V. Serg.	Kazakhstan: Zambylskaya oblast, Kurdai pass, *Susanna 2154 et al.* (BC)
Arctium minus	(Hill) Bernh.	Spain: Barcelona, Dosrius, Canyamars, dry riverbed of Canyamars, *Vilatersana 1100 & López-Vinyallonga* (BC)
Arctium umbrosum	Kuntze	Kazakhstan: Almatinskaya oblast, Alatau mt. above Almaty, *Susanna 2100 et al.* (BC)
Carduus pycnocephalus	L.	Spain: Barcelona, Montjuïc, near the Botanic Institute of Barcelona, *Garnatje & Susanna 1827* (BC)
Cirsium sairamense	O. Fedtsch. & B. Fedtsch.	Tadjikistan: Maijora canyon near Ziddi, *Susanna 2468 et al.* (BC)
Cousinia albertoregelia	C. Winkl.	Tadjikistan: Tujuntau mountains, west of Shaaftuz, *Botschantzev 166* (LE)
Cousinia armena	Takht.	Armenia: Kotayk, Abovian, *Vitek 03-07458 et al.* (BC)
Cousinia badghysi	Kult.	Turkmenistan: Badghys, Eroylanduz lake, SE –sopki" Kazan, *Kamelin 360* (LE)
Cousinia brachyptera	DC.	Armenia: Shirak province, Talin district, Arteni mountain, *Tamanian s.n.* (ERE)
Cousinia coerulea	Kult. ex Tscherneva	Tajikistan: Vorzov canyon, *Susanna 2459 et al.* (BC)
Cousinia fetissowii	C. Winkl.	Kyrgyzstan: Kirghizia, Kyrgyz Ala-Too, upper reaches of Nyldy River, *Sudnitsyna & Gorbunova s.n.* (FRU)
Cousinia franchetii	C. Winkl.	Tadjikistan: near kishlag Zimargh, *Susanna 2498 et al.* (BC)
Cousinia knorringiae	Bornm.	Kyrgyzstan: Bozbu-Too, 21.V.1970, *Sudnitsyna s.n.* (FRU, LE)
Cousinia macroptera	C. A. Mey. ex DC.	Armenia: Ararat province, Ashtarak distr., Kahtsrashen village, *Tamanian s.n.* (ERE)
Cousinia ninae	Juz.	Kyrgyzstan: Oshskaya, Torgulsky reg., Oitaya area north from Shoporovo vil., *Sultanova s.n.* (LE)
Cousinia onopordioides	Ledeb.	Iran: Tehran, between Firuzkuh and Semnan, *Susanna 1637 et al.* (BC)
Cousinia polytimetica	Tscherneva	Uzbekistan: Bukharskaya reg., Zeravshan river, to SE from Uzlishkent vil., *Kryakin s.n.* (LE)
Cousinia pusilla	C. Winkl.	Tajikistan: S Tajikistan, from Besharcha mts. to Babatag range, *Botschantzev 117* (LE)
Cousinia schischkinii	Juz.	Kyrgyzstan: Kara-Suu Lake, Nura River, 17.VI.1973, *Aydarova et al. s.n.* (FRU)
Cousinia serawschanica	C. Winkl.	Tajikistan: Voru, *Susanna 2516 et al.* (BC)
Cousinia sewertzowii	Regel	Kazakhstan: Aksu-Dzabagly reservation, *Susanna 2178 et al.* (BC)

Species	Authority	Location and voucher
Cousinia sogdiana	Bornm.	Uzbekistan: Karakalpakstan, 27 km from Nukus, *Kalibernova 5262 et al.* (LE)
Cousinia splendida	C. Winkl.	Tadjikistan: Fan mountains, road above Iskandar-Kul, *Susanna 2507 et al.* (BC)
Cousinia spryginii	Kult.	Uzbekistan: Kashkadarbinskaya reg., low mountains to SE of vil. Dekhanabad, *Botschantzev 46* (LE)
Cousinia strobilocephala	Tschern. & Vved.	Kyrgyzstan: Kirghizia, Qurama Range, Kayyndy-Say River, *Aydarova & Chypaev s.n.* (FRU)
Cousinia tenella	Fisch. & C. A. Mey.	Iran: Golestan Nat. Park, between Sharlegh and Cheshmeh Khan, *Akhani 243* (MJG)
Cousinia tianschanica	Kult.	Kazakhstan: Shimkientskaya oblast, Aksu Dzabagly reservation, Aksu canyon, *Susanna 2191 et al.* (BC)
Cynara cardunculus	L.	United States of America: Greenhouse grown seed, collected UW Medicinal Plant Garden, *Mandel s.n.* (GA 135)
Jurinea abramowii	Regel & Herder	Tadjikistan: Hissar Mt., *Smirnova 224 et al.* (DUSH)
Jurinea alata	(Desf.) Cass.	Culta in Horto Botanico Barcinonense (BC)
Jurinea algida	Iljin	Kyrgyzstan: Kok-Suu River, 16.VIII.2006, *Lazkov s.n.* (FRU)
Jurinea atropurpurea	C. Winkl. ex Iljin	Tadjikistan: sine loc, *Kotehkariova & Zhogolieva 16094* (DUSH)
Jurinea baldschuanica	C. Winkl.	Tadjikistan: mountains above Kara-Chuiráá, *Susanna 2561 et al.* (BC)
Jurinea bucarica	C. Winkl.	*Sine loc.,* 22.IV.1975, *sine col. 10387* (DUSH)
Jurinea caespitans	Iljin	Kyrgyzstan: north of Kara-Jygach village, 09.VII.2016, *Sennikov s.n.* (H)
Jurinea capusii	Franch.	Kyrgyzstan: Chapchyma-Say, 14.VII.2016, *Sennikov s.n.* (H)
Jurinea carduiformis	(Jaub. & Spach) Boiss.	Iran: Tehran, near Sorkhehesar, *Susanna 1631 et al.* (BC)
Jurinea ferganica	(Iljin) Iljin	Kyrgyzstan: near Kadamzhay village, 18.VII.2016, *Sennikov s.n.* (H)
Jurinea fontqueri	Cuatrec.	Spain: Jaén, cerro Cárceles, Mágina, *Martínez Lirola s.n.* (GDA 44615)
Jurinea kokanica	Iljin	Kyrgyzstan: 15 km E of Kosh-Bulak village, 09.V.2007, *Ganybaeva s.n.* (FRU)
Jurinea kyzylkyrensis	Kamelin & Tscherneva	Kyrgyzstan: left side of Naryn River, Kyzyl-Kyr, 12.VIII.1979, *Botschantzev et al. s.n.* (FRU)
Jurinea lanipes	Rupr.	Kyrgyzstan: Boom ravine, *Sennikov 428a* (H) [locus classicus of *Jurinea abolinii* Iljin]
Jurinea leptoloba	DC.	Iran: 30 km N from Tabriz, *Susanna 1654 et al.* (BC)
Jurinea macrocephala	DC.	Iran: 20 Km N of Qarabchaman, *Susanna 1650 et al.* (BC)
Jurinea narynensis	Kamelin & Tscherneva	Kyrgyzstan: 8 km from Tash-Kumyr to Jangi-Jol, *Lazkov & Omuralieva 11* (FRU)
Jurinea olgae	Ragel & Schmalh.	Tadjikistan: slopes over kishlag Voru, *Susanna 2517 et al.* (BC)
Jurinea orientalis	(Iljin) Iljin	Kyrgyzstan: near Shekoftar village, 13.VII.2016, *Sennikov s.n.* (H)
Jurinea pinnata	DC.	Morocco: Meknès-Tafilalt, Middle-Atlas, from Midelt to Timahdite, col du Zad, *Calleja & Hipold 20103091* (BC)
Jurinea popovii	Iljin	Tadjikistan: *sine loc*., *Chukavina et al. 163(86)* (DUSH)
Jurinea schachimardanica	Iljin	Kyrgyzstan: *sine loc*., 2016, *Sennikov s.n.* (H)
Jurinea stenophylla	Iljin	Kyrgyzstan: Kasan-Say River near Terek-Say village, 14.VI.1996, *Pimenov et al. s.n.* (FRU)
Jurinea stoechadifolia	(M. Bieb.) DC.	Ukraine: Crimea, *Romo 10321 et al.* (BC)
Jurinea suffruticosa	Regel	Kyrgyzstan: Kasan-Say River, 14.VII.2016, *Sennikov s.n.* (H)
Jurinea thianschanica	Regel & Schmalh.	Kyrgyzstan: between Kochkor and Ottuk, near Orto-Tokoy village, 03.VII.2016, *Sennikov s.n.* (H)
Jurinea trautvetteriana	Regel & Schmalh.	Tadjikistan: *sine loc*., *Ovczinnikov 16305 & Zaprjagaeva* (DUSH)
Jurinea winkleri	Iljin	Kyrgyzstan: east of Uch-Korgon village, 16.VII.2016, *Sennikov s.n.* (H)
Modestia darwasica	(C. Winkl.) Kharadze & Tamamsch.	Kyrgyzstan: 20 km NW of Samarkandyk, Kyzyl-Suu, 10.V.1978, *Aidarova & Ubukeeva s.n.* (FRU)
Jurinea xeranthemoides	Iljin	Kyrgyzstan: near Uch-Korgon village, 16.VII.2016, *Sennikov s.n.* (H)
Olgaea petriprimi	B. A. Sharipova	Tadjikistan: Kishlag Selandi, *Susanna 2539 et al.* (BC)
Saussurea carduicephala	(Iljin) Iljin	Tajikistan: Gorno-Badakhshan, Shughnon, Shughnonskii Ridge, *Semakov & Dengubenko s.n.* (LE 8428)
Saussurea controversa	DC.	Russia: Krasnoyarsk Krai, Sharypovsky, village Bolshoe ozero, *Cazzolla Gatti 10005 et al.* (TK t-01-2016)
Saussurea davurica	Adams	Russia: Altai, Kosh-Agachsky, Kuraiskiy Ridge, village Chagan-Usun, *A. Pyak & E. Pyak 11049* (TK a-067-2016)
Saussurea elegans	Ledeb.	Tadjikistan: Iskandar valley, Fan mountains, *Susanna 2505 et al.* (BC)

Species	Authority	Location and voucher
Saussurea foliosa	Ledeb.	Russia: Khakassia, Tashtypsky, Sayanskii Mountain Pass, *Cazzolla Gatti 10025 et al.* (TK t-30-2016)
Saussurea glacialis	Herder	Russia: Altai, Kosh-Agachsky, Kuraiskiy Ridge, *A. Pyak & E. Pyak 11021* (TK a-043-2016)
Saussurea jadrinzevii	Krylov	Russia: Altai, Ongudaysky, the Mount Belyy Bom, *A. Pyak & E. Pyak 11005* (TK a-023-2016)
Saussurea krylovii	Schischk. & Serg.	Russia: Altai, Kosh-Agachsky, Juzhno-Chuysky Ridge, the Jazator River Valley, *A. Pyak & E. Pyak 11079* (TK a-108-2016)
Saussurea larionowii	C. Winkl.	Kyrgyzstan: *sine loc*., *Ovczinnikov 16* (DUSH)
Saussurea latifolia	Ledeb.	Russia: Krasnoyarsk Krai, Yermakovsky, Ergaki Ridge, *A. Pyak & E. Pyak 10009* (TK t-02-2016)
Saussurea leptophylla	Hemsl.	Afghanistan: Kapisa, *Podlech 12500* (W)
Saussurea leucophylla	Schrenk	Russia: Altai, Kosh-Agachsky, northern spurs of the Mount Tjepliy Kljuch, *A. Pyak & E. Pyak 11073* (TK a-102-2016)
Saussurea manshurica	Kom.	Russia: Amur province, 02.VIII.1979, *Boyko & Starchenko s.n.* (LE)
Saussurea orgaadayi	Khanm. & Krasnob.	Russia: Altai, Kosh-Agachsky, Kuraiskiy Ridge, the Kokorja River Valley, *A. Pyak & E. Pyak 11083* (TK a-119-2016)
Saussurea sp.	N. D. Simpson	Russia: Altai, Kosh-Agachsky, Kuraiskiy Ridge, village Chagan-Usun, near Lake Balhash, *A. Pyak & E. Pyak 11044* (TK a-065-2016)
Saussurea pseudoalpina	Simps.	Russia: Altai, Kosh-Agachsky, Kuraiskiy Ridge, the Ortolyk River, *A. Pyak & E. Pyak 11032* (TK a-048-2016)
Saussurea salicifolia	(L.) DC.	Russia: Tyva, Kaa-Khemsky, the Mount Ondum, the Kaa-Khem River, *A. Pyak & E. Pyak 10014* (TK t-12-2016)
Saussurea salsa	(Pall. ex Pall.) Spreng.	Russia: Altai, Kosh-Agachsky, Chuya Steppe, village Aktal, *A. Pyak & E. Pyak 11087* (TK a-120-2016)
Saussurea schanginiana	Fisch. ex Herder	Russia: Khakassia, Tashtypsky, Sayanskii Ridge, Sayanskii Mountain Pass, *A. Pyak & E. Pyak 10057* (TK t-24-2016)
Saussurea stubendorffii	Herder	Russia: Tyva, Barun-Khemchiksky, Sayanskii Ridge, Ak-sug River Valley, *A. Pyak & E. Pyak 10057* (TK t-24-2016)
Saussurea subacaulis	(Ledeb.) Serg.	Russia: Altai, Kosh-Agachsky, Kuraiskiy Ridge, Ortolyk River, *A. Pyak & E. Pyak 11026* (TK a-046-2016)

Table S2. Species sampled and their corresponding number of raw reads, and number of informative and missing loci recovered with the PHYLUCE and the HybPiper target extraction methods.

		PHYLUCE method		HybPiper method	
Species	**Nº of raw reads**	**Nº of recovered COS loci**	**Nº of missing COS loci**	**Nº of recovered COS loci**	**Nº of missing COS loci**
Alfredia acantholepis	8,217,881	337	338	510	545
Arctium abolinii	1,853,731	300	375	1003	52
Arctium arctioides	4,754,092	350	325	1006	49
Arctium aureum	4,565,685	294	381	1018	37
Arctium egregium	1,570,769	342	333	1008	47
Arctium eriophorum	2,984,349	327	348	1008	47
Arctium fedtschenkoanum	3,550,984	289	386	1006	49
Arctium grandifolium	2,300,488	324	351	1004	51
Arctium karatavicum	2,976,529	276	399	1010	45
Arctium leiospermum	3,462,345	336	339	1006	49
Arctium minus	10,007,019	350	325	1008	47
Arctium umbrosum	4,663,613	324	351	1012	43
Carduus pycnocephalus	741,845	336	339	667	388
Cirsium sairamense	5,389,901	370	305	988	67
Cousinia albertoregelia	2,524,381	305	370	999	56
Cousinia armena	2,341,432	307	368	1003	52
Cousinia badghysi	2,650,319	208	467	1017	38
Cousinia brachyptera	2,620,531	310	365	1007	48
Cousinia coerulea	2,171,772	336	339	1004	51
Cousinia fetissowii	6,097,776	332	343	1009	46
Cousinia franchetii	3,678,564	303	372	1008	47
Cousinia knorringiae	3,129,866	338	337	1007	48
Cousinia macroptera	2,298,897	295	380	1013	42
Cousinia ninae	3,280,858	303	372	1006	49
Cousinia onopordioides	2,369,328	243	432	1008	47
Cousinia polytimetica	2,905,323	246	429	1002	53

		PHYLUCE method		HybPiper method	
Cousinia pusilla	3,383,791	264	411	1008	47
Cousinia schischkinii	2,694,939	296	379	1013	42
Cousinia serawschanica	4,345,972	266	409	1016	39
Cousinia sewertzowii	4,077,710	327	348	1006	49
Cousinia sogdiana	3,949,050	315	360	1000	55
Cousinia splendida	3,595,227	310	365	1013	42
Cousinia spryginii	3,305,209	337	338	1007	48
Cousinia strobilocephala	4,105,769	325	350	1014	41
Cousinia tenella	5,279,465	412	263	1005	50
Cousinia tianschanica	2,113,147	342	333	1003	52
Cynara cardunculus	454,885	424	251	796	259
Jurinea abramowii	4,803,672	381	294	993	62
Jurinea alata	5,069,639	386	289	1002	53
Jurinea algida	3,743,171	375	300	991	64
Jurinea atropurpurea	4,316,866	354	321	1002	53
Jurinea baldschuanica	5,113,980	351	324	999	56
Jurinea caespitans	4,407,313	351	324	992	63
Jurinea capusii	4,726,638	374	301	991	64
Jurinea carduiformis	5,200,789	383	292	983	72
Jurinea ferganica	5,170,117	348	327	999	56
Jurinea fontqueri	5,240,423	386	289	993	62
Jurinea kokanica	4,531,178	376	299	992	63
Jurinea kyzylkyrensis	5,561,006	362	313	998	57
Jurinea lanipes	4,601,775	375	300	995	60
Jurinea leptoloba	5,487,798	378	297	996	59
Jurinea macrocephala	4,093,061	374	301	985	70
Jurinea narynensis	4,564,064	374	301	989	66
Jurinea olgae	4,941,133	369	306	992	63
Jurinea orientalis	3,155,790	361	314	990	65
Jurinea pinnata	2,996,426	368	307	996	59
Jurinea popovii	3,304,462	367	308	999	56
Jurinea schachimardanica	3,568,519	368	307	994	61
Jurinea stenophylla	3,240,161	370	305	999	56
Jurinea stoechadifolia	4,403,856	323	352	1002	53
Jurinea suffruticosa	2,658,663	362	313	1000	55
Jurinea trautvetteriana	2,087,532	377	298	993	62
Modestia darwasica	5,083,617	380	295	993	62
Olgaea petriprimi	5,310,933	339	336	1001	54
Saussurea carduicephala	7,948,211	348	327	1016	39
Saussurea controversa	8,091,449	349	326	1013	42
Saussurea davurica	11,202,023	376	299	994	61
Saussurea elegans	2,784,084	359	316	997	58
Saussurea foliosa	4,089,960	340	335	1014	41
Saussurea glacialis	4,072,633	368	307	1006	49
Saussurea jadrinzevii	9,091,105	351	324	1010	45
Saussurea krylovii	3,576,809	356	319	1006	49
Saussurea larionowii	4,733,404	344	331	1001	54
Saussurea latifolia	5,065,459	335	340	1007	48
Saussurea leptophylla	6,055,256	366	309	1013	42
Saussurea leucophylla	5,597,695	352	323	1010	45
Saussurea manshurica	4,417,126	348	327	1010	45
Saussurea orgaadayi	3,578,510	378	297	1002	53
Saussurea sp.	3,884,640	364	311	998	57
Saussurea pseudoalpina	3,887,786	334	341	1012	43
Saussurea salicifolia	4,799,838	312	363	1000	55
Saussurea salsa	2,458,299	377	298	996	59
Saussurea schanginiana	4,568,611	366	309	1008	47
Saussurea stubendorffii	5,329,546	309	366	1016	39
Saussurea subacaulis	8,252,488	343	332	1013	42
Average (± standard deviation)	4,263,196 (± 1,822,355)	341.2 (± 37.4)	333.8 (± 37.4)	991.1 (± 67.9)	63.9 (± 67.9)

Table S3. Pairwise comparisons of tree topologies obtained with both the concatenation and coalescence approaches using the Robinson-Foulds (RF) distance among trees and the adjusted RF showed in brackets, ranging from 0 (identical topology) to 1 (completely discordant) calculated from RFadj = RF = RF/(2*n* – 6) being *n* the number of tree nodes.

		Concatenation approach												Coalescence approach									
		Phyluce_627	Phyluce_627_5	Phyluce_627_2.5	Phyluce_627_1	Phyluce_304	HybPiper_1051	HybPiper_1051_5	HybPiper_1051_2.5	HybPiper_1051_1	HybPiper_570	Phyluce_627	Phyluce_627_5	Phyluce_627_2.5	Phyluce_627_1	Phyluce_304	HybPiper_1051	HybPiper_1051_5	HybPiper_1051_2.5	HybPiper_1051_1	HybPiper_570
Concatenation approach	Phyluce 627	0	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
	Phyluce 627_5	0 (0)	0	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
	Phyluce 627_2.5	2 (0.01)	2 (0.01)	0	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
	Phyluce 627_1	42 (0.26)	42 (0.26)	42 (0.26)	0	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
	Phyluce 304	72 (0.44)	72 (0.44)	72 (0.44)	78 (0.48)	0	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
	HybPiper_1051	64 (0.40)	64 (0.40)	64 (0.40)	60 (0.37)	88 (0.54)	0	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
	HybPiper_1051_5	64 (0.40)	64 (0.40)	64 (0.40)	60 (0.37)	88 (0.54)	0 (0)	0	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
	HybPiper_1051_2.5	64 (0.40)	64 (0.40)	64 (0.40)	60 (0.37)	88 (0.54)	0 (0)	0 (0)	0	–	–	–	–	–	–	–	–	–	–	–	–	–	–
	HybPiper_1051_1	64 (0.40)	64 (0.40)	64 (0.40)	58 (0.36)	88 (0.54)	2 (0.01)	2 (0.01)	2 (0.01)	0	–	–	–	–	–	–	–	–	–	–	–	–	–
	HybPiper_570	64 (0.40)	64 (0.40)	64 (0.40)	64 (0.40)	88 (0.54)	24 (0.15)	24 (0.15)	24 (0.15)	24 (0.15)	0	–	–	–	–	–	–	–	–	–	–	–	–
Coalescence approach	Phyluce 627	59 (0.36)	59 (0.36)	59 (0.36)	53 (0.33)	85 (0.52)	63 (0.39)	63 (0.39)	63 (0.39)	63 (0.39)	61 (0.38)	0	–	–	–	–	–	–	–	–	–	–	–
	Phyluce 627_5	59 (0.36)	59 (0.36)	59 (0.36)	53 (0.33)	85 (0.52)	63 (0.39)	63 (0.39)	63 (0.39)	63 (0.39)	61 (0.38)	0 (0)	0	–	–	–	–	–	–	–	–	–	–
	Phyluce 627_2.5	59 (0.36)	59 (0.36)	59 (0.36)	53 (0.33)	85 (0.52)	63 (0.39)	63 (0.39)	63 (0.39)	63 (0.39)	61 (0.38)	0 (0)	0 (0)	0	–	–	–	–	–	–	–	–	–
	Phyluce 627_1	59 (0.36)	59 (0.36)	59 (0.36)	53 (0.33)	85 (0.52)	65 (0.40)	65 (0.40)	65 (0.40)	65 (0.40)	61 (0.38)	4 (0.02)	4 (0.02)	4 (0.02)	0	–	–	–	–	–	–	–	–
	Phyluce 304	75 (0.46)	75 (0.46)	75 (0.46)	67 (0.41)	91 (0.56)	77 (0.48)	77 (0.48)	77 (0.48)	77 (0.48)	81 (0.50)	44 (0.27)	44 (0.27)	44 (0.27)	48 (0.30)	0	–	–	–	–	–	–	–
	HybPiper_1051	77 (0.48)	77 (0.48)	77 (0.48)	71 (0.44)	95 (0.59)	85 (0.52)	85 (0.52)	85 (0.52)	85 (0.52)	83 (0.51)	60 (0.37)	60 (0.37)	60 (0.37)	62 (0.38)	50 (0.31)	0	–	–	–	–	–	–
	HybPiper_1051_5	67 (0.41)	67 (0.41)	67 (0.41)	59 (0.36)	85 (0.52)	73 (0.45)	73 (0.45)	73 (0.45)	73 (0.45)	71 (0.44)	50 (0.31)	50 (0.31)	50 (0.31)	52 (0.32)	56 (0.35)	18 (0.11)	0	–	–	–	–	–
	HybPiper_1051_2.5	69 (0.43)	69 (0.43)	69 (0.43)	61 (0.38)	89 (0.55)	75 (0.46)	75 (0.46)	75 (0.46)	75 (0.46)	73 (0.45)	52 (0.32)	52 (0.32)	52 (0.32)	54 (0.33)	66 (0.41)	24 (0.15)	16 (0.10)	0	–	–	–	–
	HybPiper_1051_1	73 (0.45)	73 (0.45)	73 (0.45)	65 (0.40)	89 (0.55)	75 (0.46)	75 (0.46)	75 (0.46)	75 (0.46)	73 (0.45)	56 (0.35)	56 (0.35)	56 (0.35)	58 (0.36)	54 (0.33)	22 (0.14)	20 (0.12)	28 (0.17)	0	–	–	–
	HybPiper_570	79 (0.49)	79 (0.49)	79 (0.49)	71 (0.44)	95 (0.59)	83 (0.51)	83 (0.51)	83 (0.51)	83 (0.51)	73 (0.45)	60 (0.37)	60 (0.37)	60 (0.37)	62 (0.38)	66 (0.41)	48 (0.30)	46 (0.28)	54 (0.33)	44 (0.27)	0	–	–

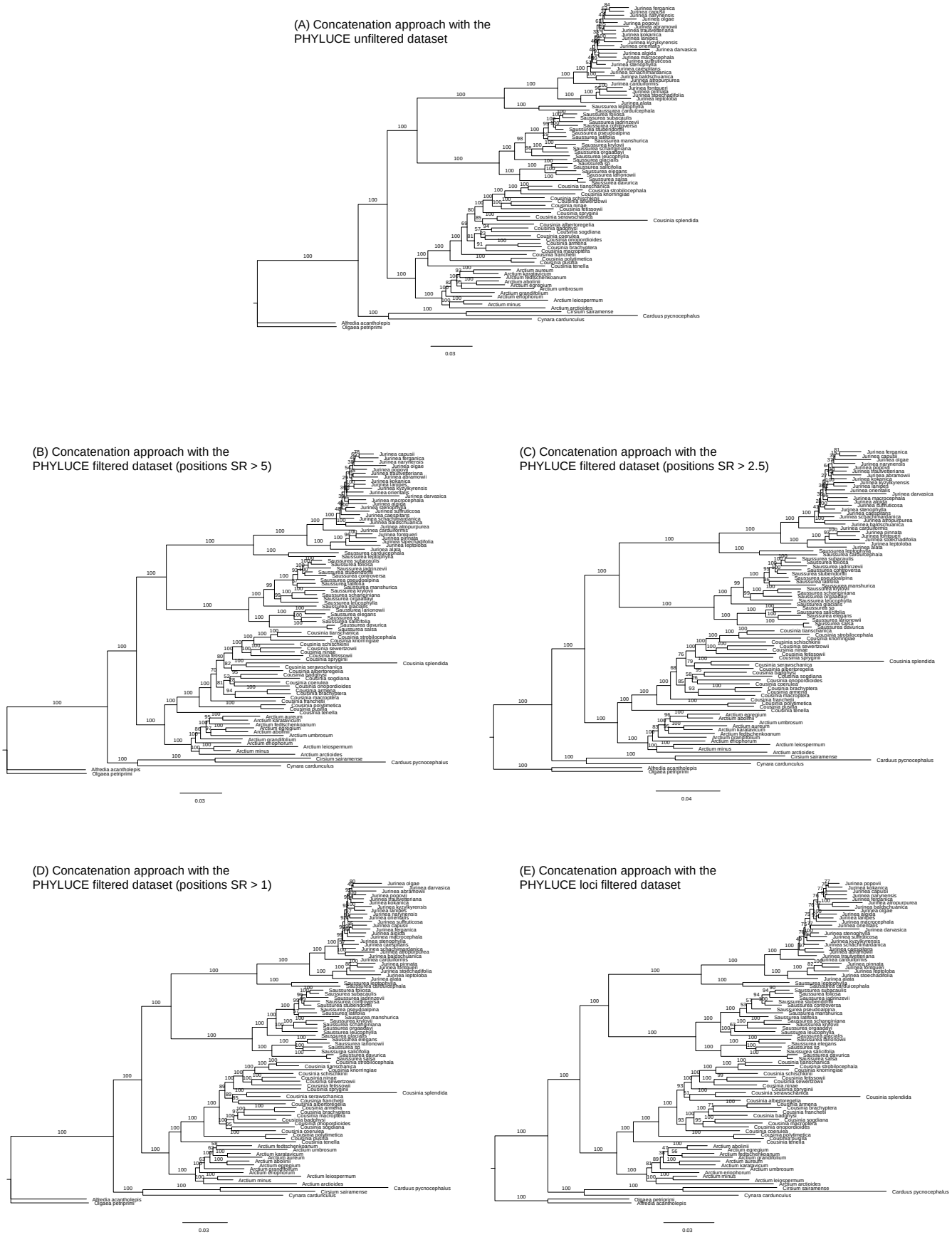


Fig. S1. Phylogenetic trees estimated from the datasets obtained with the PHYLUCE method and the concatenation approach (see main text for details). **(A)** Using the unfiltered alignment, **(B)** using the filtered alignment, removing the positions with substitution rates (SR) higher than 5, **(C)** using the filtered alignment, removing the positions with substitution rates (SR) higher than 2.5, **(D)** using the filtered alignment, removing the positions with substitution rates (SR) higher than 1, and **(E)** using the loci filtered alignment, containing only the best informative loci selected under the criteria explained in the main text.

Fig. S2. Phylogenetic trees estimated from the datasets obtained with the HybPiper method and the concatenation approach (see main text for details). **(A)** Using the unfiltered alignment, **(B)** using the filtered alignment, removing the positions with substitution rates (SR) higher than 5, **(C)** using the filtered alignment, removing the positions with substitution rates (SR) higher than 2.5, **(D)** using the filtered alignment, removing the positions with substitution rates (SR) higher than 1, and **(E)** using the loci filtered alignment, containing only the best informative loci selected under the criteria explained in the main text.

Fig. S3. Phylogenetic trees estimated from the datasets obtained with the PHYLUCE method and the coalescence approach (see main text for details). **(A)** Using the unfiltered alignments, **(B)** using the filtered alignments, removing the positions with substitution rates (SR) higher than 5, **(C)** using the filtered alignments, removing the positions with substitution rates (SR) higher than 2.5, **(D)** using the filtered alignments, removing the positions with substitution rates (SR) higher than 1, and **(E)** using the loci filtered alignments, containing only the best informative loci selected under the criteria explained in the main text.

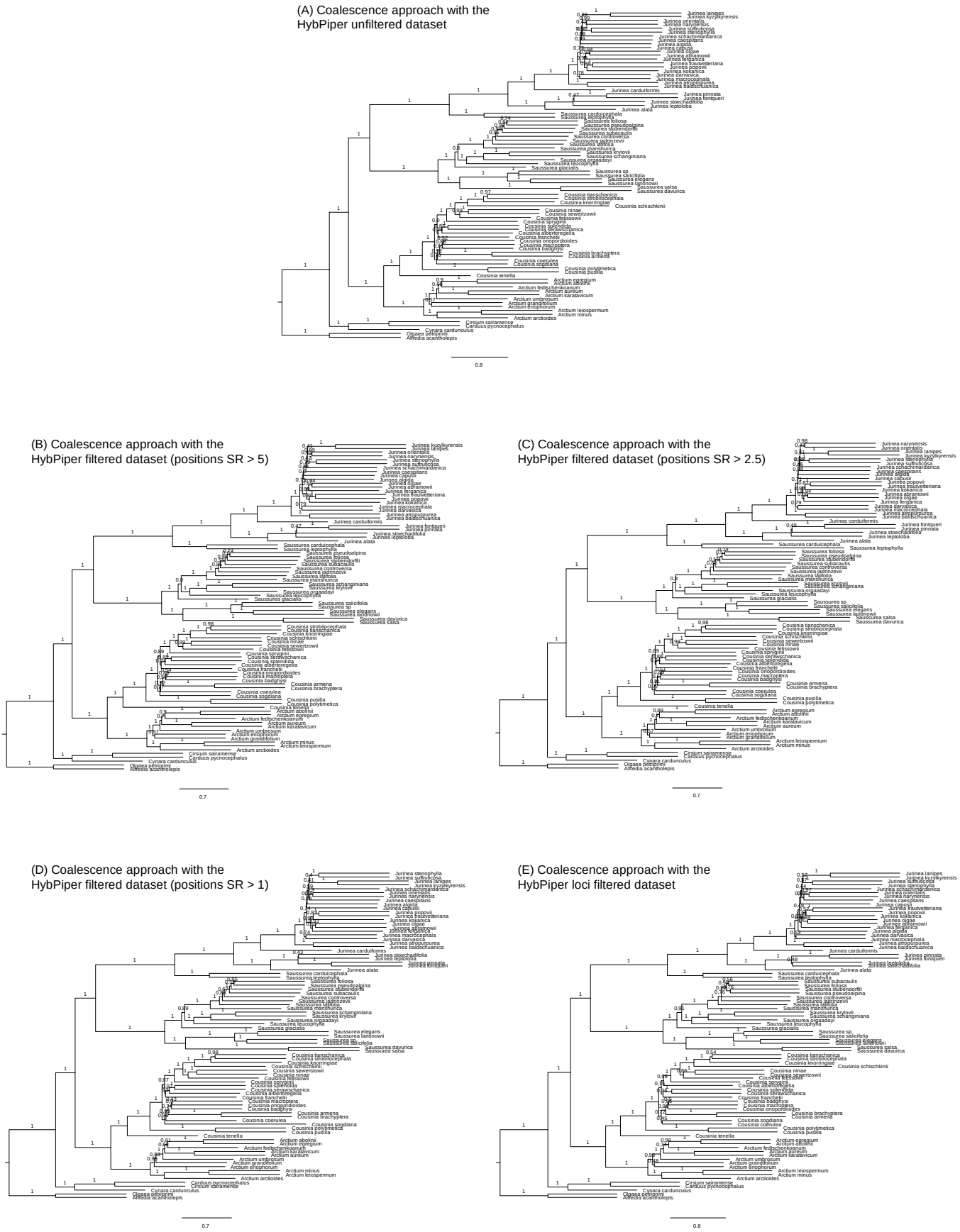


Fig. S4. Phylogenetic trees estimated from the datasets obtained with the HybPiper method and the coalescence approach (see main text for details). **(A)** Using the unfiltered alignments, **(B)** using the filtered alignments, removing the positions with substitution rates (SR) higher than 5, **(C)** using the filtered alignments, removing the positions with substitution rates (SR) higher than 2.5, **(D)** using the filtered alignments, removing the positions with substitution rates (SR) higher than 1, and **(E)** using the loci filtered alignments, containing only the best informative loci selected under the criteria explained in the main text.

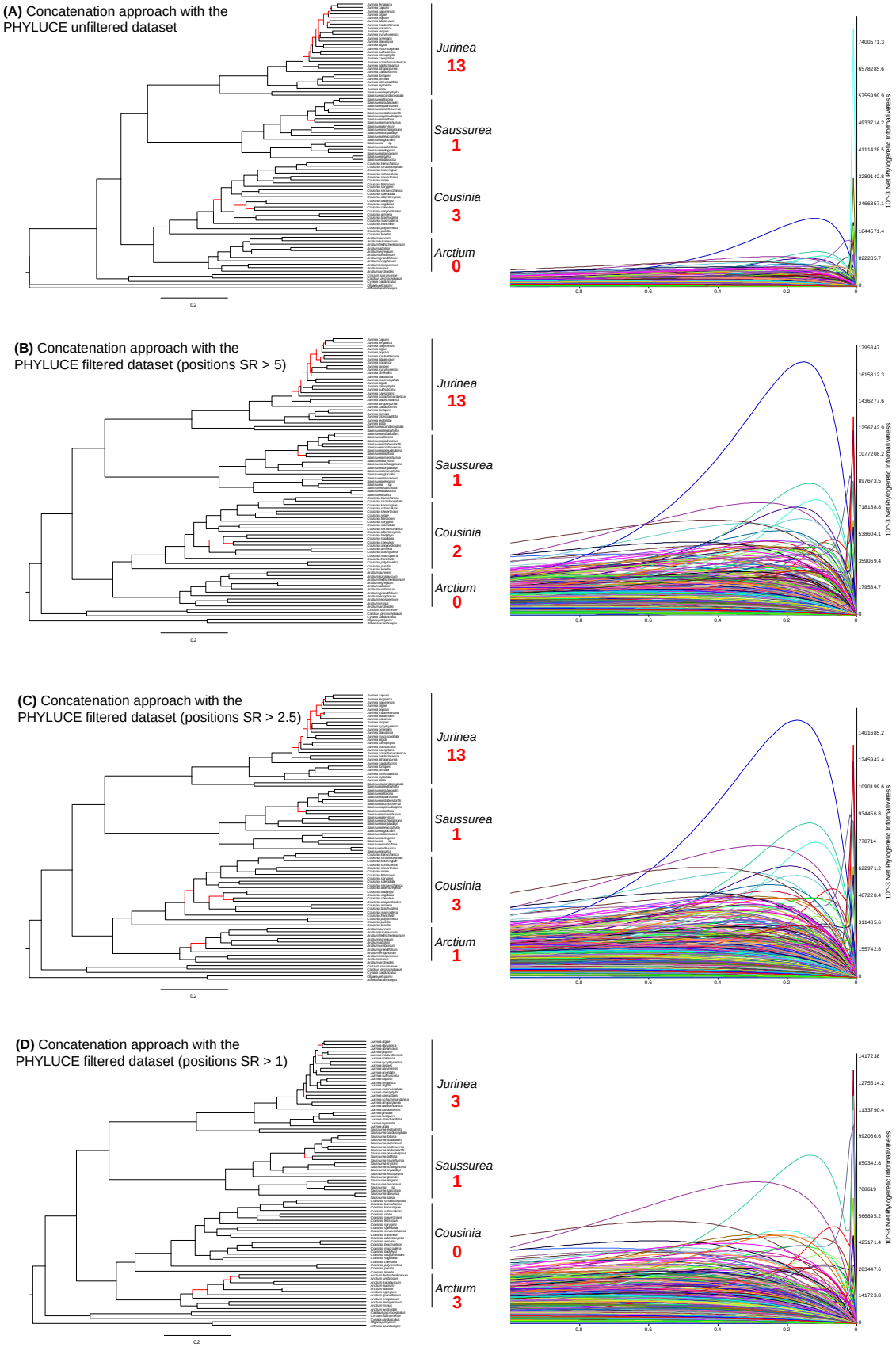


Fig. S5. Phylogenetic Informativeness (PI) analyses showing the ultrametric trees rescaled from 1 to 0 obtained from the maximum likelihood analyses of the PHYLUCE dataset and the concatenation approach, and net phylogenetic informativeness curves representing the profiles for each locus displayed in different colors. The analyses of PI were done with **(A)** the unfiltered alignment, **(B)** the filtered alignment, removing the positions with substitution rates (SR) > 5, **(C)** the filtered alignment, removing the positions with SR > 2.5, and **(D)** the filtered alignment, removing the positions with SR > 1. Branches with bootstrap support values below 70 are highlighted in red. For each of the four genera, the number unsupported nodes are shown at the right of the trees.

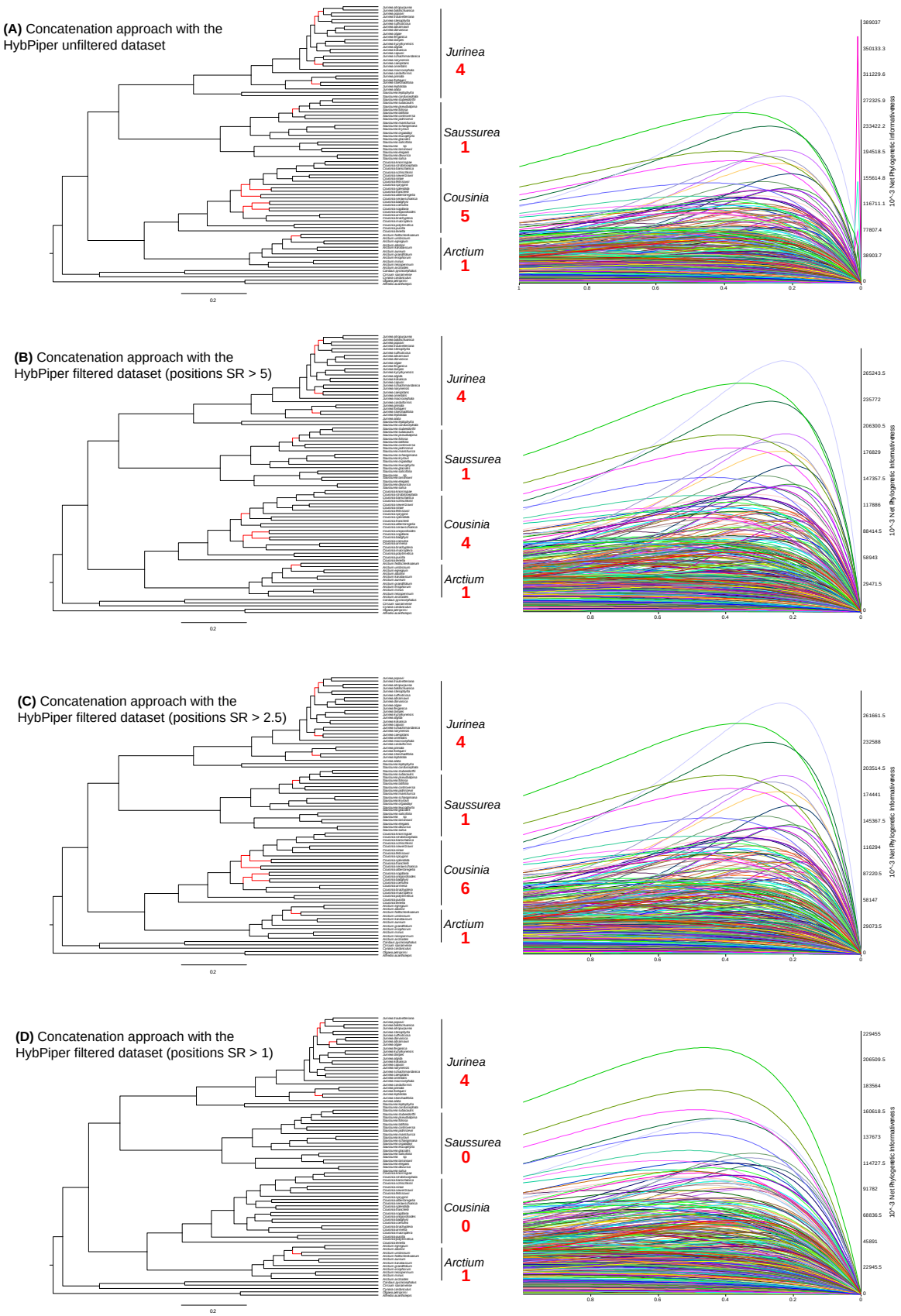


Fig. S6. Phylogenetic Informativeness (PI) analyses showing the ultrametric trees rescaled from 1 to 0 obtained from maximum likelihood analyses of the HybPiper dataset and the concatenation approach, and net phylogenetic informativeness curves representing the profiles for each locus displayed in different colors. The analyses of PI were done with **(A)** the unfiltered alignment, **(B)** the filtered alignment, removing the positions with substitution rates (SR) > 5, **(C)** the filtered alignment, removing the positions with SR > 2.5, and **(D)** the filtered alignment, removing the positions with SR > 1. Branches with bootstrap support values below 70 are highlighted in red. For each of the four genera, the number unsupported nodes are shown at the right of the trees.

Chapter 2

Nuclear and plastid DNA phylogeny of tribe Cardueae (Compositae) with Hyb-Seq data: A new subtribal classification and a temporal diversification framework

Sonia Herrando-Moraira[a*] and The Cardueae Radiations Group
In alphabetical order:

Juan Antonio Calleja[b]
Mercè Galbany-Casals[b]
Núria Garcia-Jacas[a]
Jian-Quan Liu[c]
Javier López-Alvarado[b]
Jordi López-Pujol[a]
Jennifer R. Mandel[d]
Sergi Massó[a,b]
Noemí Montes-Moreno[a]
Cristina Roquet[b,e]
Llorenç Sáez[b]
Alexander Sennikov[f]
Alfonso Susanna[a]
Roser Vilatersana[a]

[a] Botanic Institute of Barcelona (IBB, CSIC-ICUB), Pg. del Migdia, s.n., ES-08038 Barcelona, Spain
[b] Systematics and Evolution of Vascular Plants (UAB) – Associated Unit to CSIC, Departament de Biologia Animal, Biologia Vegetal i Ecologia, Facultat de Biociències, Universitat Autònoma de Barcelona, ES-08193 Bellaterra, Spain
[c] Key Laboratory for Bio-Resources and Eco-Environment, College of Life Sciences, Sichuan University, Chengdu, China
[d] Department of Biological Sciences, University of Memphis, Memphis, TN 38152, USA
[e] Univ. Grenoble Alpes, Univ. Savoie Mont Blanc, CNRS, LECA (Laboratoire d'Ecologie Alpine), FR-38000 Grenoble, France
[f] Botanical Museum, Finnish Museum of Natural History, PO Box 7, FI-00014 University of Helsinki, Finland; and Herbarium, Komarov Botanical Institute of Russian Academy of Sciences, Prof. Popov str. 2, 197376 St. Petersburg, Russia

*Corresponding author at: Botanic Institute of Barcelona (IBB, CSIC-ICUB), Pg. del Migdia, s. n., ES-08038 Barcelona, Spain. E-mail address: sonia.herrando@gmail.com (S. Herrando-Moraira).

Journal, volume, pages (year): Molecular Phylogenetics and Evolution, 137, 313–332 (2019).
DOI: https://doi.org/10.1016/j.ympev.2019.05.001

Impact factor year 2019: 3.496 (Clarivate Analytics, 2020)
Ranking ISI Journal Citation Reports ©:
- 2019: 123/297 Biochemistry & Molecular Biology
- 2019: 16/50 Evolutionary Biology
- 2019: 60/177 Genetics Heredity

Received: 23 February 2019; Received in revised form: 4 April 2019; Accepted: 1 May 2019

Abstract

Classification of tribe Cardueae in natural subtribes has always been a challenge due to the lack of support of some critical branches in previous phylogenies based on traditional Sanger markers. With the aim to propose a new subtribal delimitation, we applied a Hyb-Seq approach to a set of 76 Cardueae species representing all subtribes and informal groups defined in the tribe, targeting 1061 nuclear conserved orthology loci (COS) designed for Compositae and obtaining chloroplast coding regions as by-product of off-target reads. For the extraction of the target nuclear data, we used two strategies, PHYLUCE and HybPiper, and 776 and 1055 COS loci were recovered with each of them, respectively. Additionally, 87 chloroplast genes were assembled and annotated. With three datasets, phylogenetic relationships were reconstructed using both concatenation and coalescent approaches. Phylogenetic analyses of the nuclear datasets fully resolved virtually all nodes with very high support. Nuclear and plastid tree topologies are mostly congruent with a very limited number of incongruent nodes. Based on the well-solved phylogenies obtained, we propose a new taxonomic scheme of 12 monophyletic and morphologically consistent subtribes: Carlininae, Cardopatiinae, Echinopsinae, Dipterocominae (new), Xerantheminae (new), Berardiinae (new), Staehelininae (new), Onopordinae (new), Carduinae (redelimited), Arctiinae (new), Saussureinae (new), and Centaureinae. In addition, we further updated the temporal framework for origin and diversification of these subtribes. Our results highlight the power of Hyb-Seq over Sanger sequencing of a few DNA markers in solving phylogenetic relationships of traditionally difficult groups.

Keywords
Asteraceae
COS targets
Subtribes
Systematics
Phylogenomics
Target enrichment

Index

1. Introduction

Cardueae are one of the largest tribes of the 43 described in Compositae (Asteraceae; Funk et al., 2009), with almost 10% of the species of the whole family: 2400 species in 72 genera (Susanna and Garcia-Jacas, 2009). The tribe belongs to subfamily Carduoideae, which is composed of four tribes (Funk et al., 2009; Ortiz et al., 2009): Dicomeae (97 species), Tarchonantheae (13 species), Oldenburgieae (4 species), and Cardueae, the latter representing the 95% of subfamily's diversity (Susanna and Garcia-Jacas, 2009). Carduoideae are a very successful evolutionary lineage within Compositae, which was estimated by Panero and Crozier (2016) to have the second highest diversification rate in the family and a moderate rate of extinction. The number of species is not uniform across the genera of Cardueae. For example, six of them are highly diversified (ca. 200–600 species) and with a high endemism rate (*Carduus*, *Cirsium*, *Centaurea*, *Cousinia*, *Jurinea*, and *Saussurea*), while, on the other extreme, 22 (ca. 30%) are monotypic. Geographically, Cardueae are distributed mainly in the Mediterranean and the Irano-Turanian regions, but they are reported from all continents except the Antarctica. The ecosystems where Cardueae species inhabit are very variable, e.g. Mediterranean landscapes, steppes, semiarid areas, deserts, alpine meadows, or tropical savannahs (Susanna and Garcia-Jacas, 2009). Indeed, in many aspects Cardueae are an extremely heterogeneous group in terms of habit and life-form (perennial, biennial, monocarpic or annual herbs, shrubs, treelets, often spiny), karyology (high variability in chromosome numbers from $x = 6$ to $x = 18$, often with disploidy), or pollen structure (caveate or ecaveate, smooth, scabrate, or spiny). This complexity summed to its high diversity has greatly contributed to the turbulent taxonomic history of Cardueae at several ranks, e.g. tribal and subtribal (see Table 1 for a historical overview) or even misclassifications of some genera in different tribes or subtribes (see examples in Table 2).

A recapitulation of the troublesome history of the systematics of Cardueae was offered by Susanna and Garcia-Jacas (2009). In the first tribal classification of the family, Cassini (1819) divided the present Cardueae into three tribes: Echinopseae, Carlineae, and Cardueae, the latter comprising the subtribes Carduinae and Centaureinae. Years later, Bentham (1873) and Hoffmann (1894) proposed grouping the three former tribes in a single tribe Cardueae, with four subtribes: Carlininae, Echinopsinae [–Echinopinae"], Carduinae, and Centaureinae. This conservative approach was accepted for many years, until Wagenitz (1976) reinstated the tribe Echinopseae. Soon after, Dittrich (1977) returned to Cassini's earlier views, proposing the restoration of tribes Echinopseae, Carlineae, and Cardueae. Bremer (1994) favored the conservative classification of a single tribe Cardueae, which is now generally accepted (Susanna and Garcia-Jacas, 2007, 2009). The only exception is subtribe Cardopatiinae, originally described by Lessing (1832) and only recently recovered 174 years later (Susanna et al., 2006).

Not surprisingly, Carlininae and Echinopsinae have been considered at some time as independent tribes, mainly due to their clear diagnostic characters with respect to the rest of Cardueae assembly (Susanna and Garcia-Jacas, 2009). However, Cardueae are morphologically consistent as a whole entity at tribal level, sharing a unique synapomorphic morphological character within Compositae: style with a papillose-pilose thickening below the branches and the stigmatic areas confined to the inner surface (Susanna and Garcia-Jacas, 2009). Additionally, Cardueae as a tribe has also been broadly confirmed empirically as monophyletic, first by cladistic analyses based on morphology (Bremer, 1987, 1994; Karis et al., 1992), and later by molecular phylogenies (Jansen et al., 1990, 1991; Kim et al., 1992; Susanna et al., 1995, 2006; Garcia-Jacas et al., 2002; Barres et al., 2013).

Table 1. Historical outline of the classification of tribe Cardueae.

Classification	Taxonomic treatment references
Tribe Echinopeae Tribe Carlineae Tribe Cardueae Subtribe Carduinae Subtribe Centaureinae	Cassini (1819), Dittrich (1977)
Tribe Cardueae Subtribe Echinopsinae Subtribe Carlininae Subtribe Carduinae Subtribe Centaureinae	Bentham (1873), Hoffmann (1894), Bremer (1994)
Tribe Echinopeae Tribe Cardueae Subtribe Carlininae Subtribe Carduinae-Centaureinae	Wagenitz (1976), Petit (1988, 1997)
Tribe Cardueae Subtribe Cardopatiinae Subtribe Carlininae Subtribe Echinopsinae Subtribe Carduinae Subtribe Centaureinae	Susanna et al. (2006), Susanna and Garcia-Jacas (2007, 2009)
Tribe Cardueae Subtribe Carlininae Subtribe Cardopatiinae Subtribe Echinopsinae Subtribe Dipterocominae Subtribe Xerantheminae Subtribe Berardiinae Subtribe Staehelininae Subtribe Onopordinae Subtribe Carduinae Subtribe Arctiinae Subtribe Saussureinae Subtribe Centaureinae	New treatment proposed here

In the most recent and accepted treatments (Susanna et al., 2006; Susanna and Garcia-Jacas, 2007, 2009), the authors pointed out that four subtribes are natural groups with clear limits (Cardopatiinae, Carlininae, Centaureinae, and Echinopsinae); however, subtribe Carduinae (with total 1700 species, near 70% of the whole tribe diversity) is an unnatural, artificial, and problematic group, which has represented a dumping ground of genera that do not fit morphologically in any of the other subtribes (Susanna and Garcia-Jacas, 2007). The fact that the group is a questionable and heterogeneous assemblage of genera

Table 2. Some cases of genera with historical conflicts of adscription to subtribes (or even tribes) of tribe Cardueae resolved with phylogenies based on Sanger sequencing data. The tribes and treatment references in bold represent the most recent accepted taxonomic proposal and their corresponding source, respectively. The last column indicates the taxonomic treatment proposed in the present study for the traditional taxonomic conflict cases based on new subtribal delimitation. **Xeranthemum* group, including *Amphoricarpos*, *Chardinia*, *Siebera*, and *Xeranthemum*.

Genus	Taxonomic rank of conflict	Included in	References of treatments	Taxonomic treatment proposed here
Berardia	tribal	Mutisiae **Cardueae** (Carduinae)	Bremer (1994), Dittrich (1996a) **Garcia-Jacas et al. (2002)**	Cardueae (Berardiinae)
Dipterocome	tribal	Cichorieae Calenduleae **Cardueae** (Carduinae)	Boissier (1849) Jaubert & Spach (1850), Bentham (1873), Hoffmann (1894) **Susanna and Garcia-Jacas (2009)**	Cardueae (Dipterocominae)
Cardopatium	subtribal	Carlininae Echinopsidinae **Cardopatiinae**	Dittrich (1977, 1996b), Bremer (1994), Petit (1997) **Susanna et al. (2006)**	Cardopatiinae
Cousiniopsis	subtribal	Carlininae Echinopsidinae **Cardopatiinae**	Dittrich (1977, 1996b), Bremer (1994), Petit (1997) **Susanna et al. (2006)**	Cardopatiinae
Femeniasia	subtribal	Carduinae **Centaureinae**	Susanna (1987) **Susanna et al. (1995)**	Centaureinae
Myopordon	subtribal	Carduinae **Centaureinae**	Wagenitz (1958), Dittrich (1977) **Susanna et al. (2006)**	Centaureinae
Nikitinia	subtribal	Carduinae **Centaureinae**	Dittrich (1977), Bremer (1994) **Susanna et al. (2002, 2006)**	Centaureinae
Staehelina	subtribal	Carlininae **Carduinae**	Bentham (1873), Hoffmann (1894), Dittrich (1977, 1996b), Bremer (1994) Petit (1997), **Susanna et al. (2006)**	Staehelininae
Synurus	subtribal	Centaureinae **Cardueae**	Dittrich (1977), Bremer (1994) **Häffner and Hellwig (1999)**	Onopordinae
Syreitschikovia	subtribal	Centaureinae **Carduinae**	Dittrich (1977), Bremer (1994) **Susanna et al. (2002)**	Onopordinae
Xeranthemum (all the group)*	subtribal	Carlininae **Carduinae**	Dittrich (1977, 1996b), Bremer (1994), Petit (1997), **Susanna et al. (2006)**	Xerantheminae

was also reflected in several phylogenetic studies, which have reported the subtribe as paraphyletic because Centaureinae are nested on Carduinae (Susanna et al., 1995, 2006; Häffner and Hellwig, 1999; Garcia-Jacas et al., 2002; Barres et al., 2013). The alternate solution of combining subtribes Carduinae and Centaureinae in one enormous subtribe was discarded, owing to the impractical constitution of a huge subtribe encompassing 2300 species and 90% of the species of the whole tribe (Susanna and Garcia-Jacas, 2007).

With the intention of addressing the unresolved diversity within subtribe Carduinae, some informal morphological groups were described within it (Susanna and Garcia-Jacas, 2007), which could be considered subtribes for a more natural classification: ~~[~~the alternative is] splitting present Carduinae in at least seven new subtribes (many of them presently unsupported): Xerantheminae, Staehelininae, Berardininae, Onopordinae, Carduinae, Arctiinae, and Saussureinae […]" (Susanna and Garcia-Jacas, 2009). In view of the lack of support for the monophyly of all segregate subtribes, especially for the most important in terms of species number (core Carduinae, Arctiinae, and Saussureinae), no formal proposal has previously been performed. A well-resolved and supported phylogenetic hypothesis for the major lineages within Carduinae is still lacking, and it would be very useful to confirm the morphological alliances or generic complexes proposed by Susanna and Garcia-Jacas (2009).

An additional taxonomic problem within Cardueae is that some genera or complex of genera have been classified within

one subtribe or another depending on the classification followed (see some of the cases in Table 2). Moreover, in extreme cases, some genera have been independently classified in Cardueae or in other tribes of Compositae (e.g. *Berardia* in tribe Mutisiae, or *Dipterocome* in tribe Calenduleae). A great proportion of subtribal misplacements were reported for uncertain cases between subtribes Carduinae and Centaureinae, which could be attributed to the inconspicuous morphological differences between both subtribes (based on microcharacters of achene and pappus) that are sometimes lacking in herbarium specimens, or are immature structures during field collections (Susanna et al., 2006). Even though many cases have been successfully resolved with the aid of the molecular phylogenies obtained with Sanger sequencing (see references in Table 2), more changes in the adscription of genera could occur in a new subtribal framework for Cardueae.

To date, all previous molecular studies of Cardueae have been based on Sanger sequencing data (Susanna et al., 1995, 2006; Häffner and Hellwig, 1999; Garcia-Jacas et al., 2002; Barres et al., 2013), with the largest dataset constructed with four chloroplast regions (*trnL-trnF*, *matK*, *ndhF*, *rbcL*) and one nuclear ribosomal marker (ITS; Barres et al., 2013). Although subtribal clades of Carlininae, Echinopsinae, Cardopatiinae, and Centaureinae are well supported, Carduinae remains paraphyletic when Centaureinae are removed, and not all the informal groups have been recovered as monophyletic. Moreover, relationships among subtribes have not been resolved, and the backbone of the phylogenetic tree remains undefined without a supported dichotomous bifurcating pattern. Accordingly, the available divergence time estimation of Cardueae is based on a partially resolved phylogenetic tree resulting from the analysis of a combined dataset of chloroplast markers (Barres et al., 2013). This dating could be significantly improved in terms of the methodological approach and the sequence data used.

In the last decade, next generation sequencing (NGS) has emerged as an important methodological advance for solving phylogenetic problems (see the review of Harrison and Kidner, 2011). Plant phylogenies of historical taxonomically complex groups are becoming resolved at different taxonomic levels, e.g. order (Ranunculales, cf. Lane et al., 2018; or Zingiberales, cf. Carlsen et al., 2018), family (Goodeniaceae, cf. Gardner et al., 2016), tribe (Shoreeae, cf. Heckenhauer et al., 2018), or species complexes as *Claytonia* (Stoughton et al., 2018) or *Amaranthus* (Viljoen et al., 2018). For Compositae, Mandel et al. (2014) designed a probe set that hybridizes with 1061 nuclear conserved orthology loci (hereafter COS), which in combination with genome skimming allowed also to recover other parts of the genome such as chloroplast regions (defined as Hyb-Seq technique; Weitemier et al., 2014). This methodological workflow has proved successful in Compositae-wide studies (Mandel et al., 2015, 2017) and was tested in a recent research focused on highly radiated genera within Cardueae (Herrando-Moraira et al., 2018). However, the COS locus set has not been yet applied to taxonomic delimitation within the family.

In this study, we apply the Hyb-Seq method to a sample of 76 species representing all subtribes and informal suprageneric groups of Cardueae with the main goals to: (1) obtain a well-defined phylogeny of the high-level lineages in the tribe with nuclear and chloroplast data; (2) propose a new subtribal classification, especially focused on testing the splitting of subtribe Carduinae into smaller and more practical natural subtribes; (3) examine previously unresolved phylogenetic relationships between subtribal lineages at the backbone of the tree; and (4) update the temporal framework of the tribe and subtribes origin and diversification.

2. Materials and methods

2.1. Taxon sampling

To obtain a complete sampling of Cardueae, we included representatives of all described subtribes based on the taxonomic treatment of Susanna and Garcia-Jacas (2009): (1) Carlininae, 6 species; (2) Cardopatinae, 2 species; (3) Echinopsinae, 3 species; (4) Carduinae, 34 species; and (5) Centaureinae, 31 species. The sampling strategy was based on maximizing the number of species of unresolved informal groups, especially within Carduinae and Centaureinae, and was proportional to the total number of genera and species classified in each group. In total, 76 different species of Cardueae were included. We did not include all genera within the tribe because the adscription of many of those with taxonomic problems had already been confirmed in previous works (Table 2). In addition, we incorporated 5 species as representatives of subfamily Carduoideae outside of Cardueae; 5 species from other subfamilies within Compositae (2 from Mutisioideae, 2 from Barnadesioideae, 1 from Famatinanthoideae); and finally, *Nastanthus patagonicus* from Calyceraceae, which has been recognized as the sister family to Compositae (Mandel et al., 2017). Overall, 58 species were newly sequenced for this study, and the data of the remaining 29 were directly obtained from raw reads from Mandel et al. (2014, 2017) or Herrando-Moraira et al. (2018). See Supplementary Table S1 for details of each sampled species.

2.2. DNA extraction, library and capture preparation, and sequencing

Approximately 10–30 mg of dried plant material per sample was weighed and homogenized with a Mixer Mill MM 301 (Retsch®, Haan, Germany). The DNA was extracted with the DNeasy plant mini kit (Qiagen, Valencia, CA, USA) or the E.N.Z.A SP Plant DNA Mini Kit (Omega Bio-Tek Inc., Norcross, GA, USA) according to the manufacturer's protocol. The total genomic DNA quantity was measured with the Qubit™ 3.0 Fluorometer (Thermo Scientific, Waltham, MA, USA). The standardized DNA (1 µg in 70 µl) was sheared in the Genomics Unit of the Centre for Genomic Regulation (CRG, Barcelona, Spain) using a Covaris S2 System (Covaris, Woburn, MA, USA) in microTUBEs with a sample volume of 50 µl and a target peak set to 400 bp. Sequencing libraries and subsequent sequence capture were conducted as specified in Herrando-Moraira et al. (2018). Additionally, for the species newly sequenced for this study (see Supplementary Table S1), we conducted a library spiking with the following proportions: 40% of unenriched solution and 60% of enriched libraries. The final spiked library pools were sequenced (pair-end 100 bp) in the DNA Sequencing Core CGRC/ICBR of the University of Florida on one lane of an Illumina HiSeq 3000 (Illumina, USA).

2.3. Raw data processing

Demultiplexed raw sequence reads were checked with FastQC v.0.10.1 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) in order to perform a first quality control step. Subsequently, raw reads were trimmed by a quality based assessment (sliding-window set to 5:20), and were also subjected to an adapter trimming using Trimmomatic v.0.36 (Bolger et al., 2014). In addition, the software was programmed to remove those cleaned reads that after trimming were less than 36 bp, and those with a missing forward or reverse pair.

2.4. Nuclear data extraction

Following the methodological workflow proposed in Herrando-Moraira et al. (2018), the 1061 target COS loci were extracted with two orthology-detection pipeline packages: PHYLUCE v.1.5 (Faircloth, 2015) and HybPiper v.1.1 (Johnson et al., 2016). We used both approaches to be sure that the distinct paralog processing strategy was not affecting phylogenetic reconstructions in terms of topology and branch support values (Herrando-Moraira et al., 2018).

For the PHYLUCE method, the trimmed reads were *de novo* assembled into contigs with the software SPAdes v.3.9.0 (Bankevich et al., 2012), specifying different predefined k-mer lengths (21, 33, 55, and 77). Then, the recovered contigs were mapped back to target references using LASTZ (Harris, 2007), and were extracted following the methodological specifications detailed in Herrando-Moraira et al. (2018). For the HybPiper method, the trimmed reads were firstly mapped to the targets using BWA (Li and Durbin, 2009), and secondly were assembled into contigs also using SPAdes.

The multi-fasta files obtained for each target locus both with PHYLUCE and HybPiper were aligned with MAFFT v.7.266 (Katoh and Standley, 2013) using the *auto* setting mode. The ambiguously aligned regions were removed with trimAl v.14 (Capella-Gutiérrez et al., 2009) applying the *automated1* flag. The loci extracted for less than three species were removed, in the case of PHYLUCE 126 from the total of 902 loci, and in HybPiper 1 locus from the total 1057 loci. One additional locus was removed from the HybPiper dataset due to its short length recovered after the alignment trimming (3 bp). Finally, for each target extraction method, two datasets were constructed in order to perform two different phylogenetic inference analyses. One consisted of the trimmed aligned sequences of each locus separately, and the other consisted of a supermatrix obtained by concatenating all trimmed aligned loci with FASconCAT-G v.1.02 (Kück and Longo, 2014). The summary statistics of the supermatrices were calculated with AMAS (Borowiec, 2016).

2.5. Chloroplast data extraction

For the reconstruction of the chloroplast genomes, we used the off-target reads also recovered with the Hyb-Seq approach. The chloroplast extraction was conducted using the pipeline package MITObim v.1.9 (Hahn et al., 2013) with the default conditions. The package has incorporated the module MIRA v.4.0.2 (Chevreux et al., 1999), which is used in mapping mode. In the first step, it identifies the more conserved regions between the total readpool and a phylogenetically related reference genome (in this case, *Cirsium arvense* NCBI accession number NC_036965.1) by mapping the trimmed interleaved reads to this initial reference. Then, the mapped reads are assembled into contigs and a new reference sequence or bait is created. To extend the reference and close the gaps, two more steps are performed. The second consists of separating from the total read pool the reads that overlap with the reference, and the third recovers the reads separated to map them back to the bait, thus building a new more extended reference sequence through an assembly process of new mapped reads. These two last steps are iteratively repeated until a stationary stage of the mapped and assembled reads is reached (Hahn et al., 2013).

The chloroplast genomes recovered were annotated with the web tool application GeSeq (Tillich et al., 2017), which uses a customizable reference database to annotate the genomes using BLAT-driven best-match approach. As a database, we selected all the available annotations in NCBI of Cardueae representatives (*Carthamus tinctorius* NC_030783.1, *Centaurea diffusa* NC_024286.1, *Cirsium arvense* NC_036965.1, *Cirsium eriophorum* NC_036966.1, *Cirsium vulgare* NC_036967.1, *Cynara baetica* NC_028005.1, *Cynara cornigera* NC_028006.1, *Cynara humilis* NC_027113.1, *Saussurea chabyoungsanica* NC_036677.1, *Saussurea involucrata* NC_029465.1, *Saussurea polylephis* NC_036490.1, and *Silybum marianum* NC_028027.1). To extract the coding regions of genes (CDS) in separate files from the global multi-fasta matrix generated with GeSeq, we used the script "–Phyluce_assembly_explode_get_fastas_file" of PHYLUCE package with slight modifications. The recovered CDS were individually aligned with the MACSE codon aligner (Ranwez et al., 2011). The alignments were visualized in SeaView v.4.7 (Gouy et al., 2009), and 10 species from the initial sampling were excluded due to the high presence of frameshifts and preceding codon stops. To remove the poorly aligned regions from the alignments, we used the program Gblocks v.0.91b (Castresana, 2000) with the option "–codon" to trim entire codon sets. A total of 87 CDS regions were recovered. We concatenated all CDS in a single supermatrix (chloroplast dataset) with FASconCAT-G v.1.02, assuming that the chloroplast genome is considered to be haploid, nonrecombinant, generally uniparentally inherited, and "–single copy" (Small et al., 1998, 2004).

2.6. Phylogenetic inference analyses

The phylogenetic reconstruction of Cardueae was performed under two different approaches for nuclear data: the concatenation approach (using the supermatrix dataset) and the coalescence approach (using the individual gene trees for inferring the species tree). For chloroplast data, the phylogenetic tree inference was only performed under the concatenation approach.

For the concatenation approach, Maximum Likelihood (ML) analyses were run using the software RAxML v.8.2.9 (Stamatakis, 2014), which were conducted on XSEDE in the CIPRES Science Gateway v.3.1 (Miller et al., 2010). We selected the algorithm of simultaneously searching the best ML tree (from 10 randomized maximum parsimony starting trees) and performing a rapid bootstrapping (1000 replicates). Branches were considered as statistically supported for bootstrap (BS)

values > 70% (Hillis and Bull, 1993). In relation to the partition scheme, each locus (in the nuclear datasets) or gene (in the chloroplast dataset) was considered a unit of partition, using the evolution model of GTRGAMMA following recommendations of Stamatakis (2006). Output trees were visualized and exported in FigTree v.1.4.3 (Rambaut, 2016).

For the coalescence approach, gene trees were inferred also using RAxML, under the same conditions as for the concatenation based analyses but with a bootstrap resampling of 200 replicates. Resulting unrooted gene trees were inputted into ASTRAL-III v.5.5.3 (Zhang et al., 2018) to infer the species tree. The support values were calculated using local posterior probabilities (LPP; Sayyari and Mirarab, 2016). Branches with LPP > 0.95 were considered as strongly supported.

2.7. Gene trees concordance analyses

Topological conflicts among gene trees and the species tree were explored using the software Phyparts (Smith et al., 2015) on the nuclear HybPyper dataset. This methodological approach consists of mapping, for each node or bipartition of interest, the level of concordance/discordance between the different individual gene trees on supporting the reference tree topology, in our case the ASTRAL species tree. First, gene and species trees were rooted with the online tool STRAW (Shaw et al., 2013). The outgroup selection was based on the species divergence rank recovered by the species ML tree inferred under concatenation. When the first preferred outgroup was missing, STRAW searches the next specified species. Then, Phyparts was run, considering only the branches of gene trees with more than 33% BS support. Finally, the script phypartspiecharts.py (https://github.com/mossmatters/MJPythonNotebooks) was used to summarize and map the Phyparts output on the ASTRAL species tree. Owing to computing limitations of Phyparts, we could not run the analysis for the entire set of 1055 nuclear gene trees of HybPiper dataset, therefore a subset of 750 random gene trees was selected.

2.8. Divergence time analysis

The divergence time analysis was performed on the phylogenetic tree obtained under concatenation from the nuclear HybPiper dataset. This tree was time-calibrated rescaling the branch lengths using the penalized likelihood method (Sanderson, 2002), implemented in the software treePL (Smith and O'Meara, 2012). This approach yields similar results than the software BEAST (e.g. Lagomarsino et al., 2016; Stubbs et al., 2018), but it runs faster on larger datasets. The dating procedure was divided into two main stages, which consisted on: (1) selection of the optimal model parameters; and (2) running the analysis with the optimal parameters selected and, additionally, accounting for the uncertainty in calibration points to obtain confidence intervals (95% CI) in the estimated node ages.

For both stages, five calibration points (CP) were used as node age constrains. One was a secondary dated node corresponding to the origin of Compositae family (69.56 Myr with a 95% CI 59.02–80.17; coded here as CP1) reported in Panero and Crozier (2016). The other four points were based on fossil records: (CP2) the capitulescence of *Raiguenrayun cura* that was dated at 47.5 Myr (Barreda et al., 2012), which was used to constrain the clade of subfamily Mutisioideae + subfamily Carduoideae as in the most recent interpretation by Panero and Crozier (2016), instead of placing it at the crown Compositae as in Barres et al. (2013); (CP3) the achenes identified as belonging to *Cirsium* with an age of 14 Myr (Mai, 1995), which were placed at the stem node of the *Carduus-Cirsium* clade following Barres et al. (2013); (CP4) the achenes assigned to *Arctium* (López-Vinyallonga et al., 2009) and dated at 8 Myr (Mai, 2001), which were placed at the stem node of *Arctium*, correcting the misplacement of this fossil in Barres et al. (2013), where it was used at the split of *A. minus* and *A. lappa*; and (CP5) the pollen of *Centaurea* type *Cyanus* dated at 6 Myr (Wagenitz, 1955; Ivanov et al., 2007), which was placed at the stem node of *Centaurea* subgenus *Cyanus*. This is a new, more precise and better resolved *Centaurea* fossil that the pollen and achene records of Mai (1995) and Popescu (2002) that were used in Barres et al. (2013), where they were placed at stem age of *Centaurea* due to their uncertain placement within the genus.

In the first dating stage, an initial "priming" run (prime command) was carried out to detect the optimal parameter settings (opt = 4, optad = 2, and optcvad = 2). Subsequently, the best smoothing rate was also evaluated through a random subsample and replicate cross-validation procedure (selecting thorough and randomcv commands), allowing varying it from 0.001 (cvstart) to 1000 (cvstop). The best smoothing value resulted in 0.1 after cross-validation analysis, which was selected from the lowest value of a Chi-Square test. The two runs of the first stage were run using the range of 95% CI value for the CP1 (minimum = 59.02 and maximum = 80.17) and the lower bound of fossil age estimations (minimum = fossil age) for CP2-CP5.

In the second stage, we used our own R (R Core Team, 2014) script (Appendix A) to obtain 10,000 random values for each CP in order to account for the uncertainty of the node age estimations. Specifically, the random values were generated under a normal distribution for CP1 (based on 95% CI estimations of Panero and Crozier, 2016; mean = 69.56 and SD = 3) and a lognormal distribution for CP2-CP5 (mean = fossil age and SD = 1.1). Once values were extracted, a new data frame was created in a spreadsheet (Appendix B), which was then used to randomly select 100 unique combinations of five numbers, i.e. one for each CP from the random pool of 10,000. Finally, 100 independent treePL analyses, each one with CP constrained to a single value (minimum = maximum), were ran using the same phylogenetic input tree and the optimized parameters found in the first stage. The resultant 100 dated tree files were modified to fit the format of TreeAnnotator v.1.7.5 (Drummond et al., 2012), and were introduced in this software to obtain a maximum clade credibility tree chronogram with median node heights and corresponding CIs.

3. Results and discussion

3.1. Performance of Hyb-Seq at subtribal level within tribe Cardueae

The Hyb-Seq NGS technique designed for Compositae (Mandel et al., 2014) is confirmed here as a powerful tool to reconstruct highly resolved phylogenies at deep taxonomic levels, in line with previous studies more focused on the methods

Table 3. Overview and results from previous phylogenetic studies (performed with Sanger sequencing data) and present study (performed with NGS data) focused on tribe Cardueae with emphasis in the subtribal classification. Subtribes are divided following the classification proposed here. In the subtribe rows, the number of species sampled in the corresponding study is indicated. Statistically supported subtribes are highlighted in green, considering supported subtribes those with a monophyletic pattern and bootstrap support values ≥ 70% (in ML and MP analyses) and Bayesian posterior probabilities ≥ 0.95 (in BI analysis). In case that the study used two phylogenetic inference methods, supported subtribes are considered when one or both methods showed a significantly supported monophyly. The nuclear dataset summarized corresponds to nuclear HybPiper dataset. Abbreviations used: BI = Bayesian Inference; C = chloroplast marker; H, H = Häffner and Hellwig; ML = Maximum Likelihood; MP = Maximum Parsimony; N = nuclear marker; Nº = number; S = Susanna et al.

Systematic studies of the tribe Cardueae at subtribal level										
	S (1995)	H, H (1999)	Garcia-Jacas et al. (2002)		Susanna et al. (2006)		Barres et al. (2013)		Present study	
Nº DNA markers	1	1	1	1	1	1+2	1	4	1055	87
DNA marker type	N	N	N	C	N	N+C	N	C	N	C
Inference method	MP	MP	MP	MP	BI	MP, BI	MP, BI	MP, BI	ML	ML
Total sampling	36	32	58	42	187	108	116	124	76	67
Carlininae	1	1	**10**	**9**	11	9	**1**	**9**	**6**	**5**
Echinopsidinae	1	1	**5**	3	**7**	**5**	**7**	**7**	**3**	**2**
Cardopatiinae	-	-	2	1	**2**	**2**	**2**	**2**	**2**	**2**
Dipterocominae	-	-	-	-	-	-	1	**1**	**1**	**1**
Berardiinae	-	1	1	1	1	1	1	1	**1**	**1**
Staehelininae	-	-	1	-	**5**	**4**	**3**	**3**	**4**	**4**
Xerantheminae	-	1	**3**	**3**	**8**	**7**	**7**	**7**	**4**	**3**
Onopordiinae	-	**5**	**3**	4	**11**	**9**	**11**	**11**	**4**	**4**
Carduinae	3	14	15	11	**19**	**15**	**24**	24	**7**	**6**
Saussureinae	1	3	8	2	**25**	5	5	5	**7**	6
Arctiinae	1	**2**	5	3	28	13	10	10	**6**	**2**
Centaureinae	**29**	**4**	**5**	5	**70**	**38**	**44**	**44**	**31**	**31**

(Mandel et al., 2015, 2017; Herrando-Moraira et al., 2018). As far as we know, this is the first study in Compositae that uses the COS loci workflow to reach taxonomic conclusions, specifically aimed to clarify the entangled delimitation of subtribes within Cardueae. The transition from traditional Sanger datasets, up to 5 markers, to NGS datasets, with 1142 markers (Table 3), has resulted in obtaining the most helpful dichotomous and confidently supported divergence pattern reported to date for Cardueae (Figs. 1–3, Supplementary Figs. S1 and S2).

On average 5,420,504 reads (SD = ± 3,461,361) were sequenced per species (Supplementary Table S2). From the total 1061 COS loci, 776 were finally recovered with PHYLUCE and 1055 with HybPiper, which resulted in matrices of trimmed aligned sequences of 492,549 bp (189,716 parsimony informative [PI] sites) and 332,260 bp (123,731 PI sites), respectively (Table 4). When parameters related to missing data were compared between PHYLUCE and HybPiper (see parameters 4 and 11 in Table 4), it was remarkable that PHYLUCE is less efficient than HybPiper in the sequence extraction process from on-target reads, as documented in Herrando-Moraira et al. (2018). However, phylogenetic trees did not differ in topology or branch supports between both extraction methods at subtribal level (Figs. 1 and 3, Supplementary Figs. S1 and S2), even though PHYLUCE uses a more restrictive procedure to remove potential paralogs than HybPiper (see Fig. 1 in Herrando-Moraira et al., 2018). As no differences were observed, results from HybPiper method were analyzed in deep to support taxonomic proposals. Interestingly, this lack of differences could be reflecting that paralogy incidence is relatively low in Cardueae. In agreement, the study of Jones et al. (in prep.), which tests the COS loci applicability for seven tribes of Compositae, found that Cardueae was the tribe with the lowest number of paralogs (average = 140, similar to the value found here, average = 130), in compared to others like Vernonieae (average = 256). Also in concordance, Huang et al. (2016) did not detect any recent whole genome duplication event in Cardueae, in contrast to other Compositae tribes, which could involve paralogy issues.

As other studies have shown, Hyb-Seq can provide nearly complete datasets of chloroplast genomes (e.g. Weitemier et al., 2014; Folk et al., 2015; Carlsen et al., 2018). Regarding the present study, we were able to retrieve 87 protein coding regions of the chloroplast genome from off-target reads, which after concatenation resulted in a matrix of 78,531 bp (4290 PI sites; Table 4). Indeed, the increase of phylogenetically informative characters is responsible of notably higher topological resolution and clade support in comparison with the previous results reported with Sanger sequencing data for Cardueae (Susanna et al., 1995, 2006; Häffner and Hellwig, 1999; Garcia-Jacas et al., 2002; Barres et al., 2013; Table 3). In a similar way, this effect has been already documented for other plant groups (e.g. *Pinus*, Parks et al., 2009; *Inga*, Nicholls et al., 2015), which have benefited tremendously from the advent of high-throughput sequencing techniques.

3.2. The new subtribal classification of tribe Cardueae

As it could be expected with our sampling, we recovered a monophyletic pattern for subfamily Carduoideae as currently defined (tribes Dicomeae, Oldenburgieae, Tarchonantheae, and Cardueae; cf. Funk et al., 2009). Tribe Cardueae results again a

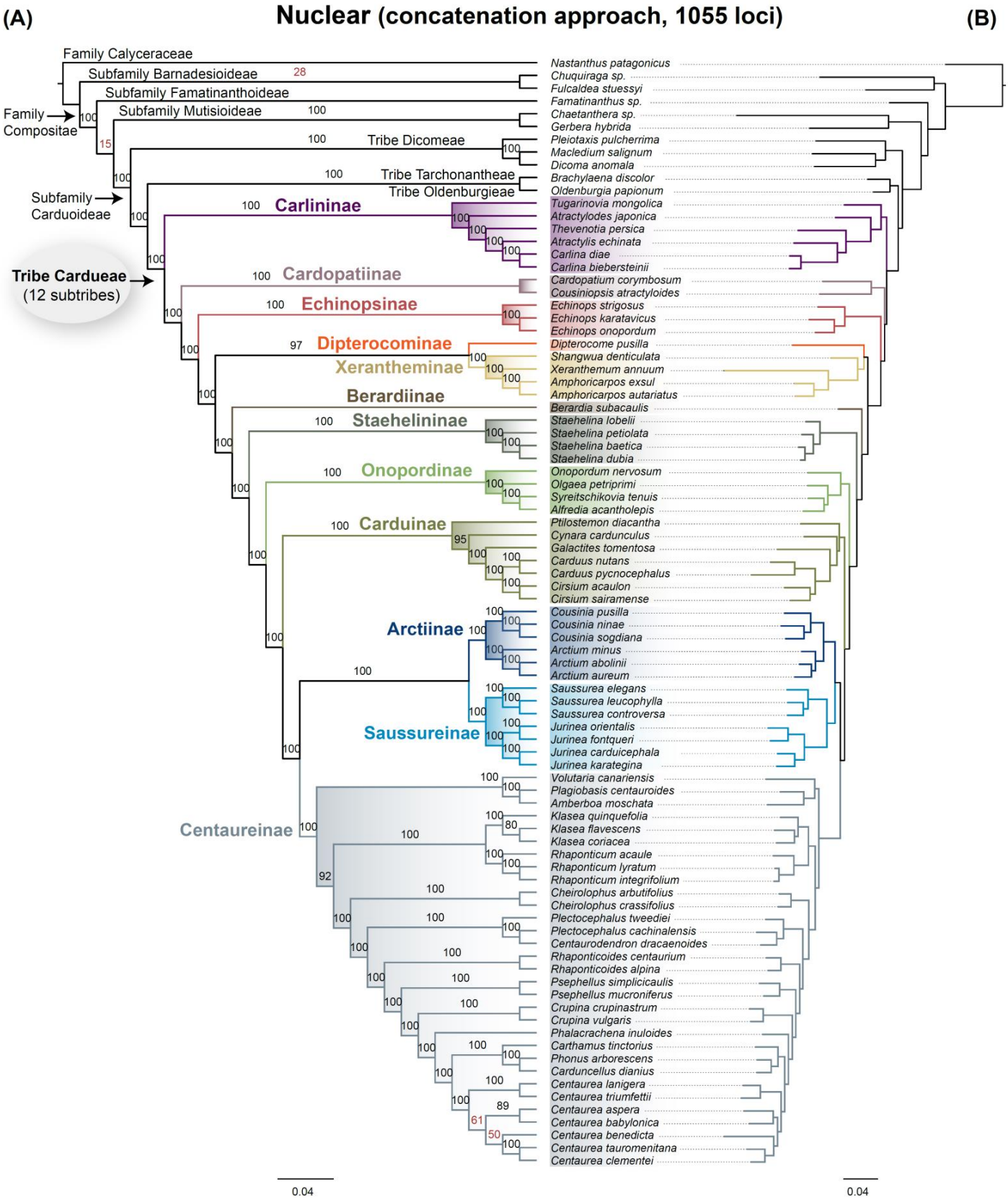


Fig. 1. Phylogenetic reconstruction obtained with HybPiper target extraction method and under the concatenation approach (maximum likelihood analysis performed with RAxML), showing the evolutionary relationships among 12 newly defined subtribes within tribe Cardueae. (A) cladogram, (B) phylogram. Branch labels indicate bootstrap (BS) support values, those < 70% (in red) are considered statistically unsupported.

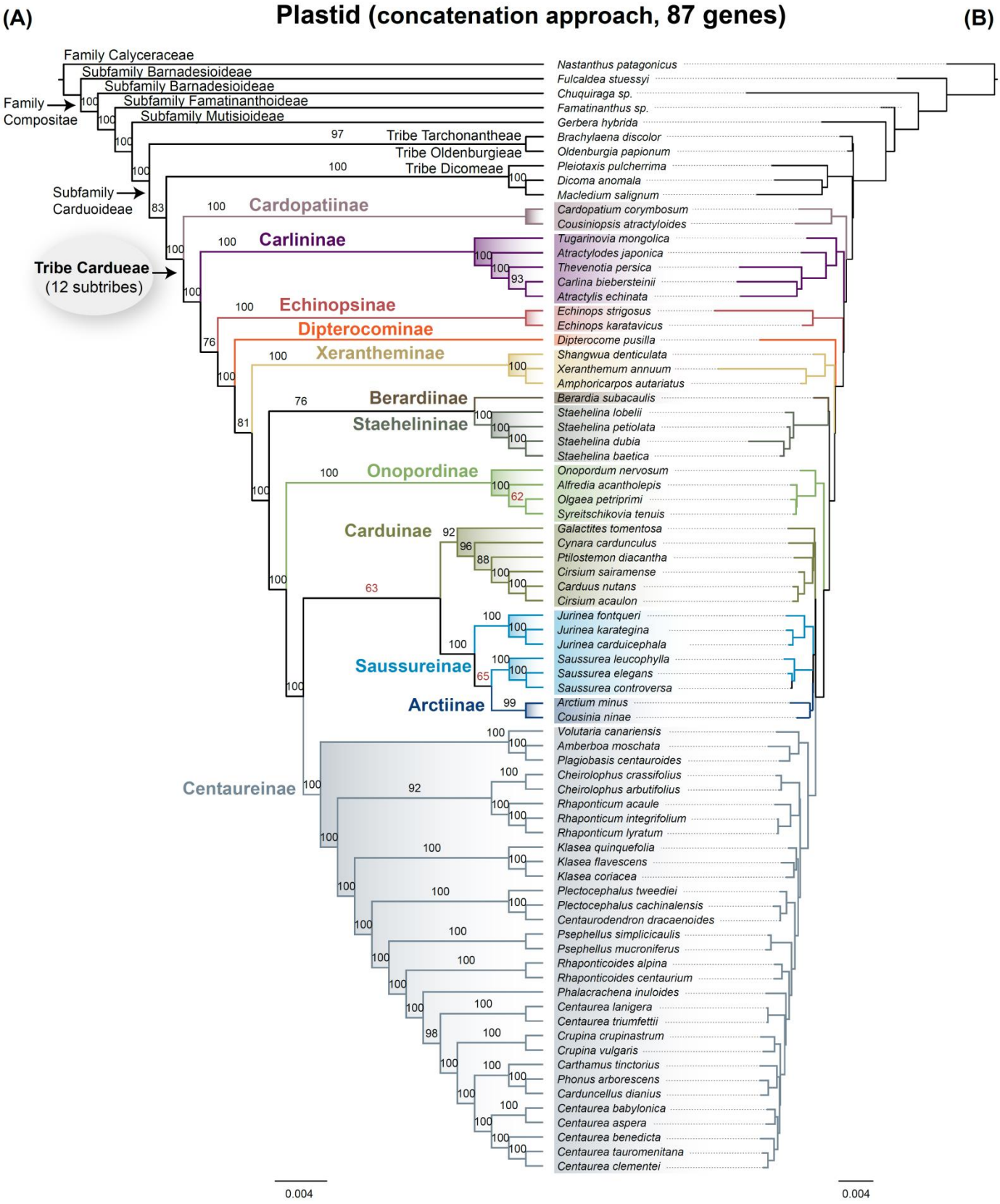


Fig. 2. Phylogenetic reconstruction inferred from the chloroplast dataset (87 coding regions) under the concatenation approach (maximum likelihood analysis performed with RAxML), showing the evolutionary relationships among 12 newly defined subtribes within tribe Cardueae. (A) cladogram, (B) phylogram. Branch labels indicate bootstrap (BS) support values, those < 70% (in red) are considered statistically unsupported

robust taxonomic unit, in agreement with former studies based on morphology and molecular data (Bremer, 1987, 1994; Jansen et al., 1990, 1991; Karis et al., 1992; Kim et al., 1992; Susanna et al., 1995, 2006; Garcia-Jacas et al., 2002; Barres et al., 2013). Specifically, the clade of Cardueae presented high support values in all the analyses (Figs. 1–3). The alternative option of rising up some subtribes to a tribal rank (see Table 1) is once again discarded. As suggested by Susanna et al. (2006), it is unpractical to fragment a natural group that is strongly phylogenetically consistent and can be easily recognized by macromorphological characters. In this way, the classification criteria for the definition of Compositae tribes are also maintained homogeneously along the family (Funk et al., 2009).

At subtribal level, the high resolution of phylogenies provided by the Hyb-Seq data makes possible the proposal of a new classification. Subtribal limits established here are based on the application of the integrative taxonomy principles (Dayrat, 2005; Schlick-Steiner et al., 2010): firstly, considering the morphological entities to be recognized (Susanna and Garcia-Jacas, 2007, 2009), and then verifying their monophyly with molecular data. Following this model, a total of 12 subtribes are suggested. Five of them had been already previously recognized and are also confirmed with our results (Carlininae, Echinopsinae, Cardopatiinae, Carduinae, and Centaureinae; Susanna et al., 2006); seven new subtribes result from partitioning the former wide and paraphyletic subtribe Carduinae into new monophyletic subtribes based on informal morphological groups (Dipterocominae, Xeranthemeinae, Berardiinae, Staehelininae, Onopordinae, Arctiinae, and Saussureinae) most of them outlined by Susanna and Garcia-Jacas (2009). These new subtribes are highly heterogeneous in terms of species richness, encompassing from relict single lineages to highly recently radiated groups. For instance, Dipterocominae and Berardiinae are defined as monotypic subtribes, comprising only *Dipterocome pusilla* and *Berardia subacaulis*, respectively. On the other hand, Arctiinae and Saussureinae harbor ca. 600 and 550 species, respectively (Susanna and Garcia-Jacas, 2007). For a complete description of subtribes see below *Taxonomic proposal*. The new phylogenetic reconstruction of high-level lineages presented here will be used as a classification basis for future studies on the systematics of Cardueae.

Certainly, the historical overview of the tribe (Table 1) reveals that Carlininae, Echinopsinae, and Centaureinae are the best morpho-molecular defined entities, which have been long recovered with statistical support since the first phylogenetic studies (Table 3). In accordance, we obtained high clade supports for these subtribes (BS = 100 and LPP = 1; Figs. 1–3). Carlininae initially showed a monophyletic pattern in Garcia-Jacas et al. (2002). However, the incorporation of *Tugarinovia mongolica* broke the monophyly of the subtribe in Susanna et al. (2006). This was probably caused by the fact that it is a dioecious species, a trait unique in the tribe (Susanna and Garcia-Jacas, 2009) and relatively uncommon in angiosperms (6% of species; Renner and Ricklefs, 1995). It has been suggested that dioecy has the negative impact of decreasing diversification rates (Heilbuth, 2000; Kay et al., 2006) with consequent phylogenetic misplacements (Vamosi et al., 2003). In Barres et al. (2013) and in our present study, the increase in the number of molecular markers aided to recover the monophyly of Carlininae, including *Tugarinovia mongolica* within this subtribe as a sister lineage to the rest of members. Echinopsinae presented a persistent monophyly across all molecular studies (Table 3), except for the chloroplast-based phylogeny of Garcia-Jacas et al. (2002), in which the limited sequence variation found in chloroplast *mat*K gene could have been the responsible for the low branch support

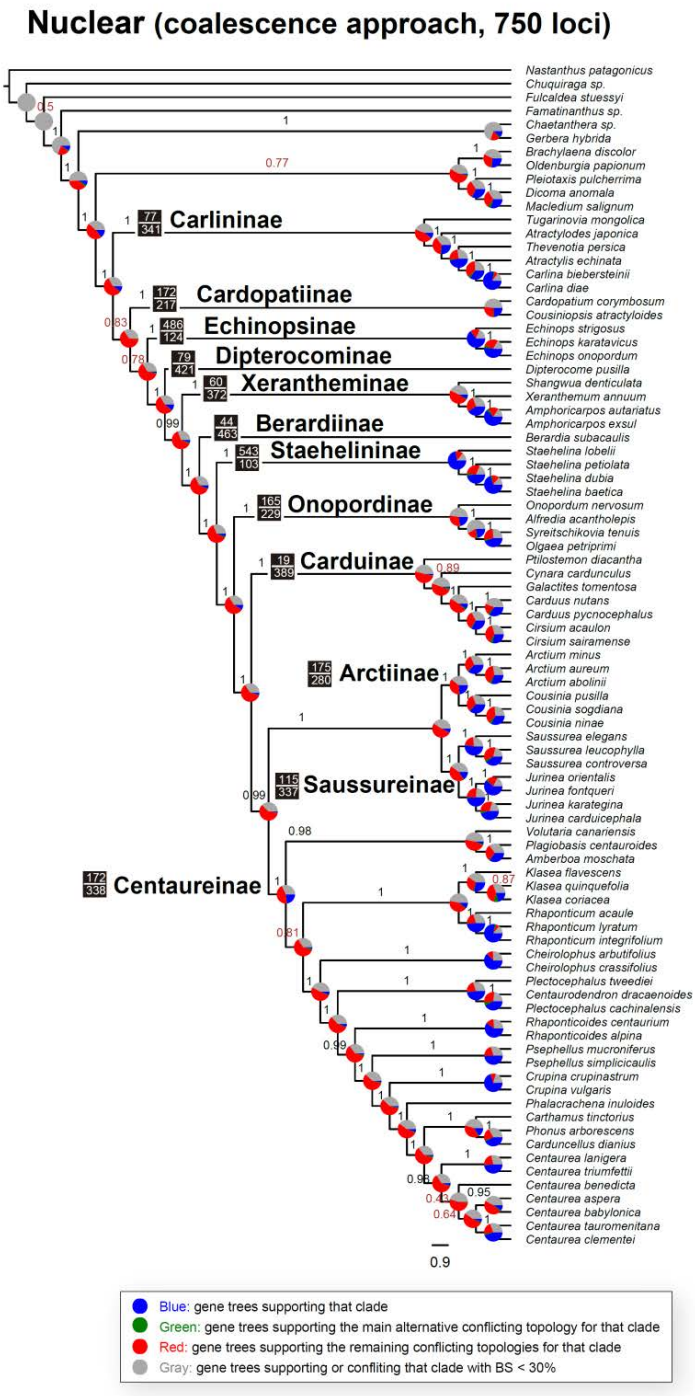


Fig. 3. Phylogenetic trees inferred under coalescence approach (estimated with ASTRAL) for the nuclear HybPiper, composed by 1055 COS loci. Branch labels indicate local posterior probabilities (LPP), values < 0.95 (in red) are considered statistically low supported. Results from PhyParts of conflicting and concordant gene trees are also outlined for the nuclear HybPiper dataset (see text for details). The pie charts represent the proportion of the four categories (blue, green, red, and gray) for each node. For the subtribal clades, the number above and below branches indicate the number of gene trees concordant and in conflict, respectively, with that clade respect the species tree.

Table 4. Summary metrics of sequence extraction performance for nuclear datasets (targeting 1061 COS targets; Mandel et al., 2014) using PHYLUCE (Faircloth, 2015) and HybPiper (Johnson et al., 2016) methods, and chloroplast dataset (targeting coding regions). Parameters from 4 to 11 are calculated based on dataset ~~3~~. Nº of used loci". Parameter values were extracted with FASconCAT-G v.1.02 (Kück and Longo, 2014) and AMAS (Borowiec, 2016) programs. Abbreviations used: bp = base pairs; max = maximum; min = minimum; Nº = number; SD = standard deviation. *loci recovered for more than three species or higher than 3 bp length (see text for details).

Parameters	Nuclear datasets		Chloroplast dataset
	PHYLUCE method	HybPiper method	
1. Nº of species included	87	87	j77
2. Nº of recovered loci (% respect COS targets)	902 (85.0)	1057 (99.6)	87
3. Nº of used loci* (% respect COS targets)	776 (73.1)	1055 (99.4)	87
4. Nº of captured loci in ≥ 90% of species (%)	13 (1.7)	729 (69.1)	85 (97.7)
5. Average of recovered loci per species (SD; min-max)	363 (59; 169–455)	924 (174; 21–1021)	84 (2; 72–86)
6. Average of species recovered per loci (SD; min-max)	41 (26; 4–82)	76 (14; 4–86)	74 (12; 4–77)
7. Mean alignment length per locus in bp (SD; min-max)	634 (360; 90–3825)	314 (184; 23–1589)	902 (1331; 89–6851)
8. Length of concatenated matrix in bp	492,549	332,260	78,531
9. Nº of variable sites in the concatenated matrix (%)	296,609 (60.2)	169,903 (51.1)	10,067 (12.8)
10. Nº of parsimony informative sites in the concatenated matrix (%)	189,716 (38.5)	123,731 (37.2)	4290 (5.5)
11. Proportion of missing data in the concatenated matrix (%)	62.6	16.6	18.2

of Echinopsinae. Similarly, the monophyly of subtribe Centaureinae has been fully supported in former phylogenetic studies (Table 3), with the exception of the chloroplast tree obtained with the *mat*K gene by Garcia-Jacas et al. (2002).

Although Cardopatiinae are a relatively newly recovered subtribe (Susanna et al., 2006), it has resulted also in a strong-supported clade in recent phylogenies with both nuclear and chloroplast sequence data (Table 3), besides its morphological singularity (Susanna and Garcia-Jacas, 2007, 2009). In agreement, we can confirm here that the two species of Cardopatiinae (*Cardopatium corymbosum* and *Cousiniopsis atractyloides*) constitute a well-defined evolutionary segregation (BS = 100 and LPP = 1; Figs. 1–3), therefore its taxonomic rank as subtribe seems appropriate.

Conversely, the new seven subtribes have not always been well delineated in Sanger-based phylogenies, at least with enough confidence to propose a formal subtribal splitting (see Table 3). From the morphological informal groups of Carduinae described in Susanna and Garcia-Jacas (2009), the *Xeranthemum* group is one of the best supported of all datasets and analyses (BS = 100 and LPP = 1; Figs. 1–3). Thus, its consideration as subtribe Xerantheminae can be corroborated. Our results are not surprising, because since the first phylogenies the Xerantheminae species were recovered as a monophyletic and statistically highly supported group, both with nuclear and chloroplast data (Table 3).

One unexpected finding was found in the former *Xeranthemum* group, which is the phylogenetic isolated position of *Dipterocome pusilla*. This species had not been classified in Cardueae until the study of Anderberg et al. (2007), confirmed by Barres et al. (2013), in which it was clustered with the rest of Xerantheminae members in the combined cpDNA (*mat*K, *ndh*F, *rbc*L, and *trn*L-*trn*F) tree, but without support in a phylogeny of the ITS nuclear region. However, in the present study we found that *Dipterocome* is placed in a different clade from Xerantheminae in four of the five reconstructed phylogenies (Figs. 2 and 3, Supplementary Figs. S1 and S2). Only in the tree reconstructed under concatenation with the nuclear HybPiper dataset *Dipterocome* was grouped with Xerantheminae (BS = 97; Fig. 1), in agreement with the results of Barres et al. (2013). This may be caused by the bias produced by concatenation analysis of NGS datasets with a large number of loci (Kubatko and Degnan, 2007; Edwards et al., 2016), which can result in long branch attraction artifacts (Liu et al., 2015). Morphologically, Susanna and Garcia-Jacas (2009) pointed out that floral characters of *Dipterocome* show close relationship with the *Xeranthemum* complex. Nevertheless, a histological analysis of the capitula structure of *Dipterocome* has shown that the purported ~~florets~~" of *Dipterocome* are actually one-flowered capitula (Susanna, unpubl. data) grouped in second-order synflorescences, an extremely infrequent structure in the tribe shared only by the genus *Echinops*. Overall, it seems that there is enough molecular and morphological evidence to classify *Dipterocome* in a separate monotypic subtribe.

Two other controversial genera considered to belong to the former Carduinae are *Berardia* and *Staehelina*. Although both genera do not present obvious morphological affinities (Susanna and Garcia-Jacas, 2009), some molecular phylogenies suggested that they could be evolutionary closely related. The genus *Staehelina* has long been reported as a genetically isolated group within the tribe, always showing a monophyletic pattern throughout all phylogenetic reconstructions (Table 3). Regarding its relationships with *Berardia*, the combined nuclear plus chloroplast tree in Susanna et al. (2006) grouped both genera in an unsupported clade; but only considering nuclear data (ITS), they appeared not directly linked but as adjacent lineages as detected in nuclear and chloroplast phylogenies of Barres et al. (2013). Here, we also identified the same two phylogenetic signals. Even though, only one of five phylogenetic analyses (chloroplast dataset analyzed under the concatenation approach) resulted in a moderately-high grouping of both genera (BS = 76; Fig. 2), and the rest separated *Berardia* and *Staehelina* (Figs. 1 and 3, Supplementary Figs. S1 and S2). As we hypothesized for the case of *Dipterocome*, *Berardia* could be misplaced in some chloroplast phylogenies by a long branch attraction effect in concatenated analyses due to its relict character without any known alive direct congener. Conclusively, *Berardia* and *Staehelina* could be considered as separate new subtribes Berardiinae and Staehelininae based on the new molecular evidence presented.

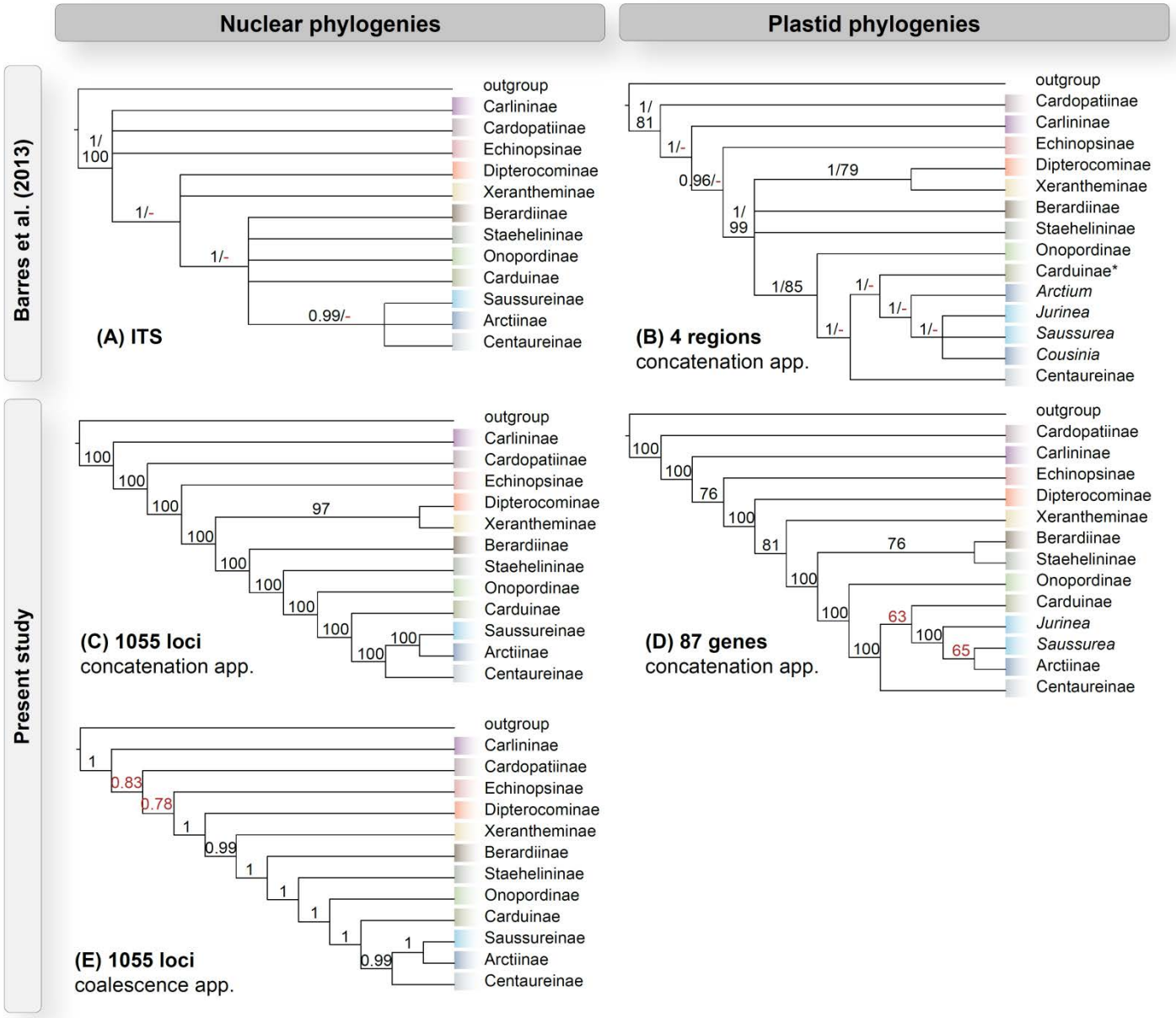


Fig. 4. Summary phylogenetic trees recovered in Barres et al. (2013) and present study showing evolutionary relationships among 12 subtribes of tribe Cardueae. On the left, we show phylogenies reconstructed with nuclear sequence data (A, C, E), and on right (B, D, F), those with plastid data. Trees redrawn from Barres et al. (2013) were obtained under maximum parsimony (MP) and Bayesian inference (BI) approaches, using a nuclear dataset (A) composed by sequences of internal transcribed spacer marker (ITS) and the chloroplast dataset (B) composed by a combination of sequence data from genes *mat*K, *ndh*F, *rbc*L and intergenic spacer *trn*L-*trn*F. Phylogenetic trees inferred here derived from the nuclear HybPiper dataset composed by 1055 COS loci (C, E) and the chloroplast dataset composed by 87 coding regions (D, F), both datasets analyzed under the concatenation approach (app.; maximum likelihood [ML] analysis performed with RAxML; C, D) and coalescence approach (estimated with ASTRAL; E). Branch labels indicate: bootstrap support (BS) values in case of MP and ML analysis; posterior probabilities (pp) in BI; and local posterior probabilities (LPP) in case of under coalescence. Values with BS < 70% and pp or LPP < 0.95 (in red) are considered statistically low supported. Note that in trees of Barres et al. (2013), the low supported branches were collapsed in the redrawn version. The genus *Saussurea* and *Jurinea* belong to Saussureinae, and *Arctium* and *Cousinia* to Arctiinae. *not monophyletic subtribe.

The *Onopordum* group is another lineage of former Carduinae with a consistent morpho-genetic historical division (Susanna and Garcia-Jacas, 2009). Our genomic phylogenies provide evidence for the monophyly of the *Onopordum* group (BS = 100 and LPP = 1; Figs. 1–3) as other previous studies (Table 3), except in Garcia-Jacas et al. (2002) in which the phylogenetic resolution power of the chloroplast *mat*K region seems limited as mentioned above for other groups. Hence, it seems reasonably to assign the subtribal rank as Onopordinae.

One of the largest and most difficult complexes to delimit in former Carduinae is the *Carduus-Cirsium* group (ca. 500 species; Susanna et al., 2006). Its complicate delimitation was mainly attributed by Susanna et al. (2006) to the fact that the group encompasses a remarkable variety of life strategies; annual or biennial species (e.g. *Carduus*, *Galactites*, *Picnomon*, *Silybum*, or *Tyrimnus*) or perennials (as many *Cirsium*, *Cynara*, *Lamyropsis*, or *Ptilostemon*). Certainly, different generation times that affect mutation rates may hinder a proper phylogenetic comparison

among species as detected in some cases in Cardueae (see López-Vinyallonga et al., 2009 and references therein). The overview of previous phylogenetic studies reflects the difficulty of reconstructing the *Carduus* group as monophyletic, for which additional sampling is necessary to resolve generic delimitations among these genera. The monophyly of the *Carduus* group was only achieved in three from the total eight Sanger-based phylogenies (Table 3). In contrast, we found a strong phylogenetic signal with Hyb-Seq data (BS = 92–100 and LPP = 1; Figs. 1–3). This supports the morphological hypothesis of Susanna et al. (2006) and subtribal delimitation of Carduinae. Susanna and Garcia-Jacas (2009) suggested a posterior division of the *Carduus* group into *Cynara* and *Carduus-Cirsium* complexes, but our phylogenies did not show a strictly bifurcating divergence pattern for these complexes (Figs. 1–3). Thus, we would reject to formalize their segregation at subtribal rank.

The *Arctium-Cousinia* and *Jurinea-Saussurea* complexes are other classically entangled groups of former Carduinae. Although both conform well defined assemblies by morphology (Susanna and Garcia-Jacas, 2009), there has not been any Sanger-based phylogeny able to correctly recover the morphological pattern as monophyletic lineages (Table 3). Only two isolated molecular studies, Häffner and Hellwig (1999) and Susanna et al. (2006), reported *Arctium-Cousinia* and *Jurinea-Saussurea* as monophyletic, respectively. In recent times, the application of Hyb-Seq technique and the COS loci set of Mandel et al. (2014) for Compositae, aided to shed light on phylogenetic connections of these four genera. Phylogenomic trees performed in Herrando-Moraira et al. (2018) showed a clear sister relationship between *Arctium-Cousinia* and *Jurinea-Saussurea*, which also formed two highly supported clades. In agreement, on a wider taxonomic sampling of the tribe, we found the same monophyletic pattern for both complexes under concatenation and coalescence approaches with nuclear data (BS = 100 and LPP = 1; Fig. 1 and 3). However, the chloroplast phylogenies analyzed under concatenation recovered an unsupported clade of *Saussurea-Cousinia-Arctium* (Fig. 2) with *Jurinea* as a separate lineage, hence leaving the complex *Jurinea-Saussurea* as a paraphyletic assembly (Fig. 2). Similarly, a morphologically incongruent grouping pattern has been also recovered in the Sanger-based chloroplast phylogenies of Barres et al (2013), where the genera *Saussurea-Jurinea-Cousinia* were clustered together with support and *Arctium* was placed as a sister clade.

Overall, it seems that even with massive Hyb-Seq data, a clear cyto-nuclear discordance exists between evolutionary histories inferred for *Arctium-Cousinia* and *Jurinea-Saussurea* groups. Two factors could explain such pattern of incongruent genealogies (Morales-Briones et al., 2018), one more related with tree inference methods (phylogenetic error), and the other with the inherent evolutionary history of the genera (incomplete lineage sorting [ILS] or hybridization). Firstly, it is stated that using gene fragments with limited phylogenetic signal may derive in an increase of both stochastic and systematic errors in reconstructed phylogenies (Lemmon et al., 2009, Lemmon and Lemmon, 2013; Carlsen et al., 2018). Hence, if we compare our nuclear vs. chloroplast datasets, a large difference exists in terms of phylogenetic informativeness, for instance, in PI sites (ca. 38% vs. 5.5%; Table 4) or the average branch length of trees (0.04 vs. 0.004). Chloroplast data from protein coding regions, in comparison with hundreds of nuclear loci, could be providing limited phylogenetic information, which may explain the differences in the resulting tree topologies.

Secondly, ILS or reticulation events may explain conflicting topologies and gene tree incongruences (García et al., 2017). The short internal branches observed at the base of genera in chloroplast tree from concatenation analysis (Fig. 2B) could match the signals left by ILS, i.e. a scenario of a rapid diversification of genera in a short timeframe (rapid bifurcating cladogenesis) or a simultaneous polychotomous divergence (hard polytomy). Past hybridization events could also have taken place as suggested by the concatenation chloroplast tree topology (Fig. 2). The chloroplast genome of the ancestral *Saussurea* lineage could have been lost due to an organelle capture (Stegemann et al., 2012), thus, the currently observed topology is that retained from an ancestral hybridization with the *Arctium-Cousinia* lineage. Cyto-nuclear discordances due to introgressive hybridization have been suggested in other plant groups like *Ficus* through chloroplast NGS phylogenies (Bruun-Lund et al., 2017). However, more explicit methods to address these processes should be performed (e.g. Folk et al., 2017; García et al., 2017; Mitchell et al., 2017; Knowles et al., 2018; Morales-Briones et al., 2018), in addition to using a wider species sampling.

Taking into account all shortcomings of chloroplast data, it would be reasonable to give more weight to nuclear results, which were also in agreement with the morphological hypothesis, and accept two new subtribes: Arctiinae (comprising *Arctium-Cousinia*) and Saussureinae (*Saussurea-Jurinea*).

3.3. Phylogenetic relationships among subtribes

Undoubtedly, one of the major improvements that we obtained using Hyb-Seq data is the high phylogenetic resolution recovered along the backbone of the Cardueae tree (Fig. 4). Although in most recent Sanger-based phylogenies, 8 of the total 12 circumscribed subtribes were already monophyletic and well-defined (see Table 3; Barres et al., 2013), their interrelationships remained unresolved until now. In Barres et al. (2013), phylogenies inferred under maximum parsimony were completely undefined for the subtribal nodes. Similarly, under Bayesian inference most of the subtribe branches were anchored in polytomic structures, especially in the nuclear ITS phylogeny (Fig. 4A). In contrast, our phylogenetic reconstructions showed a significantly supported bifurcating pattern (Figs. 4C, D, and E).

Subtribes Carlininae, Cardopatiinae, and Echinopsinae were placed as the first diverging lineages within Cardueae near the root of the tribe. Unfortunately, we could not conclusively identify the sister subtribe to the rest with confidence, due to a cyto-nuclear discordance. In the nuclear tree inferred under concatenation, Carlininae were sister to the rest of subtribes (Fig. 4C). However, individual gene trees show conflicting topologies at this part of the tree (Figs. 3 and 4E); and in the chloroplast tree the Cardopatiinae were sister to the rest of subtribes (Fig. 4D). Successive sister clades recovered with all reconstruction methods occur in the following order: Dipterocominae, Xerantheminae, Berardiinae, Staehelininae, and Onopordinae. Note that, as mentioned above, Dipterocominae appeared nested with Xerantheminae and Berardiinae with Staehelininae in the nuclear and chloroplast concatenated analyses, respectively. The grouping pattern of the rest of subtribes was slightly different on

nuclear and plastid phylogenies. In nuclear trees, Carduinae was placed as sister to Arctiinae-Saussureinae-Centaureinae clade (BS = 100 and LPP = 0.99), which was subsequently bifurcated in Arctiinae-Saussureinae (BS = 100 and LPP = 1) and Centaureinae clade (Figs. 4C, E). In the chloroplast tree, Carduinae-Arctiinae-Saussureinae-Centaureinae were clustered in a supported clade (BS = 100) that internally formed an unsupported group of Carduinae-Arctiinae-Saussureinae (BS = 63) with Centaureinae as sister to the others (Fig. 4D).

Here, despite the finding that Hyb-Seq data offered bifurcating reconstructions, we detected that branch lengths recovered were relatively short (Figs. 1B, 2B). This fact may confirm the previous suspicion of a rapid subtribal lineage divergence, which would explain the traditional difficulty to resolve this part of the tree. The Phyparts analysis allowed us to go one step further and inspect the degree of gene tree discordance along nodes of the Cardueae tree. Results of Phyparts indicated that most of gene tree topologies are in conflict with the species tree in internal subtribal nodes (great red proportion; Fig. 3), in contrast to shallow nodes of genera or species relationships that showed more congruent gene tree histories (more proportion of blue; Fig. 3). We also detected that a considerable proportion of gene trees were poorly supported and uninformative (gray color in Fig. 3 for gene tree branches with BS < 30). Considering all results, possible sources of gene tree discordance at the backbone could derive from: (1) insufficient phylogenetic signal in sampled loci (e.g. short loci length, high missing positions or taxa as a result of capture failure, or poorly variable loci; Villaverde et al., 2018); (2) extinction of critical lineages; (3) rapid splitting of major subtribal clades; or (4) gene flow among ancestral lineages or ILS effect (Morales-Briones et al., 2018). It is important to note that although individual gene tree bipartitions were mainly in conflict at subtribal level, the coalescence species tree was, in general, not weakly supported (Fig. 3). Thus, this limitation is not necessary problematic to the aim of subtribal delimitation intended here.

3.4. Dating framework for tribe Cardueae

Fig. 5 shows the divergence time estimates of Cardueae based on the nuclear tree obtained with HybPiper and the concatenation approach (1055 COS loci). Median age values and 95% CI are presented for each node in Supplementary Table S3. Our dating analysis revealed that tribe Cardueae could have originated between Late Eocene and Early Oligocene 34.12 Myr (29.97–40.04 95% CI). This time frame is in a middle range of the estimated age origin for the tribe in previous dated phylogenies that included several Cardueae members. On the one hand, the studies conducted by Kim et al. (2005) and Panero and Crozier (2016) reported a younger age of the tribe, 24–29 Myr and 22 Myr (10.61–33.35 95% CI), respectively. On the other hand, Barres et al. (2013) and Huang et al. (2016) found an older tribe median age with values of 40.15 Myr (35.80–44.26 95% CI) and ca. 42 Myr, respectively. Taxon sampling, calibration points, and molecular markers used could be responsible for dating differences between studies.

Notably, our estimation for the origin of tribe Cardueae matches particularly well the ~~"O~~i-1 Glaciation" that occurred at the transition between the Eocene and the Oligocene, when the Antarctic continental ice-shield was formed (Zachos et al., 2001). This event has been linked to a decline in atmospheric CO_2 over a period of 0.5 Myr (Pearson et al., 2009) and brought world-scale major climatic changes including cooler temperatures, increased seasonality, and aridity (Dupont-Nivet et al., 2007; Eldrett et al., 2009). Such global changes, in addition to produce large reorganization of paleo-biome distributions (e.g. Pound and Salzmann, 2017), would have acted as stimuli for plant evolution in other cold- or arid-adapted groups that also date from the Eocene/Oligocene boundary: tribe Gentianeae of Gentianaceae (crown node age = 34 Myr; Favre et al., 2016), tribes Aveneae, Poeae, Stipeae, and Triticeae of Poaceae (crown node ages = 32, 35, 37, and 33.8 Myr, respectively; Schubert et al., 2018), the entire family Brassicaceae (crown node age = 32.4 Myr; Hohmann et al., 2015), or genus *Hypericum* of Hypericaceae (Meseguer et al., 2015).

The subtribes originated in two different main time periods. Subtribes Carlininae, Cardopatiinae, Dipterocominae, Xerantheminae, and Berardiinae probably originated during the Oligocene (see Fig. 5 and Supplementary Table S3 for age of subtribes), whereas Echinopsinae, Staehelininae, Onopordinae, Carduinae, Arctiinae, Saussureinae, and Centaureinae originated during Middle-Late Miocene (Fig. 5 and Supplementary Table S3).

If we draw in parallel dating results obtained here and the ones obtained by Barres et al. (2013), which is the most comparable study in terms of taxon sampling, we can observe that tribe and subtribe age estimates were quite older in the latter than those we have found here (see Fig. 6). At tribal level, the different placement of the fossil *Raiguenrayun cura* (dated to 47.5 Myr, placed here to constrain subfamily Mutisioideae + subfamily Carduoideae; CP2 in Fig. 5) does not explain why we obtained a most recent tribal origin than Barres et al. (2013), who placed it in a more basal node as it is the crown node of Compositae family. Probably, the fact that we now have a more complete sampling at the basal part of the tree, including more representatives of subfamilies other than Carduoideae and more representatives of tribes within Carduoideae other than Cardueae, results in greater time distances between the origin of the family Compositae and tribe Cardueae. At subtribal level, the greatest differences between non-overlapping age ranges were detected for subtribes Onopordinae and Centaureinae, for which the median age differed among studies in 9.32 Myr and 15.38 Myr, respectively. The different number and nature of molecular markers used would be one of the main causes of the distinct dating results. We employed a set of 1055 biparentally inherited, low-copy nuclear loci that potentially provides more sequence variation, while Barres et al. (2013) used four chloroplast regions that are usually maternally inherited and have low mutation rates, which would explain the older age values reported in their estimations. In addition, it should be also noted that for those nodes constrained with minimum ages based on fossil calibration points, we obtained a better fitting between the minimum age assigned and the final dating result recovered in comparison with Barres et al. (2013); for instance, our CP5 constrained at 6 Myr resulted in 5.91 Myr (5.21–7.01 95% CI), and in Barres et al. (2013) the CP used at the stem node of *Centaurea* constrained at 5 Myr resulted in 15.54 Myr (12.10–19.96 95% CI).

In a general overview of the dated phylogenetic tree it is noteworthy that CI interval limits are wider and more overlapping among nodes at the base of the tree than those at the tips (Fig. 5). Accordingly, lower precision and resolution was obtained for

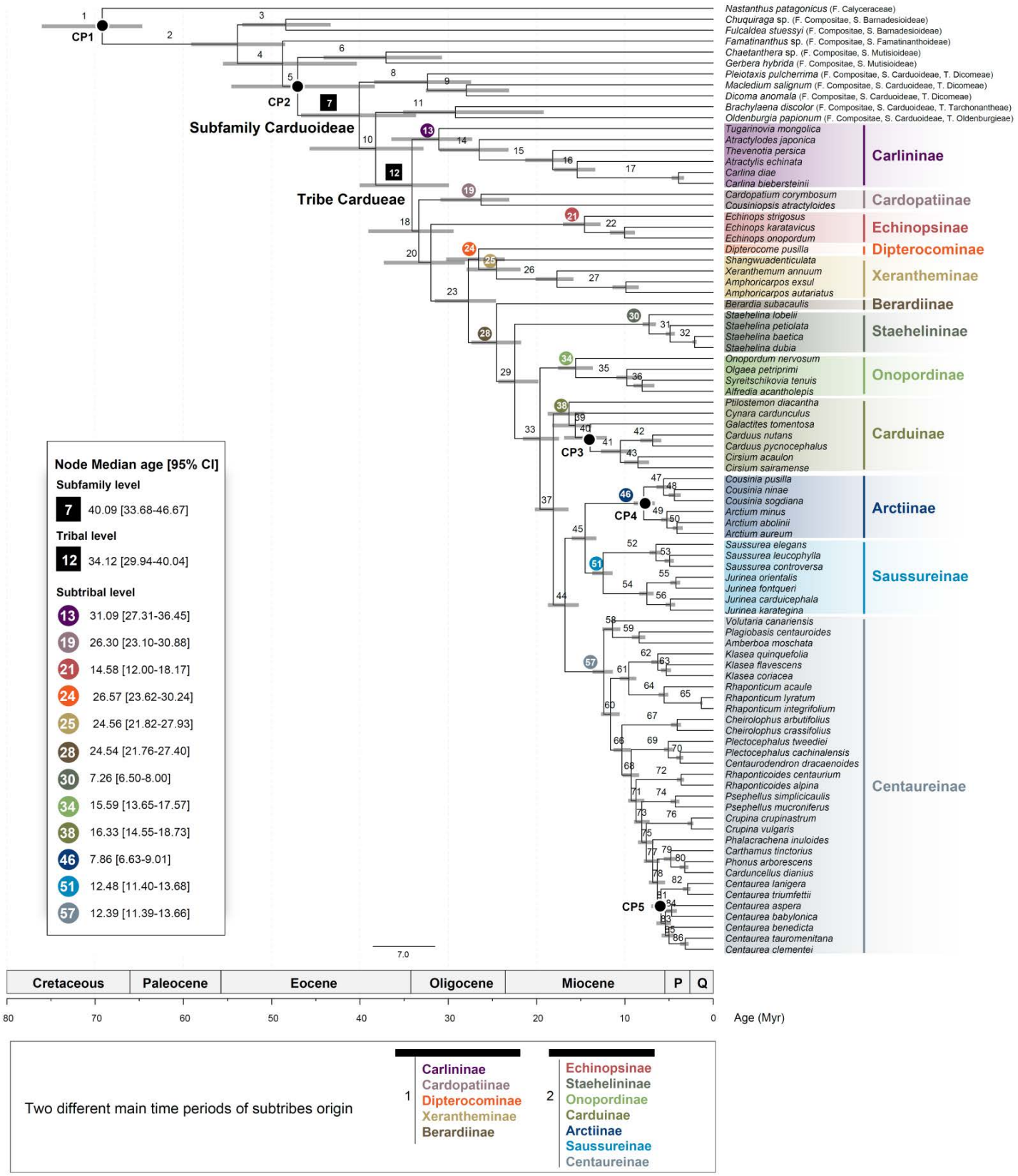


Fig. 5. Time-calibrated phylogeny of 12 newly defined subtribes within tribe Cardueae. The phylogenetic tree estimated with nuclear HybPiper dataset (1055 COS loci) and under concatenation approach (maximum likelihood analysis performed with RAxML) was used as the input tree for the penalized likelihood dating analysis using treePL. Gray bars on nodes show the 95% of confidence intervals (CI). Black circles on nodes represent the calibration points (CP) used in dating analysis. Numbers above branches are the node codes. See Supplementary Table S3 for numeric details of median estimated ages and Lower 95% CI and Upper 95% CI for each node. In the scale axis, "P" and "Q" correspond to Pliocene and Quaternary, respectively.

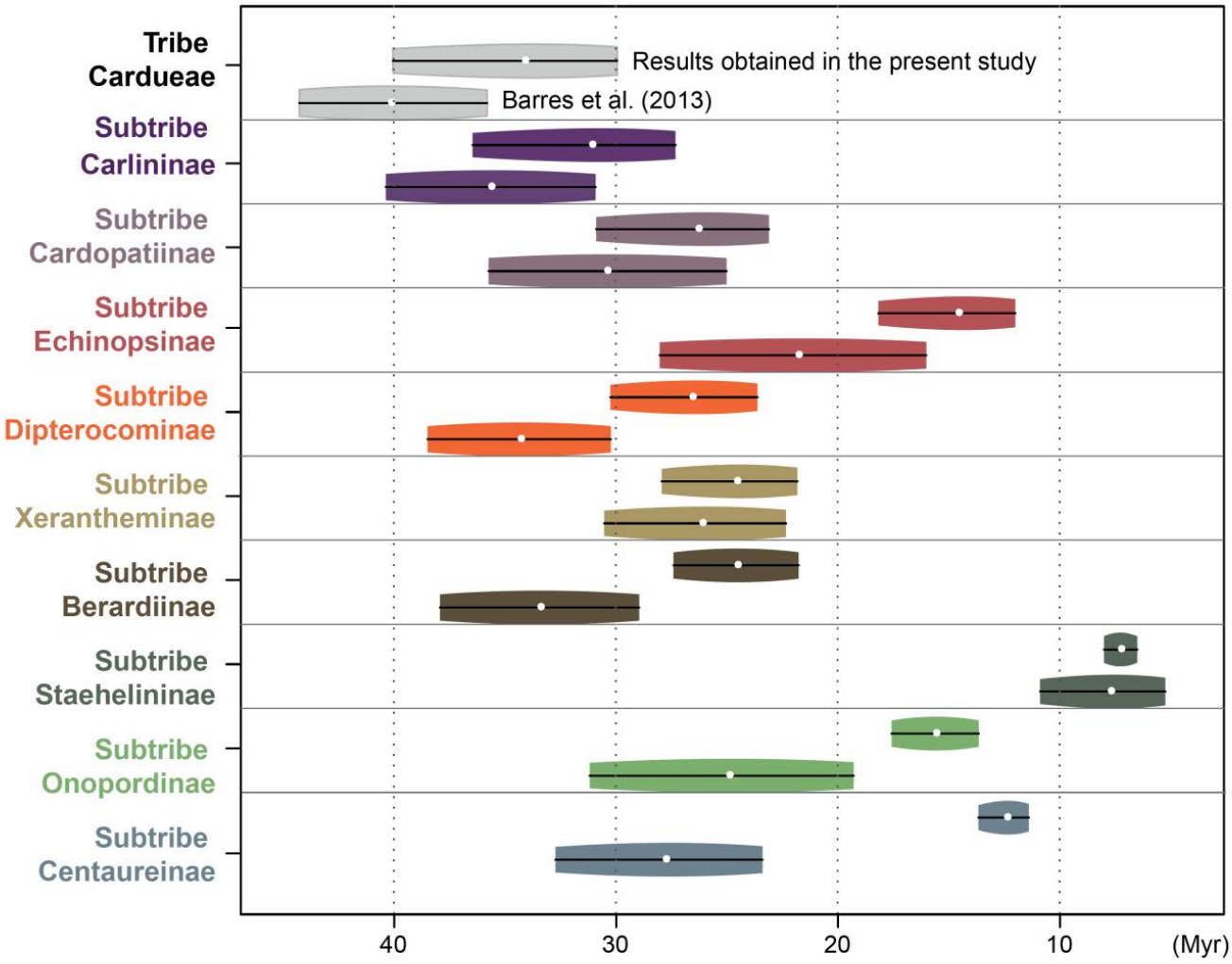


Fig. 6. Comparison of dating results of age node estimates (in Myr) for tribe Cardueae and subtribes obtained in the present study, with 1055 COS loci from nuclear genome (upper line in each subtribe), and those from Barres et al. (2013), with four chloroplast regions: *trnL-trnF*, *matK*, *ndhF*, *rbcL* (lower line in each subtribe). Colored boxes comprise the 95% of confidence interval limits and white dots represent the median age of lineages outlined. Note that subtribes Carduinae, Arctiinae, and Saussureinae are not represented due to their lack of monophyly in dated phylogeny of Barres et al. (2013).

estimated ages of non-Cardueae and early diverging lineages of Cardueae (e.g. Carlininae or Cardopatiinae). Certainly, the fact that the standard deviation of our most basal and secondary dated calibration point used (CP1) is higher than the other fossil-based CPs may cause larger uncertainty at the tree base. Moreover, the number of species included as Cardueae outgroups in our study is relatively poor, particularly compared with Panero and Crozier (2016) who used a wide representation of Compositae subfamilies but few representatives for each tribe. The different taxon sampling used could be the reason for differences in the chronograms obtained, resulting in a more uncertain dating of the poorly sampled nodes. As Linder et al. (2005) suggested, taxon sampling is one of the most influential factors in molecular clock dating analysis. For this reason, our node estimates at the tree base should be taken with caution, and undoubtedly, a wider sampling of other family members could improve lineages placement and dating resolution at the base of Cardueae.

3.5. Taxonomic proposal

Subtribe Carlininae Dumort., Fl. Belg.: 72. 1827. Type of the subtribe: *Carlina* L., Sp. Pl. 2: 828. 1753.

Perennial, spiny or spinescent herbs or shrubs, less often annual plants, monoecious or exceptionally dioecious (*Tugarinovia* Iljin). Leaves deeply pinnatisect, rarely entire. Capitula frequently subtended by pectinate leaf-like bracts, many-flowered, homogamous, rarely heterogamous with ligulate florets (Fig. 7A). Outer involucral bracts usually dissect; inner ones often showy, radiant and colored. Receptacle densely covered with large scales fused into a honeycombed structure, absent only in *Tugarinovia*. Corolla lobes very short, 1–3 mm. Anther filaments glabrous, appendages long and sericeous. Style very short. Achenes with parenchymatous pericarp, densely sericeous with twin hairs. Pappus of long, plumose bristles, often connate into stout scales, persistent or deciduous.

Comprises five genera: *Atractylodes* DC., *Atractylis*

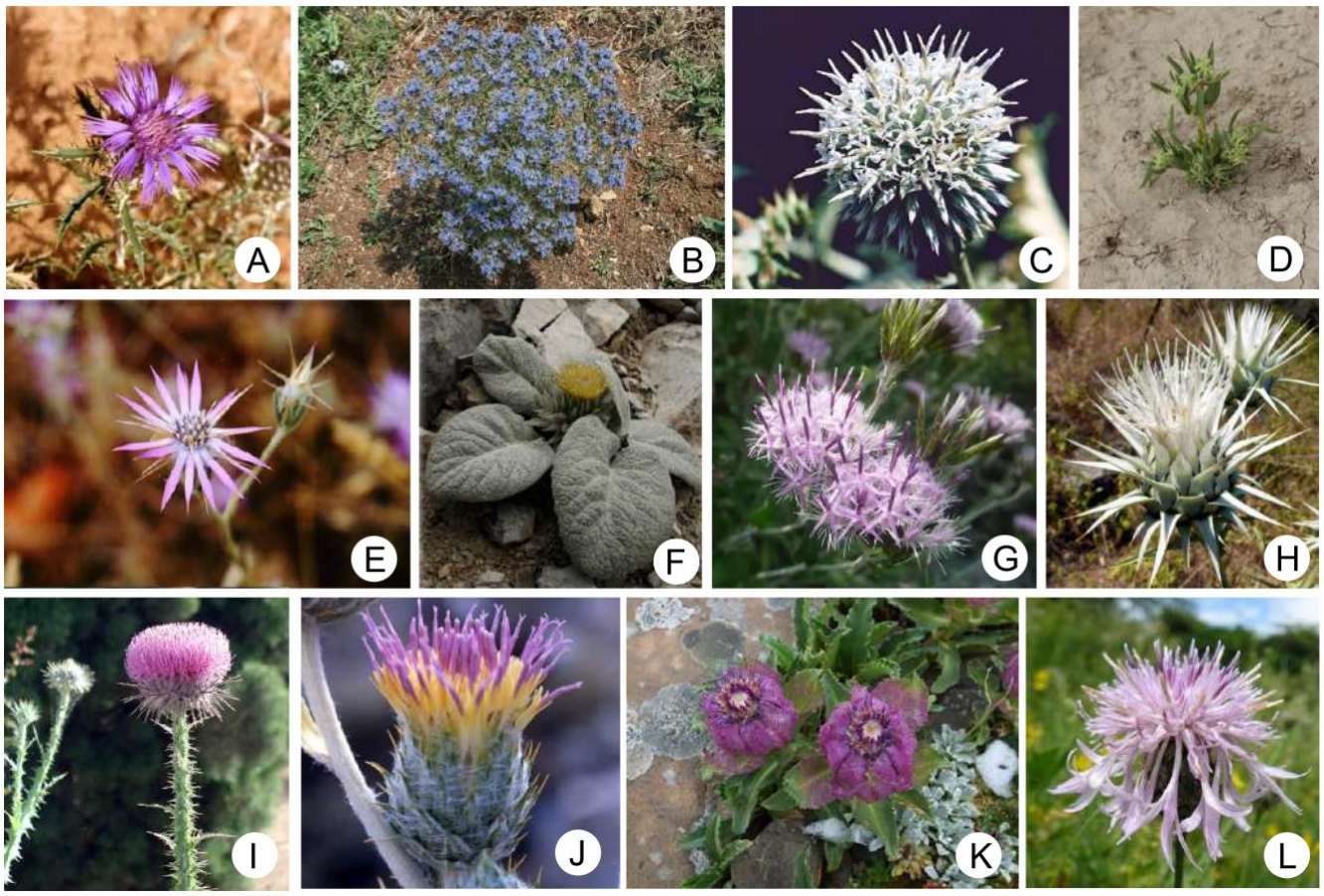


Fig. 7. Representation of capitula diversity for 12 newly subtribes recognized here within tribe Cardueae: **(A)** subtribe Carlininae, *Atractylis humilis* L, Spain; **(B)** subtribe Cardopatiinae, *Cardopatium corymbosum* Pers., Barcelona Botanical Garden; **(C)** subtribe Echinopsinae, *Echinops viscosus* Rchb., Turkey; **(D)** subtribe Dipterocominae, *Dipterocome pusilla* Fisch. & C. A. Mey., Turkmenistan; **(E)** subtribe Xerantheminae, *Siebera pungens* J. Gay, Turkey; **(F)** subtribe Berardiinae, *Berardia subacaulis*, France; **(G)** subtribe Staehelininae, *Staehelina uniflosculosa* Sm., Greece; **(H)** subtribe Onopordinae, *Onopordum nervosum* Boiss., Spain; **(I)** subtribe Carduinae, *Cynara baetica* (Spreng.) Pau, Spain; **(J)** subtribe Arctiinae, *Cousinia lanata* C. Winkl., Tajikistan; **(K)** subtribe Saussureinae, *Saussurea bracteata* Decne., China; and **(L)** subtribe Centaureinae, *Centaurea cephalariifolia* Willk., Spain. (A) (B) (C) (E) (G) (H) (I) (J) (L) Photos by A. Susanna; (D) Photo by A. Pavlenko; (F) Photo by C. Roquet; (K) Photo by J. López-Pujol.

L., *Carlina* L., *Thevenotia* DC., and *Tugarinovia* Iljin.

Geographic distribution (Fig. 8A): *Atractylis*, *Carlina*, and *Thevenotia* grow in Eurasia, North Africa, and the Irano-Turanian region; *Tugarinovia* is endemic to the deserts of Mongolia and NW China; *Atractylodes* is distributed in East Asia (China, Japan, and Korea).

Subtribe Cardopatiinae Less., Syn. Gen. Compos.: 14. 1832. Type of the subtribe: *Cardopatium* Juss., Ann. Mus. Natl. Hist. Nat. 6: 324. 1805.

Spiny perennial or annual herbs. Leaves dentate or pinnatisect. Capitula few-flowered, homogamous, densely (*Cardopatium*) or loosely (*Cousiniopsis* Nevski) clustered in corymbs (Fig. 7B). Involucral bracts with spiny pectinate-fimbriate appendages. Florets bright blue below, deep blue above; corolla lobes linear. Anther filaments glabrous. Style with very short lobes. Achenes with parenchymatous pericarp, densely sericeous. Pappus of two equal rows of short, lanceolate scales directly attached to the apical plate, persistent.

Comprises two genera: *Cardopatium* Juss. and *Cousiniopsis* Nevski.

Geographic distribution (Fig. 8B): disjunct area; *Cardopatium* is endemic in the south of the Mediterranean Region, from Algeria to Greece and Turkey, and *Cousiniopsis* grows in the deserts of Middle Asia.

Subtribe Echinopsinae Cass. ex Dumort., Anal. Fam. Pl.: 1829. Type of the subtribe: *Echinops* L., Sp. Pl. 2: 814. 1753.

Perennial herbs, rarely annuals, usually very spiny. Leaves pinnatisect. Capitula one-flowered, aggregated in spherical (rarely hemispherical) secondary inflorescences (Fig. 7C). Bracts spiny. Corolla lobes often apically scarious, densely denticulate, blue or white. Anther filaments glabrous, basal appendages short, laciniate. Style with short lobes. Achenes with parenchymatous pericarp, densely sericeous. Pappus of broad, short scales directly attached to the apical plate, persistent.

This subtribe tribe includes only the genus *Echinops*.

Geographic distribution (Fig. 8C): the genus grows in the Mediterranean and the Irano-Turanian regions, with the main center of speciation in Iran. It is one of the genera that extends more to the south, reaching tropical Africa.

Subtribe Dipterocominae Garcia-Jacas & Susanna, **new subtribe**. Type of the subtribe: *Dipterocome* Fisch. & C. A. Mey., Index Seminum Horti Petropolitani 1: 26 1835.

Annual unarmed desert plants to 20 cm. Leaves entire, linear-spatulate. Capitula one-flowered and unisexual, clustered in second-order synflorescences in the axilles of leaf verticils. Bracts of the synflorescence in several rows, lanceolate, with a membranous, shortly ciliate margin. Female capitula 6–8 in the periphery of the synflorescence. Involucral bracts fused into a

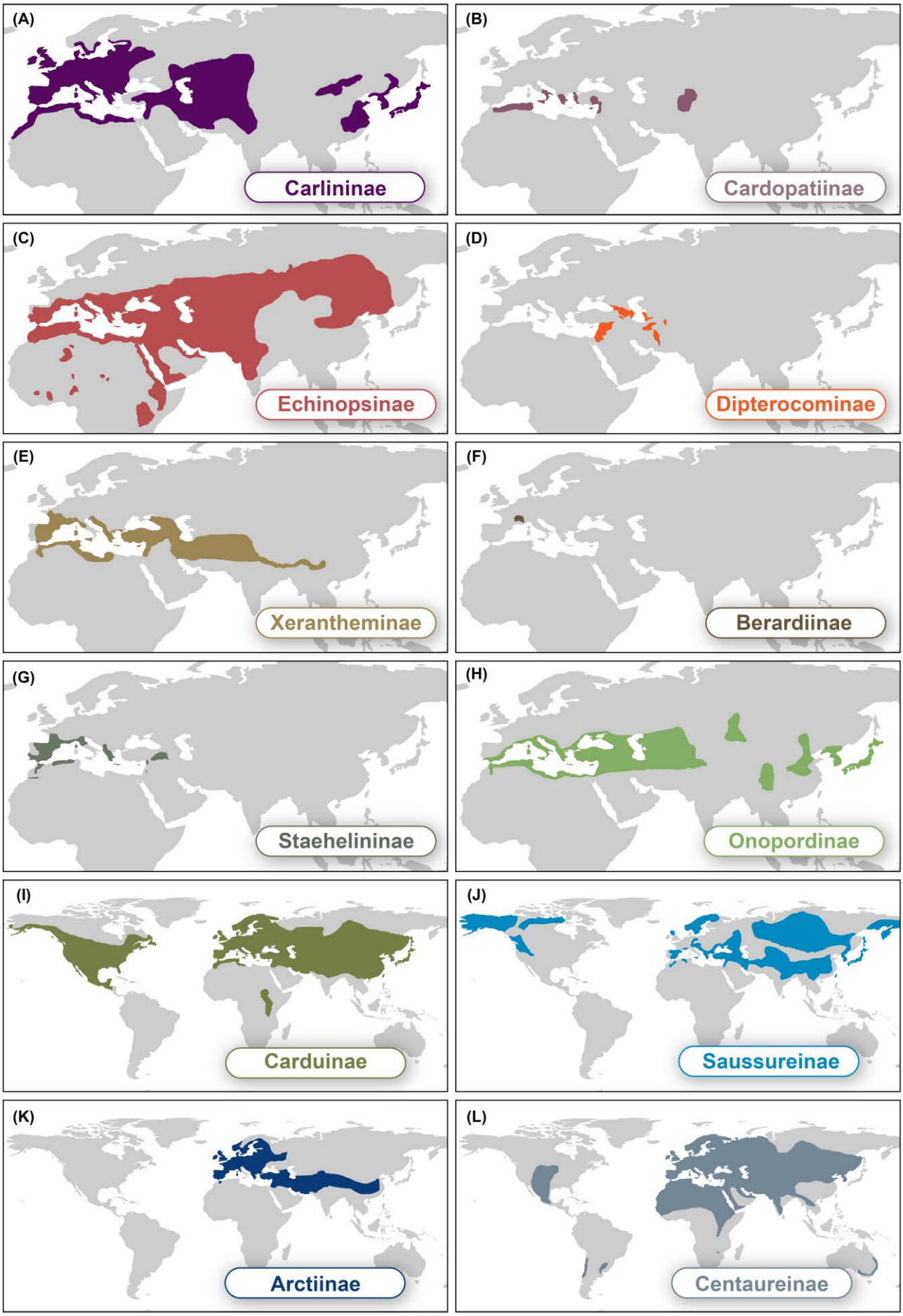


Fig. 8. Geographical representation of the main distribution ranges of 12 subtribes of tribe Cardueae defined in the present study.

spiny-horned structure acrescent with the achene. Corollas shortly bilabiate. Style with very short lobes. Achenes enclosed in the spiny-horned bracts. Pappus of few, deciduous bristles attached inside the horns. Male capitula ca. 4 in the center of the synflorescence. Bracts of the involucre fused into a basal sheath. Corolla tubulose with very short, equal lobes. Anthers shortly caudate.

Comprises only the monotypic genus *Dipterocome* Fisch. & C. A. Mey.

Geographic distribution (Fig. 8D): Deserts in the East Mediterranean and Irano-Turanian regions: Middle East (Israel, Jordan, Lebanon, and Syria), Transcaucasia (Armenia, Azerbaijan, and Georgia) and Iran.

We have interpreted the minute heads of *Dipterocome pusilla* as second-order inflorescences, formed by female and male one-flowered capitula (Fig. 7D). The bracts of the involucre of the female florets are fused into a spiny-horned structure, and the achenes are enclosed in the arcuated structure at maturity.

Subtribe Xeranthemeinae Cass. ex Dumort., Anal. Fam. Pl.: 32. 1829. Type of the subtribe: *Xeranthemum* L., Sp. Pl. 2: 857. 1753

Unarmed plants, usually annual herbs (*Chardinia* Desf., *Siebera* J. Gay, *Xeranthemum* L.), rarely dwarf shrubs (*Amphoricarpos* Vis.), exceptionally rhizomatous perennials (*Shangwua* Yu J. Wang, Raab-Straube, Susanna & J. Quan Liu). Leaves always entire, velvety underneath. Capitula usually heterogamous (with the exception of *Shangwua*), central florets discoid, peripheral florets shortly bilabiate, fertile, bisexual or female. Involucral bracts usually unarmed, rarely spiny (*Siebera*). Receptacle with large scarious scales. Corolla lobes very short. Anther filaments glabrous, appendages short, laciniate. Achenes often dimorphic. Pappus of 5–15 long tapering or subulate scales, rarely reduced to a corona in *Chardinia*, sparsely sericeous with twin hairs, persistent (Fig. 7E).

Comprises five genera: *Amphoricarpos*, *Chardinia*, *Shangwua*, *Siebera* and *Xeranthemum*.

Geographic distribution (Fig. 8E): extremely diverse and complicated. The genus *Xeranthemum* is a widespread in waste places in Eurasia, from Middle Asia to the Iberian Peninsula, more frequently in the Mediterranean region. Two other closely related genera, *Chardinia* and *Siebera*, occupy similar habitats in the Irano-Turanian region, from the Tian Shan to Turkey. *Amphoricarpos* shows a disjunct distribution in the mountains of the East Mediterranean: two species grow in the Balkans, two more in Anatolia, and one in the Caucasus. *Shangwua* is endemic to the mountains of Middle Asia and the Himalayas, from Tajikistan to China.

Subtribe Berardiinae Garcia-Jacas & Susanna, **new subtribe**. Type of the subtribe: *Berardia* Vill., Prosp. Hist. Pl. Dauphiné: 27. 1779.

Acaulescent, unarmed perennial herb (Fig. 7F). Leaves rounded, entire or denticulate, densely woolly with veins prominent beneath, green-white above. Capitula solitary, sessile, homogamous. Involucral bracts subulate, scarious, woolly, ending in a slender flattened point. Receptacle areolate. Florets yellowish or pinkish. Staminal connective very long, apiculate. Styles very long with short stigmas. Achenes oblong, glabrous, slightly sulcate. Pericarp straw-colored. Pappus of scabrid cylindric bristles retrorsely twisted, directly attached to the apical plate, persistent.

Comprises only the monotypic genus *Berardia*.

Geographic distribution (Fig. 8F): *Berardia* is a relict genus endemic to the southern Alps (SE France and NW Italy).

Subtribe Staehelininae Garcia-Jacas & Susanna, **new subtribe**. Type of the subtribe: *Staehelina* L., Sp. Pl. 2: 840. 1753.

Unarmed dwarf shrubs or subshrubs. Leaves entire or dentate-pinnatifid, linear to obovate, dark green above, white-woolly beneath. Capitula corymbose or rarely solitary, homogamous. Involucral bracts ovate to lanceolate, mucronate, sometimes minutely hirsute, unarmed. Receptacle with wide, basally connate scales. Florets whitish or pink-purple. Corolla lobes narrowly triangular, very long. Anther filaments glabrous; basal appendages very long, sericeous. Styles long, lobes straight. Achenes linear-oblong, glabrous or sericeous, with minute apical coronula. Pappus in one row of bristles basally connate into broader paleae, more or less divided apically into pinnulate fimbriae (into capillary hairs in *Staehelina dubia* L. and *S. baetica* DC.), always overtopping involucre, sometimes deciduous in a ring.

Comprises only the genus *Staehelina* (Fig. 7G).

Geographic distribution (Fig. 8G): Mediterranean region, from the Iberian Peninsula to the limits between Turkey and Iraq (*Staehelina kurdica* Merxm. & Rech. f.). Most diverse in the Eastern Mediterranean region (six species, in contrast with only two species in the Western Mediterranean).

Subtribe Onopordinae Garcia-Jacas & Susanna, **new subtribe**. Type of the subtribe: *Onopordum* L., Sp. Pl. 2: 827. 1753.

Robust biennial or perennial herbs, usually spiny-toothed, rarely weakly unarmed. Leaves lanceolate to oblong, variously dissected. Capitula cupuliform or broadly ovoid, homogamous, discoid, most often solitary or on long peduncles. Involucral bracts usually spiny, innermost sometimes scarious. Receptacle foveolate, margins of foveoles with short scales, rarely setose; receptacular bracts very often absent. Corolla lobes triangular. Anther filaments slightly papillose or glabrous. Styles very long with straight lobes. Achenes obovoid or fusiform, surface usually transversely rugulose, wrinkled or foveolate; apical plate without caruncle; insertion areole straight or lateral-abaxial. Pappus bristles in a few rows, differentiated or not, basally connate in a ring, always deciduous as a single piece (Fig. 7H).

Comprises eight genera: *Alfredia* Cass., *Ancathia* DC., *Synurus* Iljin, *Syreitschikovia* Pavlov, *Lamyropappus* Knorring & Tamamsch., *Olgaea* Iljin, *Onopordum*, and *Xanthopappus* C. Winkl.

Geographic distribution (Fig. 8H): the genus *Onopordum* is native in the Mediterranean and Irano-Turanian regions, from the Iberian Peninsula to Middle Asia, and it has been introduced as a weed in USA (California), Argentina, Chile, South Africa, and Australia. The remaining genera have narrow distributions in Middle Asia, the Tian Shan, and west China, with the exception of *Synurus*, endemic in east China, Japan, and Korea.

Subtribe Carduinae Dumort., Fl. Belg.: 73. 1827. Type of the subtribe: *Carduus* L., Sp. Pl. 2: 820. 1753.

Perennial, biennial or annual spiny herbs or subshrubs, rarely unarmed. Leaves mostly lanceolate to oblong, entire or variously dissected. Capitula globose or cupuliform, homogamous, discoid, very rarely peripheral florets sterile and radiant (*Galactites* Moench). Involucral bracts usually spiny, innermost

exappendiculate or with rudimentary appendages. Receptacle densely setose. Anther filaments papillose, rarely glabrous. Styles very long with straight lobes. Achenes obovoid-fusiform, laterally compressed, with insertion areole straight, basal or lateral-abaxial; apical plate very often slanted, inclined adaxially, usually with caruncle, missing only in *Notobasis*. Pappus inserted on a parenchymatous ring in the apical plate, with bristles in many weakly differentiated or undifferentiated rows, deciduous as a single piece (Fig. 7I).

Comprises ten genera: *Carduus*, *Cirsium* Mill., *Cynara* L., *Galactites*, *Lamyropsis* (Kharadze) Dittrich, *Notobasis* Cass., *Picnomon* Adans. *Ptilostemon* Cass., *Silybum* Vaill., and *Tyrimnus* Cass.

Geographic distribution (Fig. 8I): Widespread in all Eurasia, especially in the Mediterranean region, extending southwards to Tropical Africa, introduced elsewhere as very noxious weeds. One genus, *Cirsium*, native in America, from Canada to Guatemala.

Subtribe Saussureinae Garcia-Jacas & Susanna, **new subtribe**. Type of the subtribe: *Saussurea* DC., Ann. Mus. Natl. Hist. Nat. 16: 156. 1810.

Unarmed perennial herbs or subshrubs, very rarely annual herbs. Leaves entire or pinnatisect, often silver-white below and glabrous above, sometimes hirsute-scabrid. Capitula cylindrical or globose, often paniculate, homogamous, discoid. Involucral bracts with short appendages, unarmed. Receptacle densely setose. Anther filaments glabrous, rarely papillose. Styles long, with reflexed lobes. Achenes cylindrical, slightly obconical or obpyramidal, indistinctly ribbed to costate, surface smooth or transversally rugose, very rarely with spines or scales, with or without a crown; apical plate with a persisting style base, without caruncle. Pappus of very long (overtopping involucral bracts), showy, scabrid or plumose, strongly differentiated bristles, the inner ones basally connate in a ring, persistent or detachable as a single piece, outer bristles shorter and freely deciduous (Fig. 7J).

Comprises four genera: *Dolomiaea* DC., *Jurinea* Cass., *Polytaxis* C. Winkl., and *Saussurea* DC.

Geographic distribution (Fig. 8J): Native in Eurasia, *Jurinea* grows especially in the eastern Mediterranean and West Asia with three species reaching the mountains of the Iberian Peninsula and Morocco; *Saussurea* has its most important center of speciation in the Himalayas and the Hengduan mountains (more than 300 species); two species in Western Europe (Alps and Pyrenees), and six species in North America. *Dolomiaea* is circumscribed to the Himalayas and adjacent mountains. *Polytaxis* is limited to the Tian Shan in Middle Asia.

Subtribe Arctiinae Garcia-Jacas & Susanna, **new subtribe**. Type of the subtribe: *Arctium* L., Sp. Pl. 2: 816. 1753.

Biennial or perennial herbs, polycarpic or rarely monocarpic, exceptionally annuals, very spiny, less often unarmed. Leaves entire, lyrate or pinnatisect, usually with arachnoid pubescence. Capitula homogamous, cylindrical or globose, discoid. Outer and middle involucral bracts spiny or hooked, inner bracts ended in a scarious, unarmed or spiny appendage. Receptacle with long, twisted fimbrillae. Anther filaments slightly papillose or glabrous. Styles with short, straight lobes. Achenes cylindric to narrowly obovoid, usually laterally compressed or 4-costate, rarely shortly winged, usually longitudinally inconspicuously ribbed and stripped, apex glabrous or coronulate; apical plate truncate, without caruncle; insertion areole lateral-adaxial. Pappus in three rows of freely deciduous undifferentiated bristles, rarely reduced (Fig. 7K).

Comprises two genera: *Arctium* L. and *Cousinia* Cass.

Geographic distribution (Fig. 8K): some species of *Arctium* are widespread in Eurasia, but most of the species are endemic to the mountains of Middle Asia, especially the Tian Shan. *Cousinia* is confined to the Irano-Turanian floristic region, where it has radiated explosively (600 species!).

Subtribe Centaureinae Dumort., Fl. Belg.: 72. 1827. Type of the subtribe: *Centaurea* L., Sp. Pl. 2: 909. 1753.

Perennial, biennial or annual unarmed herbs, shrubs or very rarely treelets, rarely spiny. Leaves oblong to linear, entire or variously dissected. Capitula ovoid, cupuliform or cylindric, often heterogamous with sterile radiant florets, less often homogamous, discoid. Involucral bracts usually ended in a diversely scarious, fimbriate, pectinate, spiny or unarmed appendage; innermost bracts always with a scarious appendage. Receptacle densely setose. Anther filaments papillose, rarely glabrous; basal appendages long, laciniate. Style and stigma very variable. Achenes obovoid, laterally compressed, with sclerified pericarp, usually glabrous and smooth, less often hirsute, rugose, pitted, or ridged; insertion areole concave, lateral-adaxial, very rarely (*Crupina*) straight, often with an elaiosome; apical plate straight, without caruncle. Pappus inserted on a parenchymatous ring in the apical plate, biseriate, outer series formed by several rows of long, differently pinnulate bristles or rarely scales, basally connate in a ring or free, rarely deciduous as a whole of separately, often persistent, inner series of some short or longer bristles, sometimes reduced; rarely pappus single or missing by abortion (Fig. 7L).

Comprises 32 genera: *Amberboa* Vaill., *Callicephalus* C. A. Mey. *Carduncellus* Adans., *Carthamus* L., *Centaurea* L., *Centaurodendron* Johow, *Centaurothamnus* Wagenitz & Dittrich, *Cheirolophus* Cass., *Crocodilium* Vaill., *Crupina* (Pers.) DC., *Femeniasia* Susanna, *Goniocaulon* Cass., *Karvandarina* Rech. f., *Klasea* Cass., *Mantisalca* Cass., *Myopordon* Boiss., *Ochrocephala* Dittrich, *Oligochaeta* (DC.) K. Koch, *Phalacrachena* Iljin, *Phonus* Hill, *Plagiobasis* Schrenk, *Plectocephalus* D. Don, *Psephellus* Cass., *Rhaponticoides* Vaill., *Rhaponticum* Vaill., *Russowia* C. Winkl., *Schischkinia* Iljin, *Serratula* L., *Stizolophus* Cass., *Tricholepis* DC., *Volutaria* Cass., and *Zoegea* L.

Geographic distribution (Fig. 8L): mainly Mediterranean and Irano-Turanian regions, with some species of *Centaurea* extended widely in north Eurasia as far as Scotland, and the easternmost representatives reaching Middle Asia; exceptionally, *Tricholepis* has gone beyond the mountains of Middle Asia and reaches Burma; one species of *Rhaponticum* grows in Australia. In the south of the area, a few species grow in subtropical and tropical Africa, from Senegal to Somalia and southwards to Zimbabwe. *Plectocephalus* has a disjunct distribution in Ethiopia, U.S.A., and South America (Argentina, Brazil, and Chile). Some relict genera grow far away from the center of the distribution of the subtribe, namely *Centaurodendron* in the Juan Fernández archipelago (South Pacific Ocean), *Centaurothamnus* in Yemen, *Goniocaulon* in India, and *Ochrocephala* in Ethiopia and Sudan; all of them but *Centaurodendron* are monotypic. Some noxious weeds are widespread throughout the world, especially in regions with a Mediterranean-type climate (Australia, California, Chile, and South Africa): *Carthamus lanatus*, *Centaurea diffusa*, *C. solstitialis*, *C. stoebe*, *Rhaponticum repens*, and *Volutaria tubuliflora*.

4. Key to subtribes of tribe *Cardueae*

1.	Heads always one-flowered, unisexual, clustered in the axilles of leaf verticils. Achenes enclosed in an arcuate, spiny-toothed diaspore.	Dipterocominae
	– Heads rarely one-flowered, always bisexual, not clustered in the axilles of leaf verticils Achenes free, not enclosed in a spiny-toothed diaspore	2
2.	Capitula one-flowered, clustered in second-order spherical or hemispherical compound synflorescences	Echinopsinae
	– Capitula not one-flowered, sometimes clustered in flat-topped corymbs but never forming second-order heads	3
3.	Acaulescent perennials; achenes with the pappus bristles retrorsely twisted	Berardiinae
	– Caulescent or acaulescent annuals or perennials; achenes with pappus of straight bristles	4
4.	Leaves always entire. Achenes with pappus of 5-15 subulate, dentate or plumose scales, rarely reduced to a short corona	Xeranthemineae
	– Leaves entire or divided. Achenes with pappus of more than 15 bristles	5
5.	Capitula usually subtended by pectinate-pinnatisect leaf-like bracts; corolla lobes usually 1–3 mm long; pappus of long, plumose bristles usually connate at the base into broader, robust scales	Carlininae
	– Capitula subtended by entire or dentate bracts or on leafless pedicels; corolla lobes longer than 3 mm; pappus of free bristles or scales not connate at the base into broader scales	6
6.	Capitula with 8–12 florets, rarely many-flowered; corollas bright blue below, deep blue above, with linear lobes; pappus of two equal rows of short lanceolate scales	Cardopatiinae
	– Capitula with more than 12 florets; corollas not bright blue below and deep blue above with lobes not linear; pappus of one or several rows of pinnulate or plumose long scales or bristles	7
7.	Achenes rugulose, pitted or faintly velvety, never smooth, without apical caruncle or basal elaiosome	8
	– Achenes usually smooth or rarely ridged, either with apical caruncle or basal elaiosome	9
8.	Receptacle with long, twisted fimbrillae; achenes longitudinally stripped; pappus bristles individually deciduous	Arctiinae
	– Receptacle usually naked and foveolate with short scales, rarely with straight bristles; achenes transversally rugulose; pappus bristles basally connate in a ring, detachable as a single piece	Onopordinae
9.	Plants unarmed; leaves always white-woolly below; pappus very long, often overtopping involucral bracts at anthesis	10
	– Plants unarmed or spiny; leaves woolly or glabrous below; pappus not overtopping involucral bracts	11
10.	Shrubs or subshrubs. Corolla lobes very long, narrowly triangular; style very short, straight lobes; achenes linear-oblong, faintly sulcate; pappus in one row	Staehelininae
	– Perennial herbs, rarely annuals; corolla lobes short, broadly triangular; style lobes long, reflexed; achenes cylindrical or slightly obconical or obpyramidal, smooth, echinate or rugulose; pappus in several rows	Saussureinae
11.	Insertion areole of the achenes usually straight, basal or basal-abaxial. Apical plate of the achenes slanted, with caruncle (except. Notobasis); pappus undifferentiated, deciduous as a whole	Carduinae
	– Insertion areole of the achenes usually concave, lateral-adaxial, often with an elaiosome; apical plate straight, without caruncle; pappus usually differentiated in two rows, persistent or rarely deciduous as a single piece, rarely absent by abortion	Centaureinae

5. Acknowledgments

Authors thank the herbaria that provided material for the study: B, BC, BCN, BEOU, DUSH, E, ERE, FRU, GA, GDA, H, LE, LP, MEM, MJG, MU, OS, SANT, TARI, TK, US, and W; and A. Pavlenko for kindly supplying the photograph of *Dipterocome pusilla*. Phylogenetic analyses were performed on the high performance cluster (HPC) of the computing facilities of University of Memphis, which is greatly acknowledged. Financial support from the Ministerio de Ciencia e Innovación (Project CGL2015-66703-P and Ph.D. grant to Sonia Herrando-Moraira) and the Catalan government (–Ajuts a grups consolidats" 2017-SGR1116) is also gratefully acknowledged.

6. References

Anderberg, A., Ghahremaninejad, F., Källersjö, M., 2007. The enigmatic genus *Dipterocome*. Compositae Newslett. 45, 23–36.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin V.M., Nikolenko, S.I., Pham, S., Prjibelski A.D., Pyshkin, A.V., Sirotkin A.V., Vyahhi, N., Tesler, G., Alekseyev, M.A., Pyshkin, A.V., 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J. Comput. Biol. 19, 455–477. https://doi.org/10.1089/cmb.2012.0021

Barreda, V.D., Palazzesi, L., Katinas, L., Crisci, J.V., Tellería, M.C., Bremer, K., Passala, M.G., Bechis, F., Corsolini, R., 2012. An extinct Eocene taxon of the daisy family (Asteraceae): evolutionary, ecological and biogeographical implications. Ann. Bot., London, 109, 127–134. https://doi.org/10.1093/aob/mcr240

Barres, L., Sanmartín, I., Anderson, C.L., Susanna, A., Buerki, S., Galbany-Casals, M., Vilatersana, R., 2013. Reconstructing the evolution and biogeographic history of tribe Cardueae (Compositae). Am. J. Bot. 100, 867–882. https://doi.org/10.3732/ajb.1200058

Bentham, G., 1873. Compositae. In: Bentham, G., Hooker, J.D. (Eds.), *Genera plantarum*, vol. 2. Spottiswoode & Co., London, UK, pp. 163–533.

Boissier, P.E., 1849. *Diagnoses plantarum orientalium novarum*, series 1, vol. 11. Ducloux, Paris, France.

Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Borowiec, M.L., 2016. AMAS: a fast tool for alignment manipulation and computing of summary statistics. PeerJ 4, e1660. https://doi.org/10.7717/peerj.1660

Bremer, K., 1987. Tribal interrelationships of the Asteraceae. Cladistics 3, 210–253. https://doi.org/10.1111/j.1096-0031.1987.tb00509.x

Bremer, K., 1994. Compositae: Cladistics and classification. Timber Press, Portland, Oregon, USA.

Bruun-Lund, S., Clement, W. L., Kjellberg, F., Rønsted, N., 2017. First plastid phylogenomic study reveals potential cyto-nuclear discordance in the evolutionary history of *Ficus* L. (Moraceae). Mol. Phylogenet. Evol. 109, 93–104. https://doi.org/10.1016/j.ympev.2016.12.031

Capella-Gutiérrez, S., Silla-Martínez, J.M., Gabaldón, T., 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25, 1972–1973. https://doi.org/10.1093/bioinformatics/btp348

Carlsen, M.M., Fér, T., Schmickl, R., Leong-Škorničková, J., Newman, M., Kress, W.J., 2018. Resolving the rapid plant radiation of early diverging lineages in the tropical Zingiberales: pushing the limits of genomic data. Mol. Phylogenet. Evol. 128, 55–68. https://doi.org/10.1016/j.ympev.2018.07.020

Cassini, H., 1819. [Different articles]. Dictionnaire de Sciences Naturelles. Paris. In: King, R., Dawson, H.W. (Eds., 1975), Cassini on Compositae. Oriole, New York, New York, USA.

Castresana, J., 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol. Biol. Evol. 17, 540–552. https://doi.org/10.1093/oxfordjournals.molbev.a026334

Chevreux, B., Wetter, T., Suhai, S., 1999. Genome sequence assembly using trace signals and additional sequence information. In: Computer Science and Biology: Proceedings of the German Conference on Bioinformatics (GCB), vol. 99, pp. 45–56.

Dayrat, B., 2005. Towards integrative taxonomy. Biol. J. Linn. Soc. 85, 407–415. https://doi.org/10.1111/j.1095-8312.2005.00503.x

Dittrich, M., 1977. Cynareae–systematic review. In: Heywood, V.H., Harborne, J.B., Turner, B.L. (Eds.), The biology and chemistry of the Compositae. Academic Press, London–New York–San Francisco, UK and USA, pp. 999–1015.

Dittrich, M., 1996a. Bemerkungen zur Tribuszugehörigkeit von *Berardia*. Ann. Naturhist. Mus. Wien 98 (B, suppl.), 329–342.

Dittrich, M., 1996b. Die Bedeutung morphologischer and anatomischer Achänen—Merkmale für die Systematik der Tribus Echinopeae Cass. and Carlineae Cass. Boissiera 51, 9–102.

Drummond, A.J., Suchard, M.A., Xie, D., Rambaut, A., 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. Mol. Biol. Evol. 29, 1969–1973. https://doi.org/10.1093/molbev/mss075

Dupont-Nivet, G., Krijgsman, W., Langereis C.G., Abels, H.A., Dai, S., Fang, X., 2007. Tibetan Plateau aridification linked to global cooling at the Eocene–Oligocene transition. Nature 445, 635–638. https://doi.org/10.1038/nature05516

Edwards, S.V., Xi, Z., Janke, A., Faircloth, B.C., McCormack, J.E., Glenn, T.C., Zhong, B., Wu, S., Lemmon, E.M., Lemmon, A.R., Leache, A.D., Liu, L., Davis, C.C., 2016. Implementing and testing the multispecies coalescent model: A valuable paradigm for phylogenomics. Mol. Phylogenet. Evol. 94, 447–462. https://doi.org/10.1016/j.ympev.2015.10.027

Eldrett, J.S., Greenwood, D.R., Harding, I.C., Huber, M., 2009. Increased seasonality through the Eocene to Oligocene transition in northern high latitudes. Nature 459, 969–974. https://doi.org/10.1038/nature08069

Faircloth, B.C., 2015. PHYLUCE is a software package for the analysis of conserved genomic loci. Bioinformatics 32, 786–788. https://doi.org/10.1093/bioinformatics/btv646.

Favre, A., Michalak, I., Chen, C.-H., Wang, J.-C., Pringle, J.S., Matuszak, S., Sun, H., Yuan, Y.-M., Struwe L., Muellner-Riehl, A.N., 2016. Out-of-Tibet: the spatio-temporal evolution of *Gentiana* (Gentianaceae). J. Biogeogr. 43, 1967–1978. https://doi.org/10.1111/jbi.12840

Folk, R.A., Mandel, J.R., Freudenstein, J.V., 2015. A protocol for targeted enrichment of intron-containing sequence markers for recent radiations: A phylogenomic example from *Heuchera* (Saxifragaceae). Appl. Plant Sci. 3, 1500039. https://doi.org/10.3732/apps.1500039

Folk, R.A., Mandel, J.R., Freudenstein, J.V., 2017. Ancestral gene flow and parallel organellar genome capture result in extreme phylogenomic discord in a lineage of angiosperms. Syst. Biol. 66, 320–337. https://doi.org/10.1093/sysbio/syw083

Funk, V. A., Susanna, A., Stuessy, T.F., Robinson, H., 2009. Classification of Compositae. In: Funk, V.A., Susanna, A., Stuessy, T.F., Bayer, R.J. (Eds.), 2009. Systematics, evolution, and biogeography of Compositae. International Association for Plant Taxonomy, Vienna, pp. 171–189.

García, N., Folk, R.A., Meerow, A.W., Chamala, S., Gitzendanner, M.A., de Oliveira, R.S., Soltis, D.E., Soltis, P.S., 2017. Deep reticulation and incomplete lineage sorting obscure the diploid phylogeny of rain-lilies and allies (Amaryllidaceae tribe Hippeastreae). Mol. Phylogenet. Evol. 111, 231–247. https://doi.org/10.1016/j.ympev.2017.04.003

Garcia-Jacas, N., Garnatje, T., Susanna, A., Vilatersana, R., 2002. Tribal and subtribal delimitation and phylogeny of the Cardueae (Asteraceae): a combined nuclear and chloroplast DNA analysis. Mol. Phylogenet. Evol. 22, 51– 64. https://doi.org/10.1006/mpev.2001.1038

Gardner, A.G., Sessa, E.B., Michener, P., Johnson, E., Shepherd, K.A., Howarth, D.G., Jabaily, R.S., 2016. Utilizing next-generation sequencing to resolve the backbone of the Core Goodeniaceae and inform future taxonomic and floral form studies. Molec. Phylogenet. Evol. 94, 605–617. https://doi.org/10.1016/j.ympev.2015.10.003

Gouy, M., Guindon, S., Gascuel, O., 2009. SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. Mol. Biol. Evol. 27, 221–224. https://doi.org/10.1093/molbev/msp259

Häffner, E., Hellwig, F.H., 1999. Phylogeny of the tribe Cardueae (Compositae) with emphasis on the subtribe Carduinae: an analysis based on ITS sequence data. Willdenowia 29, 27–39. https://doi.org/10.3372/wi.29.2902

Hahn, C., Bachmann, L., Chevreux, B., 2013. Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads—a baiting and iterative mapping approach. Nucleic Acids Res. 41, e129. https://doi.org/10.1093/nar/gkt371

Harris, R.S., 2007. Improved pairwise alignment of genomic DNA. PhD Thesis, Pennsylvania State University, Pennsylvania.

Harrison, N., Kidner, C.A., 2011. Next-generation sequencing and systematics: what can a billion base pairs of DNA sequence data do for you? Taxon 60, 1552–1566.

Heckenhauer, J., Samuel, R., Ashton, P.S., Salim, K.A., Paun, O., 2018. Phylogenomics resolves evolutionary relationships and provides insights into floral evolution in the tribe Shoreeae (Dipterocarpaceae). Mol. Phylogenet. Evol. 127, 1–13. https://doi.org/10.1016/j.ympev.2018.05.010

Heilbuth, J.C., 2000. Lower species richness in dioecious clades. Am. Nat. 156, 221–241. https://doi.org/10.1086/303389

Herrando-Moraira, S., The Cardueae Radiations Group (Calleja,

J.A., Carnicero-Campmany, P., Fujikawa, K., Galbany-Casals, M., Garcia-Jacas, N., Im, H.-T., Kim, S.-C., Liu, J.-Q., López-Alvarado, J., López-Pujol, J., Mandel, J.R., Massó, S., Mehregan, I., Montes-Moreno, N., Pyak, E., Roquet, C., Sáez, L., Sennikov, A., Susanna, A., Vilatersana, R.), 2018. Exploring data processing strategies in NGS target enrichment to disentangle radiations in the tribe Cardueae (Compositae). Mol. Phylogenet. Evol. 128, 69–87. https://doi.org/10.1016/j.ympev.2018.07.012

Hillis, D.M., Bull, J.J., 1993. An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. Syst. Biol. 42, 182–192. https://doi.org/10.1093/sysbio/42.2.182

Hoffmann, O., 1894. Compositae. In: Engler, A., Prantl, K. (Eds.), Die natürlichen Pflanzenfamilien, vol. 5, part 4. Wilhelm Engelmann, Leipzig, Germany, pp. 24–333.

Hohmann, N., Wolf, E.M., Lysak, M.A., Koch, M.A., 2015. A time-calibrated road map of Brassicaceae species radiation and evolutionary history. Plant Cell 27, 2770–2784. https://doi.org/10.1105/tpc.15.00482

Huang, C.H., Zhang, C., Liu, M., Hu, Y., Gao, T., Qi, J., Ma, H., 2016. Multiple polyploidization events across Asteraceae with two nested events in the early history revealed by nuclear phylogenomics. Mol. Biol. Evol. 33, 2820–2835. https://doi.org/10.1093/molbev/msw157

Ivanov, D., Ashraf, A.R., Utescher, T., Mosbrugger, V., Slavomirova, E., 2007. Late Miocene vegetation and climate of the Balkan region, palynology of the Beli Breg Coal Basin sediments. Geol. Carpath. 58, 367–381.

Jansen, R.K., Holsinger, K.E., Michaels, H.J., Palmer, J.D., 1990. Phylogenetic analysis of chloroplast DNA restriction site data at higher taxonomic levels: an example from the Asteraceae. Evolution 44, 2089–2105. https://doi.org/10.1111/j.1558-5646.1990.tb04314.x

Jansen, R.K., Michaels, H.J., Palmer, J.D., 1991. Phylogeny and character evolution in the Asteraceae based on chloroplast DNA restriction site mapping. Syst. Bot. 16, 98–115. https://doi.org/10.2307/2418976

Jaubert, H.F., Spach, E., 1850. *Illustrationes plantarum orientalium*, vol. 3. Roret, Paris, France.

Johnson, M.G., Gardner, E.M., Liu, Y., Medina, R., Goffinet, B., Shaw, A.J., Zerega, N.J.C., Wickett, N.J., 2016. HybPiper: Extracting coding sequence and introns for phylogenetics from high-throughput sequencing reads using target enrichment. Appl. Plant Sci. 4, 1600016. https://doi.org/10.3732/apps.1600016.

Jones, K.E., Fér, T., Schmickl, R., Dikow, R.B., Funk, V.J., Herrando-Moraira, S., Johnston, P.R., Siniscalchi, C.M., Susanna, A., Slovák, M., Thapa, R., Watson, L., Mandel, J.R., in preparation. An empirical assessment of a single family-wide target capture locus set at multiple evolutionary timescales within the largest family of flowering plants.

Karis, P.O., Källersjö, M., Bremer, K., 1992. Phylogenetic analysis of the Cichorioideae (Asteraceae), with emphasis on the Mutisieae. Ann. Mo. Bot. Gard. 79, 416–427. https://doi.org/10.2307/2399778

Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. 30, 772–780. https://doi.org/10.1093/molbev/mst010

Kay, K.M., Voelckel, C., Yang, J.Y., Hufford, K.M., Kaska, D.D., Hodges, S.A., 2006. Floral characters and species diversification. In: Harder, L.D., Barrett, S.C.H. (Eds.), Ecology and evolution of flowers. Oxford University Press, Oxford, UK, pp. 311–325.

Kim, K.-J., Choi, K.-S., Jansen, R.K., 2005. Two chloroplast DNA inversions originated simultaneously during the early evolution of the sunflower family (Asteraceae). Mol. Biol. Ecol. 22, 1783–1792. https://doi.org/10.1093/molbev/msi174

Kim, K.J., Jansen, R.K., Wallace, R.S., Michaels, H.J., Palmer, J.D., 1992. Phylogenetic implications of rbcL sequence variation in the Asteraceae. Ann. Mo. Bot. Gard. 79, 428–445. https://doi.org/10.2307/2399779

Knowles, L.L., Huang, H., Sukumaran, J., Smith, S.A., 2018. A matter of phylogenetic scale: Distinguishing incomplete lineage sorting from lateral gene transfer as the cause of gene tree discord in recent versus deep diversification histories. Am. J. Bot. 105, 376–384. https://doi.org/10.1002/ajb2.1064

Kubatko, L.S., Degnan, J.H., 2007. Inconsistency of phylogenetic estimates from concatenated data under coalescence. Syst. Biol. 56, 17–24. https://doi.org/10.1080/10635150601146041.

Kück, P., Longo, G.C., 2014. FASconCAT-G: extensive functions for multiple sequence alignment preparations concerning phylogenetic studies. Front. Zool. 11, 81. https://doi.org/10.1186/s12983-014-0081-x

Lagomarsino, L.P., Condamine, F.L., Antonelli, A., Mulch, A., Davis, C.C., 2016. The abiotic and biotic drivers of rapid diversification in Andean bellflowers (Campanulaceae). New Phytol. 210, 1430–1442. https://doi.org/10.1111/nph.13920

Lane, A.K., Augustin, M.M., Ayyampalayam, S., Plant, A., Gleissberg, S., Di Stilio, V.S., Depamphilis C.W., Wong, G.K.-S., Kutchan, T.M., Leebens-Mack, J.H., 2018. Phylogenomic analysis of Ranunculales resolves branching events across the order. Bot. J. Linn. Soc. 187, 157–166. https://doi.org/10.1093/botlinnean/boy015

Lemmon, A.R., Brown, J.M., Stanger-Hall, K., Lemmon, E.M., 2009. The effect of ambiguous data on phylogenetic estimates obtained by maximum likelihood and Bayesian inference. Syst. Biol. 58, 130–145. https://doi.org/10.1093/sysbio/syp017

Lemmon, E.M., Lemmon, A.R., 2013. High-throughput genomic data in systematics and phylogenetics. Ann. Rev. Ecol. Evol. Syst. 44, 99–121. https://doi.org/10.1146/annurev-ecolsys-110512-135822

Lessing, C.F., 1832. *Synopsis generum Compositarum*. Duncker & Humboldt, Berlin.

Li, H., Durbin, R., 2009. Fast and accurate short read alignment with burrowswheeler transform. Bioinformatics 25, 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Linder, H.P., Hardy, C.R., Rutschmann, F., 2005. Taxon sampling effects in molecular clock dating: an example from the African Restionaceae. Mol. Phylogenet. Evol. 35, 569–582. https://doi.org/10.1016/j.ympev.2004.12.006

Liu, L., Xi, Z., Davis, C.C., 2015. Coalescent methods are robust to the simultaneous effects of long branches and incomplete lineage sorting. Mol. Biol. Evol. 32, 791–805. https://doi.org/10.1093/molbev/msu331

López-Vinyallonga, S., Mehregan, I., Garcia-Jacas, N., Tscherneva, O., Susanna, A., Kadereit, J.K., 2009. Phylogeny and evolution of the *Arctium-Cousinia* complex (Compositae, Cardueae-Carduinae). Taxon 58, 153–171.

Mai, D.H., 1995. Tertiäre Vegetationsgeschichte Europas: Methoden und Ergebnisse. Gustav Fischer Verlag, Jena, Germany.

Mai, D.H., 2001. Die mittelmiozänen und obermiozänen Floren aus der Meuroer und Raunoer Folge in der Lausitz, Teil II, Dicotyledonen. Palaeontogr. Abt. B 257, 35–174.

Mandel, J.R., Barker, M.S., Bayer, R.J., Dikow, R.B., Gao, T., Jones, K.E., Keely, S., Kilian, N., Ma, H., Siniscalchi, C.M., Susanna, A., Thapa, R., Watson, L., Funk, V.A., 2017. The Compositae Tree of Life in the age of phylogenomics. J. System. Evol. 55, 405–410. https://doi.org/10.1111/jse.12265.

Mandel, J.R., Dikow, R.B., Funk, V.A., 2015. Using phylogenomics to resolve mega-families: an example from Compositae. J. System. Evol. 53, 391–402. https://doi.org/10.1111/jse.12167.

Mandel, J.R., Dikow, R.B., Funk, V.A., Masalia, R.R., Staton, S.E., Kozik, A., Michelmore, R.W., Rieseberg, L.H., Burke, J.M., 2014. A target enrichment method for gathering phylogenetic information from hundreds of loci: an example from the Compositae. Appl. Plant Sci. 2, 1300085. https://doi.org/10.3732/apps.1300085.

Meseguer, A.S., Lobo, J.M., Ree, R., Beerling., D.J., Sanmartín, I., 2015. Integrating fossils, phylogenies, and niche models into biogeography to reveal ancient evolutionary history: the case of *Hypericum* (Hypericaceae). Syst. Biol. 64, 215–232, https://doi.org/10.1093/sysbio/syu088

Miller, M.A., Pfeiffer, W., Schwartz, T., 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In: Proceedings of the Gateway Computing Environments Workshop (GCE), 14 Nov. 2010. New Orleans, pp. 1–8. https://doi.org/10.1109/GCE.2010.5676129.

Mitchell, N., Lewis, P.O., Lemmon, E.M., Lemmon, A.R., Holsinger, K.E., 2017. Anchored phylogenomics improves the resolution of evolutionary relationships in the rapid radiation of *Protea* L. Am. J. Bot. 104, 102–115. https://doi.org/10.3732/ajb.1600227

Morales-Briones, D.F., Liston, A., Tank, D.C., 2018. Phylogenomic analyses reveal a deep history of hybridization and polyploidy in the Neotropical genus *Lachemilla* (Rosaceae). New Phytol. 218, 1668–1684. https://doi.org/10.1111/nph.15099

Nicholls, J.A., Pennington, R.T., Koenen, E.J., Hughes, C.E., Hearn, J., Bunnefeld, L., Dexter, K.G., Stone, G.N., Kidner, C.A., 2015. Using targeted enrichment of nuclear genes to increase phylogenetic resolution in the neotropical rain forest genus *Inga* (Leguminosae: Mimosoideae). Front. Plant Sci. 6, 710. https://doi.org/10.3389/fpls.2015.00710.

Ortiz, S., Bonifacino, J.M., Crisci, J.V., Funk, V.A., Hansen, H.V., Hind, D.J.N., Katinas, L., Roque, N., Sancho, G., Susanna, A., Tellería, M.C., 2009. The basal grade of the Compositae: Mutisieae (sensu Cabrera) and Carduoideae. In: Funk, V.A., Susanna, A., Stuessy, T.F., Bayer, R.J. (Eds.), Systematics, evolution, and biogeography of Compositae. International Association for Plant Taxonomy, Vienna, Austria, pp. 193–213.

Panero, J.L., Crozier, B.S., 2016. Macroevolutionary dynamics in the early diversification of Asteraceae. Mol. Phylogenet. Evol. 99, 116–132. https://doi.org/10.1016/j.ympev.2016.03.007

Panero, J.L., Funk, V.A., 2008. The value of sampling anomalous taxa in phylogenetic studies: Major clades of the Asteraceae revealed. Mol. Phylogenet. Evol. 47, 757–782.

Parks, M., Cronn, R., Liston, A., 2009. Increasing phylogenetic resolution at low taxonomic levels using massively parallel sequencing of chloroplast genomes. BMC Biol. 7, 84. https://doi.org/10.1186/1741-7007-7-84

Pearson, P.N., Foster, G.L., Wade, B.S., 2009. Atmospheric carbon dioxide through the Eocene–Oligocene climate transition Nature 461, 1110–1114. https://doi.org/10.1038/nature08447

Petit, D.P., 1988. Le genre *Echinops* L. (Compositae, Cardueae). 1. Position phylétique et interprétation de l'incapitulescence. Candollea 43, 467–481.

Petit, D.P., 1997. Generic interrelationships of the Cardueae (Compositae): a cladistic analysis of morphological data. Plant Syst. Evol. 207, 173–203. https://doi.org/10.1007/BF00984388

Popescu, S.M., 2002. Repetitive changes in Early Pliocene vegetation revealed by high-resolution pollen analysis: Revised cyclostratigraphy of southwestern Romania. Rev. Palaeobot. Palynol. 120, 181–202. https://doi.org/10.1016/S0034-6667(01)00142-7

Pound M.J., Salzmann, U., 2017. Heterogeneity in global vegetation and terrestrial climate change during the late Eocene to early Oligocene transition. Sci. Rep. 7, 43386. https://doi.org/10.1038/srep43386

R Core Team, 2014. R: A language and environment for statistical computing. R Foundation for statistical Computing, Vienna.

Rambaut, A. 2016. FigTree v. 1.4.3. Department of Zoology, University of Oxford, Oxford. Available at http://tree.bio.ed.ac.uk/software/figtree/

Ranwez, V., Harispe, S., Delsuc, F., Douzery, E.J., 2011. MACSE: Multiple Alignment of Coding SEquences accounting for frameshifts and stop codons. PloS One 6: e22594. https://doi.org/10.1371/journal.pone.0022594

Renner, S.S., Ricklefs, R.E., 1995. Dioecy and its correlates in the flowering plants. Am. J. Bot. 82, 596–606. https://doi.org/10.1002/j.1537-2197.1995.tb11504.x

Sanderson, M.J., 2002. Estimating absolute rates of molecular evolution and divergence times: a penalized likelihood approach. Mol. Biol. Evol. 19, 101–109. https://doi.org/10.1093/oxfordjournals.molbev.a003974

Sayyari, E., Mirarab, S., 2016. Fast coalescent-based computation of local branch support from quartet frequencies. Mol. Biol. Evol. 33, 1654–1668. https://doi.org/10.1093/molbev/msw079

Schlick-Steiner, B.C., Steiner, F.M., Seifert, B., Stauffer, C., Christian, E., Crozier, R.H., 2010. Integrative taxonomy: A multisource approach to exploring biodiversity. Ann. Rev. Entomol. 55, 421–438. https://doi.org/10.1146/annurev-ento-112408-085432

Schubert, M., Marcussen, T., Meseguer, A.S., Fjellheim, S., 2018. The grass subfamily Pooideae: late Cretaceous origin and climate-driven Cenozoic diversification. bioRxiv. https://doi.org/10.1101/462440

Shaw, T.I., Ruan, Z., Glenn, T.C., Liu, L., 2013. STRAW: Species TRee Analysis Web server. Nucleic Acids Res. W41, W238–W241.

Small, R.L., Cronn, R.C., Wendel, J.F., 2004. Use of nuclear

genes for phylogeny reconstruction in plants. Aust. Syst. Bot. 17: 145–170.

Small, R.L., Ryburn, J.A., Cronn, R.C., Seelanan, T., Wendel, J.F., 1998. The tortoise and the hare: choosing between noncoding plastome and nuclear *Adh* sequences for phylogeny reconstruction in a recently diverged plant group. Am. J. Bot. 85, 1301–1315.

Smith, S.A., Moore, M.J., Brown, J.W., Yang, Y., 2015. Analysis of phylogenomic datasets reveals conflict, concordance, and gene duplications with examples from animals and plants. BMC Evol. Biol. 15, 150. https://doi.org/10.1186/s12862-015-0423-0

Smith, S.A., O'Meara, B.C., 2012. treePL: divergence time estimation using penalized likelihood for large phylogenies. Bioinformatics 28, 2689–2690. https://doi.org/10.1093/bioinformatics/bts492

Stamatakis, A., 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22, 2688–2690. https://doi.org/10.1093/bioinformatics/btl446.

Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30, 1312–1313. https://doi.org/10.1093/bioinformatics/btu033

Stegemann, S., Keuthe, M., Greiner, S., Bock, R., 2012. Horizontal transfer of chloroplast genomes between plant species. Proc. Natl. Acad. Sci. U.S.A., 109, 2434–2438. https://doi.org/10.1073/pnas.1114076109

Stoughton, T.R., Kriebel, R., Jolles, D.D., O'Quinn, R.L., 2018. Next-generation lineage discovery: A case study of tuberous *Claytonia* L. Am. J. Bot. 105, 536–548. https://doi.org/10.1002/ajb2.1061

Stubbs, R.L., Folk, R.A., Xiang, C.L., Soltis, D.E., Cellinese, N., 2018. Pseudo-parallel patterns of disjunctions in an Arctic-alpine plant lineage. Mol. Phylogenet. Evol. 123, 88–100. https://doi.org/10.1016/j.ympev.2018.02.016

Susanna, A., 1987. *Femeniasia*, *novus genus Carduearum*. Collect. Bot. 17, 83–88. https://doi.org/10.3989/collectbot.1988.v17.158

Susanna, A., Garcia-Jacas. N., 2007. Tribe Cardueae. In: Kadereit, J.W., Jeffrey, C. (Eds.), Flowering Plants. Eudicots. Asterales, vol. 8. In: Kubitzki, K. (Ed.), The families and genera of vascular plants. Springer Verlag, Berlin, Germany, pp. 123–146.

Susanna, A., Garcia-Jacas, N., 2009. Cardueae (Carduoideae). In: Funk, V.A., Susanna, A., Stuessy, T.F., Bayer, R.J. (Eds.), Systematics, evolution, and biogeography of Compositae. IAPT, Vienna, Austria. pp. 293–313.

Susanna, A., Garcia-Jacas, N., Hidalgo, O., Vilatersana, R. Garnatje, T., 2006. The Cardueae (Compositae) revisited: Insights from ITS, trnL-trnF, and matK nuclear and chloroplast DNA analysis. Ann. Mo. Bot. Gard. 93, 150–171. https://doi.org/10.3417/0026-6493(2006)93[150:TCCRIF]2.0.CO;2

Susanna, A., Garcia-Jacas, N., Soltis, D.E., Soltis, P.S., 1995. Phylogenetic relationships in tribe Cardueae (Asteraceae) based on ITS sequences. Am. J. Bot. 82, 1056–1068. https://doi.org/10.1002/j.1537-2197.1995.tb11571.x

Susanna, A., Garnatje, T., Garcia-Jacas, N., Vilatersana, R., 2002. On the correct subtribal placement of the genera *Syreitschikovia* and *Nikitinia* (Asteraceae, Cardueae): Carduinae or Centaureinae? Bot. J. Linn. Soc. 140, 313–319. https://doi.org/10.1046/j.1095-8339.2002.00104.x

Tillich, M., Lehwark, P., Pellizzer, T., Ulbricht-Jones, E. S., Fischer, A., Bock, R., Greiner, S., 2017. GeSeq–versatile and accurate annotation of organelle genomes. Nucleic Acids Res. 45(W1), W6–W11. https://doi.org/10.1093/nar/gkx391

Vamosi, J.C., Otto, S.P., Barrett, S.C., 2003. Phylogenetic analysis of the ecological correlates of dioecy in angiosperms. J. Evol. Biol. 16, 1006–1018. https://doi.org/10.1046/j.1420-9101.2003.00559.x

Viljoen, E., Odeny, D.A., Coetzee, M.P., Berger, D.K., Rees, D.J., 2018. Application of chloroplast phylogenomics to resolve species relationships within the plant genus *Amaranthus*. J. Mol. Evol. 86, 216–239. https://doi.org/10.1007/s00239-018-9837-9

Villaverde, T., Pokorny, L., Olsson, S., Rincón-Barrado, M., Johnson, M.G., Gardner, E.M., Wickett, N.J., Molero, J., Riina, R., Sanmartín, I., 2018. Bridging the micro- and macroevolutionary levels in phylogenomics: Hyb-Seq solves relationships from populations to species and above. New Phytol. 220, 636–650. https://doi.org/10.1111/nph.15312

Wagenitz, G., 1955. Pollenmorphologie und Systematik in der Gattung *Centaurea* L. s.l. Flora 142, 213–279.

Wagenitz, G., 1958. Die Gattung *Myopordon* Boiss. (Compositae - Cynareae). Ber. Deutsch. Bot. Ges. 71, 271–277.

Wagenitz, G., 1976. Systematics and phylogeny of the Compositae (Asteraceae). Plant Syst. Evol. 125, 29–46.

Weitemier, K., Straub, S.C., Cronn, R.C., Fishbein, M., Schmickl, R., McDonnell, A., Liston, A. 2014. Hyb-Seq: Combining target enrichment and genome skimming for plant phylogenomics. App. Plant Sci. 2, 1400042. https://doi.org/10.3732/apps.1400042

Zachos, J., Pagani, M., Sloan, L., Thomas, E., Billups, K., 2001. Trends, rhythms, and aberrations in global climate 65 Ma to present. Science 292, 686–693. https://doi.org/10.1126/science.1059412

Zhang, C., Rabiee, M., Sayyari, E., Mirarab, S., 2018. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. BMC Bioinformatics 19, 153. https://doi.org/10.1186/s12859-018-2129-y

7. Appendices

All alignments and tree files for each dataset and appendices of divergence time analyses are deposited in Mendeley Data repository (http://dx.doi.org/10.17632/bhyv6rmyt6.1).

8. Supplementary material

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.ympev.2019.05.001

Table S1. Species sampled, authority, and collection information. Groups from 1 to 12 are the Cardueae subtribes.

Taxon group	Species	Authority	Location and voucher
	Brachylaena discolor	DC.	Cultivated at the Kirstenbosch Botanical Garden.
	Chaetanthera sp.	–	Argentina, *Funk 13268* (US).
	Chuquiraga sp.	–	Chile, Arica, *Funk 13112 et al.* (US).
	Dicoma anomala	Sond.	Angola, Huambo, *S. Ortiz 853et al.* (SANT).
	Famatinanthus sp.	–	Argentina, La Rioja, *Funk 13233 et al.* (US).
	Fulcaldea stuessyi	Roque & V.A. Funk	Brazil, Bahia, *Abreu 123* (US).
	Gerbera hybrida	hort.	USA, Greenhouse grown cutting, *Mandel 105* (MEM).
	Macledium salignum	(Lawalrée) S. Ortiz	Angola, Huambo, *S. Ortiz 889 et al.* (SANT).
	Nastanthus patagonicus	Speg.	Argentina, Santa Cruz, *Bonifacino 4016 & Funk* (US).
	Oldenburgia papionum	DC.	South Africa, Western Cape, *S. Ortiz 831a* (SANT).
	Pleiotaxis pulcherrima	Klatt	Angola, Huambo, *S. Ortiz 897 et al.* (SANT).
1. Carlininae	*Atractylis echinata*	Pomel	Argelia, Naama, Djebel Antar, *Susanna s. n. et al.*, 3.VI.2010 (BC).
	Atractylodes japonica	Koidz.	Japan, Honshu, *Saito & Otomo s. n.*, 3.IX.1973 (B).
	Carlina biebersteinii	Bernh. ex Hornem.	Turkey, Konya, *Susanna 2238 et al.* (BC).
	Carlina diae	(Rech.f.) Meusel & Kästner	Cultivated at the Botanic Garden of Barcelona.
	Thevenotia persica	DC.	Iran, Tehran, *Susanna 1635B et al.* (BC).
	Tugarinovia mongolica	Iljin	Mongolia, south Gobi, *Vallès s. n. et al.*, 3.IX.2004 (BCN).
2. Cardopatiinae	*Cardopatium corymbosum*	Pers.	Greece, Macedonia: Thessaloniki, *Roché & Susanna 1951* (BC).
	Cousiniopsis atractyloides	(C.Winkl.) Nevski	Afghanistan, Kataghan, *Rechinger 33825* (E 00460517).
3. Echinopsidinae	*Echinops karatavicus*	Regel & Schmalh.	Tajikistan: Vorzov Research Station, *Susanna et al. 2458* (BC).
	Echinops onopordum	P.H. Davis	Turkey, Antalya: Tahtali Dağ, near Beycik, *Susanna 2270B et al.* (BC).
	Echinops strigosus	L.	Spain, Andalucia, *Watson 95-7A* (MU).
4. Dipterocominae	*Dipterocome pusilla*	Fisch. & C.A. Mey.	Irán, *Assadi & Maasoumi 50157* (TARI).
5. Xeranthemineae	*Amphoricarpos autariatus*	Blečić & E.Mayer	Montenegro, Tara Gorge, *Janaćković s. n.*, 20.VIII.2010 (BEOU).
	Amphoricarpos exsul	O.Schwarz	Turkey, Muğla, *Susanna 2256 et al.* (BC).
	Shangwua denticulata	(DC.) Raab-Straube & Yu J. Wang	China, Xizang, Nielamu, *J.Q. Liu 07150* (BC)
	Xeranthemum annuum	L.	Turkey, Gaziantep, *Susanna 2323 et al.* (BC).
6. Berardiinae	*Berardia subacaulis*	Vill.	France, Alpes Maritimes, *C. Roquet s. n.* (Roquet pers. herb.).
7. Staehelininae	*Staehelina baetica*	DC.	Spain, Málaga, *Vilatersana 2185 et al.* (BC).
	Staehelina dubia	L.	Spain, Huesca, *Hilpold AH20084050 & López Alvarado* (BC).
	Staehelina lobelii	DC.	Turkey, Antalya, *Susanna 2272 et al.* (BC).
	Staehelina petiolata	(L.) Hilliard & B.L.Burtt	Greece, Crete, *Vitek 081418* (BC).
8. Onopordiinae	*Alfredia acantholepis*	Kar. & Kir.	Kazakhstan, Almatinskaya oblast, *Susanna 2092 et al.* (BC).
	Olgaea petriprimi	B.A.Sharipova	Tajikistan, Selandi, *Susanna 2539 et al.* (BC).
	Onopordum nervosum	Boiss.	Cultivated at the Dijon Botanical Garden.
	Syreitschikovia tenuis	(Bunge) Botsch.	Kyrgyzstan, *Seregin 2067* (FRU).
9. Carduinae	*Carduus nutans*	L.	Tajikistan, Vorzov, *Susanna 2481 et al.* (BC).
	Carduus pycnocephalus	L.	Spain, Barcelona, *Garnatje & Susanna 1827* (BC).
	Cirsium acaulon	(L.) Scop.	Tajikistan, Voru, *Susanna 2522 et al.* (BC).
	Cirsium sairamense	O. Fedtsch. & B. Fedtsch.	Tadjikistan, Maijora, *Susanna 2468 et al.* (BC).

Taxon group	Species	Authority	Location and voucher
	Cynara cardunculus	L.	USA, UW Medicinal Plant Garden., *Mandel s.n.* (GA 135).
	Galactites tomentosa	Moench	Spain, Barcelona, *Susanna s. n.*, 8.VI.2017 (BC).
	Ptilostemon diacantha	(Labill.) Greuter	Turkey, Adana, *Susanna 2313 et al.* (BC).
10. Arctiinae	*Arctium abolinii*	(Kult. ex Tscherneva) S. López et al.	Kyrgyzstan, Jalal Abad Oblast, *Lazkov s. n.* (LE).
	Arctium aureum	(C. Winkl.) Kuntze	Tajikistan, Schtut, *Susanna 2514 et al.* (BC).
	Arctium minus	(Hill) Bernh.	Spain, Barcelona, *Vilatersana 1100 & López-Vinyallonga* (BC).
	Cousinia ninae	Juz.	Kyrgyzstan, Oshskaya, *Sultanova s. n.* (LE).
	Cousinia pusilla	C. Winkl.	Tajikistan, S Tajikistan, *Botschantzev 117* (LE).
	Cousinia sogdiana	Bornm.	Uzbekistan, Karakalpakstan, *Kalibernova 5262 et al.* (LE).
11. Saussureinae	*Jurinea fontqueri*	Cuatrec.	Spain, Jaén, Mágina, *Martínez Lirola s. n.* (GDA).
	Jurinea orientalis	(Iljin) Iljin	Kyrgyzstan, Shekoftar, *Sennikov s. n.*, 13.VII.2016 (H).
	Jurinea karategina	(Lipsky) O.Fedtsch.	Tajikistan, Surjov, *Susanna 2542 et al.* (BC).
	Jurinea carduicephala	Iljin	Tajikistan: Gorno-Badakhshan, *Semakov & Dengubenko s. n.* (LE 8428).
	Saussurea controversa	DC.	Russia, Krasnoyarsk Krai, *Cazzolla Gatti 10005 et al.* (TK).
	Saussurea elegans	Ledeb.	Tajikistan, Fan mountains, *Susanna 2505 et al.* (BC).
	Saussurea leucophylla	Schrenk	Russia, Altai, *A. Pyak & E. Pyak 11073* (TK).
12. Centaureinae	*Amberboa moschata*	(L.) DC.	Armenia, *Vitek et al. 03-1539* (BC).
	Carduncellus dianius	Webb	Cultivated at the Barcelona Botanical Garden.
	Carthamus tinctorius	L.	Greenhouse grown seed, USDA PI 592391.
	Centaurea aspera	L.	Cultivated at the Barcelona Botanical Garden.
	Centaurea babylonica	M.Bieb.	Cultivated at the Barcelona Botanical Garden.
	Centaurea benedicta	(L.) L.	Greenhouse grown seed, USDA PI 311739.
	Centaurea clementei	Boiss. ex DC.	Cultivated at the Barcelona Botanical Garden.
	Centaurea lanigera	DC.	Turkey, Niğde, *Susanna 2296 et al.* (BC).
	Centaurea tauromenitana	Guss.	Cultivated at the Barcelona Botanical Garden.
	Centaurea triumfettii	All.	Turkey, Ardahan, *Susanna 2398 et al.* (BC).
	Centaurodendron dracaenoides	Johow	Chile, Juan Fernández, Masatierra, *Baeza & Peñailillo 11314* (OS).
	Cheirolophus arbutifolius	(Svent.) G.Kunkel	Cultivated at the Barcelona Botanical Garden.
	Cheirolophus crassifolius	(Bertol.) Susanna	Cultivated at the Barcelona Botanical Garden.
	Crupina crupinastrum	(Moris) Vis.	Morocco, Xauen, *Roché & Susanna 1894* (BC).
	Crupina vulgaris	Pers. ex Cass.	Spain, Palencia, *Roché & Susanna 1928* (BC).
	Klasea coriacea	(DC.) Holub	Armenia, Ararat, *Susanna 1530 et al.* (BC).
	Klasea flavescens	(L.) Holub	Spain, Almería, *Susanna 1615 et al.* (BC).
	Klasea quinquefolia	(Willd.) Greuter & Wagenitz	Iran, Azarbaijan-e Sharqi, *Susanna 1677 et al.* (BC).
	Phalacrachena inuloides	(Fisch.) Iljin	Ukraina, *Romaschenko 402 & Didukh* (BC).
	Phonus arborescens	(L.) G.López	Cultivated at the Barcelona Botanical Garden.
	Plagiobasis centauroides	Schrenk	Kazakhstan, Almaty, *Susanna 2130 et al.* (BC).
	Plectocephalus cachinalensis	(Phil.) N.Garcia & Susanna	Cultivated at the Barcelona Botanical Garden.
	Plectocephalus tweediei	(Hook. & Arn.) Garcia-Jacas & Susanna	Argentina, Chaco, *Delucchi 1970* (LP).
	Psephellus mucroniferus	(DC.) Wagenitz	Turkey, Niğde, *Susanna 2300 et al.* (BC).
	Psephellus simplicicaulis	(Boiss. & A.Huet) Wagenitz	Turkey, Erzurum, *Susanna 2386 et al.* (BC).
	Rhaponticoides alpina	(L.) M.V. Agab. & Greuter	Spain, Segovia, *Garcia-Jacas & Susanna 2798* (BC).
	Rhaponticoides centaurium	(L.) M.V. Agab. & Greuter	Italy, Apulia, *Rosati s. n.*, 22.VI.2015 (BC).
	Rhaponticum acaule	DC.	Cultivated at the Barcelona Botanical Garden.
	Rhaponticum integrifolium	C.Winkl.	Tajikistan, Vorzov, *Susanna 2478 et al.* (BC).
	Rhaponticum lyratum	Nyman	Tajikistan, Vorzov, *Susanna 2568 et al.* (BC).
	Volutaria canariensis	Wagenitz	Spain, Gran Canaria, *Kunkel 200-2* (BC).

Table S2. Species sampled and their corresponding number of raw reads, and number of recovered loci for nuclear (nrDNA) dataset (with target extraction methods: PHYLUCE from the total of 776 loci; and HybPiper from the total of 1055 loci) and chloroplast (cpDNA) dataset comprising a total of 87 genes. Groups from 1 to 12 are the Cardueae subtribes.

Taxon group	Species	Sequencing project	Nº raw reads	nrDNA		cpDNA
				Nº loci PHYLUCE	Nº loci HybPiper	Nº loci
	Brachylaena discolor	Present study	7,353,990	295	1021	84
	Chaetanthera sp.	Mandel et al. (2017)	117,364	376	445	–
	Chuquiraga sp.	Mandel et al. (2017)	39,410	192	21	72
	Dicoma anomala	Present study	4,014,387	423	952	84
	Famatinanthus sp.	Mandel et al. (2017)	1,179,957	254	991	84
	Fulcaldea stuessyi	Mandel et al. (2014)	8,999,244	358	498	82
	Gerbera hybrida	Mandel et al. (2014)	16,000,000	359	814	83
	Macledium salignum	Present study	7,513,213	455	913	84
	Nastanthus patagonicus	Mandel et al. (2017)	7,483,947	327	644	83
	Oldenburgia papionum	Present study	3,510,938	301	1017	84
	Pleiotaxis pulcherrima	Present study	7,661,670	438	971	85
1. Carlininae	*Atractylis echinata*	Present study	11,345,804	378	940	84
	Atractylis japonica	Present study	6,048,498	335	970	84
	Carlina biebersteinii	Mandel et al. (2014)	4,997,617	265	941	84
	Carlina diae	Mandel et al. (2014)	2,292,638	333	935	–
	Thevenotia persica	Present study	4,399,025	402	897	84
	Tugarinovia mongolica	Present study	3,685,187	334	981	–
2. Cardopatiinae	*Cardopatium corymbosum*	Present study	687,001	437	896	84
	Cousiniopsis atractyloides	Present study	3,918,260	419	921	84
3. Echinopsidinae	*Echinops karatavicus*	Present study	6,383,306	343	979	84
	Echinops onopordum	Present study	898,399	415	970	–
	Echinops strigosus	Mandel et al. (2014)	631,932	409	982	80
4. Dipterocominae	*Dipterocome pusilla*	Present study	4,014,451	385	973	84
5. Xerantheminae	*Amphoricarpos autariatus*	Present study	13,954,763	318	1010	84
	Amphoricarpos exsul	Present study	350,169	391	973	–
	Shangwua denticulata	Present study	5,295,384	315	1019	84
	Xeranthemum annuum	Present study	4,526,789	412	907	84
6. Berardiinae	*Berardia subacaulis*	Present study	4,849,918	383	1006	84
7. Staehelininae	*Staehelina baetica*	Present study	11,036,138	332	1003	84
	Staehelina dubia	Present study	3,031,747	438	996	84
	Staehelina lobelii	Present study	6,287,680	407	1000	84
	Staehelina petiolata	Present study	5,695,551	438	990	84
8. Onopordiinae	*Alfredia acantholepis*	Herrando-Moraira et al. (2018)	8,217,881	338	509	84
	Olgaea petriprimi	Herrando-Moraira et al. (2018)	5,310,933	341	1000	84
	Onopordum nervosum	Present study	2,796,798	359	989	84
	Syreitschikovia tenuis	Present study	3,390,943	359	982	85
9. Carduinae	*Carduus nutans*	Present study	3,891,005	385	959	85
	Carduus pycnocephalus	Herrando-Moraira et al. (2018)	741,845	367	666	–
	Cirsium acaulon	Present study	11,858,315	391	979	84
	Cirsium sairamense	Herrando-Moraira et al. (2018)	5,389,901	371	987	85
	Cynara cardunculus	Mandel et al. (2014)	454,885	429	796	86
	Galactites tomentosa	Present study	3,680,490	387	932	85
	Ptilostemon diacantha	Present study	5,299,804	417	998	84
10. Arctiinae	*Arctium abolinii*	Herrando-Moraira et al. (2018)	1,853,731	300	1002	–

Taxon group	Species	Sequencing project	Nº raw reads	nrDNA		cpDNA
				Nº loci PHYLUCE	Nº loci HybPiper	Nº loci
	Arctium aureum	Herrando-Moraira et al. (2018)	4,565,685	294	1017	–
	Arctium minus	Herrando-Moraira et al. (2018)	10,007,019	349	1007	84
	Cousinia ninae	Herrando-Moraira et al. (2018)	3,280,858	303	1005	84
	Cousinia pusilla	Herrando-Moraira et al. (2018)	3,383,791	385	973	84
	Cousinia sogdiana	Herrando-Moraira et al. (2018)	3,949,050	315	999	–
11. Saussureinae	*Jurinea fontqueri*	Herrando-Moraira et al. (2018)	5,240,423	386	992	84
	Jurinea orientalis	Herrando-Moraira et al. (2018)	3,155,790	361	989	–
	Jurinea karategina	Herrando-Moraira et al. (2018)	7,713,700	393	995	85
	Jurinea carduicephala	Herrando-Moraira et al. (2018)	7,948,211	348	1013	85
	Saussurea controversa	Herrando-Moraira et al. (2018)	8,091,449	349	1011	85
	Saussurea elegans	Herrando-Moraira et al. (2018)	2,784,084	359	995	85
	Saussurea leucophylla	Herrando-Moraira et al. (2018)	5,597,695	352	1007	85
12. Centaureinae	*Amberboa moschata*	Present study	18,784,862	333	974	84
	Carduncellus dianius	Present study	5,111,933	403	943	85
	Carthamus tinctorius	Mandel et al. (2014)	10,436,332	436	857	85
	Centaurea aspera	Present study	3,215,383	365	938	85
	Centaurea babylonica	Present study	5,379,085	407	931	85
	Centaurea benedicta	Mandel et al. (2014)	644,420	423	54	85
	Centaurea clementei	Present study	9,975,987	394	947	85
	Centaurea lanigera	Present study	4,304,079	334	952	84
	Centaurea tauromenitana	Present study	4,193,748	312	940	85
	Centaurea triumfettii	Present study	6,675,594	215	976	85
	Centaurodendron dracaenoides	Present study	5,697,171	288	987	85
	Cheirolophus arbutifolius	Present study	5,547,944	413	972	84
	Cheirolophus crassifolius	Present study	5,823,163	402	982	84
	Crupina crupinastrum	Present study	3,880,017	439	949	84
	Crupina vulgaris	Present study	6,613,389	420	949	84
	Klasea coriacea	Present study	6,205,566	381	977	85
	Klasea flavescens	Present study	13,112,139	379	992	85
	Klasea quinquefolia	Present study	2,176,753	371	957	85
	Phalacrachena inuloides	Present study	5,920,329	373	985	84
	Phonus arborescens	Present study	1,975,604	404	959	85
	Plagiobasis centauroides	Present study	6,241,039	419	981	84
	Plectocephalus cachinalensis	Present study	5,898,634	169	985	85
	Plectocephalus tweediei	Present study	6,642,762	414	979	84
	Psephellus mucroniferus	Present study	2,573,295	267	963	84
	Psephellus simplicicaulis	Present study	5,370,737	229	982	85
	Rhaponticoides alpina	Present study	4,548,131	380	970	85
	Rhaponticoides centaurium	Present study	5,788,982	385	978	84
	Rhaponticum acaule	Present study	4,209,726	409	983	85
	Rhaponticum integrifolium	Present study	5,882,610	399	988	84
	Rhaponticum lyratum	Present study	7,393,929	397	975	84
	Volutaria canariensis	Present study	4,551,867	400	955	84
Average (± standard deviation)			5,420,504 (3,461,362)	364.0 (57.5)	923.4 (173.6)	84.1 (1.6)

Table S3. Median estimated ages and 95% of confidence intervals (CI) for the phylogenetic tree constructed with 1055 COS loci from nuclear HybPiper dataset by penalized likelihood method implemented in treePL software (see text for details). Node codes correspond to those represented in Figure 5.

Node code	Taxonomic unit studied	Median age (Myr)	Lower 95% CI	Upper 95% CI
1	–			
2	–	53.89	48.49	59.13
3	–	48.41	43.34	53.34
4	–	48.77	40.38	55.52
5	–	47.07	38.34	54.56
6	–	37.06	30.72	44.11
7	Subfamily Carduoideae	40.09	33.68	46.67
8	–	32.37	27.44	38.40
9	–	27.96	23.12	32.56
10	–	38.24	32.83	45.71
11	–	29.20	19.21	35.12
12	**Tribe Cardueae**	**34.12**	**29.94**	**40.04**
13	Subtribe Carlininae	**31.09**	**27.31**	**36.45**
14	–	26.51	23.20	30.97
15	–	18.19	15.93	21.30
16	–	15.41	13.37	18.01
17	–	3.92	3.30	4.71
18	–	33.36	29.40	39.05
19	Subtribe Cardopatiinae	**26.30**	**23.10**	**30.88**
20	–	31.98	28.12	37.31
21	Subtribe Echinopsinae	**14.58**	**12.00**	**18.17**
22	–	10.06	8.87	11.70
23	–	27.74	24.61	31.51
24	Subtribe Dipterocominae	**26.57**	**23.62**	**30.24**
25	Subtribe Xerantheminae	**24.56**	**21.82**	**27.93**
26	–	17.70	15.83	20.10
27	–	9.88	8.44	11.41
28	Subtribe Berardiinae	**24.54**	**21.76**	**27.40**
29	–	22.48	19.81	24.35
30	Subtribe Staehelininae	**7.26**	**6.50**	**8.00**
31	–	4.90	4.36	5.36
32	–	2.11	1.89	2.33
33	–	19.63	17.49	21.53
34	Subtribe Onopordinae	**15.59**	**13.65**	**17.57**
35	–	9.77	8.16	10.95
36	–	8.05	6.68	9.02
37	–	18.11	16.40	20.19
38	Subtribe Carduinae	**16.33**	**14.55**	**18.73**
39	–	15.62	13.78	8.22
40	–	13.95	12.05	16.88
41	–	10.53	9.18	12.72
42	–	6.86	5.88	8.25

Node code	Taxonomic unit studied	Median age (Myr)	Lower 95% CI	Upper 95% CI
43	–	8.53	7.29	10.08
44	–	16.80	15.23	18.72
45	–	14.50	13.23	16.02
46	Subtribe Arctiinae	**7.86**	**6.63**	**9.01**
47	–	5.62	4.77	6.42
48	–	4.39	3.70	5.04
49	–	5.25	4.46	5.89
50	–	4.10	3.47	4.57
51	Subtribe Saussureinae	**12.48**	**11.40**	**13.68**
52	–	6.48	5.89	7.20
53	–	4.93	4.46	5.51
54	–	7.52	6.75	8.38
55	–	4.22	3.75	4.85
56	–	4.87	4.33	5.38
57	Subtribe Centaureinae	**12.39**	**11.39**	**13.66**
58	–	11.42	10.52	12.55
59	–	8.39	7.71	9.21
60	–	11.63	10.60	12.71
61	–	9.55	8.70	10.58
62	–	6.29	5.65	7.01
63	–	5.31	4.79	5.92
64	–	5.57	5.10	6.18
65	–	1.34	1.23	1.50
66	–	10.33	9.37	11.29
67	–	4.08	3.67	4.78
68	–	9.29	8.39	10.26
69	–	5.09	4.60	5.58
70	–	3.74	3.40	4.19
71	–	8.74	7.82	9.64
72	–	3.63	3.30	4.12
73	–	8.09	7.22	8.96
74	–	4.28	3.85	4.83
75	–	7.58	6.76	8.52
76	–	2.48	2.26	2.90
77	–	6.85	6.09	7.85
78	–	6.32	5.46	7.27
79	–	4.81	4.25	5.59
80	–	3.23	2.82	3.80
81	–	5.91	5.21	7.01
82	–	2.89	2.58	3.43
83	–	5.44	4.82	6.43
84	–	4.70	4.13	5.50
85	–	5.00	4.36	5.83
86	–	3.17	2.79	3.75

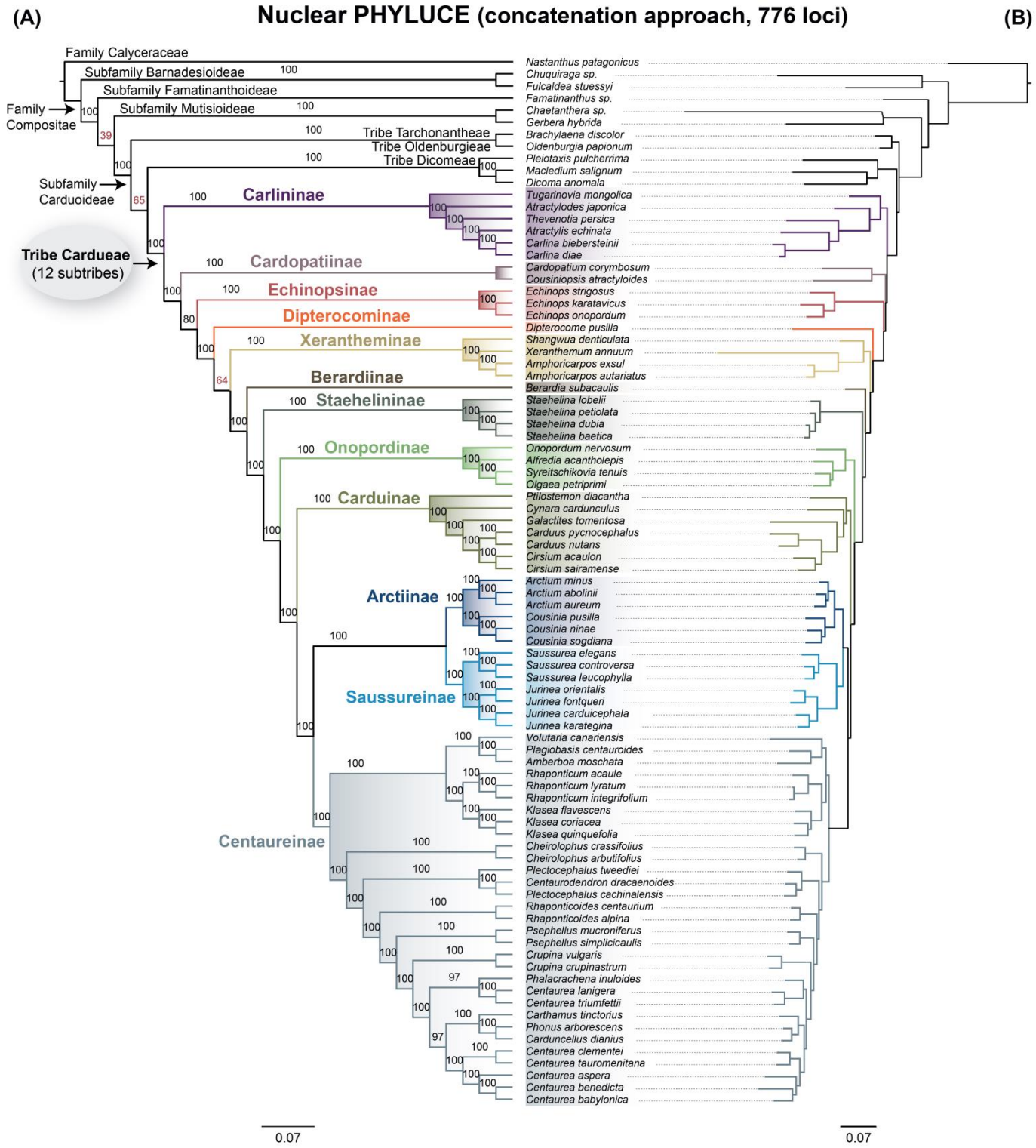


Figure S1. Phylogenetic reconstruction obtained with PHYLUCE target extraction method and under the concatenation approach (maximum likelihood analysis performed with RAxML), showing the evolutionary relationships among 12 newly defined subtribes within tribe Cardueae. (A) cladogram, (B) phylogram. Branch labels indicate bootstrap (BS) support values, those < 70% (in red) are considered statistically unsupported.

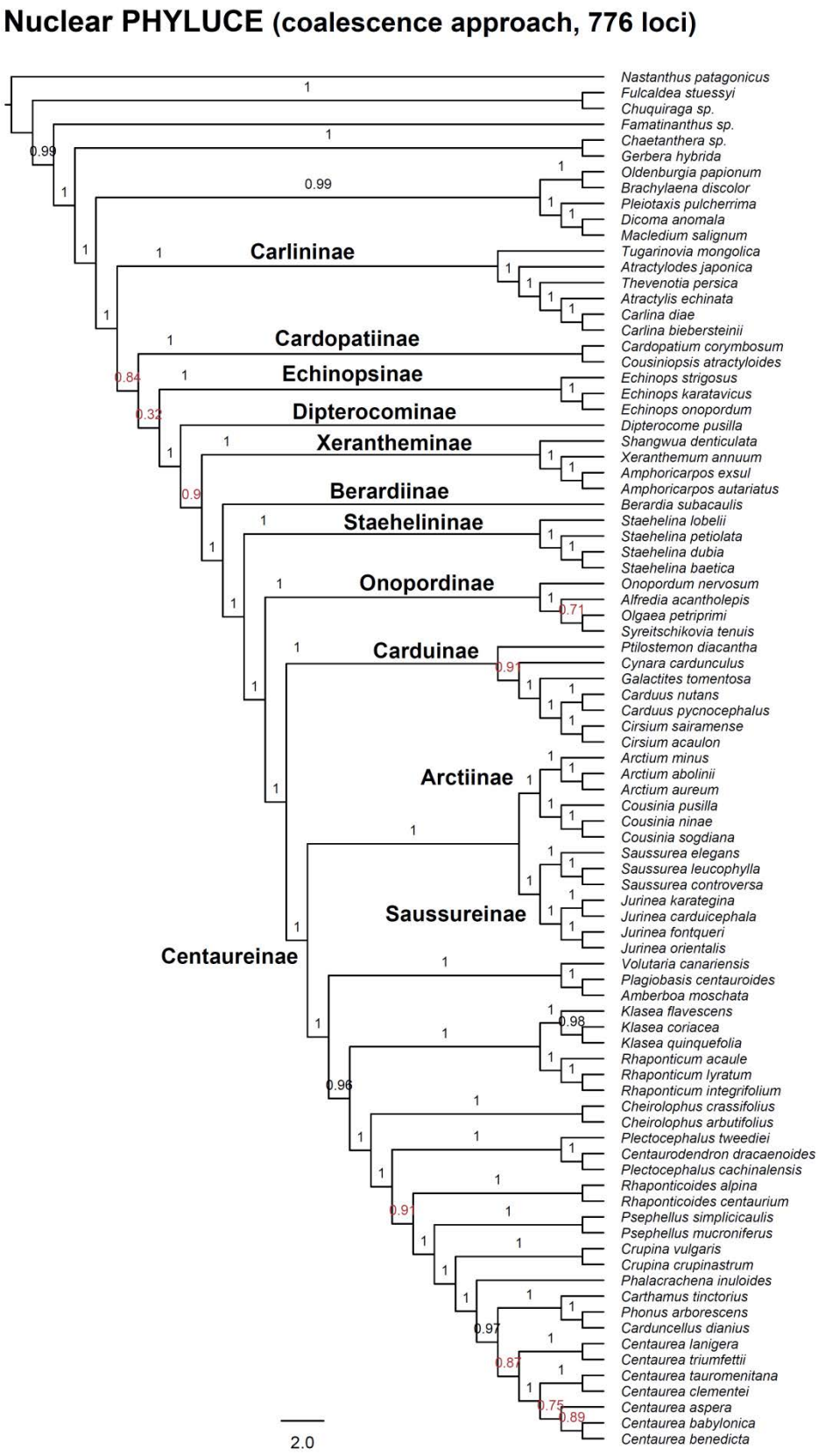


Figure S2. Phylogenetic tree inferred under coalescence approach (estimated with ASTRAL) for the nuclear PHYLUCE dataset, composed by 776 COS loci. Branch labels indicate local posterior probabilities (LPP), which values < 0.95 (in red) are considered statistically low supported.

Chapter 3

Generic boundaries in subtribe Saussureinae (Compositae: Cardueae): Insights from Hyb-Seq data

Sonia Herrando-Moraira[1] and The Cardueae Radiations Group

In alphabetical order:

Juan-Antonio Calleja[2]
You-Sheng Chen[3]
Kazumi Fujikawa[4]
Mercè Galbany-Casals[5]
Núria Garcia-Jacas[1]
Seung-Chul Kim[6]
Jian-Quan Liu[7]
Javier López-Alvarado[5]
Jordi López-Pujol[1]
Jennifer R. Mandel[8]
Iraj Mehregan[9]
Cristina Roquet[5]
Alexander N. Sennikov[10]
Alfonso Susanna[1]
Roser Vilatersana[1]
Lian-Sheng Xu[3]

[1] Botanic Institute of Barcelona (IBB, CSIC-Ajuntament de Barcelona), Pg. del Migdia, s.n., 08038 Barcelona, Spain
[2] Departament of Biology (Botany), Faculty of Sciences, Research Centre on Biodiversity and Global Change (CIBC-UAM), 28049 Madrid, Spain
[3] Key Laboratory of Plant Resources Conservation and Sustainable Utilization, South China Botanical Garden, Chinese Academy of Sciences, Guangzhou 510650, China
[4] Kochi Prefectural Makino Botanical Garden, 4200-6, Godaisan, Kochi 781-8125, Japan
[5] Systematics and Evolution of Vascular Plants (UAB) – Associated Unit to CSIC, Departament de Biologia Animal, Biologia Vegetal i Ecologia, Facultat de Biociències, Universitat Autònoma de Barcelona, 08193 Bellaterra, Spain
[6] Department of Biological Sciences, Sungkyunkwan University, Suwon 440-746, Republic of Korea
[7] Key Laboratory for Bio-Resources and Eco-Environment, College of Life Sciences, Sichuan University, Chengdu, China
[8] Department of Biological Sciences, Center for Biodiversity, University of Memphis, Memphis, Tennessee 38152, U.S.A.
[9] Department of Biology, Science and Research Branch, Islamic Azad University, Tehran, Iran
[10] Botanical Museum, Finnish Museum of Natural History, P.O. Box 7, 00014 University of Helsinki, Finland; and Herbarium, Komarov Botanical Institute of Russian Academy of Sciences, Prof. Popov str. 2, 197376 St. Petersburg, Russia

Address for correspondence: Sonia Herrando-Moraira, *sonia.herrando@gmail.com*

Journal, volume, pages (year): Taxon, 69, 694–714 (2020).
DOI: https://doi.org/10.1002/tax.12314

Impact factor year 2019: 2.817 (Clarivate Analytics, 2020)
Ranking ISI Journal Citation Reports ©:
- 2019: 55/234 (Plant Sciences)
- 2019: 22/50 (Evolutionary Biology)

Received: 27 January 2020; Received in revised form: 28 April 2020; Accepted: 27 June 2020

Abstract

The subtribe Saussureinae is a highly speciose group with more than 600 species distributed in the Northern Hemisphere and is particularly species-rich at the high mountains of central and eastern Asia. *Saussurea* and *Jurinea* are the two main genera described within the subtribe. However, up to 15 satellite genera are recognized in some recent taxonomic treatments with an analytical viewpoint. For the first time, we carried out a complete sampling to clarify generic boundaries based on a well-resolved phylogeny of Saussureinae. We employed a Hyb-Seq technique that targets 1061 nuclear conserved ortholog loci designed for Compositae. After a filtering of potential paralogs, 588 loci were retained to infer phylogenetic trees under concatenation and coalescence approaches. High branch support resolution was recovered at the generic level, but a non-monophyletic pattern was detected for most of the genera as they are currently circumscribed. Accordingly, we propose a new generic delimitation based on the three main clades recovered in the backbone tree, which are also in agreement with morphological evidence: *Dolomiaea*, *Saussurea*, and *Jurinea*. Following this classification into three genera, 18 new combinations are proposed. This new genus delineation will be used as a basis for future evolutionary studies in the Saussureinae.

Keywords
Asteraceae
Dolomiaea
Jurinea
Phylogenomics
Saussurea
Taxonomy

Index

1. Introduction

Saussureinae are one of the subtribes recently described in tribe Cardueae (Compositae or Asteraceae family; Herrando-Moraira & al., 2019). It is a highly speciose group comprising around 600 species, with two of the largest genera of the Cardueae: *Saussurea* DC., with ca. 400 species, and *Jurinea* Cass., with over 200 (Lipschitz, 1979; Susanna & Garcia-Jacas, 2007, 2009; Shi & Raab-Straube, 2011; Y.-S. Chen, 2015; Raab-Straube, 2017). Saussureinae comprises unarmed perennial herbs or subshrubs with leaves silver white below and glabrous above. The capitula are cylindrical or globose, often paniculate, homogamous, with achenes with an inner pappus of very long, showy, plumose bristles, basally connate in a ring (Susanna & Garcia-Jacas, 2007). On morphological basis, the group is well-defined, and it has been traditionally accepted as a natural one (Susanna & al., 2006; Susanna & Garcia-Jacas, 2007, 2009; Shi & Raab-Straube, 2011). However, molecular confirmation has been long delayed by the difficulties in analyzing the terminal branches of the former subtribe Carduinae, but recent advances in phylogenetics using nuclear data from high-throughput sequencing (HTS) have verified the monophyly of the subtribe (Herrando-Moraira & al., 2018, 2019).

As a general geographic distribution, Saussureinae are located in the Northern Hemisphere. It is particularly species-rich in the high mountains (>1000 m above sea level) of central and eastern Asia. *Saussurea* and *Jurinea*, the main genera, represent excellent models to explore mountain radiations across the Northern Hemisphere regions (Wen & al., 2014; Xu & al., 2019). *Saussurea* has spectacularly radiated in the Himalayas, the Qinghai-Tibetan Plateau and particularly in the adjacent Hengduan mountains. *Jurinea* has radiated in the Middle Asia mountains, the Tian Shan and the Pamir-Alai.

A clarification of the generic limits is a mandatory first step to obtain independent phylogenies for the two main genera. The problem arises from the high number of small genera segregated from *Saussurea* and *Jurinea*. These are *Aegopordon* Boiss., *Anacantha* (Iljin) Soják, *Aplotaxis* DC., *Aucklandia* Falc., *Bolocephalus* Hand.-Mazz., *Diplazoptilon* Y.Ling, *Dolomiaea* DC., *Frolovia* (DC.) Lipsch., *Hemisteptia* Fisch. & C.A.Mey., *Himalaiella* Raab-Straube, *Hyalochaete* Dittrich & Rech.f., *Jurinella* Jaub. & Spach, *Lagurostemon* Cass., *Lipschitziella* Kamelin, *Mazzettia* Iljin, *Modestia* Kharadze & Tamamsch., *Outreya* Jaub. & Spach, *Perplexia* Iljin, *Pilostemon* Iljin, *Polytaxis* Bunge, *Theodorea* Cass., and *Vladimiria* Iljin. Differences in recent treatments are striking (Table 1) and include from a wide concept of *Saussurea* and *Jurinea* (Susanna & Garcia-Jacas, 2009) to the extreme analytical view of Shi & Raab-Straube (2011). Traditional syntheses of Compositae (Bentham, 1873; Hoffmann, 1894; Dittrich, 1977; Bremer, 1994) also differ significantly and without agreement in the classification. Some of the described genera are no longer accepted in major Floras or taxonomic treatments (Table 1), and they are synonymized under *Saussurea* (*Aplotaxis*, *Lagurostemon*, *Theodorea*), under *Jurinea* (*Perplexia*), or under *Dolomiaea* (*Mazzettia*, *Vladimiria*). To delineate an updated generic classification, the affinities of the rest of genera should be examined.

One of the main reasons for this proliferation of segregates is that almost all the morphological comparisons were made to *Saussurea*, not to *Dolomiaea* or *Jurinea*, and most of them were based on achene characters. The smooth achenes of *Saussurea* s.str. are quite uniform, but the achenes of *Dolomiaea* and *Jurinea* are highly variable: they are usually wrinkled, sulcate, pitted or diversely ridged, and prone to develop protrusions, horns or coronules (see especially Häffner, 2000) that have been overinterpreted. An example of this is the traditional but polyphyletic circumscription of *Saussurea* by Lipschitz (1954, 1979), the source of a considerable part of segregate genera (e.g., from *S.* subg. *Jurinocera* to *Lipschitziella* [Kamelin, 1993], from *S.* sect. *Frolovia* to *Frolovia* [Lipschitz, 1954], or *S.* sect. *Elatae* to *Himalaiella* [Raab-Straube, 2003]).

Another challenge within the study of Saussureinae's systematics is that most of the species are adapted to the extreme environments like those found at higher elevations (even beyond 5000 m) or dry habitats (Shi & Raab-Straube, 2011; Raab-Straube, 2017; Szukala & al., 2019). As a consequence, taxonomists could have interpreted as synapomorphies some traits resulting from convergent/parallel evolution (homoplastic characters) that have emerged multiple times across the subtribe. Wang & al. (2009) already pointed out that the special "glasshouse" habit of some species of *Saussurea* appeared several times in different sections of the genus. Certainly, molecular characters and phylogenetic evidence would aid to discard misclassifications based on morphological adaptations (Huang & al., 2015).

Under a phylogenetic approach, some attempts have been made, but the phylogenetic studies with a wider taxon sampling were generally focused only on *Saussurea* (Raab-Straube, 2003; Kita & al., 2004; Wang & al., 2009; Xu & al., 2019) or only on *Jurinea* (Szukala & al., 2019). At the moment, none of the 12 existing phylogenetic studies has reunited a comprehensive sampling of the segregate genera (see suppl. Table S1 for a summary of phylogenetic studies performed and genus sampling). The most complete ones gathered a generic coverage of 71% in Susanna & al. (2006) and 65% in Szukala & al. (2019) (see suppl. Table S1). The monophyly of the non-monotypic segregate genera (>1 sp.) has not been evaluated because published studies only included one species (for *Diplazoptilon*, *Modestia*, and

Table 1. Recent generic taxonomic treatments for subtribe Saussureinae.

Häffner (2000)	Raab-Straube (2003, 2017), Shi & Raab-Straube (2011), Szukala & al. (2019)	Susanna & Garcia-Jacas (2007)	Susanna & Garcia-Jacas (2009)
Dolomiaea	*Dolomiaea* *Aucklandia* *Bolocephalus* *Frolovia*	*Dolomiaea*	*Dolomiaea*
Jurinea *Aegopordon* *Anacantha** *Hyalochaete* *Jurinella* *Outreya* *Pilostemon*	*Jurinea* *Himalaiella* *Lipschitziella*	*Jurinea*	*Jurinea*
Saussurea *Diplazoptilon* *Hemisteptia* *Polytaxis*	*Saussurea* *Diplazoptilon* *Hemisteptia* *Polytaxis*	*Saussurea* *Polytaxis*	*Saussurea*

* Note that *Anacantha* is a nomenclatural synonym of *Modestia*.

Pilostemon). When several species have been included, most genera were paraphyletic or polyphyletic (*Aegopordon*, *Dolomiaea*, *Frolovia*, *Himalaiella*, *Jurinella*, *Lipschitziella*). One notable exception is *Polytaxis*, from which the two sampled species formed a strongly supported monophyletic group (Susanna & al., 2006; Yuan & al., 2015).

Despite these limitations, several clades have been recovered by molecular data (see an overview in Fig. 1): (1) the "Saussurea clade": it includes *Hemisteptia* and *Polytaxis*, which are clustered as sister to the Saussurea s.str. clade, in which is deeply anchored *Diplazoptilon picridifolium* (Hand.-Mazz.) Y.Ling; (2) the "Dolomiaea clade": it includes *Dolomiaea*, *Aucklandia*, *Bolocephalus*, and *Frolovia*; (3) the "Himalaiella clade": it includes *Himalaiella*, *Lipschitziella*, *Pilostemon*, and *Diplazoptilon cooperi* (J.Anthony) C.Shih; and (4) the "Jurinea clade": it includes *Aegopordon*, *Hyalochaete*, *Jurinella*, *Modestia*, and *Outreya*, which are closely related to, or nested in, the *Jurinea* s.str. clade. Among these four clades, the only supported sister relationship was found between *Himalaiella* (clade 3) and *Jurinea* (clade 4). The relationships of this group (clades 3+4) with *Dolomiaea* (clade 2) and *Saussurea* (clade 1) remain unresolved (Fig. 1), although as a general trend *Saussurea* (clade 1) is recovered sister to the others (clades 2+3+4).

The exploration of Saussureinae phylogeny has also been difficult due to the lack of informative phylogenetic signal provided by traditional Sanger sequence markers (see references in suppl. Table S1). Previous studies failed to recover a supported dichotomous bifurcating pattern, and most of the species analyzed were anchored into large polytomies especially on trees performed with plastid datasets (Kita & al., 2004; Wang & al., 2009). Rapid species diversification has been proposed as the most plausible reason for the resulting unresolved trees. However, the emergence of HTS techniques represents a promising tool for the clarification of Saussureinae radiation (Herrando-Moraira & al., 2018; Xu & al., 2019; Zhang & al., 2019). As an example, in a comparison of two *Saussurea*-focused studies, the phylogenetic informative sites vary from 2%, in that performed with *trnL-F* and *psbA-trnH* markers (Wang & al., 2009), to 38%, in that including whole chloroplast genomes (Xu & al., 2019).

Here, in view of the well-resolved phylogenies obtained in preliminary works on tribe Cardueae (Herrando-Moraira & al., 2018, 2019), we used the same Hyb-Seq HTS approach based on a probe set of 1061 nuclear conserved ortholog loci (hereafter COS, for Conserved Ortholog Set; Mandel & al., 2014) with the following goals: (1) to obtain a well-resolved phylogeny of subtribe Saussureinae with a complete genus sampling; and (2) to test the validity of the genera described or accepted in recent taxonomic treatments based on phylogenetic evidence.

2. Materials and methods

2.1. Taxon sampling

Our sampling comprised for the first time a full generic representation of Saussureinae, including all the genera currently accepted in the analytical treatments (Häffner, 2000; Raab-Straube, 2003; Shi & Raab-Straube, 2011). We analyzed in total 17 genera, represented by 112 different species and 138 individuals (indiv.; note that more than one specimen per species was included when possible): *Aegopordon* (2 sp., 3 indiv.), *Aucklandia* (1 sp.), *Bolocephalus* (1 sp.), *Diplazoptilon* (2 sp.), *Dolomiaea* (9 sp., 15 indiv.), *Frolovia* (5 sp.), *Hemisteptia* (1 sp., 2 indiv.), *Hyalochaete* (1 sp.), *Himalaiella* (11 sp., 19 indiv.), *Jurinea* (31 sp., 32 indiv.), *Jurinella* (4 sp., 10 indiv.), *Lipschitziella* (2 sp., 3 indiv.), *Modestia* (= *Anacantha*, 1 sp.), *Outreya* (1 sp.), *Pilostemon* (1 sp.), *Polytaxis* (2 sp.), and *Saussurea* (36 sp., 39 indiv.). Two species of *Jurinea* (*J. cartilaginea* Mozaff., *J. gedrosiaca* Bornm.) that have been moved to genera *Karvandarina* Rech.f. and *Tricholepis* DC., respectively from Centaureinae (Mirtadzadini & al., 2018, 2019; Szukala & al., 2019) were included to confirm their new generic

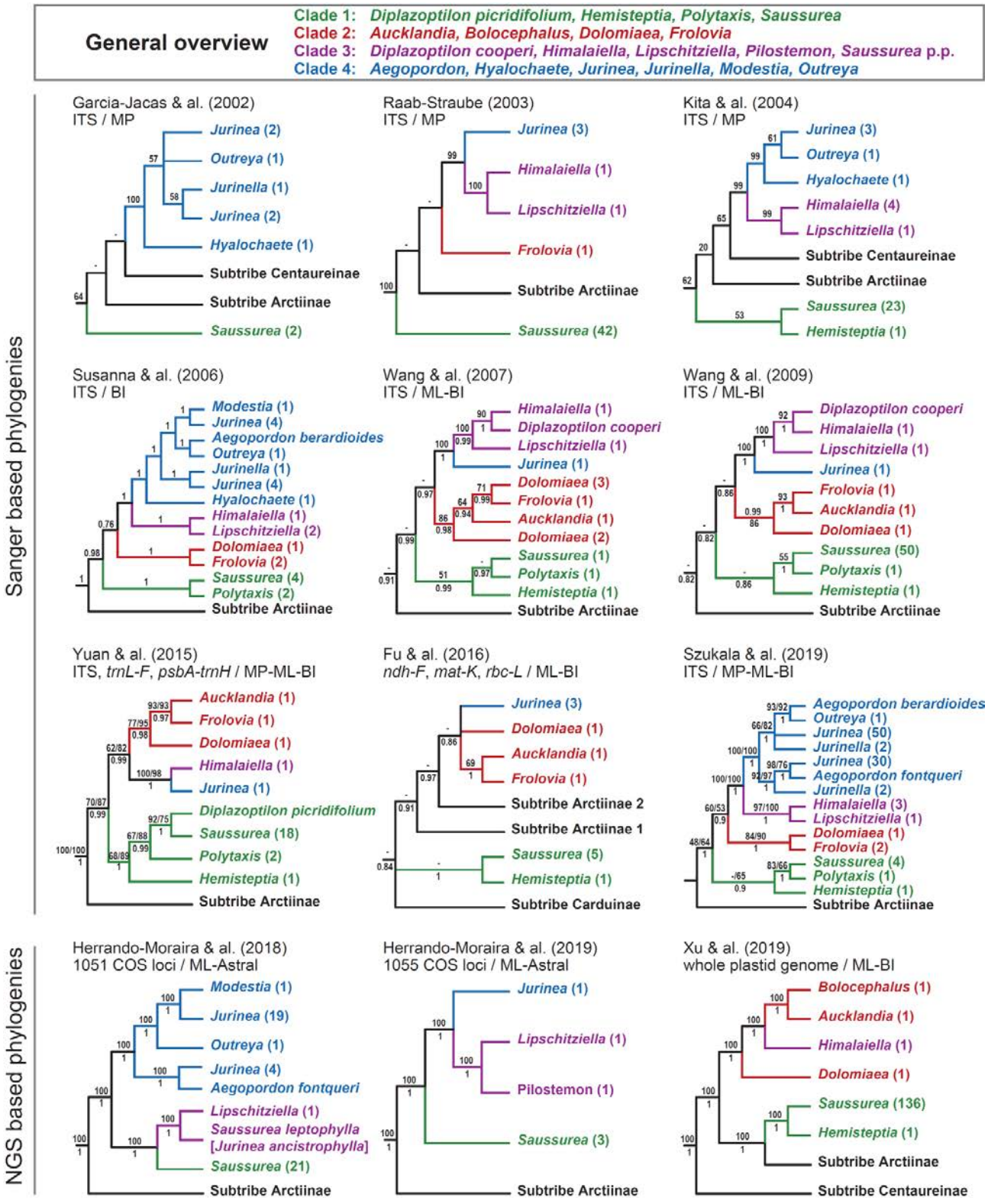


Fig. 1. Comparison of previous phylogenies performed on subtribe Saussureinae with considerable generic sampling. The number of species for each genus included in the phylogenies is indicated in brackets. Below the study citation are specified the molecular markers and the phylogenetic inference method used. The branch support values correspond to bootstrap values in case of maximum parsimony (MP) and maximum likelihood (ML) methods, posterior probabilities in Bayesian inference (BI), and local posterior probabilities in coalescence Astral method. A dash (-) indicates a bootstrap support value below 50%. Note that *Diplazoptilon cooperi* is synonymized as *Himalaiella yakla* by Shi & Raab-Straube (2011).

identity. Based on Herrando-Moraira & al. (2018, 2019), we included 13 representatives of the genera *Arctium* L. and *Cousinia* Cass. from the sister subtribe Arctiinae. We also added 6 members of Centaureinae, 4 of Carduinae, 2 of Onopordiinae, and 2 of Staehelininae. Vouchers and accession numbers are detailed in Appendix 1. From the total sampling (167), 97

individuals were newly sequenced in this study, and the remaining 70 were obtained from raw reads sequenced on previous studies (Mandel & al., 2014, 2019; Herrando-Moraira & al., 2018, 2019; Jones & al., 2019) (see suppl. Table S2).

2.2. DNA extraction, library preparation, sequence capture, and sequencing

Total genomic DNA was obtained from dried leaf material from herbarium collections for most of the species (Appendix 1). From the leaves, around 10–30 mg per sample were selected and homogenized with a Mixer Mill MM 301 (Retsch, Haan, Germany). Then, the E.N.Z.A SP Plant DNA Mini Kit (Omega Bio-Tek, Norcross, Georgia, U.S.A.) was used to extract the DNA following manufacturer's specifications. The quantity of the extracted DNA was measured with the Qubit 3.0 Fluorometer (Thermo Scientific, Waltham, Massachusetts, U.S.A.), which was used to standardize the amount of DNA in all the samples (1 µg in 70 µl). The DNA was fragmented in microTUBEs using a Covaris S2 Biodisruptor from the Genomics Unit of the Centre for Genomic Regulation (CRG, Barcelona, Spain). For the shearing step, a target size of 400 bp on average was selected. The libraries were constructed using the NEBNext Ultra II DNA Library Prep kit for Illumina (New England Biolabs, Ipswich, Massachusetts, U.S.A.), which then were pooled in groups of 4 samples and the target COS loci enriched with MyBaits COS 1Kv1 (MYcroarray, Ann Arbor, Michigan, U.S.A.). For protocol details of library and capture preparation see Herrando-Moraira & al. (2018). Before sequencing, we additionally conducted a spiking procedure combining 40% of pre-capture libraries with 60% of post-capture library solution. The final spiked samples were sequenced in the DNA Sequencing Core CGRC/ICBR of the University of Florida using one lane of an Illumina HiSeq 3000 with a 100 bp paired-end format.

2.3. Extraction of target-enriched sequences

The quality control evaluation of raw reads was performed with FastQC v.0.10.1 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Then, we used the software Trimmomatic v.0.36 (Bolger & al., 2014) to obtain a cleaned read set: trimming the Illumina adapters, removing reads that did not pass the defined quality threshold (sliding-window set to 5:20), and finally discarding short cleaned reads (<36 bp length) or with a missing forward or reverse pair.

The extraction of the 1061 target COS loci was conducted with the easy-to-use workflow of HybPiper pipeline v.1.1 (Johnson & al., 2016). With this method, a first mapping step of cleaned reads to targets is conducted with BWA mapper (Li & Durbin, 2009). Secondly, the assembler SPAdes (Bankevich & al., 2012) was used to obtain the contigs, performing a *de novo* assembly of the reads previously mapped to targets. In the case of several coexisting contigs for a same target locus (potential paralogs), HybPiper is designed to finally retain only one contig through a series of hierarchical decisions (for details see Johnson & al., 2016), but flagging these loci as potential paralogs. To prevent the problems of inclusion of loci potentially affected by paralogy, we did not include them in downstream analyses.

The individual multi-fasta files obtained for each COS locus were aligned using the *auto* setting mode of MAFFT v.7.266 (Katoh & Standley, 2013). To remove the ambiguously aligned regions from the aligned sequence files, we applied the *automated1* flag of trimAl v.14 (Capella-Gutiérrez & al., 2009). A short length locus (6 bp) that was recovered after the alignment trimming and two loci recovered for less than three species were not included in the final used matrices. In order to conduct two different phylogenetic inference methods (see below), we created two sequence datasets. One consisted in the separate trimmed alignments for each locus, and the other consisted in a single supermatrix file obtained from the concatenation of all trimmed aligned loci with FASconCAT-G v.1.02 (Kück & Longo, 2014). For the supermatrix, a set of summary statistics were calculated using AMAS (Borowiec, 2016).

2.4. Phylogenetic inference analyses

To reconstruct the phylogeny of Saussureinae, two complementary approaches were employed: the concatenation approach (using the supermatrix dataset as input) and the coalescence approach (using the separate matrices of each locus).

For the concatenation approach, the sequence data were analyzed under maximum likelihood (ML) assumptions using RAxML v.8.2.9 (Stamatakis, 2014). The method selected was a simultaneous rapid bootstrapping of 1000 replicates to assess branch support, and best ML tree search with 10 randomized maximum parsimony starting trees. The branches with bootstrap support (BS) >70% were considered statistically well supported (Hillis & Bull, 1993). As partition scheme, each locus was treated as a unit evolving under the model GTRGAMMA following the recommendations of Stamatakis (2006). The ML analysis was conducted on XSEDE in the CIPRES Science Gateway v.3.1 (Miller & al., 2010).

For the coalescence approach, we first estimated the individual gene trees for each locus using RAxML under the same conditions as specified above, but this time with a bootstrap resampling of 200 replicates. Second, a summary statistic method as implemented in ASTRAL-III v.5.5.3 (Zhang & al., 2018) was used to estimate the species tree from the previous set of gene trees. Default parameters were selected for the ASTRAL running. Branch support values were calculated using local posterior probabilities (LPP; Sayyari & Mirarab, 2016), and values of LPP > 0.95 were considered as branches strongly supported. Resulting trees from both methods were visualized in FigTree v.1.4.3 (Rambaut, 2016). Alignments and tree files for each dataset are deposited in Mendeley Data repository (https://doi.org/10.17632/hrmyfhpb5c.2) with initial-tested and definitive-proposed final taxon names.

3. Results

3.1. Sequencing efficiency and target loci recovery

On average 6,450,918 (SD = ±4,420,030) of raw sequence reads were sequenced per species (see details in suppl. Table S2). The sequence efficiency was notably different among some species (e.g., the ones with lowest and highest number of raw reads were *Cynara cardunculus* L. with 454,885 and *Dolomiaea baltalensis* Dar & Naqshi with 42,600,816, respectively). From

Table 2. Summary metrics of target recovering performance using HybPiper (Johnson & al., 2016) method.

Parameters	Extraction performance values
1. Number of species included	167
2. Number of recovered loci (% respect COS targets)	1054 (99.3)
3. Number of used loci (% respect COS targets)	588 (55.4)
4. Number of captured loci in ≥90% of species (%)	509 (86.6)
5. Average of recovered loci per species (SD; min–max)	550 (13; 457–565)
6. Average of species recovered per loci (SD; min–max)	156 (29; 7–167)
7. Mean alignment length per locus in bp (SD; min–max)	336 (188; 54–1191)
8. Length of concatenated matrix in bp	197,784
9. Number of variable sites in the concatenated matrix (%)	92,301 (46.7)
10. Number of parsimony-informative sites in the concatenated matrix (%)	61,990 (31.3)
11. Proportion (%) of missing data in the concatenated matrix	12.6

Values of parameters were calculated with FASconCAT-G v.1.02 (Kück & Longo, 2014) and AMAS (Borowiec, 2016) programs. Parameters from 4 to 11 are calculated based on dataset "3. Nº of used loci", which includes those loci (1) not detected as potential paralogs; (2) recovered for more than three species; and (3) more than 6 bp length (see text for details). Abbreviations: bp = base pairs; max = maximum; min = minimum; SD = standard deviation.

the total 1061 target COS loci, we were able to recover 1054, which represents 99.3% of targets (Table 2).

After a step of loci filtering (by potential paralogy, short length, and poor species recovery; see Methods for details), 588 loci were retained in our final dataset, representing approximately half of the targets (55.4%; Table 2). From these loci finally used, 86.6% were captured in ≥90% of species and were on average 336 bp of alignment length (Table 2). The concatenation of the 588 loci resulted in a supermatrix of 197,784 bp with 46.7% of variable sites, 31.3% of parsimony-informative sites, and 12.6% of missing data (Table 2). Raw sequence data are deposited at the National Center for Biotechnology Information (NCBI) Sequence Read Archive (see suppl. Table S2 for BioProjects codes for each species).

3.2. Phylogenetic analyses

The two species recently excluded from Saussureinae, *Jurinea cartilaginea* and *J. gedrosiaca*, and currently considered as *Karvandarina cartilaginea* (Mozaff.) Parishani & al. and *Tricholepis edmondsonii* Rech.f., respectively, were placed with high confidence within Centaureinae (Figs. 2, 3). Considering this exclusion, Saussureinae formed a highly supported and monophyletic group itself (BS = 100 and LPP = 1), which was recovered as sister to Arctiinae.

Deep tree branches of Saussureinae phylogenies resulted in highly supported topologies (Figs. 2, 3). Otherwise, branch support resolution decreased gradually in internal nodes closer to the tips, especially at close species relationships. All genera with more than three species (*Dolomiaea*, *Frolovia*, *Himalaiella*, *Jurinea*, *Jurinella*, *Saussurea*) were recovered as non-monophyletic assemblies (Figs. 2, 3) as currently circumscribed (Table 1). Some of the genera with less than four species (*Aegopordon*, *Aucklandia*, *Bolocephalus*, *Diplazoptilon*, *Modestia*, *Outreya*, *Pilostemon*) appeared nested within one of the large genera mentioned above, while the others emerged as separate distant lineages sister to more speciose clades (*Hemisteptia*, *Hyalochaete*, *Lipschitziella*, *Polytaxis*; Figs. 2, 3).

Comparing the trees inferred with the two approaches (concatenation and coalescence), we found that both showed almost identical topologies under a generic-level focus (Figs. 2, 3). Exceptions regarding the positions of *Lipschitziella*, *Hemisteptia*, and *Polytaxis* were detected. *Lipschitziella* was recovered as sister to a diverse clade composed mainly of *Himalaiella* under the concatenation approach (Fig. 2), while under coalescence it was embedded in that clade but not statistically supported as sister to all other members (Fig. 3). *Hemisteptia* was resolved as sister to the *Polytaxis*+*Saussurea* clade under concatenation (Fig. 2), and conversely, *Polytaxis* was sister to *Hemisteptia*+*Saussurea* under coalescence but without support (LPP = 0.46; Fig. 3). In general, the tree inferred under the coalescence approach (Fig. 3) resulted in lower resolution and support than the one obtained under the concatenation approach (Fig. 2), particularly for shallow nodes close to the tree tips.

Overall, three main lineages were recovered at the backbone of the trees with maximum statistical support (BS = 100 and LPP = 1; Figs. 2, 3): (1) the Dolomiaea clade, composed of *Dolomiaea* and some segregates, namely *Aucklandia*, *Bolocephalus*, and most *Frolovia*; (2) the Saussurea clade, including the bulk of *Saussurea*, *Diplazoptilon*, *Polytaxis*, and *Hemisteptia*; and (3) the Jurinea clade, comprising *Jurinea*, *Aegopordon*, some *Frolovia*, *Himalaiella*, *Hyalochaete*, *Jurinella*, *Lipschitziella*, *Modestia*, *Outreya*, *Pilostemon*, and some species of *Saussurea*. The first diverging lineage within the subtribe was the Dolomiaea clade that emerged as sister to the remainder. The next two diverging lineages were the Saussurea clade and the Jurinea clade that formed a sister-group relationship.

Within the Dolomiaea clade, two highly supported groups were found. One was primarily composed of *Dolomiaea* species and *Bolocephalus* (BS = 100 and LPP = 1), and the other by two *Dolomiaea* species, *Frolovia*, and *Aucklandia* (BS = 100 and LPP = 0.97; Figs. 2 and 3). *Frolovia* appeared in three separate clades, highly supported themselves (Figs. 2, 3): (1) three species of *Frolovia* were clustered with two of *Dolomiaea* (BS = 100 and LPP = 1); (2) *Frolovia frolowii* (Ledeb.) Raab-Straube was clustered with *Aucklandia* (BS = 100 and LPP = 1); and (3) *Frolovia gilesii* (Hemsl.) B.A.Sharipova was placed within the Jurinea clade with *J. chondrilloides* (C.Winkl.) O.Fedtsch., *J. karategina* (Lipsky) O.Fedtsch. [*Pilostemon*], and *Saussurea leptophylla* Hemsl. [*Jurinea ancistrophylla* Boiss.] (BS = 94 and PP = 0.86).

Regarding the Saussurea clade, we found that *Diplazoptilon* (*D. picridifolium*) was completely embedded within the

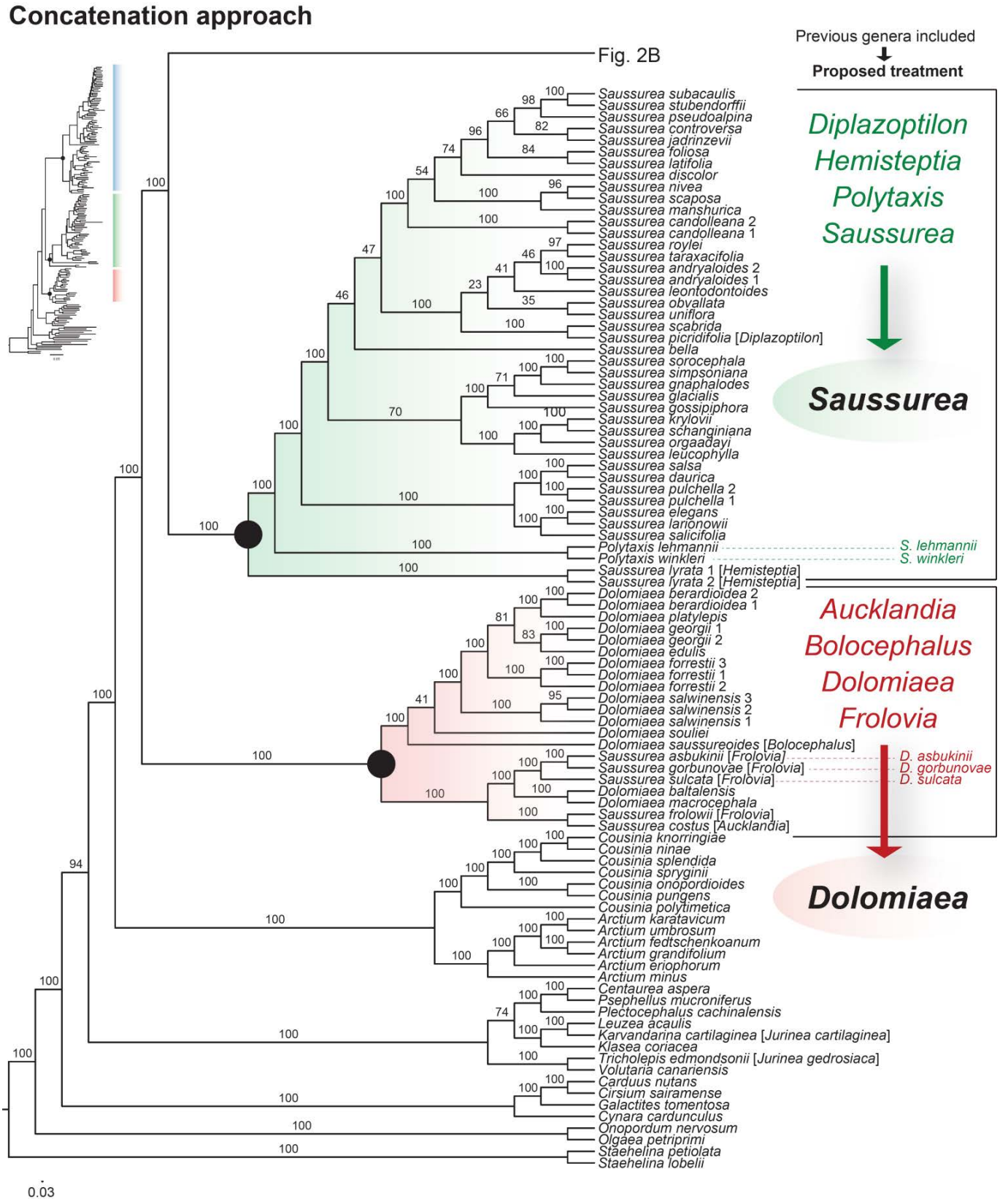


Fig. 2. Phylogenetic reconstruction of subtribe Saussureinae genera inferred with 588 nuclear conserved ortholog loci under the concatenation approach (maximum likelihood analysis performed with RAxML). Branch labels indicate bootstrap support values. Between claudators are specified the segregate genera, when the species have been synonymized as *Jurinea* or *Saussurea*. The new generic delimitation and the species newly combined are shown on the right. **Anacantha* is a synonym of *Modestia*. According to *Flora of China* (Shi & Raab-Straube, 2011), ***Saussurea radiata* is a synonym of *Himalaiella deltoidea* and ****Saussurea yakla* [*Himalaiella*] is a synonym of *Diplazoptilon cooperi*.

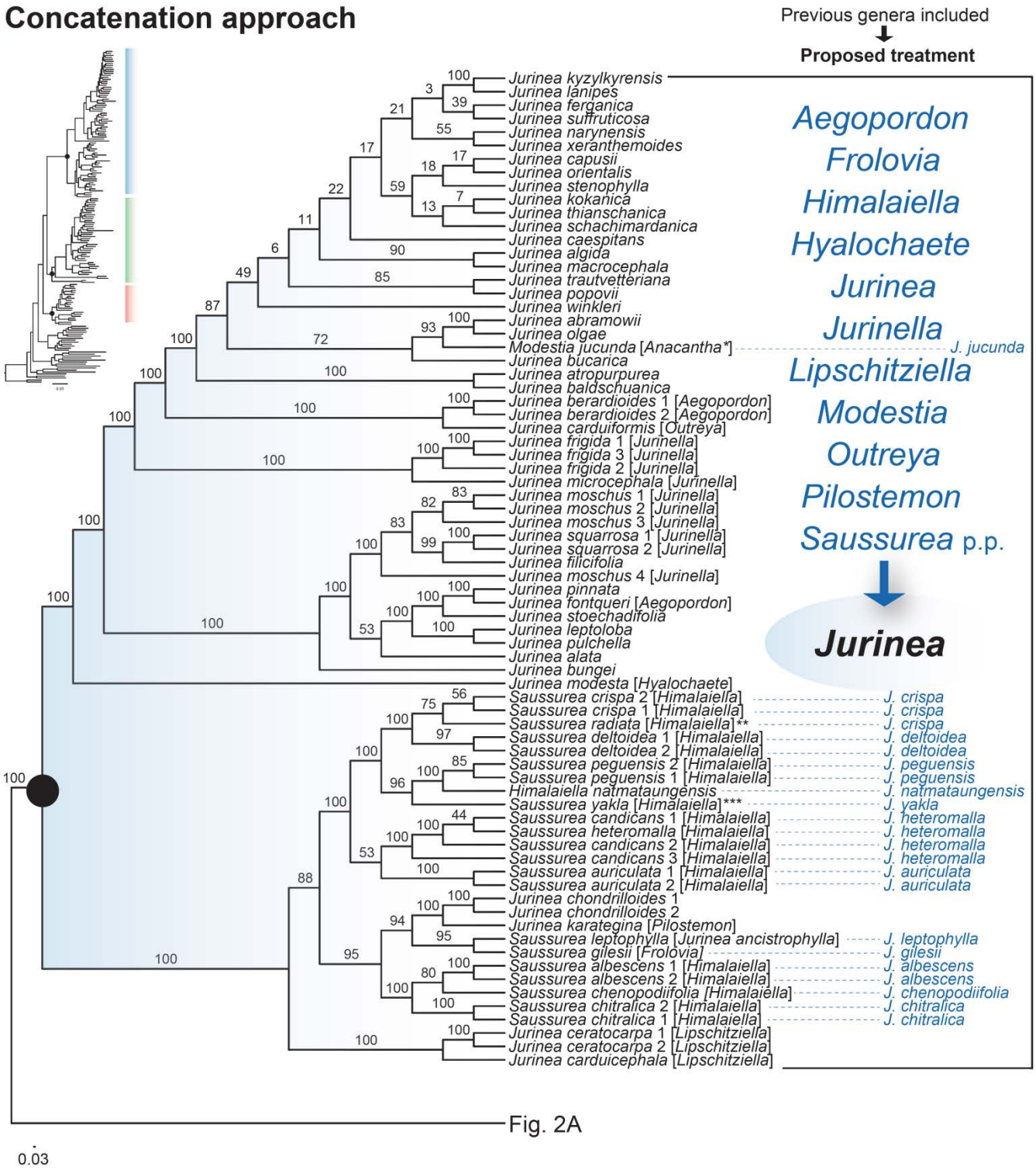


Fig. 2. Continued

Saussurea genus (Figs. 2, 3). Conversely, *Polytaxis* was placed as sister to *Saussurea*, and *Hemisteptia* as sister to *Polytaxis+Saussurea*. The two species of *Polytaxis* were clustered together with the maximum support (BS = 100 and LPP = 1; Figs. 2, 3), as well as the two individuals of *Saussurea lyrata* (Bunge) Sch.Bip [*Hemisteptia*].

The Jurinea clade presented two main highly supported groups (BS = 100 and LPP = 1; Figs. 2, 3), a first one including *Jurinea* s.str., *Aegopordon*, *Hyalochaete*, *Jurinella*, *Modestia*, and *Outreya*; and the second one including *Himalaiella*, *Lipschitziella*, *Pilostemon*, and *Frolovia*, in addition to *Jurinea chondrilloides* and *Diplazoptilon cooperi* (outlined as *Saussurea yakla* C.B.Clarke on the Figures). Within the first group, *Hyalochaete* was recovered as the sister lineage. *Aegopordon* was

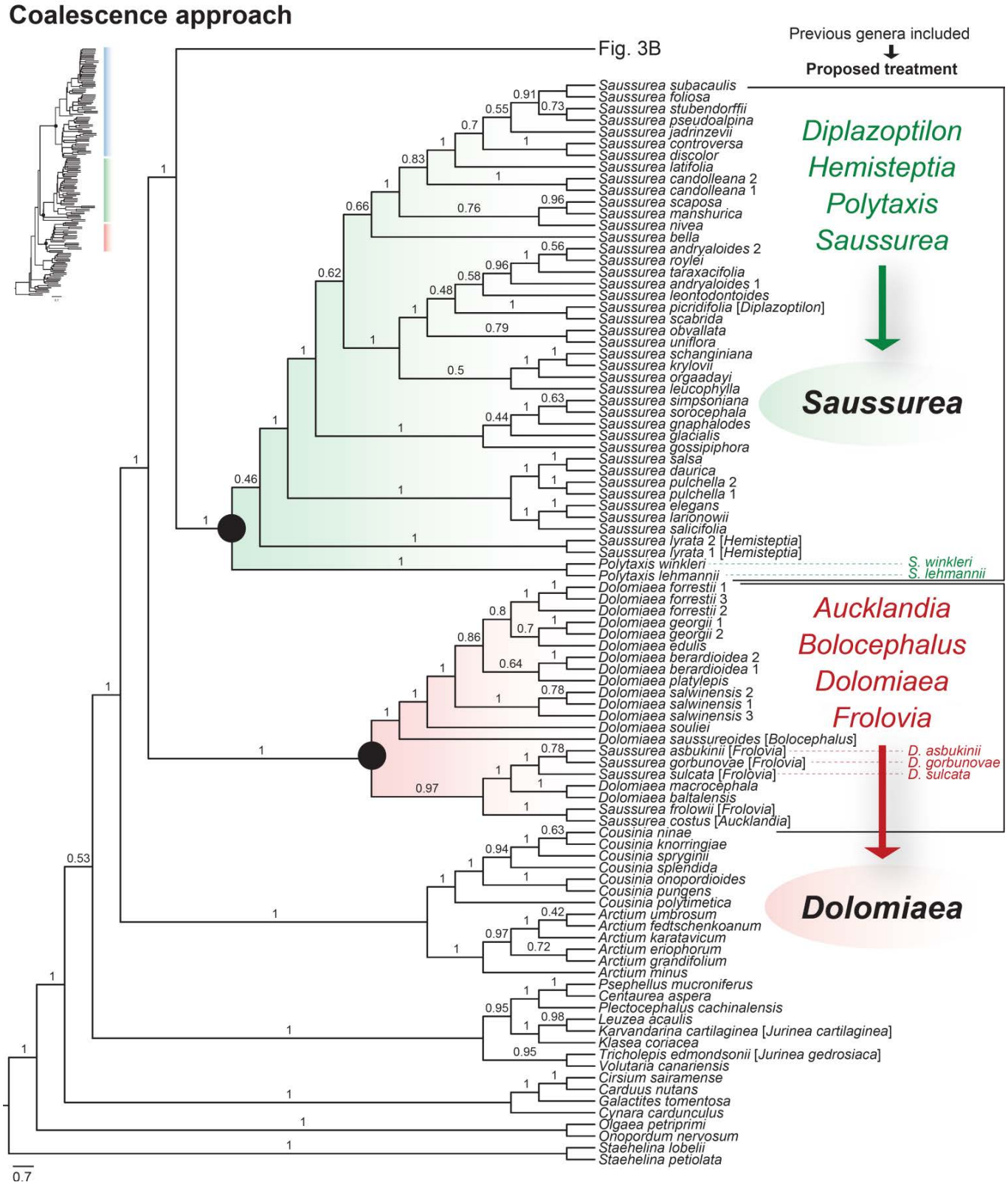


Fig. 3. Phylogenetic reconstruction of subtribe Saussureinae genera inferred with 588 nuclear conserved ortholog loci under the coalescence approach (individual gene trees obtained with RAxML and the species tree with Astral). Branch labels indicate support values of local posterior probabilities (LPP). Between claudators are specified the segregate genera, when the species have been synonymized as *Jurinea* or *Saussurea*. The new generic delimitation and the species newly combined are shown on the right. **Anacantha* is a synonym of *Modestia*. According to *Flora of China* (Shi & Raab-Straube, 2011), ***Saussurea radiata* is a synonym of *Himalaiella deltoidea* and ****Saussurea yakla* [*Himalaiella*] is a synonym of *Diplazoptilon cooperi*.

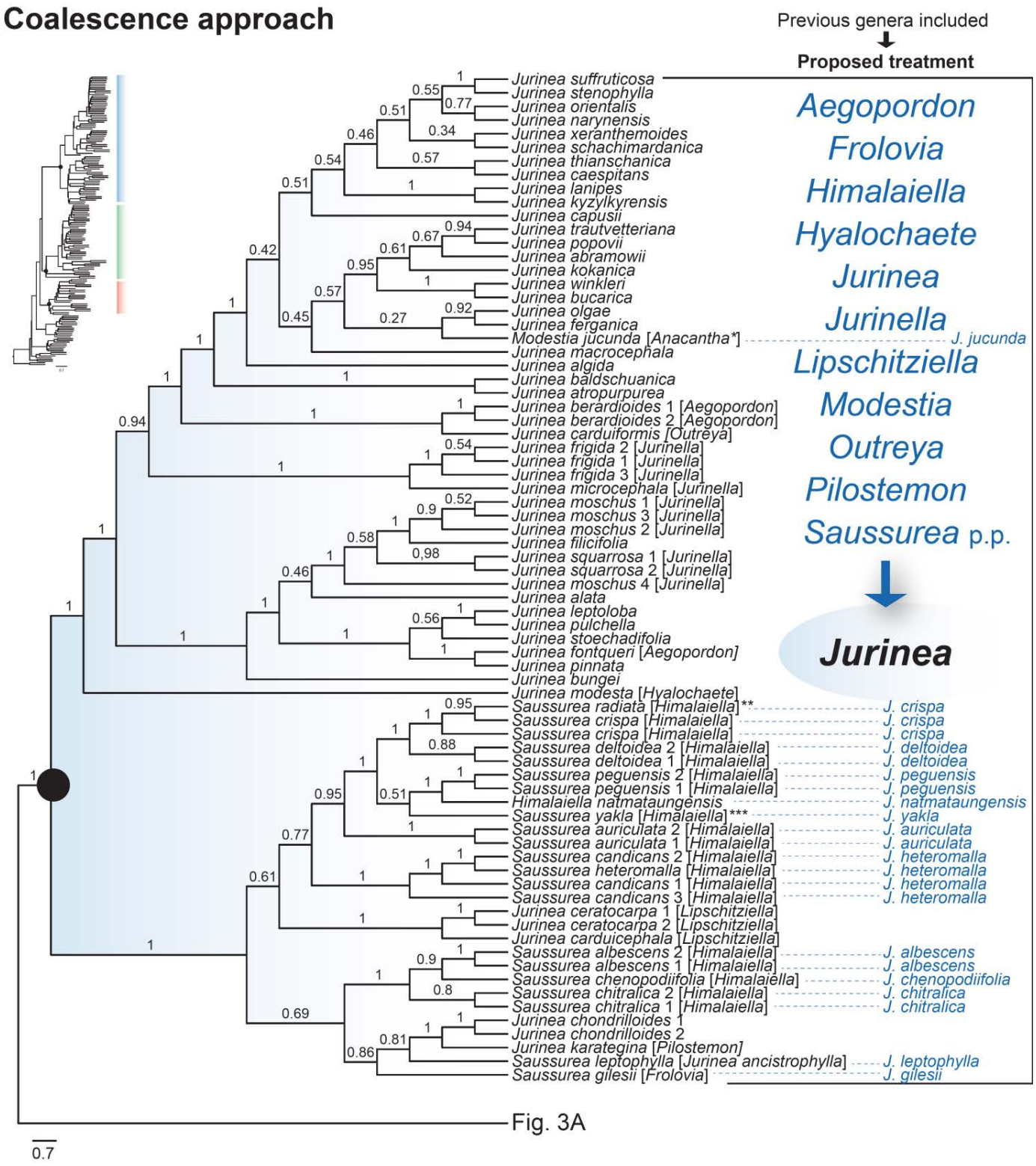


Fig. 3. Continued.

placed at the core of *Jurinea* species, specifically *A. berardioides* Boiss. nested with *J. carduiformis* (Jaub. & Spach) Boiss. [*Outreya*] (BS = 100 and LPP = 1) and *A. fontqueri* (Cuatrec.) Tscherneva with *J. pinnata* (Pers.) DC. (BS = 100 and LPP = 1). *Modestia* (= *Anacantha*) was also deeply embedded within the Jurinea s.str. clade. *Jurinella* was recovered in two well-defined and separate clades with a clear polyphyletic pattern. Within the second group of the Jurinea clade, *Himalaiella* also resulted as a polyphyletic assembly, occurring in several supported clades. *Pilostemon* appeared highly related to *Jurinea chondrilloides* (BS = 100 and LPP = 1). Under the coalescence approach, we did not obtain enough branch support resolution at the internal nodes of this group, the included species emerging in a considerable polytomy with some supported small groups of species (Fig. 3).

4. Discussion

4.1. Applicability of Hyb-Seq on Saussureinae phylogeny

The present study includes for the first time a complete genus-level sampling of Saussureinae in a phylogenetic framework (suppl. Table S1). This sampling improvement was increased using herbarium material, in which the DNA is usually highly degraded and sometimes can be difficult to amplify with Sanger sequencing technologies. With the second generation of sequencing methods, the efficiency of target loci capture is remarkably high even with old herbarium samples. For example, here we recovered in total 99.3% of the initial target set, and 91.2% from a sample from 1899 (*Saussurea lyrata* [*Hemisteptia*] individual 2). Therefore, as also pointed out by other Hyb-Seq studies (Hart & al., 2016; Villaverde & al., 2018; Brewer & al., 2019; Viruel & al., 2019), the use of herbarium material represents a step forward to sort out plant systematics on groups highly speciose, distributed on remote or politically unstable regions, such as Saussureinae.

In recent years, most phylogenomic studies conducted on Compositae have succesfully employed the Hyb-Seq technique based on 1061 COS loci, at family (Mandel & al., 2014, 2015, 2017, 2019; Jones & al., 2019), tribe (Herrando-Moraira & al., 2019; Siniscalchi & al., 2019), and genus level (Herrando-Moraira & al., 2018; Thapa & al., 2019). For Saussureinae, this method also allowed to recover a well-supported phylogeny. Sequencing of whole plastid genomes is becoming another widely used method within HTS techniques. However, we detected considerable differences in generic placements between the present study and another performed with plastid genomes by Xu & al. (2019). The plastid phylogeny recovered *Himalaiella* as a part of the *Dolomiaea-Aucklandia-Bolocephalus* clade. This grouping has never been found previously (see Fig. 1), not even in Sanger plastid phylogenies (e.g., Wang & al., 2007). Moreover, in Xu & al. (2019) and also in a study with 20 whole plastomes of *Saussurea* (Zhang & al., 2019), the three main clades recovered showed incongruent topologies when the whole plastid genome and the protein-coding regions from the chloroplast were analyzed separately. Although these results could be due to a poor taxon sampling effect, plastid HTS datasets could tend to produce phylogenetic misplacements but with maximum support values in rapidly radiated lineages like Saussureinae species (e.g., *Oxalis* L., Schmickl & al., 2016). Actually, Herrando-Moraira & al. (2019) already detected incongruence between nuclear and plastid Cardueae phylogenies and discussed the possible causes, which could be added to the differences in taxon sampling between the present work and that of Xu & al. (2019): (1) phylogenetic informativeness much higher in nuclear than in plastid phylogenies; (2) incomplete lineage sorting; and (3) past or current hybridization events. Further studies are needed to explore these differences on nuclear/plastid HTS phylogenies and under concatenation/coalescence approaches.

4.2. Generic limits delineation

Our results call for a reclassification of Saussureinae genera. Taxonomic entities should be ideally based on monophyletic clades. Considering an extremely analytical circumscription that splits the subtribe in 17 genera, we did not find monophyly for almost any of them (see Fig. 4, tree on the left). Along the tree backbone (Figs. 2, 3), three major fully supported lineages emerged that have also been already retrieved in previous studies (Fig. 1): *Dolomiaea*, *Saussurea*, and *Jurinea*. A new generic delimitation based on these three lineages is proposed here under a synthetic point of view (Fig. 4). This treatment seems well-justified according to both molecular and morphological characters, as discussed below. In general terms, the diagnostic morphological characters of segregate genera fall into the broad variability of the three large lineages *Dolomiaea*, *Saussurea*, or *Jurinea* as previously pointed out by Susanna & Garcia-Jacas (2007). A question that our work leaves open is the subgeneric and sectional classification of the three resulting genera. This classification is especially needed for the two largest ones, *Jurinea* and *Saussurea*. However, this is an extremely complicated issue (see Y.-S. Chen, 2015 and Szukala & al., 2019 for the artificial classification of *Saussurea* and *Jurinea*, respectively) that further work will try to resolve.

4.3. Dolomiaea clade

Besides *Dolomiaea*, three other genera fall deeply nested within this clade: *Aucklandia*, *Bolocephalus*, and *Frolovia* (Figs. 2, 3). All published phylogenies including these genera have found the same grouping pattern (see references and trees in Fig. 1). Within this clade, two highly supported groups emerged, one composed of *Dolomiaea* and *Bolocephalus* and the other of *Dolomiaea*, *Frolovia* and *Aucklandia* (Figs. 2, 3), which included two clades: some *Dolomiaea* + some *Frolovia* and one *Frolovia* + *Aucklandia* (Figs. 2, 3). This points out that both *Dolomiaea* and *Frolovia* as currently described are polyphyletic. In addition, one member of *Frolovia* (*F. gilesii*), which was sequenced for the first time, was placed in the distant Jurinea clade (Figs. 2, 3; see more discussion in the Jurinea clade section). Undoubtedly, the description of *Frolovia* and *Aucklandia* as sections within *Saussurea* (*S.* sect. *Frolovia* and *S.* sect. *Aucklandia*, respectively) reflects that they do not fit into *Saussurea* s.str. (Lipschitz, 1954, 1961, 1962, 1979; Kamelin, 1993; Kamelin & Kovalevskaja, 1993). *Frolovia* has also been compared to *Jurinea*, but Raab-Straube (2003) claimed that differences in pappus structure prevented its inclusion in this genus. *Aucklandia* has experienced multiple taxonomic changes, being combined as *Aplotaxis*, *Saussurea*, and *Theodorea* (Kasana & al., 2018). However, from the initial phylogenetic trees (Fig. 1) to the current phylogenomic ones (Figs. 2, 3), both genera are nested within *Dolomiaea*. Accordingly, Susanna & al. (2006) and Wang & al. (2007) already called for a broad redefinition of *Dolomiaea* including *Frolovia*, but without making the formal changes. In our attempt to define monophyletic groups and considering molecular evidence, we also propose that *Frolovia* and *Aucklandia* should be synonymized under *Dolomiaea*. The monotypic *Bolocephalus* was already formally transferred to *Dolomiaea* by Y.-L. Chen & Shih (1981), and also marked as doubtful genus by Shi & Raab-Straube (2011). Our molecular data and the plastid phylogeny reported in Xu & al. (2019) confirm its inclusion within the Dolomiaea clade.

4.4. Saussurea clade

The second clade recovered is mainly composed of *Saussurea* s.str. together with the genera *Diplazoptilon*, *Hemisteptia*, and *Polytaxis* (BS = 100 and LPP = 1; Figs. 2, 3). All previous phylogenies have unanimously identified this generic grouping (Fig. 1). *Diplazoptilon* was already transferred back to *Saussurea* by Yuan & al. (2015), and we will comment only the cases of *Hemisteptia* and *Polytaxis*. Both genera have always emerged as a sister group to *Saussurea* s.str. (Figs. 1–3). However, which genus is sister to *Saussurea* remains an open question since the tree topology differed between both inference approaches.

Hemisteptia has been merged into *Saussurea* several times (Susanna & Garcia-Jacas, 2007, 2009). On the contrary, *Polytaxis* has never been transferred to any other genus since its description. Among the three genera, micromorphological differences in pappus and achene structure have been proposed as diagnostic traits (Häffner, 2000). Additionally, the life-history trait has been assigned as another distinctive character, *Hemisteptia* and *Polytaxis* being annual plants and *Saussurea* mainly perennials. Despite this, phylogenies of other groups within tribe Cardueae also found that annual species are often resolved as sister to the remaining perennial ones (in *Cousinia* in López-Vinyallonga & al., 2009; in *Echinops* L. in Garnatje & al., 2005). To what extent the inclusion of annual/perennial species in a single phylogeny would bias tree topologies due in part to generation time effect of neutral molecular markers is still open to discussion and requires further methodological exploration (Andreasen & Baldwin, 2001; Yue & al., 2010; Gaut & al., 2011). Furthermore, at least one species of *Saussurea* (*S. kingii* J.R.Drumm. ex C.E.C.Fisch.) is an annual herb (Shi & Raab-Straube, 2011; Y.-S. Chen, 2015), and in a recent plastid phylogenomic study it appeared as sister to the species of the informal "Clade 3" defined by Xu & al. (2019). These facts point out that life form alone would not be a useful diagnostic character to segregate *Hemisteptia-Polytaxis* from *Saussurea*.

The only other character alleged by Shi & Raab-Straube (2011) for segregating *Hemisteptia* from *Saussurea* is the pappus structure: double with an incomplete outer row of very short scales and an inner row of plumose bristles in *Hemisteptia*; double with an outer row of long, scabrid bristles and plumose inner bristles in *Saussurea*. However, some species of *Saussurea* have a very reduced or even missing outer pappus and, according to the key of the genera by Shi & Raab-Straube (2011), all species of *Saussurea* have an inner pappus of plumose bristles, as pointed out repeatedly by Häffner (2000) and Susanna & Garcia-Jacas (2007). Moreover, Kita & al. (2004) made an extensive revision of the achenes of *Saussurea* and *Hemisteptia* and did not find any character separating both genera. In the case of *Polytaxis*, the genus shows two autoapomorphic characters not found elsewhere in the whole Cardueae: presence in the pericarp of resin-ducts and long, slender hairs according to Häffner (2000), who very accurately pointed out: "The possibility of *Polytaxis* being an advanced and morphologically strongly deviant representative of e.g. the *Jurinea* or *Saussurea* group can not be excluded." The deviating achene of *Polytaxis*, when compared to the typical one of *Saussurea*, is the same case as in the purported genera *Cnicus* L. or *Stephanochilus* Coss. & Durieu ex Benth. & Hook.f., described on the basis of very peculiar achenes and now merged into *Centaurea* L. (Font & al., 2002; Garcia-Jacas & al., 2006). In summary, considering phylogenetic evidence and the morphological characters shared by the three genera (floral morphology and pappus structure), we favor the inclusion of *Hemisteptia* and *Polytaxis* in *Saussurea*, which remains a monophyletic and well-defined genus.

4.5. Jurinea clade

Firstly, our results confirm the exclusion of *Jurinea cartilaginea* (currently classified as *Karvandarina cartilaginea*) and *Jurinea gedrosiaca* (considered a synonym of *Tricholepis edmondsonii*) from *Jurinea* and even from Saussureinae. We shall now review the fate of the genera that appear nested in the Jurinea clade. Some of them (*Aegopordon*, *Anacantha*, *Hyalochaete*, *Jurinella*, *Outreya*, *Pilostemon*) were already shown to be part of *Jurinea* s.str. (Susanna & al., 2006), and we will focus on the most disputed ones, i.e., *Himalaiella* and *Lipschitziella*.

When examining the resurrection of both genera by Raab-Straube (2003), we can see that all the comparisons were made to *Saussurea, and not to Jurinea.* Species of *Himalaiella* and *Lipschitziella* would be aberrant in *Saussurea* because their achenes are very different from *Saussurea* (for a detailed and well-illustrated description, see Häffner, 2000). However, the achenes of *Himalaiella* or *Lipschitziella* would not be so discording in *Jurinea* as already pointed out as early as 1873 (!) by Bentham (1873: 474), who suggested that *Saussurea ceratocarpa* Decne. (*Lipschitziella ceratocarpa* (Decne.) Kamelin in Raab-Straube, 2003) should be included in *Jurinea*. We have confirmed this suggestion, because *L. ceratocarpa* was nested in *Jurinea* in Herrando-Moraira & al. (2018), and our current results also support it. As for the very characteristic habit of *Himalaiella*, described as having "large, nodding capitula" (Raab-Straube, 2003), this is obviously a case of convergent evolution, an adaptation to mountain conditions that has appeared several times in the Cardueae and in other tribes of Compositae. Species of the unrelated genera *Alfredia* Cass. and *Synurus* Iljin (Onopordinae) and some species of *Cirsium* Mill. (Carduinae) show this habit, as well as *Saussurea qinghaiensis* S.W.Liu & T.N.Ho (Y.-S. Chen, 2015). Subalpine species with large, nodding heads appear in *Lasiocephalus* of tribe Senecioneae (Silva-Moure & al., 2013). In fact, *Himalaiella* is not monophyletic (Fig. 2) because species of the genus are scattered in two different clades. In one of the clades, *Himalaiella albescens* (DC.) Raab-Straube, *H. chenopodiifolia* (Klatt) Raab-Straube and *H. chitralica* (Duthie) Raab-Straube are grouped with species that share the typical morphology of *Jurinea* s.str., like *J. chondrilloides*, *J. karategina*, and *Saussurea leptophylla* [*Jurinea ancystrophylla*], and even with *Saussurea gilesii* Hemsl., a species wrongly included in *Frolovia* (Fig. 2). It was already suggested that *Saussurea gilesii* is closer to *Jurinea*: a note by Rechinger points out that *Jurinea kerstanii* Bornm. is the same as *Saussurea gilesii* (K.H. Rechinger in schedis, herbarium W). The rest of species of *Himalaiella*, including the type, *H. heteromalla* (D.Don) Raab-Straube, form a robust clade. However, segregating *Himalaiella* would leave *Jurinea* paraphyletic and, in view of the adaptive character of habit and the remarkably close similarities of the achenes to those of *Jurinea*, we strongly favor keeping it in a broadly defined *Jurinea*.

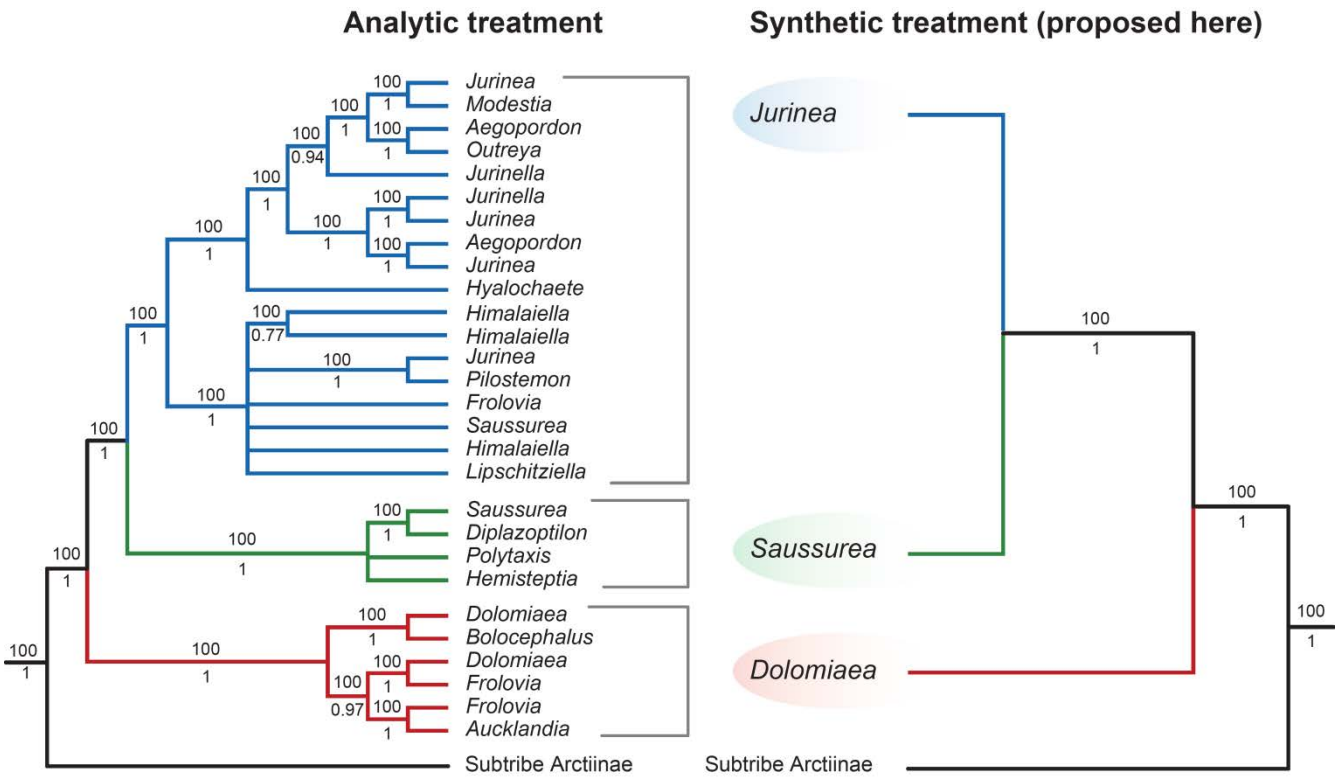


Fig. 4. Schematic representation of the phylogenetic trees inferred in this study under two different treatments (analytic vs. synthetic). The analytic treatment considers the genera currently accepted in most recent analytical classifications (see Table 1). The tree topology of the analytic treatment is outlined following the consensus clades between concatenation and coalescence approaches. Branch labels indicate support values of bootstrap (above branches) and local posterior probabilities (below branches).

5. Taxonomic implications

Subtr. **Saussureinae** N.Garcia & Susanna in Molec. Phylogen. Evol. 137: 329. 2019

Unarmed perennial herbs or subshrubs, very rarely annual herbs. Leaves entire or pinnatisect, often silver-white below and glabrous above, sometimes hirsute-scabrid. Capitula cylindrical or globose, often paniculate, homogamous, discoid. Involucral bracts with short appendages, unarmed. Receptacle densely setose or rarely naked. Anther filaments glabrous, rarely papillose. Styles long, with reflexed lobes. Achenes cylindrical, slightly obconical or obpyramidal, indistinctly ribbed to costate, surface smooth or transversally rugose, very rarely with spines or scales, with or without a crown; apical plate with a persisting style base, without caruncle. Pappus of very long (overtopping involucral bracts), showy, scabrid or plumose bristles in one or more usually several rows, often dimorphic; inner bristles basally connate in a ring, persistent or detachable as a single piece, outer ones shorter, bristle-like or scaly, rarely missing, freely deciduous. Genera included in the subtribe: *Dolomiaea*, *Jurinea* and *Saussurea*.

Dolomiaea DC. in Arch. Bot. (Paris) 2: 330. 1833 – Type: *Dolomiaea macrocephala* DC.

= *Aucklandia* Falc. in Ann. Mag. Nat. Hist. 6: 475. 1841 – Type: *Dolomiaea costus* (Falc.) Kasana & A.K.Pandey.

= *Aplotaxis* [unranked] *Frolovia* DC., Prodr. 6: 538. 1868 ≡ *Frolovia* (DC.) Lipsch. in Bot. Mater. Gerb. Bot. Inst. Komarova Akad. Nauk S.S.S.R. 16: 461. 1954, **syn. nov.** – Type: *Dolomiaea frolowii* (Ledeb.) Kasana & A.K.Pandey.

= *Bolocephalus* Hand.-Mazz. in J. Bot. 76: 291. 1938 – Type: *Dolomiaea saussureoides* (Hand.-Mazz.) Y.L.Chen & C.Shih.

= *Vladimiria* Iljin in Sovetsk. Bot. 1939(8): 55. 1939 ≡ *Mazzettia* Iljin in Bot. Mater. Gerb. Bot. Inst. Komarova Akad. Nauk S.S.S.R. 17: 443. 1955 – Type: *Dolomiaea salwinensis* (Hand.-Mazz.) C.Shih.

Perennial herbs, often rosulate and acaulescent or shortly caulescent and subscapose, very rarely long-stemmed to 0.6–1(2) m. Leaves usually dentate or lobed, rarely entire, hirsute-spinulose above, often denselly wooly below. Capitula solitary clustered in the center of a rosette or rarely terminal and then often nodding, large, 2–6(8) cm wide, homogamous. Involucre broadly campanulate, rarely almost sphaeric. Phyllaries in several rows, usually coriaceous, very often broadly triangular with narrow, undulate, black margin, rarely scariose or herbaceous. Receptacle usually naked and pitted, rarely setose. Florets purple or purple-bluish. Achenes obpyramidal or broadly obconic, 4–6-ribbed, rarely cylindric, usually rugulose-squamulose, straw-colored often with black wavy fringes, with a narrow apical rim. Pappus in one or several rows, subequal or outer row shorter; all bristles from scabrid to shortly plumose, basally connate into a ring, detachable as a single piece.

Jurinea Cass. in Bull. Sci. Soc. Philom. Paris 1821: 140. 1821 – Type: *Jurinea alata* Cass.

= *Outreya* Jaub. & Spach, Ill. Pl. Orient. 1: 131 [= ad. t. 68]. 1843 – Type: *Jurinea carduiformis* (Jaub. & Spach) Boiss.
= *Aegopordon* Boiss., Diagn. Pl. Orient. ser. 1, 6: 112. 1846 – Type: *Jurinea berardioides* (Boiss.) Diels.
= *Jurinella* Jaub. & Spach, Ill. Pl. Orient. 2: 101 [= ad t. 183]. 1847 – **Type (designated here):** *Jurinella aucheri* Jaub. & Spach (= *Jurinea moschus* (Hablitz) Bobrov).
= *Cirsium* sect. *Anacantha* Iljin in Bot. Mat. Gerb. Glavn. Bot. Sada 3: 57. 1922 ≡ *Modestia* Kharadze & Tamamsch. in Zametki Sist. Geogr. Rast. 19: 40. 1956, nom. illeg. (non *Modesta* Raf. 1838) ≡ *Anacantha* (Iljin) Soják in Sborn. Nár. Muz. Praze, Řada B, Přír. Vědy 1982: 108. 1982 – Type: *Cnicus darwasicus* C.Winkl. (≡ *Jurinea darwasica* (C.Winkl.) Sennikov).
= *Pilostemon* Iljin in Bot. Mater. Gerb. Bot. Inst. Komarova Akad. Nauk S.S.S.R. 21: 391. 1961 – Type: *Jurinea karategina* (Lipsky) O.Fedtsch.
= *Hyalochaete* Dittrich & Rech.f. in Rechinger, Fl. Iranica 139a: 215. 1979 – Type: *Jurinea modesta* Boiss.
= *Lipschitziella* Kamelin in Kamelin, Opred. Rast. Sred. Azii 10: 632. 1993 – Type: *Jurinea carduicephala* Iljin.
= *Himalaiella* Raab-Straube in Willdenowia 33(2): 290. 2003, **syn. nov.** – Type: *Jurinea heteromalla* (D.Don) N.Garcia, Herrando & Susanna.
– "*Perplexia* Iljin" in Komarov, Fl. U.S.S.R. 27: 727. 1962, not validly published (Art. 40.1).

Unarmed dwarf shrubs, shrublets or perennial herbs, sometimes acaulescent, exceptionally annual herbs. Leaves dentate to pinnatifid, less often entire, usually white-woolly beneath. Capitula on subscapose leafless pedicels or sessile in the center of a rosette, homogamous, rarely outer florets bent outwards mimicking rays. Involucral bracts narrowly triangular, usually herbaceous and unarmed, very rarely rigid and spiny, without distinct appendages. Florets pink or purple, rarely whitish. Achenes obpyramidal, tetrangular and 4-ribbed; pericarp squamulose, verrucate, tuberculate, ridged or pitted, especially on the upper part. Apical rim usually present, patent, often crenate, rarely horned or toothed, or absent. Pappus of scabrid, barbellate or plumose bristles, pluriseriate, rarely uniseriate; inner row usually longer, basally enlarged and broader, rarely similar to the outer ones, usually attached to a hemisphaeric cupula, deciduous as a single piece with the cupula or persistent.

Saussurea DC. in Ann. Mus. Natl. Hist. Nat. 16: 156. 1810, nom. cons. – Type: *Saussurea alpina* (L.) DC.
= *Saussurea* subg. *Theodorea* Cass. in Bull. Sci. Soc. Philom. Paris 1818: 168. 1818 ≡ *Theodorea* (Cass.) Cass. in Cuvier, Dict. Sci. Nat. 35: 13. 1819 – Type: *Saussurea amara* (L.) DC.
= *Heterotrichum* M.Bieb., Fl. Taur.-Caucas. 3: 551. 1819 – Type: *Saussurea salsa* (Pall.) Spreng.
= *Lagurostemon* Cass. in Cuvier, Dict. Sci. Nat. 53: 466. 1828 – Type: *Saussurea pygmaea* (Jacq.) Spreng.
= *Eriostemon* Less., Syn. Gen. Compos.: 12. 1832, nom. illeg., non Sm. 1798 ≡ *Aplotaxis* DC. in Arch. Bot. (Paris) 2: 330. 1833 – Type: *Centaurea taraxacifolia* D.Don = *Saussurea eriostemon* Wall. ex C.B.Clarke.
= *Cyathidium* Lindl. ex Royle, Ill. Bot. Himal. Mts. 7: t. 56, fig. 2. 1835 – Type: *Saussurea taraxacifolia* (Lindl. ex Royle) Wall. ex DC.
= *Hemisteptia* Bunge ex Fisch. & C.A.Mey., Index Sem. Hort. Petrop. 2: 38. 1836 – Type: *Saussurea lyrata* (Bunge) Sch.Bip.
= *Polytaxis* Bunge in Bot. Zeitung (Berlin) 1: 256. 14 Apr 1843 [Bunge, Del. Sem. Hort. Dorpat.: VIII. 30 Nov 1843, isonym], **syn. nov.** – Type: *Saussurea lehmannii* (Bunge) N.Garcia, Herrando & Susanna.
= *Diplazoptilon* Y.Ling in Acta Phytotax. Sin. 10(1): 85. 1965 – Type: *Saussurea picridifolia* (Hand.-Mazz.) Y.S.Chen & Qian Yuan.

Unarmed perennial herbs, rarely subshrubby, exceptionally annuals, caulescent or acaulescent. Leaves entire to pinnatisect, sometimes decurrent. Capitula solitary, corymbose or paniculate, homogamous, sometimes enclosed into colored and foliaceous or translucent and scariose bracts, or intermixed with very dense woolly tomentum. Involucral bracts in several rows, entire, from lanceolate to ovate, apically triangular-subulate, not spiny, middle ones sometimes with an appendage. Receptacle with large scales apically divided into very narrow, twisted bristles, sometimes squamulose, rarely naked. Florets purple or pink, sometimes blue-tinged, rarely white. Achenes narrowly obconic, laterally or rarely dorsiventrally compressed, faintly 4- or 6-ridged, exceptionally with six broad longitudinal ribs, rarely smooth, less often glandulose, papillose or hairy, usually with a small apical rim. Pappus biseriate; outer row of short, scabrid, free, easily deciduous bristles, exceptionally of a few flattened scales at the abaxial side of the achene, rarely absent; inner row of much longer, plumose, laterally connate (up to 5 mm) bristles, persistent or deciduous.

5.1. Key to the genera

1. Pappus in one row.....2
1. Pappus in two or more rows.....3
2. Achenes with an indistinct, crenate to minutely denticulate apical rim ..... ***Dolomiaea [Aucklandia + Frolovia]***
2. Achenes with a distinct, prominently dentate apical rim..... ..... ***Jurinea [Himalaiella]***
3. Pappus very distinctly dimorphic, in two prominently unequal rows; inner pappus long, with plumose bristles laterally connate at the base (sometimes for up to 5 mm); outer pappus much shorter, with scabrid, fragile scales 4
3. Pappus monomorphic or only undistinctly dimorphic, in several (often more than two) equal or slightly unequal rows; outer bristles not fragile 5
4. Achenes narrowly oblanceolate, usually striate or smooth, seldom rugulose, without a dentate apical rim ..... ***Saussurea***
4. Achenes obconical, apically aculeate, with an apical rim of four prominent teeth..... ***Jurinea [Lipschitziella]***
5. Pappus bristles inserted on a conic cupule, dimorphic (innermost row with 2–5 longer and broader bristles) or monomorphic, scabrid or plumose..... ***Jurinea***
5. Pappus bristles not inserted on a conic cupule, monomorphic, plumose..... ***Dolomiaea***

5.2. New combinations in *Dolomiaea*

The following species from *Frolovia* and *Aucklandia* are newly transferred here.

Dolomiaea asbukinii (Iljin) N.Garcia, Herrando & Susanna, **comb. nov.** ≡ *Saussurea asbukini* Iljin in Bot. Zhurn. S.S.S.R.

27(6): 144. 1942 ≡ *Frolovia asbukinii* (Iljin) Lipsch. in Bot. Mater. Gerb. Bot. Inst. Komarova Akad. Nauk S.S.S.R. 16: 462. 1954.

Dolomiaea gorbunovae (Kamelin) N.Garcia, Herrando & Susanna, **comb. nov.** ≡ *Saussurea gorbunovae* Kamelin in Turczaninowia 2(4): 25. 1999 ≡ *Frolovia gorbunovae* (Kamelin) Raab-Straube in Willdenowia 33: 392. 2003.

Dolomiaea sulcata (Iljin) N.Garcia, Herrando & Susanna, **comb. nov.** ≡ *Saussurea sulcata* Iljin in Bot. Mater. Gerb. Glavn. Bot. Sada R.S.F.S.R. 3: 101. 1922 ≡ *Frolovia sulcata* (Iljin) Lipsch. in Bot. Mater. Gerb. Bot. Inst. Komarova Akad. Nauk S.S.S.R. 16: 462. 1954.

5.3. New combinations and correct names in *Saussurea*

Two new transfers are required to accommodate the species of *Polytaxis*, newly synonymised with *Saussurea* here.

Saussurea lehmannii (Bunge) N.Garcia, Herrando & Susanna, **comb. nov.** ≡ *Polytaxis lehmannii* Bunge, Del. Sem. Hort. Dorpat.: VIII. 1843.

Saussurea winkleri (Iljin) N.Garcia, Herrando & Susanna, **comb. nov.** ≡ *Polytaxis winkleri* Iljin in Bot. Mater. Gerb. Bot. Inst. Komarova Akad. Nauk S.S.S.R. 7: 52. 1937.

5.4. New combinations in *Jurinea*

The synonymization of *Himalaiella* and *Modestia* to *Jurinea* requires several new species transfers effected here.

Jurinea albescens (DC.) N.Garcia, Herrando & Susanna, **comb. nov.** ≡ *Aplotaxis albescens* DC., Prodr. 6: 540. 1838 ≡ *Saussurea albescens* (DC.) Sch.Bip. in Linnaea 19: 330. 1846 [(DC.) Hook.f. & Thomson ex C.B.Clarke, Compos. Ind. 233. 1876, isonym] ≡ *Theodorea albescens* (DC.) Kuntze, Revis. Gen. Pl. 1: 368. 1891 ≡ *Himalaiella albescens* (DC.) Raab-Straube in Willdenowia 33: 390. 2003.

Jurinea auriculata (DC.) N.Garcia, Herrando & Susanna, **comb. nov.** ≡ *Aplotaxis auriculata* DC., Prodr. 6: 541. 1838 ≡ *Saussurea auriculata* (DC.) Sch.Bip. in Linnaea 19: 331. 1846 ≡ *Theodorea auriculata* (DC.) Kuntze, Revis. Gen. Pl. 1: 367. 1891 ≡ *Himalaiella auriculata* (DC.) Raab-Straube in Willdenowia 33: 390. 2003.

Jurinea chenopodiifolia (Klatt) N.Garcia, Herrando & Susanna, **comb. nov.** ≡ *Saussurea chenopodiifolia* Klatt in Sitzungsber. Math.-Phys. Cl. Königl. Bayer. Akad. Wiss. München 8: 92. 1878 ≡ *Himalaiella chenopodiifolia* (Klatt) Raab-Straube in Willdenowia 33: 390. 2003.

Jurinea chitralica (Duthie) N.Garcia, Herrando & Susanna, **comb. nov.** ≡ *Saussurea chitralica* Duthie in Ann. Roy. Bot. Gard., Calcutta 9: 45, t. 57. 1901 ≡ *Himalaiella chitralica* (Duthie) Raab-Straube in Willdenowia 33: 390. 2003.

Jurinea crispa (Vaniot) N.Garcia, Herrando & Susanna, **comb. nov.** ≡ *Saussurea crispa* Vaniot in Bull. Acad. Int. Géogr. Bot., sér. 3, 12: 21. 1902.

= *Aplotaxis nivea* DC., Prodr. 6: 541. 1838, non *Jurinea nivea* C.Winkl. 1890 ≡ *Saussurea nivea* (DC.) Sch.Bip. in Linnaea 19: 331. 1846, nom. illeg., non Turcz. 1837 ≡ *Himalaiella nivea* (DC.) Raab-Straube in Willdenowia 33: 391. 2003.

Jurinea deltoidea (DC.) N.Garcia, Herrando & Susanna, **comb. nov.** ≡ *Aplotaxis deltoidea* DC., Prodr. 6: 541. 1838 ≡ *Saussurea deltoidea* (DC.) Sch.Bip. in Linnaea 19: 331. 1846 ≡ *Theodorea deltoidea* (C.B.Clarke) Kuntze, Revis. Gen. Pl. 1: 367. 1891 ≡ *Himalaiella deltoidea* (DC.) Raab-Straube in Willdenowia 33: 391. 2003.

Jurinea gilesii (Hemsl.) N.Garcia, Herrando & Susanna, **comb. nov.** ≡ *Saussurea gilesii* Hemsl. in Hooker's Icon. Pl. 18: t. 1736. 1888 ≡ *Frolovia gilesii* (Hemsl.) B.A.Sharipova in Rasulova, Fl. Tadzhikskoĭ SSR 10: 161. 1991.

Jurinea heteromalla (D.Don) N.Garcia, Herrando & Susanna, **comb. nov.** ≡ *Cnicus heteromallus* D.Don, Prodr. Fl. Nepal.: 166. 1825 ≡ *Theodorea heteromalla* (D.Don) Kuntze, Revis. Gen. Pl. 1: 367. 1891 ≡ *Saussurea heteromalla* (D.Don) Hand.-Mazz., Symb. Sin. 7: 1152. 1936 ≡ *Himalaiella heteromalla* (D.Don) Raab-Straube in Willdenowia 33: 391. 2003.

Jurinea jucunda (C.Winkl.) Sennikov, **comb. nov.** ≡ *Cnicus jucundus* C.Winkl. in Trudy Imp. S.-Peterburgsk. Bot. Sada 9: 427. 1886 ≡ *Saussurea jucunda* (C.Winkl.) O.Fedtsch. & B.Fedtsch. in Izv. Turkestansk. Otd. Imp. Russk. Geogr. Obshch. 6(Suppl. 4): 234. 1911 ≡ *Modestia jucunda* (C.Winkl.) Kharadze & Tamamsch. in Zametki Sist. Geogr. Rast. 19: 42. 1956 ≡ *Anacantha jucunda* (C.Winkl.) Soják in Sborn. Nár. Muz. Praze, Řada B, Přír. Vědy 1982: 108. 1982.

Jurinea leptophylla (Hemsl.) N.Garcia, Herrando & Susanna, **comb. nov.** ≡ *Saussurea leptophylla* Hemsl. in Hooker's Icon. Pl. 18: t. 1734. Mar 1888.

= *Jurinea ancistrophylla* Boiss., Fl. Orient. Suppl.: 310. Oct 1888.

Jurinea natmataungensis (Fujikawa) Fujikawa, **comb. nov.** ≡ *Himalaiella natmataungensis* Fujikawa in Makinoa, n.s., 10: 168. 2012.

Jurinea peguensis (C.B.Clarke) N.Garcia, Herrando & Susanna, **comb. nov.** ≡ *Saussurea peguensis* C.B.Clarke, Compos. Ind.: 235. 1876 ≡ *Himalaiella peguensis* (C.B.Clarke) Raab-Straube in Willdenowia 33: 391. 2003.

Jurinea yakla (C.B.Clarke) N.Garcia, Herrando & Susanna, **comb. nov.** ≡ *Saussurea yakla* C.B.Clarke, Compos. Ind.: 227. 1876 ≡ *Himalaiella yakla* (C.B.Clarke) Fujikawa & H.Ohba in J. Jap. Bot. 82(3): 133. 2007.

6. Note added in proof

With this paper closed, Kasana & al. (2020) have published a phylogeny of the subtribe. They agree with us in merging *Aucklandia* and *Frolovia* into *Dolomiaea*. They propose, however, to combine *Himalaiella* and *Lipschitziella* under a widely defined *Lipschitziella*. The study is flawed by an insufficient sampling of *Jurinea* and this solution cannot be accepted by two reasons: Firstly, both genera are compared to *Saussurea* and not to *Jurinea*, and achene morphology of the redefined *Lipschitziella* is compatible with *Jurinea*. Secondly, the presence of species that belong unambiguously to *Jurinea* nested in the *Lipschitziella*/*Himalaiella* clade show that the morphological characters alleged for sustaining a different genus are inadequate.

7. Author contributions

AS, SHM, NGJ, and MGC designed and outlined the study, with contributions from LSX and YSC. JRM provided sequences from previous works and supervised the methods and the analyses together with JAC and CR. ANS reviewed and rewrote part of the nomenclature section. ANS, AS, MGC, NGJ, and SHM interpreted the results and outlined the new taxonomic layout. KF critically read the manuscript and suggested nomenclatural changes. SCK, JQL, JLA, JLP, RV, and IM provided materials and critically read the manuscript and contributed to the discussion. The submitted draft has been reviewed and accepted by all the authors.

SHM, http://orcid.org/0000-0002-0488-5112;
JAC, http://orcid.org/0000-0002-6586-0939;
MGC,https://orcid.org/0000-0002-7267-3330;
NGJ, https://orcid.org/0000-0003-1893-5122;
JLA, https://orcid.org/0000-0003-0568-4125;
JLP,https://orcid.org/0000-0002-2091-6222;
JRM, https://orcid.org/0000-0003-3539-2991;
IM,https://orcid.org/0000-0002-5108-2558;
CR, https://orcid.org/0000-0001-8748-3743;
ANS,https://orcid.org/0000-0001-6664-7657;
AS, https://orcid.org/0000-0003-4717-9063;
RV, https://orcid.org/0000-0002-5106-8764

8. Acknowledgments

Authors thank the herbaria that provided material for the study: BC, DUSH, E, FRU, GA, GDA, H, LE, MBK, SKK, TAD, TARI, TI, TK, W, and WU. We performed the phylogenetic analyses on the high-performance cluster (HPC) of the computing facilities of the University of Memphis, which is gratefully acknowledged. Financial support from the Ministerio de Ciencia e Innovación (Project CGL2015-66703-P MINECO/FEDER, UE and Ph.D. grant to Sonia Herrando-Moraira) and the Catalan government ("Ajuts a grups consolidats" 2017-SGR1116) is also greatly acknowledged. This study has been performed under the Ph.D. program "Plant Biology and Biotechnology" of the Autonomous University of Barcelona (UAB). This paper honours the memory of Vicki Ann Funk (1947–2019), who inspired all synantherologists and contributed enormously to the classification of Compositae.

9. Literature cited

Andreasen, K. & Baldwin, B.G. 2001. Unequal evolutionary rates between annual and perennial lineages of checker mallows (*Sidalcea*, Malvaceae): Evidence from 18S-26S rDNA internal and external transcribed spacers. *Molec. Biol. Evol.* 18: 936–944. https://doi.org/10.1093/oxfordjournals.molbev.a003894

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M., Nikolenko, S.I., Pham, S., Prjibelski, A.D., Pyshkin, A.V., Sirotkin, A.V., Vyahhi, N., Tesler, G., Alekseyev, M.A. & Pevzner, P.A. 2012. SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *J. Computat. Biol.* 19: 455–477. https://doi.org/10.1089/cmb.2012.0021

Bentham, G. 1873. Compositae. Pp. 162–533 in: Bentham, G. & Hooker, J.D. (eds.), *Genera plantarum*, vol. 2(1). London: Spottiswoode & Co. https://doi.org/10.5962/bhl.title.747

Bolger, A.M., Lohse, M. & Usadel, B. 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* 30: 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Borowiec, M.L. 2016. AMAS: A fast tool for alignment manipulation and computing of summary statistics. *PeerJ* 4: e1660. https://doi.org/10.7717/peerj.1660

Bremer, K. 1994. *Asteraceae: Cladistics and classification*. Portland: Timber Press.

Brewer, G.E., Clarkson, J.J., Maurin, O., Zuntini, A.R., Barber, V., Bellot, S., Biggs, N., Cowan, R.S., Davies, N.M.J., Dodsworth, S., Edwards, S.L., Eiserhardt, W.L., Epitawalage, N., Frisby, S., Grall, A., Kersey, P.J., Pokorny, L., Leitch, I.J., Forest, F. & Baker, W.J. 2019. Factors affecting targeted sequencing of 353 nuclear genes from herbarium specimens spanning the diversity of angiosperms. *Frontiers Pl. Sci.* 10: 1102. https://doi.org/10.3389/fpls.2019.01102

Capella-Gutiérrez, S., Silla-Martínez, J.M. & Gabaldón, T. 2009. trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25: 1972–1973. https://doi.org/10.1093/bioinformatics/btp348

Chen, Y.-L. & Shih, C. 1981. Geological and ecological studies of Qinghai-Xizang plateau. *Proceedings of Symposium on Qinghai-Xizang (Tibet) Plateau (Beijing, China)*, vol. 2. Beijing: Science Press. Titel is title of vol. 1?

Chen, Y.-S. 2015. *Flora of Pan-Himalaya*, vol. 48(2), *Asteraceae II: Saussurea.* Cambridge: Cambridge University Press.

Dittrich, M. 1977. Cynareae – Systematic review. Pp. 999–

1015 in: Heywood, V.H., Harborne, J.B. & Turner, B.L. (eds.), *The biology and chemistry of the Compositae*, vol. 2. London: Academic Press.

Font, M., Garnatje, T., Garcia-Jacas, N. & Susanna, A. 2002. Delineation and phylogeny of *Centaurea* sect. *Acrocentron* based on DNA sequences: A restoration of the genus *Crocodylium* and indirect evidence of introgression. *Pl. Syst. Evol.* 234: 15–26. https://doi.org/10.1007/s00606-002-0203-3

Fu, Z.X., Jiao, B.H., Nie, B., Zhang, G.J., Gao, T.G. & China Phylogeny Consortium 2016. A comprehensive generic-level phylogeny of the sunflower family: Implications for the systematics of Chinese Asteraceae. *J. Syst. Evol.* 54: 416–437. https://doi.org/10.1111/jse.12216

Garcia-Jacas, N., Garnatje, T., Susanna, A. & Vilatersana, R. 2002. Tribal and subtribal delimitation and phylogeny of the Cardueae (Asteraceae): A combined nuclear and chloroplast DNA analysis. *Molec. Phylogen. Evol.* 22: 51–64. https://doi.org/10.1006/mpev.2001.1038

Garcia-Jacas, N., Uysal, T., Romashchenko, K., Suárez-Santiago, V.N., Ertuğrul, K. & Susanna, A. 2006. *Centaurea* revisited: A molecular survey of the *Jacea* group. *Ann. Bot. (Oxford)* 98: 741–753. https://doi.org/10.1093/aob/mcl157

Garnatje, T., Susanna, A., Garcia-Jacas, N., Vilatersana, R. & Vallès, J. 2005. A first approach to the molecular phylogeny of the genus *Echinops* (Asteraceae): Sectional delimitation and relationships with the genus *Acantholepis* Less. *Folia Geobot.* 40: 407–419. https://doi.org/10.1007/bf02804288

Gaut, B., Yang, L., Takuno, S. & Eguiarte, L.E. 2011. The patterns and causes of variation in plant nucleotide substitution rates. *Annual Rev. Ecol. Evol. Syst.* 42: 245–266. https://doi.org/10.1146/annurev-ecolsys-102710-145119

Häffner, E. 2000. On the phylogeny of the subtribe Carduinae (tribe Cardueae, Compositae). *Englera* 21: 3–208. https://doi.org/10.2307/3776757

Hart, M.L., Forrest, L.L., Nicholls, J.A. & Kidner, C.A. 2016. Retrieval of hundreds of nuclear loci from herbarium specimens. *Taxon* 65: 1081–1092. https://doi.org/10.12705/655.9

Herrando-Moraira, S., The Cardueae Radiations Group (Calleja, J.A., Carnicero-Campmany, P., Fujikawa, K., Galbany-Casals, M., Garcia-Jacas, N., Im, H.-T., Kim, S.-C., Liu, J.-Q., López-Alvarado, J., López-Pujol, J., Mandel, J.R., Massó, S., Mehregan, I., Montes-Moreno, N., Pyak, E., Roquet, C., Sáez, L., Sennikov, A., Susanna, A. & Vilatersana, R.) 2018. Exploring data processing strategies in NGS target enrichment to disentangle radiations in the tribe Cardueae (Compositae). *Molec. Phylogen. Evol.* 128: 69–87. https://doi.org/10.1016/j.ympev.2018.07.012

Herrando-Moraira, S., The Cardueae Radiations Group (Calleja, J.A., Galbany-Casals, M., Garcia-Jacas, N., Liu, J.-Q., López-Alvarado, J., López-Pujol, J., Mandel, J.R., Massó, S., Montes-Moreno, N., Roquet, C., Sáez, L., Sennikov, A., Susanna, A. & Vilatersana, R.) 2019. Nuclear and plastid DNA phylogeny of tribe Cardueae (Compositae) with Hyb-Seq data: A new subtribal classification and a temporal diversification framework. *Molec. Phylogen. Evol.* 137: 313–332. https://doi.org/10.1016/j.ympev.2019.05.001

Hillis, D.M. & Bull, J.J. 1993. An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. *Syst. Biol.* 42: 182–192. https://doi.org/10.1093/sysbio/42.2.182

Hoffmann, O. 1894. Compositae. Pp. 324–333 in: Engler, A. & Prantl, K. (eds.), *Die natürlichen Pflanzenfamilien*. Leipzig: Engelmann.

Huang, C.-H., Sun, R., Hu, Y., Zeng, L., Zhang, N., Cai, L., Zhang, Q., Koch, M.A., Al-Shehbaz, I., Edger, P.P., Pires, J.C., Tan, D.-Y., Zhong, Y. & Ma, H. 2015. Resolution of Brassicaceae phylogeny using nuclear genes uncovers nested radiations and supports convergent morphological evolution. *Molec. Biol. Evol.* 33: 394–412. https://doi.org/10.1093/molbev/msv226

Johnson, M.G., Gardner, E.M., Liu, Y., Medina, R., Goffinet, B., Shaw, A.J., Zerega, N.J.C. & Wickett, N.J. 2016. HybPiper: Extracting coding sequence and introns for phylogenetics from high-throughput sequencing reads using target enrichment. *Applic. Pl. Sci.* 4: 1600016. https://doi.org/10.3732/apps.1600016

Jones, K.E., Fér, T., Schmickl, R.E., Dikow, R.B., Funk, V.A., Herrando-Moraira, S., Johnston, P.R., Kilian, N., Siniscalchi, C.M., Susanna, A., Slovák, M., Thapa, R., Watson, L.E. & Mandel, J.R. 2019. An empirical assessment of a single family-wide hybrid capture locus set at multiple evolutionary timescales in Asteraceae. *Applic. Pl. Sci.* 7: e11295. https://doi.org/10.1002/aps3.11295

Kamelin, R.V. 1993. Rod *Lipschitziella* R. Kam. gen. nov. Pp. 371, 632 in: Kamelin, R.V. (ed.), *Opreditel' rastenij Srednej Azii: Kriticeskij konspekt flory* [*Conspectus florae Asiae mediae*], vol. 10. Tashkent: Izdatel'stvo "FAN" Uzbekskoj S.S.R. [in Russian]

Kamelin, R.V. & Kovalevskaja, S.S. 1993. Rod *Frolovia* (DC.) Lipsch. Pp. 353–354 in: Kamelin, R.V. (ed.), *Opreditel' rastenij Srednej Azii: Kriticeskij konspekt flory* [Conspectus florae Asiae mediae], vol. 10. Tashkent: Izdatel'stvo "FAN" Uzbekskoj S.S.R. [in Russian]

Kasana, S., Dwivedi, M.D. & Uniyal, P.L. 2018. Taxonomic status of *Saussurea costus* (Falc.) Lipsch. (Asteraceae: Cardueae): A critically endangered species from Himalaya, India. *Pleione* 12: 81–84. https://doi.org/10.26679/Pleione.12.1.2018.081-084

Kasana, S., Dwivedi, M.D., Uniyal, P.L. & Pandey, A.K. 2020. An updated circumscription of *Saussurea* (Cardueae, Asteraceae) and allied genera based on morphological and molecular data. *Phytotaxa* 450: 173–187. https://doi.org/10.11646/phytotaxa.450.2.3

Katoh, K. & Standley, D.M. 2013. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molec. Biol. Evol.* 30: 772–780. https://doi.org/10.1093/molbev/mst010

Kita, Y., Fujikawa, K., Ito, M., Ohba, H. & Kato, M. 2004. Molecular phylogenetic analyses and systematics of the genus *Saussurea* and related genera (Asteraceae, Cardueae). *Taxon* 53: 679–690. https://doi.org/10.2307/4135443

Kück, P. & Longo, G.C. 2014. FASconCAT-G: Extensive functions for multiple sequence alignment preparations concerning phylogenetic studies. *Frontiers Zool.* 11: 81. https://doi.org/10.1186/s12983-014-0081-x

Li, H. & Durbin, R. 2009. Fast and accurate short read alignment with burrowswheeler transform. *Bioinformatics* 25: 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Lipschitz, S.J. 1954. O rode *Frolovia* (DC.) Lipsch. [De genere novo *Frolovia* (DC.) Lipsch.]. *Bot. Mater. Gerb. Bot. Inst. Komarova Akad. Nauk S.S.S.R.* 16: 461–462. [in Russian]

Lipschitz, S.J. 1961. K poznaniju roda *Saussurea* DC. flory SSSR. [Ad cognitionem generis *Saussurea* DC. florae URSS]. *Bot. Mater. Gerb. Bot. Inst. Komarova Akad. Nauk S.S.S.R.* 21: 369–381. [in Russian]

Lipschitz, S.J. 1962. Rod *Saussurea* DC. Pp. 356–530 in: Schischkin, B.K. & Bobrov, E.G. (eds.), *Flora S.S.S.R.*, vol. 27. Moscow & Leningrad: Academy of Sciences of the USSR. [in Russian]

Lipschitz, S.J. 1979. *Rod Saussurea DC. (Asteraceae)* [*Genus Saussurea DC. (Asteraceae)*]. Leningrad: Nauka, Leningradskoje Otdelenie. [in Russian]

López-Vinyallonga, S., Mehregan, I., Garcia-Jacas, N., Tscherneva, O., Susanna, A. & Kadereit, J.W. 2009. Phylogeny and evolution of the *Arctium-Cousinia* complex (Compositae, Cardueae–Carduinae). *Taxon* 58: 153–171. https://doi.org/10.1002/tax.581016

Mandel, J.R., Dikow, R.B., Funk, V.A., Masalia, R.R., Staton, S.E., Kozik, A., Michelmore, R.W., Rieseberg, L.H. & Burke, J.M. 2014. A target enrichment method for gathering phylogenetic information from hundreds of loci: An example from the Compositae. *Applic. Pl. Sci.* 2: 1300085. https://doi.org/10.3732/apps.1300085

Mandel, J.R., Dikow, R.B. & Funk, V.A. 2015. Using phylogenomics to resolve mega-families: An example from Compositae. *J. Syst. Evol.* 53: 391–402. https://doi.org/10.1111/jse.12167

Mandel, J.R., Barker, M.S., Bayer, R.J., Dikow, R.B., Gao, T.-G., Jones, K.E., Keeley, S., Kilian, N., Ma, H., Siniscalchi, C.M., Susanna, A., Thapa, R., Watson, L.E. & Funk, V.A. 2017. The Compositae Tree of Life in the age of phylogenomics. *J. Syst. Evol.* 55: 405–410. https://doi.org/10.1111/jse.12265

Mandel, J.R., Dikow, R.B., Siniscalchi, C.M., Thapa, R., Watson, L.E. & Funk, V.A. 2019. A fully resolved backbone phylogeny reveals numerous dispersals and explosive diversifications throughout the history of Asteraceae. *Proc. Natl. Acad. Sci. U.S.A.* 116: 14083–14088. https://doi.org/10.1073/pnas.1903871116

Miller, M.A., Pfeiffer, W. & Schwartz, T. 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. Pp. 45–52 in: *Gateway Computing Environments Workshop (GCE 2010)*, New Orleans, Lousiana, U.S.A., 14 Nov. 2010. New Orleans: IEEE. https://doi.org/10.1109/gce.2010.5676129

Mirtadzadini, M., Bordbar, F., Parishani, M.R., Vitek, E. & Rahiminejad, M.R. 2018. Novelties in Cardueae (Asteraceae). *Nova Biol. Reperta* 5: 337–348. https://doi.org/10.29252/nbr.5.3.336

Mirtadzadini, M., Bordbar, F., Parishani, M.R., Rahiminejad, M.R. & Vitek, E. 2019. *Jurinea gedrosiaca* (Asteraceae, Cardueae), correctly belongs to *Tricholepis*. *Ann. Naturhist. Mus. Wien, B* 121: 263–270. https://doi.org/10.29252/nbr.5.3.336

Raab-Straube, E. von 2003. Phylogenetic relationships in *Saussurea* (Compositae, Cardueae) sensu lato, inferred from morphological, ITS and *trnL-trnF* sequence data, with a synopsis of *Himalaiella* gen. nov., *Lipschitziella* and *Frolovia*. *Willdenowia* 33: 379–402. https://doi.org/10.3372/wi.33.33214

Raab-Straube, E. von 2017. *Taxonomic revision of Saussurea subgenus Amphilaena (Compositae, Cardueae)*. Englera 34. Berlin: Botanic Garden and Botanical Museum Berlin, Freie Universität Berlin.

Rambaut, A. 2016. FigTree, version 1.4.3 [computer program]. http:// tree.bio.ed.ac.uk

Sayyari, E. & Mirarab, S. 2016. Fast coalescent-based computation of local branch support from quartet frequencies. *Molec. Biol. Evol.* 33: 1654–1668. https://doi.org/10.1093/molbev/msw079

Schmickl, R., Liston, A., Zeisek, V., Oberlander, K., Weitemier, K., Straub, S.C.K., Cronn, R.C., Dreyer, L.L. & Suda, J. 2016. Phylogenetic marker development for target enrichment from transcriptome and genome skim data: the pipeline and its application in southern African *Oxalis* (Oxalidaceae). *Molec. Ecol. Resources* 16: 1124–1135. https://doi.org/10.1111/1755-0998.12487

Sennikov, A.N. & Lazkov, G.A. 2013. Taxonomic corrections and new records in vascular plants of Kyrgyzstan, 2. *Memoranda Soc. Fauna Fl. Fenn.* 89: 125–138.

Shi, Z. & Raab-Straube, E. von 2011. *Saussurea* group. Pp. 42–149 in: Wu, Z.Y., Raven, P.H. & Hong, D.Y. (eds.), *Flora of China*, vol. 20–21 (Asteraceae). Beijing: Science Press; St Louis: Missouri Botanical Garden Press.

Silva-Moure, K., Torrecilla, P. & Lapp, M. 2013. Taxonomía de *Lasiocephalus* Willd. ex Schltdl. (Asteraceae) en Venezuela. *Ernstia* 23: 91–118.

Siniscalchi, C.M., Loeuille, B., Funk, V.A., Mandel, J.R. & Pirani, J.R. 2019. Phylogenomics yields new insight into relationships within Vernonieae (Asteraceae). *Frontiers Pl. Sci.* 10: 1224. https://doi.org/10.3389/fpls.2019.01224

Stamatakis, A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22: 2688–2690. https://doi.org/10.1093/bioinformatics/btl446

Stamatakis, A. 2014. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30: 1312–1313. https://doi.org/10.1093/bioinformatics/btu033

Susanna, A. & Garcia-Jacas, N. 2007. Tribe Cardueae Cass. (1819). Pp. 123–146 in: Kadereit, J.W. & Jeffrey, C. (eds.), *The families and genera of vascular plants*, vol. 8. Berlin & Heidelberg: Springer.

Susanna, A. & Garcia-Jacas, N. 2009. Cardueae (Carduoideae). Pp. 293–313 in: Funk, V.A., Susanna, A., Stuessy, T.F. & Bayer, R.J. (eds.), *Systematics, evolution, and biogeography of Compositae*. Vienna: International Association for Plant Taxonomy.

Susanna, A., Garcia-Jacas, N., Hidalgo, O., Vilatersana, R. & Garnatje, T. 2006. The Cardueae (Compositae) revisited: Insights from ITS, *trnL-trnF*, and *matK* nuclear and chloroplast DNA analysis. *Ann. Missouri Bot. Gard.* 93: 150–171. https://doi.org/10.3417/0026-6493(2006)93[150:TCCRIF]2.0.CO;2

Szukala, A., Korotkova, N., Gruenstaeudl, M., Sennikov, A.N., Lazkov, G.A., Litvinskaya, S.A., Gabrielian, E., Borsch, T. & Raab-Straube, E. von 2019. Phylogeny of the Eurasian genus *Jurinea* (Asteraceae: Cardueae): Support for a monophyletic genus concept and a first hypothesis on overall species relationships. *Taxon* 68: 112–131. https://doi.org/10.1002/tax.12027

Thapa, R., Bayer, R.J. & Mandel, J.R. 2019. Phylogenomics resolves the relationships within *Antennaria* (Asteraceae, Gnaphalieae) and yields new insights into its morphological character evolution and biogeography. *Syst. Bot.* In Press.

Villaverde, T., Pokorny, L., Olsson, S., Rincón-Barrado, M., Johnson, M.G., Gardner, E.M., Wickett, N.J., Molero, J., Riina, R. & Sanmartín, I. 2018. Bridging the micro- and macroevolutionary levels in phylogenomics: Hyb-Seq solves relationships from populations to species and above. *New Phytol.* 220: 636–650. https://doi.org/10.1111/nph.15312

Viruel, J., Conejero, M., Hidalgo, O., Pokorny, L., Powell, R.F., Forest, F., Kantar, M.B., Soto Gomez, M., Graham, S.W., Gravendeel, B., Wilkin, P. & Leitch, I.J. 2019. A target capture-based method to estimate ploidy from herbarium specimens. *Frontiers Pl. Sci.* 10: 937. https://doi.org/10.3389/fpls.2019.00937

Wang, Y.-J., Liu, J.-Q. & Miehe, G. 2007. Phylogenetic origins of the Himalayan endemic *Dolomiaea*, *Diplazoptilon* and *Xanthopappus* (Asteraceae: Cardueae) based on three DNA regions. *Ann. Bot. (Oxford)* 99: 311–322. https://doi.org/10.1093/aob/mcl259

Wang, Y.-J., Susanna, A., Raab-Straube, E. von, Milne, R. & Liu, J.-Q. 2009. Island-like radiation of *Saussurea* (Asteraceae: Cardueae) triggered by uplifts of the Qinghai–Tibetan Plateau. *Biol. J. Linn. Soc.* 97: 893–903. https://doi.org/10.1111/j.1095-8312.2009.01225.x

Wen, J., Zhang, J., Nie, Z.-L., Zhong, Y. & Sun, H. 2014. Evolutionary diversifications of plants on the Qinghai-Tibetan Plateau. *Frontiers Genet.* 5: 4. https://doi.org/10.3389/fgene.2014.00004

Xu, L.-S., Herrando-Moraira, S., Susanna, A., Galbany-Casals, M. & Chen, Y.-S. 2019. Phylogeny, origin and dispersal of *Saussurea* (Asteraceae) based on chloroplast genome data. *Molec. Phylogen. Evol.* 141: 106613. https://doi.org/10.1016/j.ympev.2019.106613

Yuan, Q., Bi, Y.-C. & Chen, Y.-S. 2015. *Diplazoptilon* (Asteraceae) is merged with *Saussurea* based on evidence from morphology and molecular systematics. *Phytotaxa* 236: 53–61. https://doi.org/10.11646/phytotaxa.236.1.4

Yue, J.-X., Li, J., Wang, D., Araki, H., Tian, D. & Yang, S. 2010. Genome-wide investigation reveals high evolutionary rates in annual model plants. *B. M. C. Pl. Biol.* 10: 242. https://doi.org/10.1186/1471-2229-10-242

Zhang, C., Rabiee, M., Sayyari, E. & Mirarab, S. 2018. ASTRAL-III: Polynomial time species tree reconstruction from partially resolved gene trees. *B. M. C. Bioinform.* 19: 153. https://doi.org/10.1186/s12859-018-2129-y

Zhang, X., Deng, T., Moore, M.J., Ji, Y., Lin, N., Zhang, H., Meng, A., Wang, H., Sun, Y. & Sun, H. 2019. Plastome phylogenomics of *Saussurea* (Asteraceae: Cardueae). *B. M. C. Pl. Biol.* 19: 290. https://doi.org/10.1186/s12870-019-1896-6

10. Appendix 1

Voucher information of the studied material and GenBank (NCBI) accession numbers of raw sequence reads. New sequences generated in this study are indicated with an asterisk (*) after GenBank number. Taxon names follow the final proposed taxonomic treatment presented in the study. Notes: [1] In Herrando-Moraira & al. (2018) appears as *Arctium eriophorum*; [2] In Herrando-Moraira & al. (2018) it appears as *Cousinia ninae*; [3] In Herrando-Moraira & al. (2018) it appears as *Jurinea abramowii*; [4] In Herrando-Moraira & al. (2018) it appears as *Jurinea bucarica*; [5] In Herrando-Moraira & al. (2019) it appears as *Rhaponticum acaule*; [6] In Herrando-Moraira & al. (2018) it appears as *Saussurea davurica*; [7] In Herrando-Moraira & al. (2018) it appears as *Saussurea jadrinzevii*.

Arctium fedtschenkoanum (Bornm.) S.López & al., Tajikistan, *Romaschenko 632 & Susanna* (BC), SAMN11585472; ***Arctium grandifolium*** (Kult.) S.López & al., Kazakhstan, Zambylskaya oblast, Talaski Alatau, 6 km W from Il Tai, *Susanna 2181 & al.* (BC), SAMN15064103*; ***Arctium karatavicum*** (Regel & Schmalh.) Kuntze, Tadjikistan, s. loc., *Romanshenko 607 & Susanna* (BC), SAMN15064104*; ***Arctium minus*** (Hill) Bernh., Spain, Barcelona, *Vilatersana 1100 & López-Vinyallonga* (BC), SAMN15064105*; ***Arctium nidulans***[1] (Regel) Sennikov, Kazakhstan, Almatinskaya oblast, Alatau mt., above Almaty, *Susanna 2088 & al.* (BC), SAMN15064106*; ***Arctium umbrosum*** (Bunge) Kuntze, Kazakhstan, Almatinskaya oblast, Alatau mt. above Almaty, *Susanna 2100 & al.* (BC), SAMN15064107*; ***Carduus nutans*** L., Tajikistan, Vorzov, *Susanna 2481 & al.* (BC), SAMN15064108*; ***Centaurea aspera*** L., cultivated at the Barcelona Botanical Garden, SAMN15064109*; ***Cirsium sairamense*** (C.Winkl.) O.Fedtsch. & B.Fedtsch., Tadjikistan, Maijora, *Susanna 2468 & al.* (BC), SAMN11585477; ***Cousinia arachnoidea***[2] Fisch. & C.A.Mey. ex DC., Kyrgyzstan, 20–22 km E of Shakaftar, *Sultanova & al. s.n.* (FRU, LE), SAMN15064110*; ***Cousinia knorringiae*** Bornm., Kyrgyzstan, Bozbu-Too, 21 May 1970, *Sudnitsyna s.n.* (FRU, LE), SAMN15064111*; ***Cousinia onopordioides*** Ledeb., Iran, Tehran, between Firuzkuh and Semnan, *Susanna 1637 & al.* (BC), SAMN15064112*; ***Cousinia polytimetica*** Tscherneva, Uzbekistan, Bukharskaya reg., Zeravshan river, to SE from Uzlishkent vil., *Kryakin s.n.* (LE), SAMN15064113*; ***Cousinia pungens*** Juz., Iran, Khorasan, *Rechinger 51337* (W0007098), SAMN15064114*; ***Cousinia splendida*** C.Winkl., Tadjikistan, Fan mountains, road above Iskandar-Kul, *Susanna 2507 & al.* (BC), SAMN15064115*; ***Cousinia spryginii*** Kult., Uzbekistan, Kashkadarbinskaya reg., low mountains to SE of vil. Dekhanabad, *Botschantzev 46* (LE), SAMN10983402; ***Cynara cardunculus*** L., U.S.A., UW Medicinal Plant Garden., *Mandel s.n.* (GA 135), SAMN11585480; ***Dolomiaea asbukinii*** (Iljin) N.Garcia & al., Tajikistan, Sogdiiskai Oblast, Ashtskii District, Kuraminskii Ridge, the Pangaza River Valley, Suje-kuh city, 25 Jul 1943, *Kamelin s.n.* (LE), SAMN15064116*; ***Dolomiaea baltalensis*** Dar & Naqshi, India, Kangi, Ladak, Kashmir, ca. 3950 m, *Walter Koelz 2823* (E00469695), SAMN15064117*; ***Dolomiaea berardioidea*** (Franch.) C.Shih **1**, China, Yulong Shan, Gang Ho Ba valley, 3350 m, *Chungtien-Lijiang-Dali Expedition no. 717* (E00469654), SAMN15064118*; ***Dolomiaea berardioidea*** **2**, China, Yunnan bor.-occid., Prope urbem Lidjiang, 2950–3100 m, *Handel-Mazzetti 662* (WU), SAMN15064119*; ***Dolomiaea costus*** **(Falc.)** Kasana & A.K.Pandey, China, Yunnan, Lijiang Prefecture, Xin Zhu forest, between Judian and Litiping Plateau, 2690 m, *B. Aldén & al. 1756* (E00003397), SAMN15064120*; ***Dolomiaea edulis*** (Franch.) C.Shih, China, Yangbi Xian, W side of Diancang Shan mountain range, *Sino-Amer. Bot. Exped. No. 534* (E00469665), SAMN15064121*; ***Dolomiaea forrestii*** (Diels) C.Shih **1**, China, Yunnan bord.-occid., prope urbem Lidjiang, *Handel-Mazzetti 661* (W0004887), SAMN15064122*; ***Dolomiaea forrestii*** **2**, China, Yunnan, Dêqên Zang Aut., Dêqên, E flank of Bei Ma Xue Shan, 4140 m, 17 Sep 1995, *s. col.* (E00051554), SAMN15064123*; ***Dolomiaea forrestii*** **3**, China, Yunnan bor.-occid., prope urbem Lidjiang 06–09.1914, 1915, 1916, *Handel-Mazzetti 3672* (WU), SAMN15064124*; ***Dolomiaea frolowii*** (Ledeb.) Kasana & A.K.Pandey, Russia, Altai, Kosh-Agachsky, Kuraiskiy Ridge, the Ortolyk River, 2249 m, *A. Pyak & E. Pyak 11012* (TK a-034-2016), SAMN15064125*; ***Dolomiaea georgii*** (Anth.) C.Shih **1**, China, Yunnan, Yangtze watershed, Likiang, *J.F. Rock 6138* (E00469683), SAMN15064126*; ***Dolomiaea georgii*** **2**, China, Yunnan, Yangtze watershed, *J.F. Rock 6138* (E00469683), SAMN15064127*; ***Dolomiaea gorbunovae*** (Kamelin) N.Garcia & al., Kyrgyzstan, Kurama Range easternmost side, right side of Kasansay River, *Sennikov 461* (H), SAMN15064128*; ***Dolomiaea macrocephala*** Royle, Pakistan, Khaibar, upper Hunza, 3820 m, *G.S. Miehe 2627* (W0001920), SAMN15064129*; ***Dolomiaea platylepis*** (Hand.-Mazz.) C.Shih, China, Muli, SW Szechuan, between Muli Gomba and Baurong and Wa-Erh-Dje, 3700 m, Jul 1928, *J.F. Rock s.n.* (E00469687), SAMN15064130*; ***Dolomiaea salwinensis*** (Hand.-Mazz.) C.Shih **1**, China, Yunnan, bor.-occid., prope fines Tibeto-Birmanicas inter fluvios Lu-djiang (Salween) et Djiou-djiang (Irrawadi or sup.), ca. 3825 m, *Handel-Mazzetti 1853* (WU0061241), SAMN15064131*; ***Dolomiaea salwinensis*** **2**, China, Mekhong, Salween divide Yunnan, ca. 3950 m, *George Forrest 14350* (E00469690), SAMN15064132*; ***Dolomiaea salwinensis*** **3**, China, Yunnan, Mekong-Salween divide, ca. 3950 m, Aug 1917, *s. col.* (W0013379), SAMN15064133*; ***Dolomiaea saussureoides*** (Hand.-Mazz.) Y.L.Chen & C.Shih, China, SE Tibet, Kongbo Province, Tse La, Langong, ca. 4260–4570 m, *F. Ludlow & al. 5620* (E00469649), SAMN15064134*; ***Dolomiaea souliei*** (Franch.) C.Shih, China, Sikang, Kangting (Tachienlu) Distr., Chungo valley, Hsintientzü, ca. 3800 m, *Harry Smith 11630* (W0012531), SAMN15064135*; ***Dolomiaea sulcata*** (Iljin) N.Garcia & al., Kyrgyzstan, Jalal-Abad Region, Aksy District, Ferganskii Ridge, between Maili-sai and Jedde-sai cities, 3000–3200 m, 3 May 1952, *Popova & al. s.n.* (LE), SAMN15064136*; ***Galactites tomentosa*** Moench, Spain, Barcelona, 8 Jun 2017, *Susanna s.n.* (BC), SAMN15064137*; ***Jurinea abramovii***[3] Regel & Herder, Tadjikistan, Hissar Mt., *Smirnova 224 & al.* (DUSH), SAMN15064138*; ***Jurinea alata*** Cass., cultivated at the Barcelona Botanical Garden, SAMN11585483; ***Jurinea albescens*** (DC.) N.Garcia & al. **1**, India, Kashmir, Dras, 10,000 ft, *Stainton 7175* (E00466963), SAMN15064139*; ***Jurinea albescens*** **2**, India, Kashmir, Srinagar, in monte Shankaracharya, 1770–2000 m, *K.H. Rechinger 62135* (W0003315), SAMN15064140*; ***Jurinea algida*** Iljin, Kyrgyzstan, Kok-Suu River, 16 Aug 2006, *Lazkov s.n.* (FRU), SAMN15064141*; ***Jurinea atropurpurea*** C.Winkl., Tadjikistan, s. loc., *Kotehkariova & Zhogolieva 16094* (DUSH), SAMN11585484; ***Jurinea auriculata*** (DC.) N.Garcia & al. **1**, Nepal, Taplejung Dist., Mane

Bhanjyang–Jaljale Polhari, *Y. Omori & al. 9920052* (MBK0254878), SAMN15064142*; ***Jurinea auriculata* 2**, cultivated plant in Royal Botanic Garden Edinburgh, *S. Sikkim McPherson s.n.* (E00469273), SAMN15064143*; ***Jurinea baldschuanica*** C.Winkl., Tadjikistan, mountains above Kara-Chuiráá, *Susanna 2561 & al.* (BC), SAMN15064144*; ***Jurinea berardioides*** (Boiss.) O.Hoffm. **1**, Pakistan, Kalat, Kolpur to Mach, *Rechinger 28456* (W), SAMN15064145*; ***Jurinea berardioides* 2**, Iran, Kazerun, Kotal Dahli, *Stapf 1284* (WU), SAMN15064146*; ***Jurinea bucharica*[4]** C.Winkl., s. loc., 22 Apr 1975, *s. col. 10387* (DUSH), SAMN15064147*; ***Jurinea bungei*** Boiss., Iran, Persia, S Fars, 20 km S Abadeh, *F. Kasy 426* (W0016436), SAMN15064148*; ***Jurinea caespitans*** Iljin, Kyrgyzstan, north of Kara-Jygach village, 9 Jul 2016, *Sennikov s.n.* (H), SAMN15064149*; ***Jurinea capusii*** Franch., Kyrgyzstan, Chapchyma-Say, 14 Jul 2016, *Sennikov s.n.* (H), SAMN15064153*; ***Jurinea carduicephala*** Iljin, Tajikistan, Gorno-Badakhshan, Shughnon, Shughnonskii Ridge, *Semakov & Dengubenko s.n.* (LE 8428), SAMN15064154*; ***Jurinea carduiformis*** (Jaub. & Spach) Boiss., Iran, Tehran, near Sorkhehesar, *Susanna 1631 & al.* (BC), SAMN15064155*; ***Jurinea ceratocarpa*** (Decne.) Benth. ex C.B.Clarke **1**, India, Garhwal, *Polunin 14941* (E00714512), SAMN15064156*; ***Jurinea ceratocarpa* 2**, India, Himachal Pradesh, 11,600 ft, *J.D.A. Stainton s.n.* (E00158889), SAMN15064157*; ***Jurinea chenopodiifolia*** (Klatt) N.Garcia & al., Afghanistan, Parwan, Salang Tal, quelliger Hang bei der Brücke oberhalb km 85, 1800 m, *O. Anders 3800* (W0006559), SAMN15064158*; ***Jurinea chitralica*** (Duthie) N.Garcia & al. **1**, Afghanistan, Nuristan, Pashki, 2300 m, *Rechinger 1117* (W0000344), SAMN15064159*; ***Jurinea chitralica* 2**, Pakistan, Chitral, Zirrat, lowarai pass, 7000 ft, *D.A. Stainton 2546* (E00160033), SAMN15064160*; ***Jurinea chondrilloides*** (C.Winkl.) O.Fedtsch. **1**, Afghanistan, Ghorat, in faucibus septentriones versus spectantibus Mollah Allah, SW Taiwara, ca. 2200–2300 m, *K.H. Rechinger 18978* (W0006534), SAMN15064161*; ***Jurinea chondrilloides* 2**, Afghanistan, Ghorat, in faucibus septentriones versus spectantibus Mollah Allah, SW Taiwara, ca. 2200–2300 m, *K.H. Rechinger 18978* (E00466342), SAMN15064162*; ***Jurinea crispa*** (Vaniot) N.Garcia & al. **1**, Thailand, Chiang Mai, western flank of Doi Inthanond, Mae Pau, c. 2250 m, *Expedition of the Rijksherbarium, Leiden, Netherlands, and the Forest Herbarium, Bangkok, Thailand 2388* (E00469242), SAMN15064163*; ***Jurinea crispa* 2**, China, Gongshan Zizhixian, Dulongjiang Xiang, *Gaoligong Shan Biodiversity Survey 21197* (E00270054), SAMN15064164*; ***Jurinea crispa* 3**, China, Yunnan, Mekong, 2100–2500 m, *Handel-Mazzetti 1486* (W0013329), SAMN15064198*; ***Jurinea deltoidea*** (DC.) N.Garcia & al. **1**, China, Himachal Pradesh, Wegvon Phakding-Namche Bazar, Solo Khumbu, 3100 m, *M. Staudinger N8/22* (W0001056), SAMN15064165*; ***Jurinea deltoidea* 2**, India, Distr. Darjeelin, "Tigerhill", ca. 2500 m, *Ern 3428* (E00466988), SAMN15064166*; ***Jurinea ferganica*** (Iljin) Iljin, Kyrgyzstan, near Kadamzhay village, 18 Jul 2016, *Sennikov s.n.* (H), SAMN15064167*; ***Jurinea filicifolia*** Boiss., Georgia, Zazbeg, Bergkamm K'vena mt'a (Kvena mta), 1.7–3.8 km W ober der Kirche Tsminda Sameba (Cminda Sameba), 2450–2940 m, 21 Aug 1997, *M. Staudinger s.n.* (W0014044), SAMN15064168*; ***Jurinea fontqueri*** Cuatr., Spain, Jaén, cerro Cárceles, Mágina, *Martínez Lirola s.n.* (GDA 44615), SAMN15064169*; ***Jurinea frigida*** Boiss. **1**, Iran, Persia borealis, in excelsis alpinum Totschal, 3800 m s.m., *J. et A. Bornmüller 7309* (WU), SAMN15064170*; ***Jurinea frigida* 2**, Iran, Tehran, Tuchal Mts., 3700 m, *J. Noroozi 2478* (W0011550), SAMN15064171*; ***Jurinea frigida* 3**, Iran, Damavand, 4000 m, *Susanna 2625 & al.* (BC), SAMN15064172*; ***Jurinea gilesii*** (Hemsl.) N.Garcia & al., Tajikistan, East Pamir, Kara-Djilga, *Kochkariev 17* (TAD), SAMN15064173*; ***Jurinea heteromalla*** (D.Don) N.Garcia & al. **1**, India, Mussoorie, *Wavra 1498* (W0075519), SAMN15064150*; ***Jurinea heteromalla* 2**, India, Utar Pradesh, Kumaon, Supra Katghodam versus Bhimtal, 700–1000 m, *K.H. Rechinger 6242* (W0003374), SAMN15064151*; ***Jurinea heteromalla* 3**, Pakistan, Murree, 7800 ft, *Duthie 14606* (WU), SAMN15064152*; ***Jurinea heteromalla 4***, Pakistan, Baluchistan, Quetta, Coralai to Harnai, Tokhan Pass., 900–1400 m, *Jennifer Lamond 1251* (E00469094), SAMN15064174*; ***Jurinea jucunda*** (C.Winkl.) Sennikov, Kyrgyzstan, 20 km NW of Samarkandyk, Kyzyl-Suu, 10 May 1978, *Aidarova & Ubukeeva s.n.* (FRU), SAMN15064175*; ***Jurinea karategina*** (Lipsky) O.Fedtsch., Tajikistan, Surjov, *Susanna 2542 & al.* (BC), SAMN15064176*; ***Jurinea kokanica*** Iljin, Kyrgyzstan, 15 km E of Kosh-Bulak village, 9 May 2007, *Ganybaeva s.n.* (FRU), SAMN15064177*; ***Jurinea kyzylkyrensis*** Kamelin & Tscherneva, Kyrgyzstan, left side of Naryn River, Kyzyl-Kyr, 12 Aug 1979, *Botschantzev & al. s.n.* (FRU), SAMN15064178*; ***Jurinea lanipes*** Rupr., Kyrgyzstan, Boom ravine, *Sennikov 428a* (H), SAMN15064179*; ***Jurinea leptoloba*** DC., Iran, 30 km N from Tabriz, *Susanna 1654 & al.* (BC), SAMN15064180*; ***Jurinea leptophylla*** (Hemsl.) N.Garcia & al., Afghanistan, Kapisa, *Podlech 12500* (W), SAMN15064181*; ***Jurinea macrocephala*** DC., Iran, 20 Km N of Qarabchaman, *Susanna 1650 & al.* (BC), SAMN15064182*; ***Jurinea microcephala*** Boiss., Iran, Shahrud-Bustam, in declivibus australibus montium Shahvar in saxasis calc., 3500–3900 m, *K.H. et F. Rechinger 5975* (W0003394), SAMN15064183*; ***Jurinea modesta*** Boiss., Afghanistan, Nangarhar, Torkham, stony slopes, c. 700 m, 30 Apr 1969, *I. Hedge & al. s.n.* (E00467348), SAMN15064184*; ***Jurinea moschus*** Fisch. & C.A.Mey. **1**, Georgia, Mtskheta-Mtianeti, Great Caucasus, from church Tsminda Sameba in direction of Mt. Kasbek, 2650–3160 m, *G.M. Schneeweiss & al. 8684* (W0013182), SAMN15064185*; ***Jurinea moschus* 2**, Georgia, Pirikiti Khevsureti District. Dusheti District (21), gorge of Arguni river, ca. 2764 m, *Shamil Shetekauri 1073* (W0018609), SAMN15064186*; ***Jurinea moschus* 3**, Georgia, Mtskheta-Mtianeti, Great Caucasus, from church Tsminda Sameba in direction of Mt. Kasbek, 2650–3160 m, *G.M. Schneeweiss & al. 8684* (WU), SAMN15064187*; ***Jurinea moschus* 4**, Turkey, Kastamonu, Ilgaz Daglari, *Buchner B83-80-11* (W0010480), SAMN15064188*; ***Jurinea narynensis*** Kamelin & Tscherneva, Kyrgyzstan, 8 km from Tash-Kumyr to Jangi-Jol, *Lazkov & Omuralieva s.n.* (FRU), SAMN15064189*; ***Jurinea natmataungensis*** (Fujikawa) Fujikawa, Myanmar, Chin State, Natma Taung National Park, *K. Fujikawa & al. 086717* (MBK0239428), SAMN15064190*; ***Jurinea olgae*** Regel & Schmalh., Tadjikistan, slopes over kishlag Voru, *Susanna 2517 & al.* (BC), SAMN15064191*; ***Jurinea orientalis*** (Iljin) Iljin, Kyrgyzstan, near Shekoftar village, 13 Jul 2016, *Sennikov s.n.* (H), SAMN15064192*; ***Jurinea peguensis*** (C.B.Clarke) N. Garcia & al. **1**, China, Yunnan, 1916, *Cavalerie s.n.* (W0012455), SAMN15064193*; ***Jurinea peguensis* 2**, Thailand, Mae hong Son Prov., Maung Dist., *N. Tanaka & al. HN8512* (MBK0202098), SAMN15064194*; ***Jurinea pinnata*** (Lag. ex Pers.) DC., Morocco, Meknès-Tafilalt, Middle-Atlas, from Midelt to Timahdite, col du Zad, *Calleja & Hipold 20103091* (BC), SAMN15064195*; ***Jurinea popovii*** Iljin, Tadjikistan, s. loc., *Chukavina & al. 163(86)* (DUSH), SAMN15064196*; ***Jurinea pulchella*** DC., Iran, Azerbaidjan, Khoi to Schapour, *Belboudi 5462E* (W0015068), SAMN15064197*; ***Jurinea schachimardanica*** Iljin, Kyrgyzstan, W of Shahimardan, *Sennikov 472* (H), SAMN15064199*; ***Jurinea squarrosa*** Fisch. & C.A.Mey. **1**, Armenian SSR, Gugark region, Gamzachiman, Bazum Ridge, Chingiliyurt mountain, 2100 m, 19 Jun 1968, *A. Pogosyan s.n.* (WU), SAMN15064200*; ***Jurinea squarrosa* 2**, Armenia, Shirak range, left slope of

the remarkable gorge NE of Krashen, 2076 m.s.m., *Ernst Vitek 04-1513 & al.* (W004012), SAMN15064201*; ***Jurinea stenophylla*** Iljin, Kyrgyzstan, Kasan-Say River near Terek-Say village, 14 Jun 1996, *Pimenov & al. s.n.* (FRU), SAMN15064202*; ***Jurinea stoechadifolia*** DC., Ukraine, Crimea, *Romo 10321 & al.* (BC), SAMN15064203*; ***Jurinea suffruticosa*** Regel, Kyrgyzstan, Kasan-Say River, 14 Jul 2016, *Sennikov s.n.* (H), SAMN15064204*; ***Jurinea thianschanica*** Regel & Schmalh., Kyrgyzstan, between Kochkor and Ottuk, near Orto-Tokoy village, 3 Jul 2016, *Sennikov s.n.* (H), SAMN15064205*; ***Jurinea trautvetteriana*** Regel & Schmalh., Tadjikistan, s. loc., *Ovczinnikov 16305 & Zaprjagaeva* (DUSH), SAMN15064206*; ***Jurinea winkleri*** Iljin, Kyrgyzstan, east of Uch-Korgon village, 16 Jul 2016, *Sennikov s.n.* (H), SAMN15064207*; ***Jurinea xeranthemoides*** Iljin, Kyrgyzstan, near Uch-Korgon village, 16 Jul 2016, *Sennikov s.n.* (H), SAMN15064208*; ***Jurinea yakla*** (C.B.Clarke) N.Garcia & al., Nepal, Sagarmatha Zone, Solukhumbu Distr., Mosom Kharka, 3600 m, *M. Wakabayashi & al. 97-30345* (E00232078), SAMN15064209*; ***Karvandarina cartilaginea*** (Mozaff.) Parishani & al., Iran, Khuzistan, between Behbahan and Dehdasht, Tang-e Takab, 500 m, *Mozaffarian 58838* (TARI), SAMN15064210*; ***Klasea coriacea*** (DC.) Holub, Armenia, Ararat, *Susanna 1530 & al.* (BC), SAMN15064211*; ***Leuzea acaulis***[5] (L.) Holub, cultivated at the Barcelona Botanical Garden, SAMN15064212*; ***Olgaea petriprimi*** B.A.Sharipova, Tajikistan, Selandi, *Susanna 2539 & al.* (BC), SAMN15064213*; ***Onopordum nervosum*** Boiss., cultivated at the Dijon Botanical Garden, SAMN15064214*; ***Plectocephalus cachinalensis*** (Phil.) N.Garcia & Susanna, cultivated at the Barcelona Botanical Garden, SAMN15064215*; ***Psephellus mucroniferus*** (DC.) Wagenitz, Turkey, Niğde, *Susanna 2300 & al.* (BC), SAMN15064216*; ***Saussurea andryaloides*** (DC.) Sch.Bip. **1**, India, Himachal Pradesh, Jalori Pass, [unread.] (WU), SAMN15064217*; ***Saussurea andryaloides*** **2**, India, between Da and Hanle, Rupshu, Kashmir, 15,000 ft., *Walter Koelz 2280* (E00160052), SAMN15064218*; ***Saussurea bella*** Y.Ling, China, Qinghai 3660 m, 15 Aug 2002, *Liu LJQ852*, SAMN15064219*; ***Saussurea candolleana*** (DC.) Wall. ex Sch.Bip. **1**, India, Kashmir, Jilai, 9000 ft, *Duthie 13966* (WU), SAMN15064220*; ***Saussurea candolleana*** **2**, Nepal, east of Chalike Pahar, in guilles among Bush vegetation, Stainton, *Sykes & Williams 3758* (E00469165), SAMN15064221*; ***Saussurea controversa*** DC., Russia, Krasnoyarsk Krai, Sharypovsky, village Bolshoe ozero, *Cazzolla Gatti 10005 & al.* (TK t-01-2016), SAMN15064222*; ***Saussurea daurica***[6] Adams, Russia, Altai, Kosh-Agachsky, Kuraiskiy Ridge, village Chagan-Usun, *A. Pyak & E. Pyak 11049* (TK a-067-2016), SAMN15064223*; ***Saussurea discolor*** (Willd.) DC., Austria, Steiermark, 2130 m, 3 Jul 1997, *Schneeweiss s.n.* (WU), SAMN15064224*; ***Saussurea elegans*** Ledeb., Tadjikistan, Iskandar valley, Fan mountains, *Susanna 2505 & al.* (BC), SAMN15064225*; ***Saussurea foliosa*** Ledeb., Russia, Khakassia, Tashtypsky, Sayanskii Mountain Pass, *Cazzolla Gatti 10025 & al.* (TK t-30-2016), SAMN15064226*; ***Saussurea glacialis*** Herder, Russia, Altai, Kosh-Agachsky, Kuraiskiy Ridge, *A. Pyak & E. Pyak 11021* (TK a-043-2016), SAMN15064227*; ***Saussurea gnaphalodes*** (Royle ex DC.) Sch. Bip., China, Sichuan Prov., Khangding Xian, *D.E. Boufford 34778 & al.* (MBK0147538), SAMN15064228*; ***Saussurea gossipiphora*** D.Don, Nepal, Sankhuwasabha Dist., Bandhuke, *Y. Omori & al. 9920067* (TI), SAMN15064229*; ***Saussurea jadrincevi***[7] Kryl., Russia, Altai, Ongudaysky, the Mount Belyy Bom, *A. Pyak & E. Pyak 11005* (TK a-023-2016), SAMN15064230*; ***Saussurea krylovii*** Schischk. & Serg., Russia, Altai, Kosh-Agachsky, Juzhno-Chuysky Ridge, the Jazator River Valley, *A. Pyak & E. Pyak 11079* (TK a-108-2016), SAMN15064231*; ***Saussurea larionowii*** C.Winkl., Kyrgyzstan, s. loc., *Ovczinnikov 16* (DUSH), SAMN15064232*; ***Saussurea latifolia*** Ledeb., Russia, Krasnoyarsk Krai, Yermakovsky, Ergaki Ridge, *A. Pyak & E. Pyak 10009* (TK t-02-2016), SAMN15064233*; ***Saussurea lehmannii*** (Bunge) N.Garcia & al., Tajikistan, Khatlonskaja Oblast, Shahritusskii District, *Kamelin s.n.* (LE), SAMN15064234*; ***Saussurea leontodontoides*** (DC.) Sch.Bip., Nepal, between Ghunsa and Tamola, 3960 m, *Kew-Edinburgh Kathmandu expedition to NE Nepal 1989 KEKE 659* (E00466866), SAMN15064235*; ***Saussurea leucophylla*** Schrenk, Russia, Altai, Kosh-Agachsky, northern spurs of the Mount Tjepliy Kljuch, *A. Pyak & E. Pyak 11073* (TK a-102-2016), SAMN15064236*; ***Saussurea lyrata*** (Bunge) Sch.Bip. **1**, Republick of Korea, Jeollabuk-do, Jeonji-citu, Geonji-hill, *G.S. Kim & C.H. Oho 116* (WU021661), SAMN15064237*; ***Saussurea lyrata*** **2**, China, Chao-Chow-Fu, a city on the Han river, twenty-four miles north of Swatow, May 1899, *J.M. Dalziel s.n.* (E00467504), SAMN15064238*; ***Saussurea manshurica*** Kom., Russia, Amur province, 2 Aug 1979, *Boyko & Starchenko s.n.* (LE), SAMN15064239*; ***Saussurea nivea*** Turcz., China, Tsobili [?], 10 Aug 1930, *E. Licent s.n.* (W0075522), SAMN15064240*; ***Saussurea obvallata*** (DC.) Sch.Bip., China, Yunnan, 3320–4320 m, *Liu LJQ2621*, SAMN15064241*; ***Saussurea orgaadayi*** Khanm. & Krasnob., Russia, Altai, Kosh-Agachsky, Kuraiskiy Ridge, the Kokorja River Valley, *A. Pyak & E. Pyak 11083* (TK a-119-2016), SAMN15064242*; ***Saussurea picridifolia*** (Hand.-Mazz.) Y.S.Chen & Qian Yuan, China, Yunnan bor.-occid., in regione frigide temperata jugi Si-la inter fluvios Landsang-djiang (Mekong) et Lu-djiang (Salween), ca. 3800 m, *Handel-Mazzetti 9964* (W0002048), SAMN15064243*; ***Saussurea pseudoalpina*** N.D.Simpson, Russia, Altai, Kosh-Agachsky, Kuraiskiy Ridge, the Ortolyk River, *A. Pyak & E. Pyak 11032* (TK a-048-2016), SAMN15064244*; ***Saussurea pulchella*** (Fisch.) Colla **1**, Republic of Korea, Gangwon Province, Mt. Hambaek, *Yun 130926024* (SKK), SAMN15064245*; ***Saussurea pulchella*** **2**, Japan, Honshu Yamanashi Pref., Okishinhata, Yamanakako-mura, Mina-mitsuru-gun, 950 m, *M. Togashi 541* (WU), SAMN15064246*; ***Saussurea roylei*** (DC.) Sch.Bip., Nepal, Marsiandi Valley, 13,500 ft, *D.G. Lowndes L1179* (E00469184), SAMN15064247*; ***Saussurea salicifolia*** (L.) DC., Russia, Tyva, Kaa-Khemsky, the Mount Ondum, the Kaa-Khem River, *A. Pyak & E. Pyak 10014* (TK t-12-2016), SAMN15064248*; ***Saussurea salsa*** (Pall.) Spreng., Russia, Altai, Kosh-Agachsky, Chuya Steppe, village Aktal, *A. Pyak & E. Pyak 11087* (TK a-120-2016), SAMN11585485; ***Saussurea scabrida*** Franch., China, Sichuan Province, Xiangcheng Xian, between Xiangcheng and Sandui near the viallge of Riyin, S of Wuming Shan, 3550–4000 m, *D.E. Boufford & al. 28405* (E00280462), SAMN15064249*; ***Saussurea scaposa*** Franch. & Sav., Japan, Shikoku, Kochi Pref., Nanotani, Yusuhara Town, *S. Kobayashi FOK-074378* (MBK0129268), SAMN15064250*; ***Saussurea schanginiana*** (Wydler) Fisch. ex Herd., Russia, Khakassia, Tashtypsky, Sayanskii Ridge, Sayanskii Mountain Pass, *A. Pyak & E. Pyak 10057* (TK t-24-2016), SAMN15064251*; ***Saussurea simpsoniana*** (DC.) Wall. ex Sch.Bip., Nepal, Mustang Dist., around Yak Kharka, *K. Fujikawa 9920116* (MBK0254898), SAMN15064252*; ***Saussurea sorocephala*** (Schrenk) Schrenk, Pakistan, Ghareda glacier, 16,100 ft, *Polunin 6305* (E00710753), SAMN15064253*; ***Saussurea stubendorffii*** Herder, Russia, Tyva, Barun-Khemchiksky, Sayanskii Ridge, Ak-sug River Valley, *A. Pyak & E. Pyak 10057* (TK t-24-2016), SAMN15064254*; ***Saussurea subacaulis*** (Ledeb.) Serg., Russia, Altai, Kosh-Agachsky, Kuraiskiy Ridge, Ortolyk River, *A. Pyak & E. Pyak 11026* (TK a-046-2016), SAMN15064255*; ***Saussurea taraxacifolia*** (Lindl. ex Royle) Wall. ex DC., China, Yushu Xian, SW of Machang, in a side valley on S side the Baitang He

basin, 4000 m, *T.N. Ho & al. 2133* (E00064934), SAMN15064256*; ***Saussurea uniflora*** (DC.) Wall. ex Sch.Bip., Nepal, Solukhumbu Dist., Tashing Dingma–Thuli Kharka, *M. Wakabayashi & al. 9730355* (MBK0254886), SAMN15064257*; ***Saussurea winkleri*** (Iljin) N.Garcia & al., Tajikistan, Khatlonskaja Oblast, Shahritusskii District, 600 m, *Botchancev s.n. & al.* (LE), SAMN15064258*; ***Staehelina lobelii*** DC., Turkey, Antalya, *Susanna 2272 & al.* (BC), SAMN15064259*; ***Staehelina petiolata*** (L.) Hilliard & B.L.Burtt, Greece, Crete, *Vitek 081418* (BC), SAMN15064260*; ***Tricholepis edmondsonii*** Rech.f., Iran, Hormozgan, 5 km from Bandar-Abbas to Fareghan village, 1100 m, *Mozaffarian 44869* (TARI), SAMN15064261*; ***Volutaria canariensis*** Wagenitz, Spain, Gran Canaria, *Kunkel 200-2* (BC), SAMN15064262*.

11. Supplementary material

Table S1. Overview of previous phylogenetic studies and present study performed on the subtribe Saussureinae with extensive generic sampling.

Genus (no. of species)	Garcia-Jacas & al., 2002	Raab-Straube, 2003	Kita & al., 2004	Susanna & al., 2006	Wang & al., 2007	Wang & al., 2009	Yuan & al., 2015	Fu & al., 2016	Herrando-Moraira & al., 2018	Herrando-Moraira & al., 2019	Szukala & al., 2019	Xu & al., 2019	This study
Aegopordon (2)				1					1	1	2		2
Aucklandia (1)					1	1	1	1				1	1
Bolocephalus (1)												1	1
Diplazoptilon (2)*					1	1	1						2
Dolomiaea (17)				1	5	1	1	1			1	1	9
Frolovia (8)		1		2	1	1	1	1			2		5
Hemisteptia (1)			1		1	1	1	1			1	1	1
Hyalochaete (1)	1		1	1									1
Himalaiella (16)		1	4	1	1	1	1				3	1	12
Jurinea (200)	4	3	3	8	1	1	1	3	23	1	80		32
Jurinella (9)	1		1	1							4		4
Lipschitziella (2)		1	1	2	1	1			1	1	1		2
*Modestia*** (3)				1					1				1
Outreya (1)	1		1	1					1		1		1
Pilostemon (2)										1			1
Polytaxis (3)				2	1	1	2				1		2
Saussurea (400)	2	42	23	4	1	50	18	5	21	3	4	136	35
Generic coverage (%)	29	29	47	71	59	59	53	35	35	29	65	35	100

*Considering *Diplazoptilon picridifolium* and *D. cooperi*, but *D. cooperi* is considered a synonym of *Himalaiella yakla* (Shi & Raab-Straube, 2011; Yuan & al., 2015).
**Synonymized as *Anacantha* (*Modestia darwasica*, *M. jucunda*, *M. mira*). *Modestia pteroclada* is considered a synonym of *Jurinea jucunda* (Sennikov & Lazkov, 2013).

Fu, Z.X., Jiao, B.H., Nie, B., Zhang, G.J., Gao, T.G. & China Phylogeny Consortium 2016. A comprehensive generic-level phylogeny of the sunflower family: Implications for the systematics of Chinese Asteraceae. *J. Syst. Evol.* 54: 416–437. https://doi.org/10.1111/jse.12216

Garcia-Jacas, N., Garnatje, T., Susanna, A. & Vilatersana, R. 2002. Tribal and subtribal delimitation and phylogeny of the Cardueae (Asteraceae): A combined nuclear and chloroplast DNA analysis. *Molec. Phylogen. Evol.* 22: 51–64. https://doi.org/10.1006/mpev.2001.1038

Herrando-Moraira, S., The Cardueae Radiations Group (Calleja, J.A., Carnicero-Campmany, P., Fujikawa, K., Galbany-Casals, M., Garcia-Jacas, N., Im, H.-T., Kim, S.-C., Liu, J.-Q., López-Alvarado, J., López-Pujol, J., Mandel, J.R., Massó, S.,

Mehregan, I., Montes-Moreno, N., Pyak, E., Roquet, C., Sáez, L., Sennikov, A., Susanna, A. & Vilatersana, R.) 2018. Exploring data processing strategies in NGS target enrichment to disentangle radiations in the tribe Cardueae (Compositae). *Molec. Phylogen. Evol.* 128: 69–87. https://doi.org/10.1016/j.ympev.2018.07.012

Herrando-Moraira, S., The Cardueae Radiations Group (Calleja, J.A., Galbany-Casals, M., Garcia-Jacas, N., Liu, J.-Q., López-Alvarado, J., López-Pujol, J., Mandel, J.R., Massó, S., Montes-Moreno, N., Roquet, C., Sáez, L., Sennikov, A., Susanna, A. & Vilatersana, R.) 2019. Nuclear and plastid DNA phylogeny of tribe Cardueae (Compositae) with Hyb-Seq data: A new subtribal classification and a temporal diversification framework. *Molec. Phylogen. Evol.* 137: 313–332. https://doi.org/10.1016/j.ympev.2019.05.001

Kita, Y., Fujikawa, K., Ito, M., Ohba, H. & Kato, M. 2004. Molecular phylogenetic analyses and systematics of the genus *Saussurea* and related genera (Asteraceae, Cardueae). *Taxon* 53: 679–690. https://doi.org/10.2307/4135443

Raab-Straube, E. von 2003. Phylogenetic relationships in *Saussurea* (Compositae, Cardueae) sensu lato, inferred from morphological, ITS and *trnL-trnF* sequence data, with a synopsis of *Himalaiella* gen. nov., *Lipschitziella* and *Frolovia*. *Willdenowia* 33: 379–402. https://doi.org/10.3372/wi.33.33214

Sennikov, A.N. & Lazkov, G.A. 2013. Taxonomic corrections and new records in vascular plants of Kyrgyzstan, 2. *Memoranda Soc. Fauna Fl. Fenn.* 89: 125–138.

Shi, Z. & Raab-Straube, E. von 2011. *Saussurea* group. Pp. 42–149 in: Wu, Z.Y., Raven, P.H. & Hong, D.Y. (eds.), *Flora of China*, vol. 20–21 (Asteraceae). Beijing: Science Press; St Louis: Missouri Botanical Garden Press.

Susanna, A., Garcia-Jacas, N., Hidalgo, O., Vilatersana, R. & Garnatje, T. 2006. The Cardueae (Compositae) revisited: Insights from ITS, *trnL-trnF*, and *matK* nuclear and chloroplast DNA analysis. *Ann. Missouri Bot. Gard.* 93: 150–171. https://doi.org/10.3417/0026-6493(2006)93[150:TCCRIF]2.0.CO;2

Szukala, A., Korotkova, N., Gruenstaeudl, M., Sennikov, A.N., Lazkov, G.A., Litvinskaya, S.A., Gabrielian, E., Borsch, T. & Raab-Straube, E. von 2019. Phylogeny of the Eurasian genus *Jurinea* (Asteraceae: Cardueae): Support for a monophyletic genus concept and a first hypothesis on overall species relationships. *Taxon* 68: 112–131. https://doi.org/10.1002/tax.12027

Wang, Y.-J., Liu, J.-Q. & Miehe, G. 2007. Phylogenetic origins of the Himalayan endemic *Dolomiaea*, *Diplazoptilon* and *Xanthopappus* (Asteraceae: Cardueae) based on three DNA regions. *Ann. Bot. (Oxford)* 99: 311–322. https://doi.org/10.1093/aob/mcl259

Wang, Y.-J., Susanna, A., Raab-Straube, E. von, Milne, R. & Liu, J.-Q. 2009. Island-like radiation of *Saussurea* (Asteraceae: Cardueae) triggered by uplifts of the Qinghai–Tibetan Plateau. *Biol. J. Linn. Soc.* 97: 893–903. https://doi.org/10.1111/j.1095-8312.2009.01225.x

Xu, L.-S., Herrando-Moraira, S., Susanna, A., Galbany-Casals, M. & Chen, Y.-S. 2019. Phylogeny, origin and dispersal of *Saussurea* (Asteraceae) based on chloroplast genome data. *Molec. Phylogen. Evol.* 141: 106613. https://doi.org/10.1016/j.ympev.2019.106613

Yuan, Q., Bi, Y.-C. & Chen, Y.-S. 2015. *Diplazoptilon* (Asteraceae) is merged with *Saussurea* based on evidence from morphology and molecular systematics. *Phytotaxa* 236: 53–61. https://doi.org/10.11646/phytotaxa.236.1.4

Table S2. Species sampled and their corresponding number of raw reads, number of loci extracted with HybPiper (the total were 588), sequencing project, and NCBI accession number. Taxon names follow the final proposed taxonomic treatment presented in the study. New combinations are in bold. Notes: [1] In Herrando-Moraira & al. (2018) it appears as *Arctium eriophorum*; [2] In Herrando-Moraira & al. (2018) it appears as *Cousinia ninae*; [3] In Herrando-Moraira & al. (2018) it appears as *Jurinea abramowii*; [4] In Herrando-Moraira & al. (2018) it appears as *Jurinea bucarica*; [5] In Herrando-Moraira & al. (2019) it appears as *Rhaponticum acaule*; [6] In Herrando-Moraira & al. (2018) it appears as *Saussurea davurica*; [7] In Herrando-Moraira & al. (2018) it appears as *Saussurea jadrinzevii*.

Species	Nº raw reads	Nº loci extracted	Sequencing project	NCBI BioSample accession numbers
Arctium fedtschenkoanum	3,550,984	561	Herrando-Moraira & al. (2018), Mandel & al. (2019)	SAMN11585472
Arctium grandifolium	2,300,488	556	Herrando-Moraira & al. (2018)	SAMN15064103
Arctium karatavicum	2,976,529	561	Herrando-Moraira & al. (2018)	SAMN15064104
Arctium minus	10,007,019	558	Herrando-Moraira & al. (2018)	SAMN15064105
Arctium nidulans[1]	2,984,349	557	Herrando-Moraira & al. (2018)	SAMN15064106
Arctium umbrosum	4,663,613	560	Herrando-Moraira & al. (2018)	SAMN15064107
Carduus nutans	3,891,005	528	Herrando-Moraira & al. (2019)	SAMN15064108
Centaurea aspera	3,215,383	527	Herrando-Moraira & al. (2019)	SAMN15064109
Cirsium sairamense	5,389,901	544	Herrando-Moraira & al. (2018) Mandel & al. (2019)	SAMN11585477

Species	Nº raw reads	Nº loci extracted	Sequencing project	NCBI BioSample accession numbers
Cousinia arachnoidea[2]	3,280,858	552	Herrando-Moraira & al. (2018)	SAMN15064110
Cousinia knorringiae	3,129,866	557	Herrando-Moraira & al. (2018)	SAMN15064111
Cousinia onopordioides	2,369,328	554	Herrando-Moraira & al. (2018)	SAMN15064112
Cousinia polytimetica	2,905,323	556	Herrando-Moraira & al. (2018)	SAMN15064113
Cousinia pungens	13,672,179	560	Present study	SAMN15064114
Cousinia splendida	3,595,227	561	Herrando-Moraira & al. (2018)	SAMN15064115
Cousinia spryginii	3,305,209	557	Herrando-Moraira & al. (2018), Jones & al. (2019)	SAMN10983402
Cynara cardunculus	454,885	457	Mandel & al. (2014, 2019), Herrando-Moraira & al. (2018)	SAMN11585480
Dolomiaea asbukinii	5,880,400	549	Present study	SAMN15064116
Dolomiaea baltalensis	42,600,816	557	Present study	SAMN15064117
Dolomiaea berardioidea 1	6,613,635	555	Present study	SAMN15064118
Dolomiaea berardioidea 2	6,613,635	548	Present study	SAMN15064119
Dolomiaea costus	3,494,166	552	Present study	SAMN15064120
Dolomiaea edulis	6,753,360	559	Present study	SAMN15064121
Dolomiaea forrestii 1	9,965,954	559	Present study	SAMN15064122
Dolomiaea forrestii 2	10,642,455	555	Present study	SAMN15064123
Dolomiaea forrestii 3	7,744,898	557	Present study	SAMN15064124
Dolomiaea frolowii	5,353,925	562	Present study	SAMN15064125
Dolomiaea georgii 1	12,272,881	557	Present study	SAMN15064126
Dolomiaea georgii 2	8,697,941	554	Present study	SAMN15064127
Dolomiaea gorbunovae	3,365,645	540	Present study	SAMN15064128
Dolomiaea macrocephala	13,913,282	554	Present study	SAMN15064129
Dolomiaea platylepis	10,077,877	555	Present study	SAMN15064130
Dolomiaea salwinensis 1	6,830,748	558	Present study	SAMN15064131
Dolomiaea salwinensis 2	10,219,376	562	Present study	SAMN15064132
Dolomiaea salwinensis 3	1,681,110	543	Present study	SAMN15064133
Dolomiaea saussureoides	8,318,886	550	Present study	SAMN15064134
Dolomiaea souliei	6,823,494	559	Present study	SAMN15064135
Dolomiaea sulcata	9,931,761	552	Present study	SAMN15064136
Galactites tomentosa	3,680,490	516	Herrando-Moraira & al. (2019)	SAMN15064137
Jurinea abramovii[3]	4,803,672	548	Herrando-Moraira & al. (2018)	SAMN15064138
Jurinea alata	5,069,639	558	Herrando-Moraira & al. (2018), Mandel & al. (2019)	SAMN11585483
***Jurinea albescens* 1**	6,185,365	552	Present study	SAMN15064139
***Jurinea albescens* 2**	11,082,089	552	Present study	SAMN15064140
Jurinea algida	3,743,171	548	Herrando-Moraira & al. (2018)	SAMN15064141
Jurinea atropurpurea	4,316,866	556	Herrando-Moraira & al. (2018), Mandel & al. (2019)	SAMN11585484
***Jurinea auriculata* 1**	6,172,402	549	Present study	SAMN15064142

Species	Nº raw reads	Nº loci extracted	Sequencing project	NCBI BioSample accession numbers
***Jurinea auriculata* 2**	5,081,700	548	Present study	SAMN15064143
Jurinea baldschuanica	5,113,980	555	Herrando-Moraira & al. (2018)	SAMN15064144
Jurinea berardioides 1	9,192,788	547	Present study	SAMN15064145
Jurinea berardioides 2	5,074,375	543	Present study	SAMN15064146
Jurinea bucharica[4]	4,255,435	544	Herrando-Moraira & al. (2018)	SAMN15064147
Jurinea bungei	5,593,772	558	Present study	SAMN15064148
Jurinea caespitans	4,407,313	552	Herrando-Moraira & al. (2018)	SAMN15064149
Jurinea capusii	4,726,638	550	Herrando-Moraira & al. (2018)	SAMN15064153
Jurinea carduicephala	7,948,211	565	Present study	SAMN15064154
Jurinea carduiformis	5,200,789	550	Herrando-Moraira & al. (2018)	SAMN15064155
Jurinea ceratocarpa 1	6,897,455	554	Present study	SAMN15064156
Jurinea ceratocarpa 2	5,504,601	555	Present study	SAMN15064157
Jurinea chenopodiifolia	12,228,813	553	Present study	SAMN15064158
***Jurinea chitralica* 1**	3,957,539	491	Present study	SAMN15064159
***Jurinea chitralica* 2**	3,883,146	544	Present study	SAMN15064160
Jurinea chondrilloides 1	3,384,274	545	Present study	SAMN15064161
Jurinea chondrilloides 2	13,741,521	556	Present study	SAMN15064162
***Jurinea crispa* 1**	6,413,883	554	Present study	SAMN15064163
***Jurinea crispa* 2**	5,185,407	545	Present study	SAMN15064164
***Jurinea crispa* 3**	9,236,288	547	Present study	SAMN15064198
***Jurinea deltoidea* 1**	6,719,836	499	Present study	SAMN15064165
***Jurinea deltoidea* 2**	6,854,780	551	Present study	SAMN15064166
Jurinea ferganica	5,170,117	552	Herrando-Moraira & al. (2018)	SAMN15064167
Jurinea filicifolia	7,046,917	556	Present study	SAMN15064168
Jurinea fontqueri	5,240,423	552	Herrando-Moraira & al. (2018)	SAMN15064169
Jurinea frigida 1	17,424,490	563	Present study	SAMN15064170
Jurinea frigida 2	10,174,994	555	Present study	SAMN15064171
Jurinea frigida 3	16,431,602	561	Present study	SAMN15064172
Jurinea gilesii	8,514,559	558	Present study	SAMN15064173
***Jurinea heteromalla* 1**	2,007,573	545	Present study	SAMN15064150
***Jurinea heteromalla* 2**	7,869,423	524	Present study	SAMN15064151
***Jurinea heteromalla* 3**	14,178,130	548	Present study	SAMN15064152
***Jurinea heteromalla* 4**	3,342,983	554	Present study	SAMN15064174
Jurinea jucunda	5,083,617	548	Herrando-Moraira & al. (2018)	SAMN15064175
Jurinea karategina	7,713,700	553	Present study	SAMN15064176
Jurinea kokanica	4,531,178	549	Herrando-Moraira & al. (2018)	SAMN15064177
Jurinea kyzylkyrensis	5,561,006	550	Herrando-Moraira & al. (2018)	SAMN15064178
Jurinea lanipes	4,601,775	548	Herrando-Moraira & al. (2018)	SAMN15064179
Jurinea leptoloba	5,487,798	554	Herrando-Moraira & al. (2018)	SAMN15064180

Species	Nº raw reads	Nº loci extracted	Sequencing project	NCBI BioSample accession numbers
Jurinea leptophylla	6,055,256	562	Herrando-Moraira & al. (2018)	SAMN15064181
Jurinea macrocephala	4,093,061	542	Herrando-Moraira & al. (2018)	SAMN15064182
Jurinea microcephala	9,602,332	552	Present study	SAMN15064183
Jurinea modesta	7,983,429	552	Present study	SAMN15064184
Jurinea moschus 1	14,116,368	557	Present study	SAMN15064185
Jurinea moschus 2	7,778,198	553	Present study	SAMN15064186
Jurinea moschus 3	12,327,877	564	Present study	SAMN15064187
Jurinea moschus 4	12,414,230	558	Present study	SAMN15064188
Jurinea narynensis	4,564,064	544	Herrando-Moraira & al. (2018)	SAMN15064189
Jurinea natmataungensis	9,238,450	546	Present study	SAMN15064190
Jurinea olgae	4,941,133	549	Herrando-Moraira & al. (2018)	SAMN15064191
Jurinea orientalis	3,155,790	546	Herrando-Moraira & al. (2018)	SAMN15064192
***Jurinea peguensis* 1**	4,024,000	537	Present study	SAMN15064193
***Jurinea peguensis* 2**	4,555,896	544	Present study	SAMN15064194
Jurinea pinnata	2,996,426	553	Herrando-Moraira & al. (2018)	SAMN15064195
Jurinea popovii	3,304,462	554	Herrando-Moraira & al. (2018)	SAMN15064196
Jurinea pulchella	15,259,336	555	Present study	SAMN15064197
Jurinea schachimardanica	3,568,519	550	Herrando-Moraira & al. (2018)	SAMN15064199
Jurinea squarrosa 1	11,922,309	561	Present study	SAMN15064200
Jurinea squarrosa 2	3,734,717	544	Present study	SAMN15064201
Jurinea stenophylla	3,240,161	556	Herrando-Moraira & al. (2018)	SAMN15064202
Jurinea stoechadifolia	4,403,856	560	Herrando-Moraira & al. (2018)	SAMN15064203
Jurinea suffruticosa	2,658,663	554	Herrando-Moraira & al. (2018)	SAMN15064204
Jurinea thianschanica	3,111,323	550	Present study	SAMN15064205
Jurinea trautvetteriana	2,087,532	549	Herrando-Moraira & al. (2018)	SAMN15064206
Jurinea winkleri	2,061,087	543	Present study	SAMN15064207
Jurinea xeranthemoides	2,333,426	549	Present study	SAMN15064208
Jurinea yakla	4,325,189	549	Present study	SAMN15064209
Karvandarina cartilaginea	8,454,714	540	Present study	SAMN15064210
Klasea coriacea	6,205,566	537	Herrando-Moraira & al. (2019)	SAMN15064211
Leuzea acaulis[5]	4,209,726	548	Herrando-Moraira & al. (2019)	SAMN15064212
Olgaea petriprimi	5,310,933	549	Herrando-Moraira & al. (2018)	SAMN15064213
Onopordum nervosum	2,796,798	544	Herrando-Moraira & al. (2019)	SAMN15064214
Plectocephalus cachinalensis	5,898,634	548	Herrando-Moraira & al. (2019)	SAMN15064215
Psephellus mucroniferus	2,573,295	540	Herrando-Moraira & al. (2019)	SAMN15064216
Saussurea andryaloides 1	3,088,154	497	Present study	SAMN15064217
Saussurea andryaloides 2	2,463,229	547	Present study	SAMN15064218
Saussurea bella	5,899,937	546	Present study	SAMN15064219
Saussurea candolleana 1	4,599,912	547	Present study	SAMN15064220

Species	Nº raw reads	Nº loci extracted	Sequencing project	NCBI BioSample accession numbers
Saussurea candolleana 2	4,504,851	550	Present study	SAMN15064221
Saussurea controversa	8,091,449	561	Herrando-Moraira & al. (2018)	SAMN15064222
Saussurea daurica[6]	11,202,023	553	Present study	SAMN15064223
Saussurea discolor	16,539,934	541	Present study	SAMN15064224
Saussurea elegans	2,784,084	553	Present study	SAMN15064225
Saussurea foliosa	4,089,960	560	Herrando-Moraira & al. (2018)	SAMN15064226
Saussurea glacialis	4,072,633	554	Herrando-Moraira & al. (2018)	SAMN15064227
Saussurea gnaphalodes	6,179,290	550	Present study	SAMN15064228
Saussurea gossipiphora	5,563,927	551	Present study	SAMN15064229
Saussurea jadrincevi[7]	9,091,105	562	Herrando-Moraira & al. (2018)	SAMN15064230
Saussurea krylovii	3,576,809	557	Herrando-Moraira & al. (2018)	SAMN15064231
Saussurea larionowii	4,733,404	554	Herrando-Moraira & al. (2018)	SAMN15064232
Saussurea latifolia	5,065,459	556	Herrando-Moraira & al. (2018)	SAMN15064233
Saussurea lehmannii	7,854,228	541	Present study	SAMN15064234
Saussurea leontodontoides	7,209,427	560	Present study	SAMN15064235
Saussurea leucophylla	5,597,695	562	Herrando-Moraira & al. (2018)	SAMN15064236
***Saussurea lyrata* 1**	7,057,764	545	Present study	SAMN15064237
***Saussurea lyrata* 2**	3,919,740	539	Present study	SAMN15064238
Saussurea manshurica	4,417,126	559	Herrando-Moraira & al. (2018)	SAMN15064239
Saussurea nivea	7,048,980	557	Present study	SAMN15064240
Saussurea obvallata	12,608,195	558	Present study	SAMN15064241
Saussurea orgaadayi	3,578,510	554	Herrando-Moraira & al. (2018)	SAMN15064242
Saussurea picridifolia	8,055,160	558	Present study	SAMN15064243
Saussurea pseudoalpina	3,887,786	559	Herrando-Moraira & al. (2018)	SAMN15064244
Saussurea pulchella 1	11,166,564	542	Present study	SAMN15064245
Saussurea pulchella 2	16,171,554	528	Present study	SAMN15064246
Saussurea roylei	4,576,394	562	Present study	SAMN15064247
Saussurea salicifolia	4,799,838	554	Herrando-Moraira & al. (2018)	SAMN15064248
Saussurea salsa	2,458,299	552	Herrando-Moraira & al. (2018), Mandel & al. (2019)	SAMN11585485
Saussurea scabrida	4,100,343	555	Present study	SAMN15064249
Saussurea scaposa	6,065,931	553	Present study	SAMN15064250
Saussurea schanginiana	4,568,611	559	Herrando-Moraira & al. (2018)	SAMN15064251
Saussurea simpsoniana	5,378,069	548	Present study	SAMN15064252
Saussurea sorocephala	2,834,853	550	Present study	SAMN15064253
Saussurea stubendorffii	5,329,546	559	Herrando-Moraira & al. (2018)	SAMN15064254
Saussurea subacaulis	8,252,488	559	Herrando-Moraira & al. (2018)	SAMN15064255
Saussurea taraxacifolia	2,743,424	554	Present study	SAMN15064256
Saussurea uniflora	4,097,420	544	Present study	SAMN15064257

Species	Nº raw reads	Nº loci extracted	Sequencing project	NCBI BioSample accession numbers
Saussurea winkleri	5,243,142	540	Present study	SAMN15064258
Staehelina lobelii	6,287,680	550	Herrando-Moraira & al. (2019)	SAMN15064259
Staehelina petiolata	5,695,551	544	Herrando-Moraira & al. (2019)	SAMN15064260
Tricholepis edmondsonii	9,369,430	554	Present study	SAMN15064261
Volutaria canariensis	4,551,867	529	Herrando-Moraira & al. (2019)	SAMN15064262

Herrando-Moraira, S., The Cardueae Radiations Group (Calleja, J.A., Carnicero-Campmany, P., Fujikawa, K., Galbany-Casals, M., Garcia-Jacas, N., Im, H.-T., Kim, S.-C., Liu, J.-Q., López-Alvarado, J., López-Pujol, J., Mandel, J.R., Massó, S., Mehregan, I., Montes-Moreno, N., Pyak, E., Roquet, C., Sáez, L., Sennikov, A., Susanna, A. & Vilatersana, R.) 2018. Exploring data processing strategies in NGS target enrichment to disentangle radiations in the tribe Cardueae (Compositae). *Molec. Phylogen. Evol.* 128: 69–87. https://doi.org/10.1016/j.ympev.2018.07.012

Herrando-Moraira, S., The Cardueae Radiations Group (Calleja, J.A., Galbany-Casals, M., Garcia-Jacas, N., Liu, J.-Q., López-Alvarado, J., López-Pujol, J., Mandel, J.R., Massó, S., Montes-Moreno, N., Roquet, C., Sáez, L., Sennikov, A., Susanna, A. & Vilatersana, R.) 2019. Nuclear and plastid DNA phylogeny of tribe Cardueae (Compositae) with Hyb-Seq data: A new subtribal classification and a temporal diversification framework. *Molec. Phylogen. Evol.* 137: 313–332. https://doi.org/10.1016/j.ympev.2019.05.001

Jones, K.E., Fér, T., Schmickl, R.E., Dikow, R.B., Funk, V.A., Herrando-Moraira, S., Johnston, P.R., Kilian, N., Siniscalchi, C.M., Susanna, A., Slovák, M., Thapa, R., Watson, L.E. & Mandel, J.R. 2019. An empirical assessment of a single family-wide hybrid capture locus set at multiple evolutionary timescales in Asteraceae. *Applic. Pl. Sci.* 7: e11295. https://doi.org/10.1002/aps3.11295

Mandel, J.R., Dikow, R.B., Funk, V.A., Masalia, R.R., Staton, S.E., Kozik, A., Michelmore, R.W., Rieseberg, L.H. & Burke, J.M. 2014. A target enrichment method for gathering phylogenetic information from hundreds of loci: An example from the Compositae. *Applic. Pl. Sci.* 2: 1300085. https://doi.org/10.3732/apps.1300085

Mandel, J.R., Dikow, R.B., Siniscalchi, C.M., Thapa, R., Watson, L.E. & Funk, V.A. 2019. A fully resolved backbone phylogeny reveals numerous dispersals and explosive diversifications throughout the history of Asteraceae. *Proc. Natl. Acad. Sci. U.S.A.* 116: 14083–14088. https://doi.org/10.1073/pnas.1903871116

Chapter 4

Climate Stability Index maps, a global high resolution cartography of climate stability from Pliocene to 2100

Sonia Herrando-Moraira[1], Neus Nualart[1], Mercè Galbany-Casals[2], Núria Garcia-Jacas[1], Haruka Ohashi[3], Tetsuya Matsui[3,4], Alfonso Susanna[1], Cindy Q. Tang[5], Jordi López-Pujol[1]

[1] Botanic Institute of Barcelona (IBB, CSIC-Ajuntament de Barcelona), Pg. del Migdia, s.n., 08038 Barcelona, Spain.
[2] Systematics and Evolution of Vascular Plants (UAB) – Associated Unit to CSIC, Departament de Biologia Animal, Biologia Vegetal i Ecologia, Facultat de Biociències, Universitat Autònoma de Barcelona, 08193 Bellaterra, Spain.
[3] Forestry and Forest Products Research Institute, Forest Research and Management Organization, Matsunosato 1, Tsukuba-shi, Ibaraki-ken, 305-8687, Japan.
[4] Faculty of Life and Environmental Sciences, University of Tsukuba, Tennodai 1-1-1, Tsukuba, Ibaraki 305-8572, Japan.
[5] Institute of Ecology and Geobotany, College of Ecology and Environmental Science, Yunnan University, Dongwaihuan South Road, University Town, Chenggong New District, Kunming, Yunnan, 650504, China

Address for correspondence: Sonia Herrando-Moraira, *sonia.herrando@gmail.com*

Submitted to journal: Scientific data
Impact factor year 2019: 5.541
Date of submission: 1 June 2021

Abstract

Climate changes are top biodiversity shapers, both during the past and future. Mapping the most climatic stable and unstable zones on Earth could improve our understanding of biodiversity distribution and evolution. Here, we present a set of maps based on a global scale, high resolution (ca. 5 km) new Climate Stability Index (CSI). The CSI considers bioclimatic variables for two different time ranges: (1) from Pliocene (3.3 Ma) to the present (CSI-past map set), using 12 time periods of *PaleoClim* representing warm and cold cycles; and (2) from present to the year 2100 (CSI-future), using nine general circulation models of climate change of four periods available from *WorldClim*. We calculated standard deviation of the variables and selected an uncorrelated set for summing, normalizing and obtaining the CSI maps. Our approach is useful for fields such as biogeography, earth sciences, agriculture, or sociology. However, CSI is an index that can be re-calculated according to particular criteria and objectives (e.g. temperature variables); maps are, therefore, customizable to every user.

Index

1. Background & Summary

Long-term climatic variation has had an enormous impact on the evolution of biodiversity, including humans. Climatic instability linked to the Pleistocene glacial/interglacial cycles has caused active diversification[1,2,3]. In contrast, regions with relatively stable climates have often acted either as "museums" (places of persistence)[4,5] or both as "museums" and "cradles"[6,7]. In addition, recent global warming has caused even deeper changes in biodiversity in a very short space of time. Increases of temperature in just a few decades might produce large regional species turnover by increasing extinction rates and large migrations[8], and also by loss of phylogenetic diversity (= evolutionary potential)[9].

Identifying areas with high climatic stability is of enormous interest. Delimiting past trends of stability and mapping them can help to shed light into evolutionary processes that shaped current biota, including human evolution. They may also be helpful in designing new protected areas (PAs) and prioritizing or redesigning the existing ones[10], following the demand of incorporating evolutionary processes into conservation planning[11]. Mapping climatic stability for the future can be even more important, as climate patterns are predicted to change dramatically during this century, including the occurrence of extreme weather events[8,12]. Such mapping may have, thus, implications on economy and human health; e.g., when selecting new arable lands or new human settlements that will be needed as a consequence of the predicted large-scale migrations[8].

Attempts to cartography climate stability are not new. The only ones that offer data at a global scale are *StableClim*[13] and the "map of climate stability"[14] (hereinafter "MCS"), which are projections of climate variation, the former from 21,000 BP (LGM or Last Glacial Maximum) to 2100 CE[13], and the latter from 21,000 BP to 100 BP[14]. *StableClim* offers estimates of climate stability on an ideal time-scale (every 100 years) for surveying the eco-evolutionary impacts of short-term climate shifts[13], but it shares with MCS three limitations[14]: (1) coverage of only ca. 21,000 years, which hinders their utility for explaining the role of climate stability in evolution and speciation that take thousands or millions of years; (2) spatial resolution is insufficient (2.5º grid, ca. 278 km at the equator), which does not allow one to correlate climatic stability with population data on both local and regional scales; and (3) they use one or two variables (mean monthly temperatures[13] or mean monthly temperatures and precipitations[14]) that may not be the key ones for a given case of study; for example, distribution patterns of bees in South Africa are highly dependent on seasonality and not on mean rainfall values[15].

Thanks to the recent publication of *PaleoClim*[16], a free-access database of paleoclimate layers at 2.5 arc-min (~5 km) grid resolution that includes data much older than the LGM, we are presenting a new set of maps that are based on a new Climatic Stability Index (CSI). Given the markedly different speed of inferred climatic changes (see above), we divide the sets of maps into two: CSI-past and CSI-future. CSI-past is based on the 12 time periods (listed in Fig. 1A) included in *PaleoClim*[16], which span from 3.3 Myr (Pliocene) to present. CSI-future uses the average values of nine general circulation models (GCMs) along four time intervals until 2100 (Fig. 1B), available in *WorldClim*[17]. CSI-future offers maps of climate stability for four Shared Socioeconomic Pathways (SSP[18]; Fig. 1B) that will be used to produce the IPCC Sixth Assessment Report on climate change. Variables for CSI calculation (Table 1) include annual averages (e.g. mean annual temperature), extremes (e.g. minimum temperature of coldest month), and seasonality (e.g. annual range in temperatures). CSI-past takes into account 14 bioclimatic variables, while CSI-future uses five more.

In spite of the limitation of using only 12 time periods as a basis for inferring the long-term climatic stability (with only three older to 0.12 Myr), the much wider time interval, the variety of measures (means, peaks, and intra-year variability) and the finer resolution (ca. 5 km) makes our proposal a very versatile tool with a wide array of applications. With both CSI-past and CSI-future we offer maps of climatic stability adhering to FAIR principles[19], adaptable to every user and circumstance: they are very easy to use and the variables to estimate the climate stability can be selected at the user's discretion.

2. Methods

A workflow for the calculation of CSI is presented in Fig. 1C. For all the analyses, we used the R v. 4.0.3 software environment[20] implemented in RStudio v. 1.4.1103. The scripts used for each methodological step are available in Supplementary Data X. After data download from primary sources (*PaleoClim* and *WorldClim*), specifically for the CSI-future map set we performed an initial step aimed to obtain individual bioclimatic variables for each future time period for the four SSPs (Fig. 1B). To achieve this, the median values of nine GCMs were calculated in functions compiled in raster R package[21] for each individual bioclimatic variable (see a few exceptions of number of GCMs used in Table 2).

The standard deviation (SD) was estimated as a measure of the amount of variation or dispersion along time series, from which the resulting output maps showed the places where climate conditions remained constant or variable across the temporal periods considered (Figs. 1A and 1B). The SD, as a way to identify stable/unstable climatic areas, was previously used in other climatic or evolutionary studies[4,14]. To compute the SD output rasters, we applied the mosaic function setting "fun = sd" from raster R package, calculating the SD for each pixel in the 12 time period rasters for CSI-past and 5 times for CSI-future,

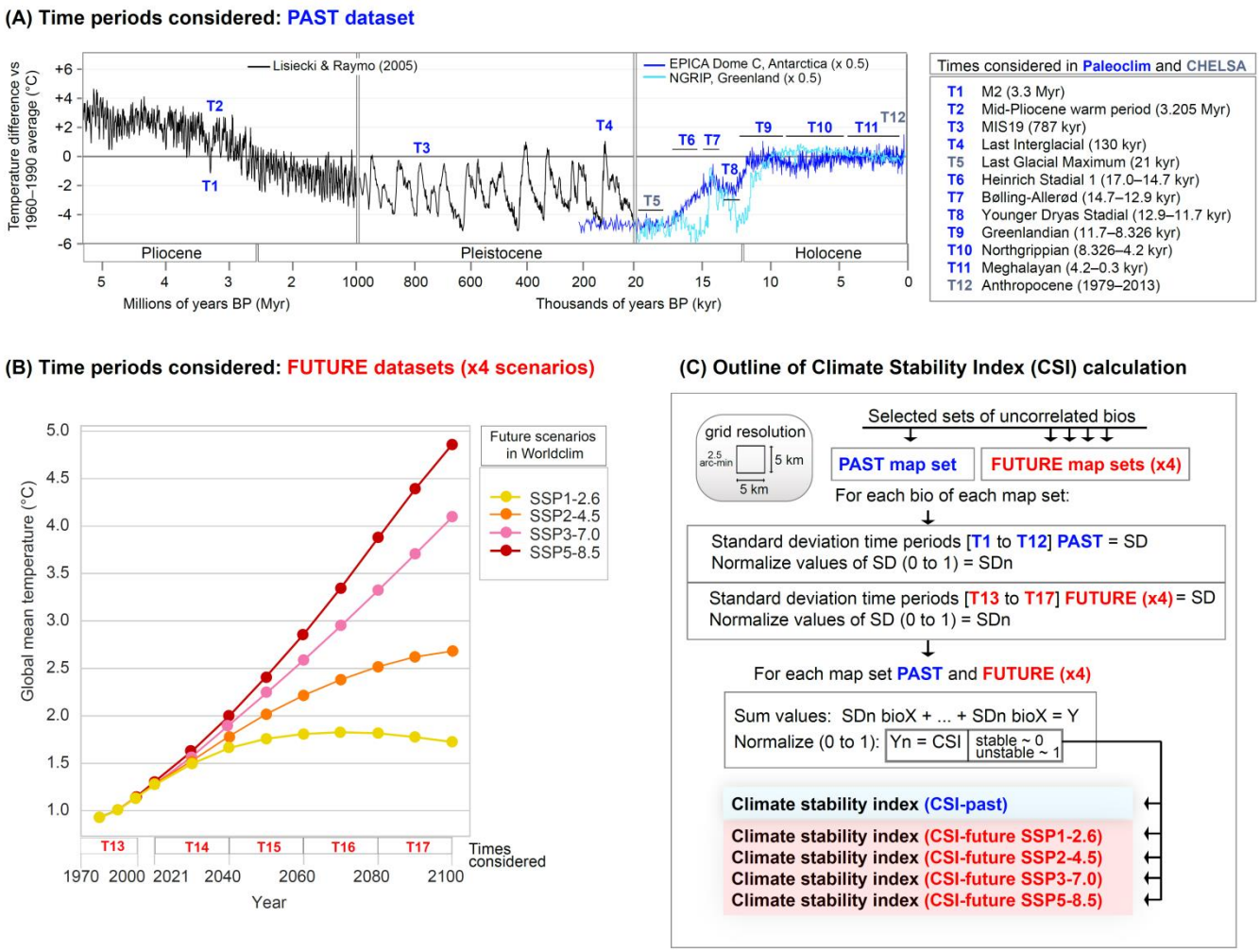


Fig. 1. General overview of time periods considered to calculate the Climate Stability Index (CSI) for past and future map sets and graphical illustration of index generation. (A) Time periods used to estimate the CSI-past and from which paleoclimate data have been taken (from PaleoClim[16] database, which includes two map sets from CHELSA[24]), represented by a Pliocene to Holocene temperature reconstruction modified from https://en.wikipedia.org/wiki/Geologic_temperature_record. Temperature reconstruction showed in a black line derived from Lisiecki and Raymo[25]. (B) Time periods and four Shared Socioeconomic Pathways (SSPs: SSP1-2.6, SSP2-4.5, SSP3-7.0, and SSP5-8.5) considered to estimate CSI-future using climate data from *WorldClim*[17], modified from Gidden et al.[26]. (C) Schematic representation of the methodology employed to calculate the CSI for each map set.

independently for each variable. The mosaic function was also used for the range calculation, with "fun = min" and "fun = max" to obtain the minimum and maximum values of input rasters, respectively, with a further step for subtracting maximum to minimum values.

Specifically, for CSI-past, as it includes several time periods with sea-level dropping below the present level (T1, T3, T5, T6, T7, T8, T9; Fig. 1A), we applied a mask of the current land surface, i.e. taking the T12 (Anthropocene) as a template. With this additional step, we were able to remove those pixels (grid cells) currently under the sea but that were once emerged. Most of these pixels, however, were only emerged during the LGM (ca. 21 ka), thus having values for bioclimatic variables for just a single time period (instead of the 12 routinely used for the variability estimation). The inclusion of these areas would result in highly climatically stable regions (low SD values; Supplementary Fig. 1), but this would be an obviously biased result. In contrast, we did not remove those areas affected by the sea-level rising periods, as only three periods contained "NoData" values (T2, T4, T10; Fig. 1A). However, to take this fact into consideration, we created a raster file in which these areas submerged during warm periods are indicated (see Supplementary Fig. 1). Finally, for both CSI-past and CSI-future, the resulting SD values were normalized to values between 0 and 1, with 0 representing completely stable areas and 1 the most unstable ones.

The next step was focused on the selection of a relatively uncorrelated set of variables for each map set. We used the removeCollinearity function from virtualspecies R package[22] that estimates the correlation value among pairs of variables from a given number of random sample points (10,000 in present case) according to a given method (Pearson for the present case) and a threshold of statistic selected ($r > 0.8$ as a cut-off value). The function removeCollinearity returns a list of uncorrelated variables according to the settings specified, randomly selecting just one variable from groups of correlated ones (see Table 1 for a

Table 1. Bioclimatic variables used to generate the Climate Stability Index (CSI) maps. For each variable we include the following information: the units, whether they are included in a climate map set (and source from where it can be downloaded), and whether they have been taken into account to calculate the CSI. Note that bio2, bio3, bio5, bio6 and bio7 were not included for the calculation of CSI-past as these variables are not available in *PaleoClim* database[16] for T1 (M2, Pliocene, ca. 3.3 Myr), T2 (mid-Pliocene warm period, Pliocene, 3.205 Myr), and T3 (MIS19, Pleistocene, ca. 787 kyr). Unit abbreviations: °C (Celsius), mm (millimetres), frt (fraction).

Variable	Unit	Map set (source)	Included in CSI maps
Bio1: Annual mean temperature	°C	Past (PaleoClim) Future (WorldClim v2)	Past Future (SSP3)
Bio2: Mean Diurnal Range (Mean of monthly (max temp - min temp))	°C	Future (WorldClim v2)	Future (SSP1, SSP2, SSP3, SSP5)
Bio3: Isothermality (BIO2/BIO7) (×100)	Frt	Future (WorldClim v2)	Future (SSP1, SSP2, SSP3, SSP5)
Bio4: Temperature Seasonality (standard deviation ×100)	°C	Past (PaleoClim) Future (WorldClim v2)	Past
Bio5: Max Temperature of Warmest Month	°C	Future (WorldClim v2)	Future (SSP3)
Bio6: Min Temperature of Coldest Month	°C	Future (WorldClim v2)	Future (SSP1, SSP2)
Bio7: Temperature Annual Range (BIO5-BIO6)	°C	Future (WorldClim v2)	Future (SSP1, SSP5)
Bio8: Mean Temperature of Wettest Quarter	°C	Past (PaleoClim) Future (WorldClim v2)	Future (SSP1, SSP2, SSP3, SSP5)
Bio9: Mean Temperature of Driest Quarter	°C	Past (PaleoClim) Future (WorldClim v2)	Past Future (SSP1, SSP2, SSP3, SSP5)
Bio10: Mean Temperature of Warmest Quarter	°C	Past (PaleoClim) Future (WorldClim v2)	Future (SSP1, SSP2, SSP5)
Bio11: Mean Temperature of Coldest Quarter	°C	Past (PaleoClim) Future (WorldClim v2)	Past
Bio12: Annual Precipitation	mm	Past (PaleoClim) Future (WorldClim v2)	Past Future (SSP1, SSP3, SSP5)
Bio13: Precipitation of Wettest Month	mm	Past (PaleoClim) Future (WorldClim v2)	Past Future (SSP3, SSP5)
Bio14: Precipitation of Driest Month	mm	Past (PaleoClim) Future (WorldClim v2)	Past Future (SSP1)
Bio15: Precipitation Seasonality (Coefficient of Variation)	Frt	Past (PaleoClim) Future (WorldClim v2)	Past Future (SSP1, SSP2, SSP3, SSP5)
Bio16: Precipitation of Wettest Quarter	mm	Past (PaleoClim) Future (WorldClim v2)	Future (SSP2)
Bio17: Precipitation of Driest Quarter	mm	Past (PaleoClim) Future (WorldClim v2)	Future (SSP2, SSP3, SSP5)
Bio18: Precipitation of Warmest Quarter	mm	Past (PaleoClim) Future (WorldClim v2)	Past Future (SSP1, SSP2, SSP3, SSP5)
Bio19: Precipitation of Coldest Quarter	mm	Past (PaleoClim) Future (WorldClim v2)	Past Future (SSP1, SSP2, SSP3, SSP5)

complete list of variables used for each map set). As we compiled estimates of variability independently for each variable and map set (e.g. SD bio1 past, SD bio2 past, etc.), each user can define his own CSI, selecting the more interesting variables according to the case of study.

The final CSI maps were obtained by summing the SD values of the variables selected and the subsequent outputs normalized (0 to 1) (Figs. 2–4). Histogram plots were represented with ggplot2 R package[23] and maps were exported with ArcGIS v.10.2.2 (Esri, Redlands, California, USA 2014). The histograms were computed for these final CSI maps, which represent the frequency and distribution of CSI values. We presented the final CSI maps with two different colour ramp schemes with ArcGIS. The first consisted of defining equal interval breaks from 0 to 1. The second was based on defining 32 categories with different value breaks for past and future map sets according to the value frequency shown by the histogram plot, i.e. the category with the highest CSI values (no. 32) was 0.71–1 in the past map set and 0.356–1 in the future map set.

3. Data Records

Our data records are available through figshare (https://doi.org/10.6084/m9.figshare.14672637) in format of map raster layers (.TIF format).

Two sets of data records are stored:

(1) SD-based maps of individual bioclimatic variables, which contain a set of raster layers (one for each variable): 14 for the case of the past map set (bio2, bio3, bio5, bio6, bio7 are missing) and 19 for the future map sets based on median calculations. These layers contain individual measures of the SD of each variable, which could be independently used according to the user's study purpose or combined according to the user's preferences to generate a customized CSI.
(2) the five CSI-based maps presented in this study, corresponding to CSI-past, CSI-future SSP1-2.6, CSI-future SSP2-4.5, CSI-future SSP3-7.0, and CSI-future SSP5-8.5.

Table 2. General circulation models (GCM) used to construct the future map sets. Those GCM marked with "NA" correspond to non-available models in *WorldClim*[17] database, and thus not used for the calculation of means and medians.

2021–2040				
GCM	**SSP1-2.6**	**SSP2-4.5**	**SSP3-7.0**	**SSP5-8.5**
BCC-CSM2-MR	✓	✓	✓	✓
CNRM-CM6-1	✓	✓	✓	✓
CNRM-ESM2-1	✓	✓	✓	✓
CanESM5	✓	✓	✓	✓
GFDL-ESM4	✓	NA	✓	NA
IPSL-CM6A-LR	✓	NA	✓	✓
MIROC-ES2L	✓	✓	✓	✓
MIROC6	✓	✓	✓	✓
MRI-ESM2-0	✓	✓	✓	NA
2041–2060				
GCM	**SSP1-2.6**	**SSP2-4.5**	**SSP3-7.0**	**SSP5-8.5**
BCC-CSM2-MR	✓	✓	✓	✓
CNRM-CM6-1	✓	✓	✓	✓
CNRM-ESM2-1	✓	✓	✓	✓
CanESM5	✓	✓	✓	✓
GFDL-ESM4	✓	NA	✓	NA
IPSL-CM6A-LR	✓	✓	✓	✓
MIROC-ES2L	✓	✓	✓	✓
MIROC6	✓	✓	✓	✓
MRI-ESM2-0	✓	✓	✓	✓
2061–2080				
GCM	SSP1-2.6	SSP2-4.5	SSP3-7.0	SSP5-8.5
BCC-CSM2-MR	✓	✓	✓	✓
CNRM-CM6-1	✓	✓	✓	✓
CNRM-ESM2-1	✓	✓	✓	✓
CanESM5	✓	✓	✓	✓
GFDL-ESM4	✓	NA	✓	NA
IPSL-CM6A-LR	✓	✓	✓	✓
MIROC-ES2L	✓	✓	✓	✓
MIROC6	✓	✓	✓	✓
MRI-ESM2-0	✓	✓	✓	✓
2081–2100				
GCM	SSP1-2.6	SSP2-4.5	SSP3-7.0	SSP5-8.5
BCC-CSM2-MR	✓	✓	✓	✓
CNRM-CM6-1	✓	✓	✓	✓
CNRM-ESM2-1	✓	✓	✓	✓
CanESM5	✓	✓	✓	✓
GFDL-ESM4	✓	NA	✓	NA
IPSL-CM6A-LR	✓	✓	✓	✓
MIROC-ES2L	✓	✓	✓	✓
MIROC6	✓	✓	✓	✓
MRI-ESM2-0	✓	✓	✓	✓

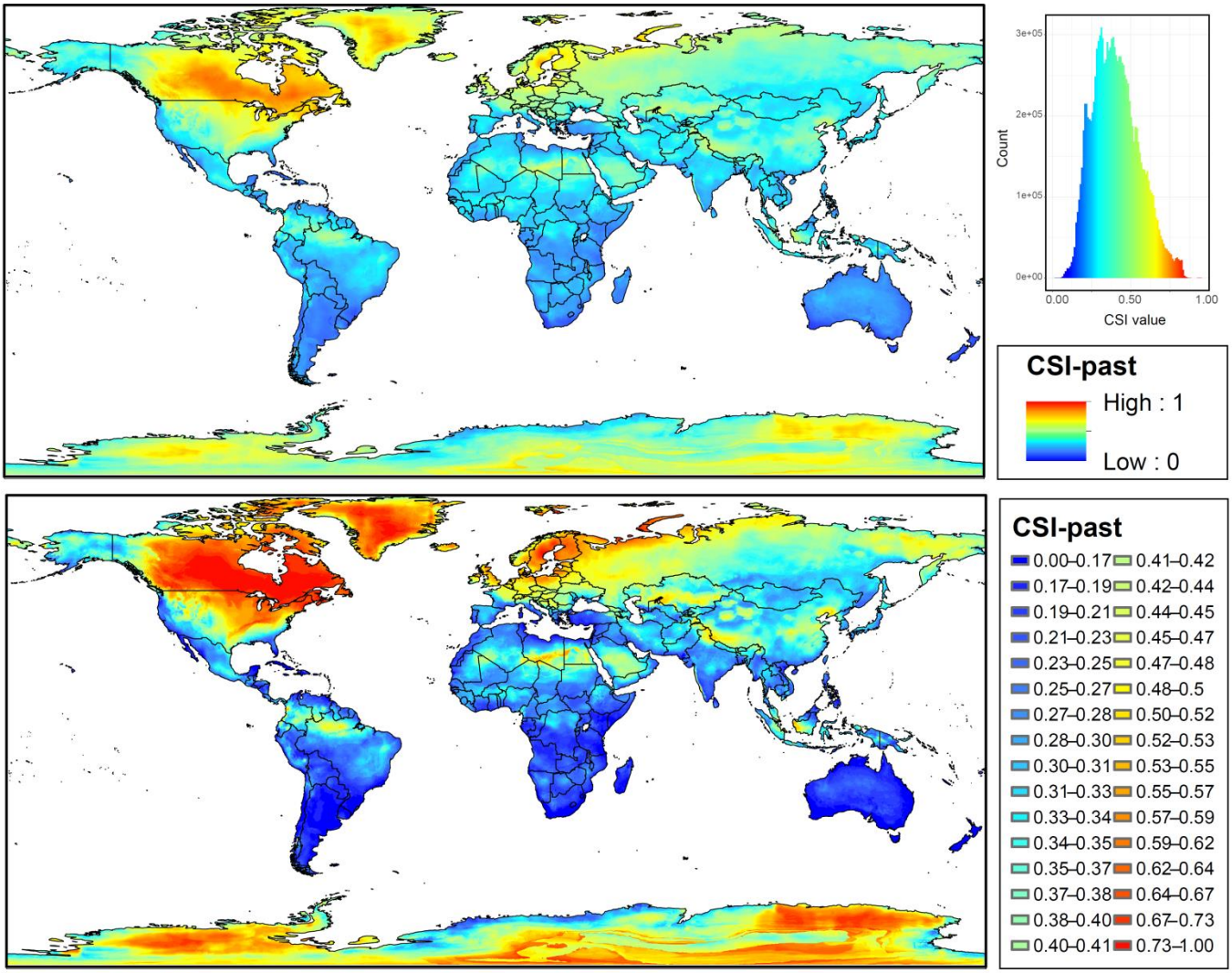


Fig. 2. Maps of Climate Stability Index (CSI) values for the past map set from Pliocene (3.3 Myr) to present (1979–2013), at 2.5 arc-min grid resolution. Colours range from blue for low standard deviation (SD) values, which represents areas with low climatic fluctuations (i.e, low values of CSI) during the period Pliocene–present, to red for high SD values, which shows areas where high climatic fluctuations would have taken place (i.e., high values of CSI). On the upper map, the colour ramp shows equal interval breaks. The histogram with frequency and distribution of CSI values is also shown. On the lower map, the colour ramp has been manually adjusted to a defined set of break values (see details in the text).

Data files share a common naming pattern:

For *Past map set* from individual variables and CSI, respectively:
"sd_past_<bioclimatic variable>.tif"
"csi_past.tif"

For *Future map sets* from individual variables and CSI, respectively:
"sd_future_<SSP scenario>_<bioclimatic variable>.tif"
"csi_future_<SSP scenario>.tif"

In addition, for the CSI-past map set, we include a map showing the areas affected by sea-level rising periods (intergl_affected.tif) and the raster map used to remove regions with landmasses currently under the sea but that were once emerged (lgm_del_mask.tif; see Supplementary Fig. 1).

4. Technical Validation

To evaluate the robustness of the CSI index, we compared its performance by varying the statistics used for its calculation (see detailed workflow in Fig. 5). For the past map set, the CSI was computed independently by means of SD and range. For the future map sets, the CSI was computed using the mean and the median of the nine GCMs to obtain individual future rasters for each time and SSPs. In parallel, the sensitivity derived from selecting a more restrictive variable correlation threshold ($r > 0.7$), an intermediate ($r > 0.8$), or a less restrictive cut-off value ($r > 0.9$) was also checked. In Supplementary Tables 1 and 2 the variables selected for each r threshold tested are specified. All pairs of CSI rasters were compared through Pearson's correlation analyses, which showed a very high r value, ranging from 0.93 to 1.00. All r values in Table 3 and all correlation scatterplots are available in Supplementary Files and some examples are included

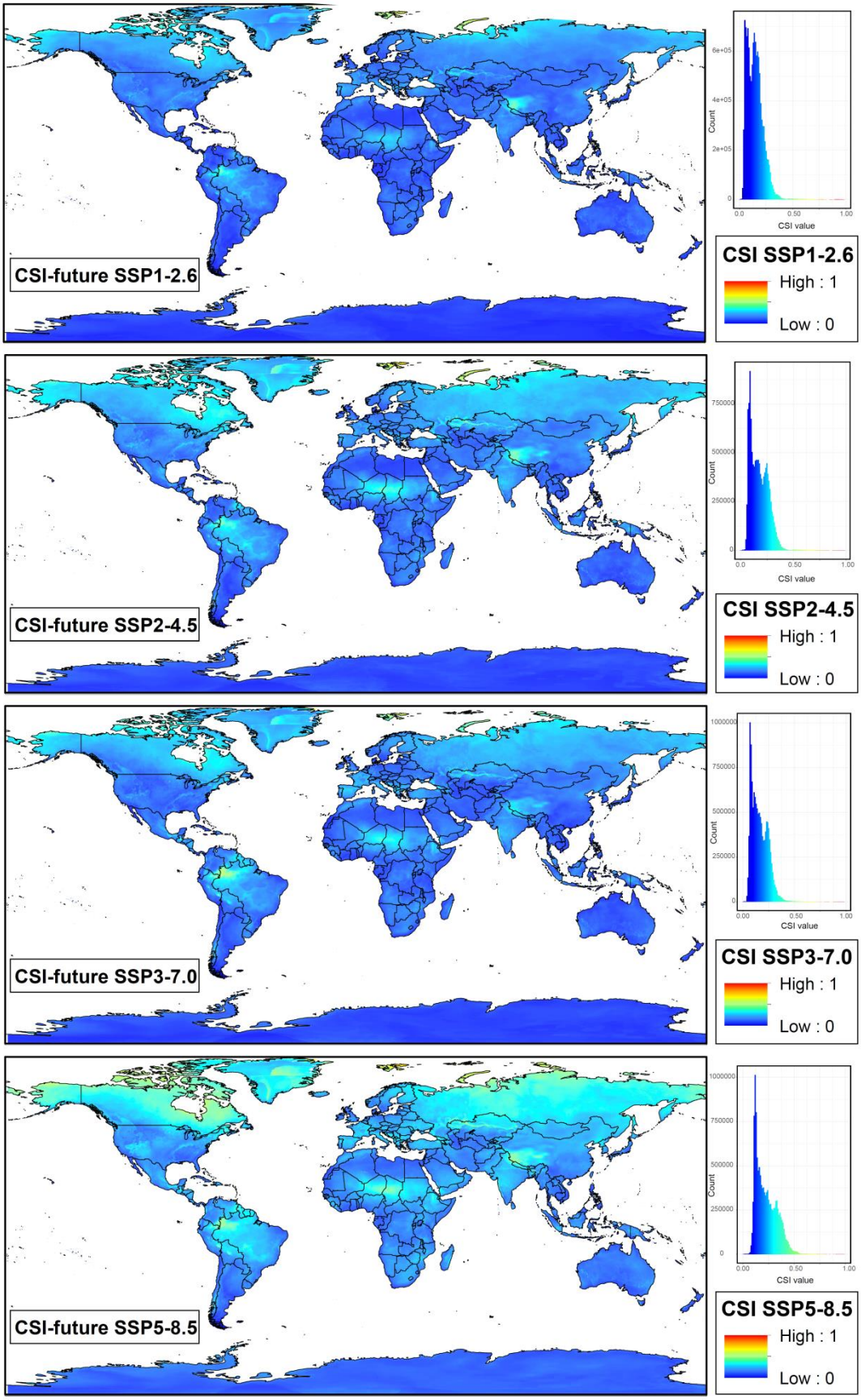


Fig. 3. Maps of Climate Stability Index (CSI) values for the future conditions (Shared Socioeconomic Pathways: SSP1-2.6, SSP2-4.5, SSP3-7.0, and SSP5-8.5) from present (1970–2000) to future (2100), at 2.5 arc-min grid resolution. Colours range from blue for low standard deviation (SD) values, which represents areas with low climatic fluctuations (i.e, low values of CSI) from present to future, to red for high SD values, which shows areas where high climatic fluctuations would have taken place (i.e., high values of CSI). The colour ramp shows equal interval breaks. The histogram with frequency and distribution of CSI values is also shown for each future scenario.

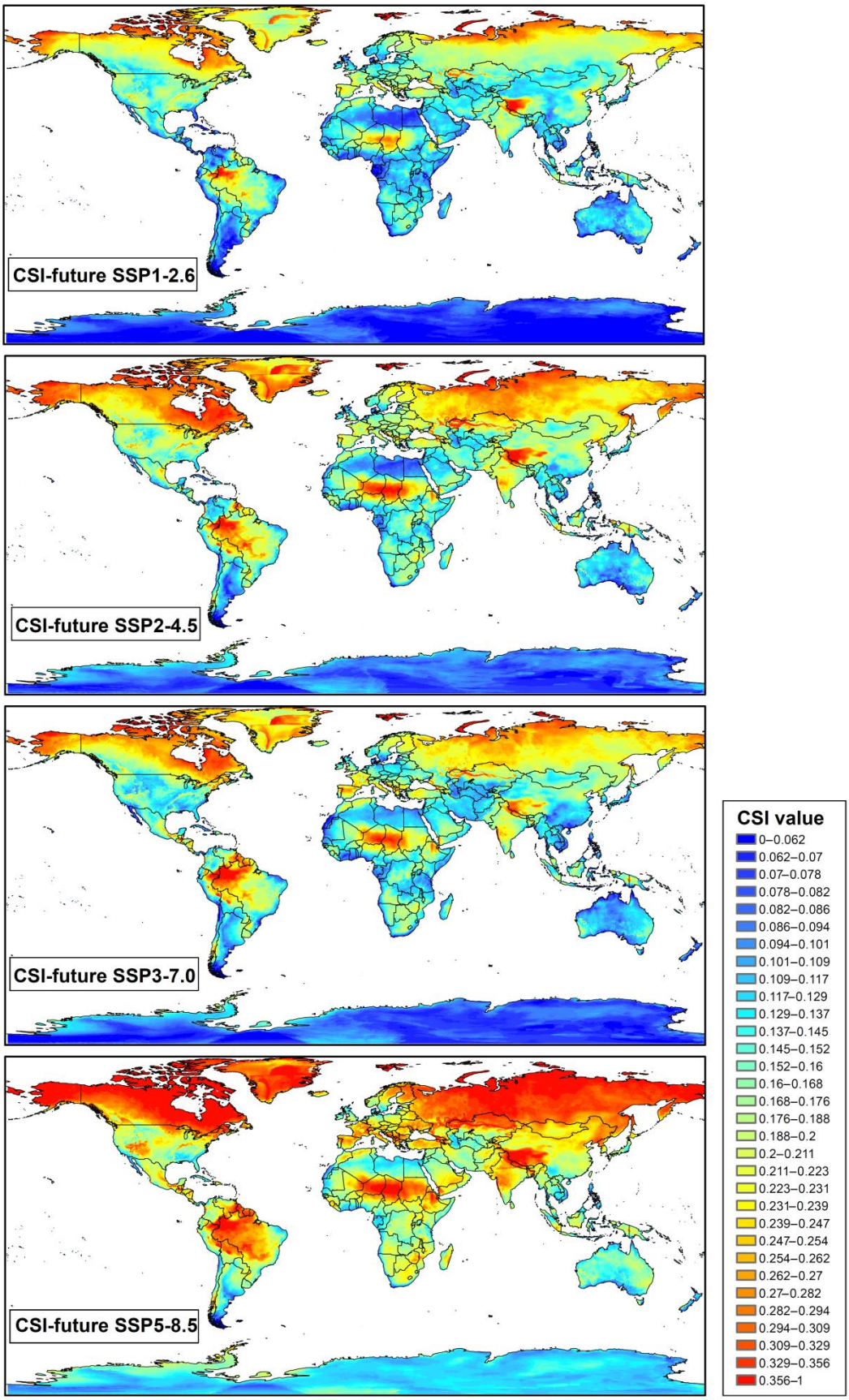


Fig. 4. Maps of Climate Stability Index (CSI) values for the future conditions (Shared Socioeconomic Pathways: SSP1-2.6, SSP2-4.5, SSP3-7.0, and SSP5-8.5) from present (1970–2000) to future (2100), at 2.5 arc-min grid resolution. Colours range from blue for low standard deviation (SD) values, which represents areas with low climatic fluctuations (i.e, low values of CSI) from present to future, to red for high SD values, which shows areas where high climatic fluctuations would have taken place (i.e., high values of CSI). The colour ramp has been manually adjusted to a defined set of break values (see details in the text).

Table 3. Correlation values (Pearson's *r*) of Technical validation procedure. Abbreviations used: SD (Standard Deviation), CSI (Climate Stability Index), th (threshold), *r* (*r* statistic of Pearson's correlation analysis).

Validation category	Correlation pairs	*r*
SD vs. range	CSI-past SD th 0.7 vs. CSI-past range th 0.7	0.94
	CSI-past SD th 0.8 vs. CSI-past range th 0.8	0.96
	CSI-past SD th 0.9 vs. CSI-past range th 0.9	0.97
mean vs. median	CSI-future mean SSP1-2.6 th 0.7 vs. CSI-future median SSP1-2.6 th 0.7	1.00
	CSI-future mean SSP1-2.6 th 0.8 vs. CSI-future median SSP1-2.6 th 0.8	1.00
	CSI-future mean SSP1-2.6 th 0.9 vs. CSI-future median SSP1-2.6 th 0.9	0.97
	CSI-future mean SSP2-4.5 th 0.7 vs. CSI-future median SSP2-4.5 th 0.7	1.00
	CSI-future mean SSP2-4.5 th 0.8 vs. CSI-future median SSP2-4.5 th 0.8	1.00
	CSI-future mean SSP2-4.5 th 0.9 vs. CSI-future median SSP2-4.5 th 0.9	1.00
	CSI-future mean SSP3-7.0 th 0.7 vs. CSI-future median SSP3-7.0 th 0.7	1.00
	CSI-future mean SSP3-7.0 th 0.8 vs. CSI-future median SSP3-7.0 th 0.8	1.00
	CSI-future mean SSP3-7.0 th 0.9 vs. CSI-future median SSP3-7.0 th 0.9	1.00
	CSI-future mean SSP5-8.5 th 0.7 vs. CSI-future median SSP5-8.5 th 0.7	1.00
	CSI-future mean SSP5-8.5 th 0.8 vs. CSI-future median SSP5-8.5 th 0.8	1.00
	CSI-future mean SSP5-8.5 th 0.9 vs. CSI-future median SSP5-8.5 th 0.9	1.00
threshold (*r*)	CSI-past SD th 0.7 vs. CSI-past SD th 0.8	0.95
	CSI-past SD th 0.7 vs. CSI-past SD th 0.9	0.97
	CSI-past SD th 0.8 vs. CSI-past SD th 0.9	0.98
	CSI-past range th 0.7 vs. CSI-past range th 0.8	0.98
	CSI-past range th 0.7 vs. CSI-past range th 0.9	0.95
	CSI-past range th 0.8 vs. CSI-past range th 0.9	0.99
	CSI-future mean SSP1-2.6 th 0.7 vs. CSI-future mean SSP1-2.6 th 0.8	0.99
	CSI-future mean SSP1-2.6 th 0.7 vs. CSI-future mean SSP1-2.6 th 0.9	0.99
	CSI-future mean SSP1-2.6 th 0.8 vs. CSI-future mean SSP1-2.6 th 0.9	0.99
	CSI-future mean SSP2-4.5 th 0.7 vs. CSI-future mean SSP2-4.5 th 0.8	0.96
	CSI-future mean SSP2-4.5 th 0.7 vs. CSI-future mean SSP2-4.5 th 0.9	0.93
	CSI-future mean SSP2-4.5 th 0.8 vs. CSI-future mean SSP2-4.5 th 0.9	0.99
	CSI-future mean SSP3-7.0 th 0.7 vs. CSI-future mean SSP3-7.0 th 0.8	0.98
	CSI-future mean SSP3-7.0 th 0.7 vs. CSI-future mean SSP3-7.0 th 0.9	0.99
	CSI-future mean SSP3-7.0 th 0.8 vs. CSI-future mean SSP3-7.0 th 0.9	0.98
	CSI-future mean SSP5-8.5 th 0.7 vs. CSI-future mean SSP5-8.5 th 0.8	0.98
	CSI-future mean SSP5-8.5 th 0.7 vs. CSI-future mean SSP5-8.5 th 0.9	0.99
	CSI-future mean SSP5-8.5 th 0.8 vs. CSI-future mean SSP5-8.5 th 0.9	0.99
	CSI-future median SSP1-2.6 th 0.7 vs. CSI-future median SSP1-2.6 th 0.8	0.99
	CSI-future median SSP1-2.6 th 0.7 vs. CSI-future median SSP1-2.6 th 0.9	0.96
	CSI-future median SSP1-2.6 th 0.8 vs. CSI-future median SSP1-2.6 th 0.9	0.96
	CSI-future median SSP2-4.5 th 0.7 vs. CSI-future median SSP2-4.5 th 0.8	0.96
	CSI-future median SSP2-4.5 th 0.7 vs. CSI-future median SSP2-4.5 th 0.9	0.93
	CSI-future median SSP2-4.5 th 0.8 vs. CSI-future median SSP2-4.5 th 0.9	0.99
	CSI-future median SSP3-7.0 th 0.7 vs. CSI-future median SSP3-7.0 th 0.8	0.98
	CSI-future median SSP3-7.0 th 0.7 vs. CSI-future median SSP3-7.0 th 0.9	0.99
	CSI-future median SSP3-7.0 th 0.8 vs. CSI-future median SSP3-7.0 th 0.9	0.98
	CSI-future median SSP5-8.5 th 0.7 vs. CSI-future median SSP5-8.5 th 0.8	0.98
	CSI-future median SSP5-8.5 th 0.7 vs. CSI-future median SSP5-8.5 th 0.9	0.99
	CSI-future median SSP5-8.5 th 0.8 vs. CSI-future median SSP5-8.5 th 0.9	0.99

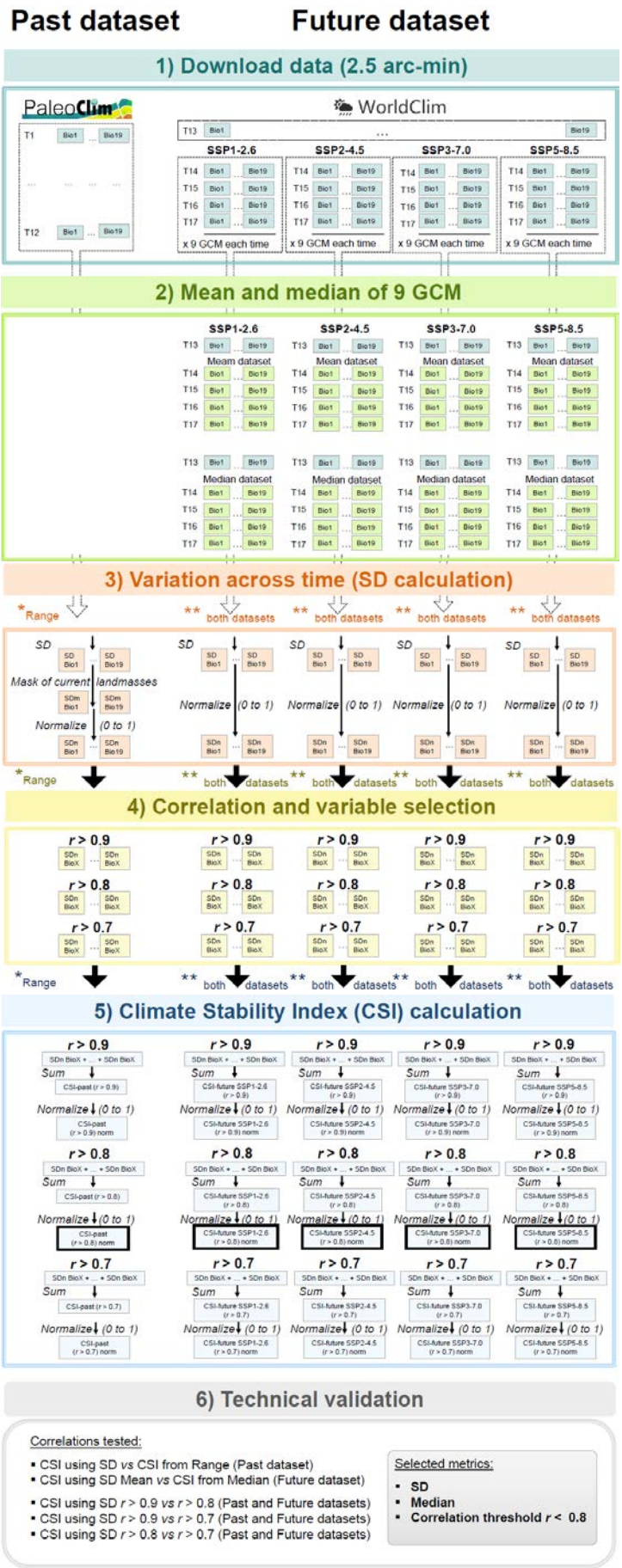


Fig. 5 Detailed workflow of the analyses employed to calculate and test the robustness of the Climate Stability Index (CSI) for past and future map sets. Final maps presented were obtained with conditions marked with wide-lined frame in the fifth step, which are: SD applying $r > 0.8$ threshold for variable correlation in past map set, and SD from median of nine GCM and applying $r > 0.8$ threshold for variable correlation in future map sets. Note that in the third step further analyses are repeated for the range statistic in the case of past map set and for both future map sets (mean and median).

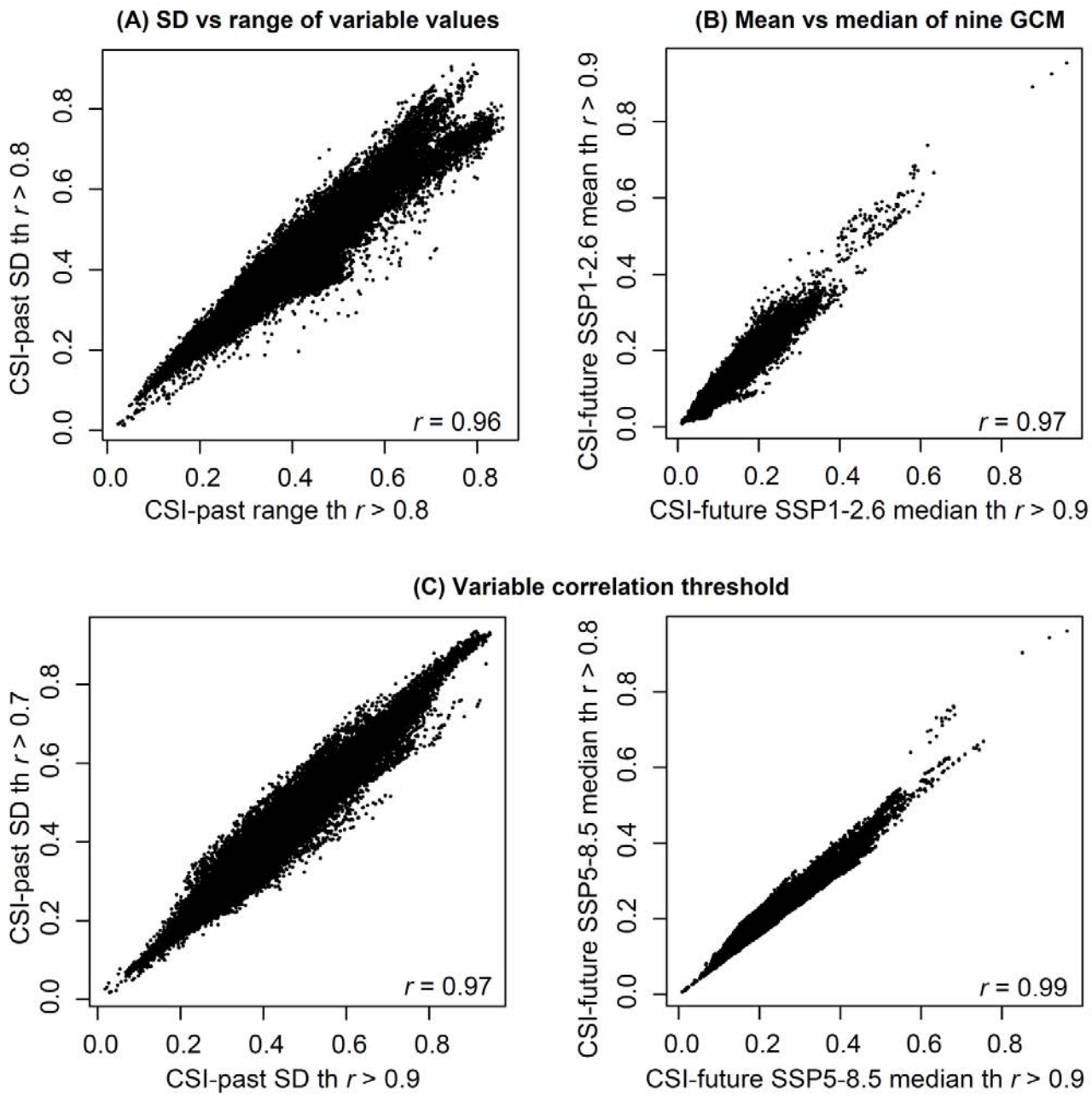


Fig. 6. Pearson's correlation coefficient (r) for examples of Climate Stability Index (CSI) pair comparisons: (A) between SD and range variable values of the past map set; (B) between mean and median of General Circulation Models of future map sets; and (C) between $r > 0.7$ and $r > 0.9$ (left) and between $r > 0.8$ and $r > 0.9$ (right) of Pearson cut-off values to remove highly correlated variables.

in Fig. 6. As a conclusion, methodological choices resulted in a non-remarkable impact of metric selected (mean vs. median, SD vs. range) or correlation threshold set ($r > 0.7$, 0.8, 0.9). Our choice was using the SD, median, and threshold $r > 0.8$, to draw the CSI-based maps.

5. Code Availability

Commented R codes used to generate CSI are available at https://doi.org/10.6084/m9.figshare.14672637

6. Acknowledgements

Financial support from the Ministerio de Ciencia e Innovación (Project CGL2015- 66703-P MINECO/FEDER, UE and Ph.D. grant to S.H.M) and the Catalan government ("Ajuts a grups consolidats" 2017-SGR1116) is also greatly acknowledged. This study has been performed under the Ph.D. program "Plant Biology and Biotechnology" of the Autonomous University of Barcelona (UAB). H.O. and T.M. were funded by the Environment Research and Technology Development Fund (JPMEERF20202002) of the Environmental Restoration and Conservation Agency of Japan.

7. Author contributions

S.H.M, N.N., and J.L.P designed and outlined the project, with contributions from A.S., C.Q.T., H.O., N.G.J., M.G.C, and T.M. The analyses were carried out by S.H.M, who wrote the manuscript together with J.L.P. N.N. suggested significant changes to the paper. All authors reviewed and agreed on the final submitted version of the manuscript.

8. Competing interests

The authors declare no competing interests.

9. References

[1] Rangel, T. F. *et al*. Modeling the ecology and evolution of biodiversity: Biogeographical cradles, museums, and graves. *Science* **361**, eaar5452, https://doi.org/10.1126/science.aar5452 (2018).

[2] Liu, Y. *et al*. Sino-Himalayan mountains act as cradles of diversity and immigration centres in the diversification of parrotbills (Paradoxornithidae). *J. Biogeogr.* **43**, 1488–1501, https://doi.org/10.1111/jbi.12738 (2016).

[3] Owen, R. B. *et al*. Progressive aridification in East Africa over the last half million years and implications for human evolution. *Proc. Natl. Acad. Sci. USA* **115**, 11174–11179, https://doi.org/10.1073/pnas.1801357115 (2018).

[4] Tang, C. Q. *et al*. Identifying long-term stable refugia for relict plant species in East Asia. *Nat. Commun.* **9**, 4488, https://doi.org/10.1038/s41467-018-06837-3 (2018).

[5] Sundaram, M. *et al*. Accumulation over evolutionary time as a major cause of biodiversity hotspots in conifers. *Proc. R. Soc. B* **286**, 20191887, http://dx.doi.org/10.1098/rspb.2019.1887 (2019).

[6] Morales-Barbero, J., Ariel Martinez, P., Ferrer-Castán, D. & Olalla-Tárraga, M. A. Quaternary refugia are associated with higher speciation rates in mammalian faunas of the Western Palaearctic. *Ecography* **41**, 607–621, https://doi.org/10.1111/ecog.02647 (2018).

[7] Cowling, R. M. *et al*. Variation in plant diversity in mediterranean-climate ecosystems: the role of climatic and topographical stability. *J. Biogeogr.* **42**, 552–564, https://doi.org/10.1111/jbi.12429 (2015).

[8] Hoegh-Guldberg, O. *et al*. Impacts of 1.5°C Global Warming on Natural and Human Systems. In: Global warming of 1.5°C. An IPCC Special Report on the impacts of global warming of 1.5°C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty [V. Masson-Delmotte, P. *et al*. (eds.)]. In Press. (2018).

[9] Park, D. S., Willis, C. G., Xi, Z., Kartesz, J. T., Davis, C. C. & Worthington S. Machine learning predicts large scale declines in native plant phylogenetic diversity. *New Phytol.* **227**, 1544–1556, https://doi.org/10.1111/nph.16621 (2020).

[10] Iwamura, T., Wilson, K. A., Venter, O. & Possingham, H. P. A climatic stability approach to prioritizing global conservation investments. *PLoS ONE* 5, e15103, https://doi.org/10.1371/journal.pone.0015103 (2010).

[11] Klein, C. *et al*. Incorporating ecological and evolutionary processes into continental-scale conservation planning. *Ecol. Appl.* 19, 206–217, https://doi.org/10.1890/07-1684.1 (2009).

[12] Brown, S. C., Wigley, T. M. L., Otto-Bliesner, B. L., Rahbek, C. & Fordham, D. A. Persistent Quaternary climate refugia are hospices for biodiversity in the Anthropocene. *Nat. Clim. Change* **10**, 244–248, https://doi.org/10.1038/s41558-019-0682-7 (2020).

[13] Brown, S. C., Wigley, T. M. L., Otto-Bliesner, B. L. & Fordham, D. A. *StableClim*, continuous projections of climate stability from 21000 BP to 2100 CE at multiple spatial scales. *Sci. Data* **7**, 335, https://doi.org/10.1038/s41597-020-00663-3 (2020).

[14] Fordham, D. A., Brown, S. C., Wigley, T. M. & Rahbek, C. Cradles of diversity are unlikely relics of regional climate stability. *Curr. Biol.* **29**, R356–R357, https://doi.org/10.1016/j.cub.2019.04.001 (2019).

[15] Kuhlmann, M., Guo, D., Veldtman, R. & Donaldson, J. Consequences of warming up a hotspot: species range shifts within a centre of bee diversity. *Divers. Distrib.* **18**, 885–897, https://doi.org/10.1111/j.1472-4642.2011.00877.x (2012).

[16] Brown, J. L., Hill, D. J., Dolan, A. M., Carnaval, A. C. & Haywood, A. M. PaleoClim, high spatial resolution paleoclimate surfaces for global land areas. *Sci. Data* **5**, 180254, https://doi.org/10.1038/sdata.2018.254 (2018).

[17] Fick, S. E. & Hijmans, R. J. WorldClim 2: new 1km spatial resolution climate surfaces for global land areas. *Int. J. Climatol.* **37**, 4302–4315, https://doi.org/10.1002/joc.5086 (2017).

[18] O'Neill, B. C. *et al*. A new scenario framework for climate change research: the concept of shared socioeconomic pathways. *Clim. Change* **122**, 387–400, https://doi.org/10.1007/s10584-013-0905-2 (2014).

[19] Wilkinson, M. D. *et al*. The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* **3**, 160018, https://doi.org/10.1038/sdata.2016.18 (2016).

[20] R Development Core Team. R: A Language and Environment for Statistical Computing. *R foundation for Statistical Computing* http://www.R-project.org/ (2005).

[21] Hijmans, R. J. *raster* R package version 2.5-8 https://cran.r-project.org/web/packages/raster/raster.pdf (2015).

[22] Leroy, B., Meynard, C.N. Bellard, C. & Courchamp, F. Virtualspecies, an R package to generate virtual species distributions. *Ecography* **39**, 599–607, https://doi.org/10.1111/ecog.01388 (2016).

[23] Wickham, H. ggplot2. Wiley Interdisciplinary Reviews. *Comput. Stat.* **3**, 180–185, https://doi.org/10.1002/wics.147 (2011).

[24] Karger, D. N. *et al*. Climatologies at high resolution for the Earth land surface areas. *Sci. Data* **4**, 170122, https://doi.org/10.1038/sdata.2017.122 (2017).

[25] Lisiecki, L. E. & Raymo, M. E. A Pliocene-Pleistocene stack of 57 globally distributed benthic δ18O records. *Paleoceanogr. Paleoclimatol.* **20**, PA1003, https://doi.org/10.1029/2004PA001071 (2005).

[26] Gidden, M. J. *et al*. Global emissions pathways under different socioeconomic scenarios for use in CMIP6: a dataset of harmonized emissions trajectories through the end of the century. *Geosci. Model Dev.* **12**, 1443–1475, https://doi.org/10.5194/gmd-12-1443-2019 (2019).

10. Supplementary material

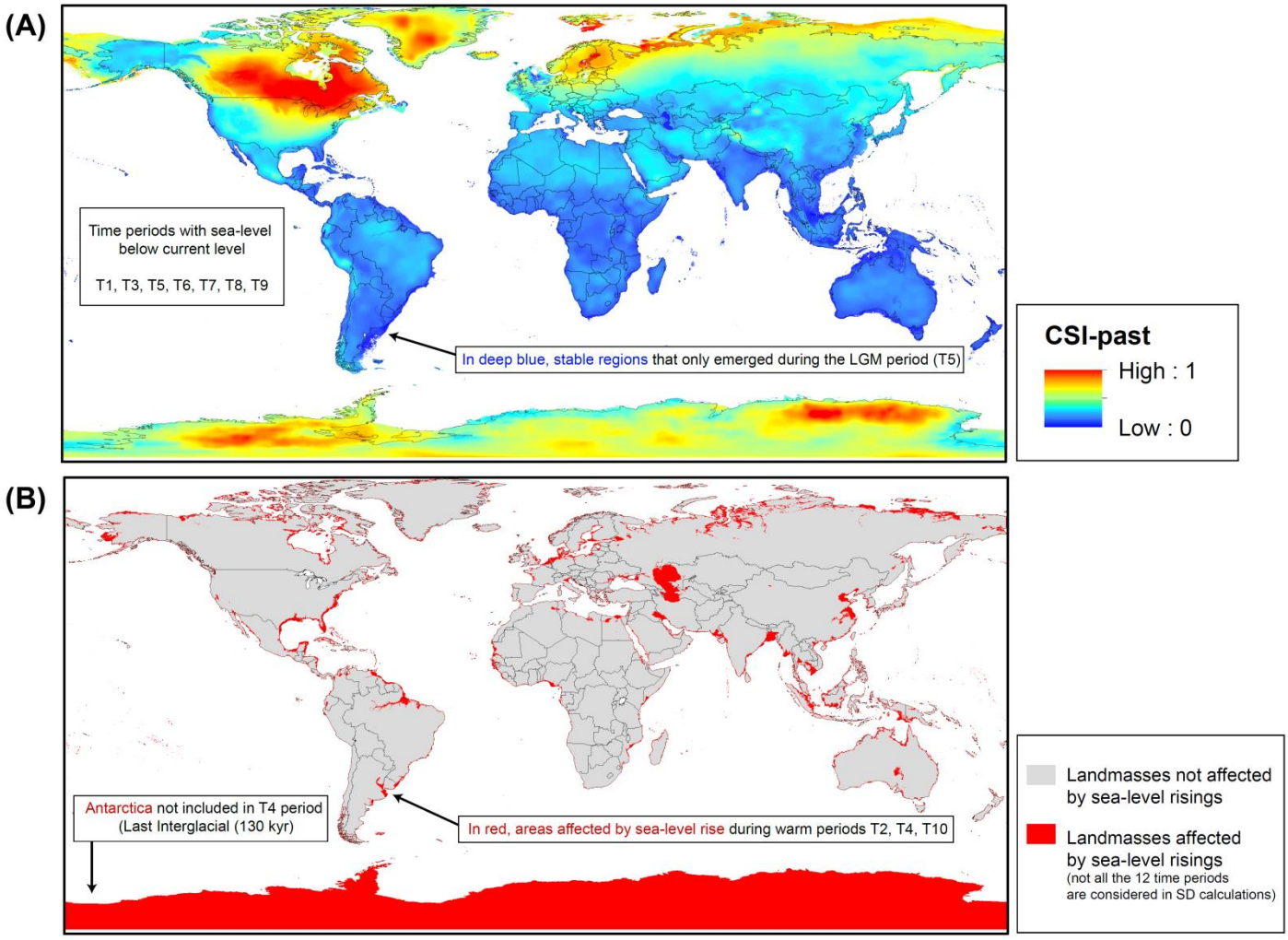


Fig. 1 (A) Map of Climate Stability Index (CSI) values for the past map set [Pliocene (3.3 Myr) to present (1979–2013)], at 2.5 arc-min grid resolution. All pixels have been considered in calculations, despite the fact that in some regions where sea-level dropped during cold periods (T1, T3, T5, T6, T7, T8, T9, see Fig. 1 for time period names) there are values for only one or few time periods, while the remaining time periods have NoData codes. (B) Map showing the areas (in red) affected by some sea-level rising periods (T2, T4, T10, see Fig. 1 for time period names). A raster layer with the affected areas is available at figshare (https://doi.org/10.6084/m9.figshare.14672637).

Table 1. Bioclimatic variables used to generate the Climate Stability Index (CSI), considering the median calculation of nine future scenarios, for each analysis of different *r* (*r* statistic of Pearson's correlation analysis; $r > 0.9$, $r > 0.8$, $r > 0.7$). For each variable we include the following information: the units, whether they are included on a climatic map set to calculate the CSI. Note that bio2, bio3, bio5, bio6 and bio7 were not included for the calculation of CSI past, as these variables are not available in *PaleoClim* database[16] for T1 (M2, Pliocene, ca. 3.3 Myr), T2 (mid-Pliocene warm period, Pliocene, 3.205 Myr), and T3 (MIS19, Pleistocene, ca. 787 kyr). Unit abbreviations: °C (Celsius), mm (millimetres), frt (fraction).

Variable	Unit	Included in CSI ($r > 0.9$)	Included in CSI ($r > 0.8$)	Included in CSI ($r > 0.7$)
Bio1: Annual mean temperature	°C	Future (SSP2)	Past Future (SSP3)	Future (SSP2, SSP5)
Bio2: Mean Diurnal Range (Mean of monthly (max temp - min temp))	°C	Future (SSP1, SSP2, SSP3, SSP5)	Future (SSP1, SSP2, SSP3, SSP5)	Future (SSP1, SSP2, SSP3, SSP5)
Bio3: Isothermality (BIO2/BIO7) (×100)	frt	Future (SSP1, SSP2, SSP3, SSP5)	Future (SSP1, SSP2, SSP3, SSP5)	Future (SSP1, SSP2, SSP3, SSP5)
Bio4: Temperature Seasonality (standard deviation ×100)	°C	Past Future (SSP2, SSP5)	Past	Past Future (SSP1)
Bio5: Max Temperature of Warmest Month	°C	Future (SSP1, SSP2, SSP3)	Future (SSP3)	Future (SSP1, SSP3)
Bio6: Min Temperature of Coldest Month	°C	Future (SSP3)	Future (SSP1, SSP2)	
Bio7: Temperature Annual Range (BIO5-BIO6)	°C	Future (SSP1, SSP3)	Future (SSP1, SSP5)	Future (SSP3)
Bio8: Mean Temperature of Wettest Quarter	°C	Past Future (SSP1, SSP2, SSP3, SSP5)	Future (SSP1, SSP2, SSP3, SSP5)	Past Future (SSP1, SSP2, SSP3, SSP5)
Bio9: Mean Temperature of Driest Quarter	°C	Past Future (SSP1, SSP2, SSP3, SSP5)	Past Future (SSP1, SSP2, SSP3, SSP5)	Past Future (SSP1, SSP2, SSP3, SSP5)
Bio10: Mean Temperature of Warmest Quarter	°C	Past Future (SSP3, SSP5)	Future (SSP1, SSP2, SSP5)	Future (SSP2, SSP5)
Bio11: Mean Temperature of Coldest Quarter	°C	Past Future (SSP1, SSP5)	Past	
Bio12: Annual Precipitation	mm	Past Future (SSP1, SSP2, SSP3, SSP5)	Past Future (SSP1, SSP3, SSP5)	Past Future (SSP1)
Bio13: Precipitation of Wettest Month	mm	Future (SSP2, SSP3)	Past Future (SSP3, SSP5)	Future (SSP3)
Bio14: Precipitation of Driest Month	mm	Future (SSP5)	Past Future (SSP1)	Future (SSP1, SSP5)
Bio15: Precipitation Seasonality (Coefficient of Variation)	frt	Past Future (SSP1, SSP2, SSP3, SSP5)	Past Future (SSP1, SSP2, SSP3, SSP5)	Past Future (SSP1, SSP2, SSP3, SSP5)
Bio16: Precipitation of Wettest Quarter	mm	Past Future (SSP1, SSP5)	Future (SSP2)	Past Future (SSP2, SSP5)
Bio17: Precipitation of Driest Quarter	mm	Past Future (SSP1, SSP2, SSP3)	Future (SSP2, SSP3, SSP5)	Past Future (SSP2, SSP3)
Bio18: Precipitation of Warmest Quarter	mm	Past Future (SSP1, SSP2, SSP3, SSP5)	Past Future (SSP1, SSP2, SSP3, SSP5)	Past Future (SSP1, SSP2, SSP3, SSP5)
Bio19: Precipitation of Coldest Quarter	mm	Past Future (SSP1, SSP2, SSP3, SSP5)	Past Future (SSP1, SSP2, SSP3, SSP5)	Past Future (SSP1, SSP2, SSP3, SSP5)

Table 2. Bioclimatic variables used to generate the Climate Stability Index (CSI), considering the mean calculation of nine future scenarios, for each analysis of different *r* (*r* statistic of Pearson's correlation analysis; $r > 0.9$, $r > 0.8$, $r > 0.7$). For each variable we include the following information: the units, whether they are included on a climatic map set to calculate the CSI. Note that bio2, bio3, bio5, bio6 and bio7 were not included for the calculation of CSI past as these variables are not available in *PaleoClim* database[16] for T1 (M2, Pliocene, ca. 3.3 Myr), T2 (mid-Pliocene warm period, Pliocene, 3.205 Myr), and T3 (MIS19, Pleistocene, ca. 787 kyr). Unit abbreviations: °C (Celsius), mm (millimetres), frt (fraction).

Variable	Unit	Included in CSI ($r > 0.9$)	Included in CSI ($r > 0.8$)	Included in CSI ($r > 0.7$)
Bio1: Annual mean temperature	°C	Future (SSP1, SSP2, SSP5)	Past	
Bio2: Mean Diurnal Range (Mean of monthly (max temp - min temp))	°C	Future (SSP1, SSP2, SSP3, SSP5)	Future (SSP1, SSP2, SSP3, SSP5)	Future (SSP1, SSP3, SSP5)
Bio3: Isothermality (BIO2/BIO7) (×100)	frt	Future (SSP1, SSP2, SSP3, SSP5)	Future (SSP1, SSP2, SSP3, SSP5)	Future (SSP1, SSP2, SSP3, SSP5)
Bio4: Temperature Seasonality (standard deviation ×100)	°C	Past Future (SSP1, SSP3)	Past Future (SSP1)	Past
Bio5: Max Temperature of Warmest Month	°C	Future (SSP3, SSP5)		Future (SSP5)
Bio6: Min Temperature of Coldest Month	°C	Future (SSP3)	Future (SSP2, SSP5)	Future (SSP2)
Bio7: Temperature Annual Range (BIO5-BIO6)	°C	Future (SSP2, SSP5)	Future (SSP3)	
Bio8: Mean Temperature of Wettest Quarter	°C	Past Future (SSP1, SSP2, SSP3, SSP5)	Future (SSP1, SSP2, SSP3, SSP5)	Past Future (SSP1, SSP2, SSP3, SSP5)
Bio9: Mean Temperature of Driest Quarter	°C	Past Future (SSP1, SSP2, SSP3, SSP5)	Past Future (SSP1, SSP2, SSP3, SSP5)	Past Future (SSP1, SSP2, SSP3, SSP5)
Bio10: Mean Temperature of Warmest Quarter	°C	Past Future (SSP1, SSP2)	Future (SSP1, SSP2, SSP3, SSP5)	Future (SSP1, SSP2, SSP3, SSP5)
Bio11: Mean Temperature of Coldest Quarter	°C	Past	Past Future (SSP1)	Future (SSP1, SSP3)
Bio12: Annual Precipitation	mm	Past Future (SSP1, SSP2, SSP3, SSP5)	Past Future (SSP2, SSP3, SSP5)	Past Future (SSP1, SSP2, SSP3, SSP5)
Bio13: Precipitation of Wettest Month	mm	Future (SSP3, SSP5)	Past Future (SSP1)	
Bio14: Precipitation of Driest Month	mm	Future (SSP3)	Past Future (SSP1, SSP3)	Future (SSP1, SSP2, SSP3, SSP5)
Bio15: Precipitation Seasonality (Coefficient of Variation)	frt	Past Future (SSP1, SSP2, SSP3, SSP5)	Past Future (SSP1, SSP2, SSP3, SSP5)	Past Future (SSP1, SSP2, SSP3, SSP5)
Bio16: Precipitation of Wettest Quarter	mm	Past Future (SSP1, SSP2)	Future (SSP3)	Past
Bio17: Precipitation of Driest Quarter	mm	Past Future (SSP1, SSP2, SSP5)	Future (SSP2, SSP5)	Past
Bio18: Precipitation of Warmest Quarter	mm	Past Future (SSP1, SSP2, SSP3, SSP5)	Past Future (SSP1, SSP2, SSP3, SSP5)	Past Future (SSP1, SSP2, SSP3, SSP5)
Bio19: Precipitation of Coldest Quarter	mm	Past Future (SSP1, SSP2, SSP3, SSP5)	Past Future (SSP1, SSP2, SSP3, SSP5)	Past Future (SSP1, SSP2, SSP3, SSP5)

Chapter 5

Impact of the climatic changes in the Pliocene-Pleistocene transition on Irano-Turanian species. The radiation of genus *Jurinea* (Compositae)

Sonia Herrando-Moraira[1*] and The Cardueae Radiations Group

In alphabetical order:

Juan Antonio Calleja[2]
You-Sheng Chen[3]
Kazumi Fujikawa[4]
Mercè Galbany-Casals[5]
Núria Garcia-Jacas[1]
Jian-Quan Liu[6]
Javier López-Alvarado[5]
Jordi López-Pujol[1]
Jennifer R. Mandel[7]
Iraj Mehregan[8]
Cristina Roquet[5]
Llorenç Sáez[5]
Alexander Sennikov[9]
Alfonso Susanna[1]
Roser Vilatersana[1]
Lian-Sheng Xu[3]

[1] Botanic Institute of Barcelona (IBB, CSIC-Ajuntament de Barcelona), Pg. del Migdia, s.n., 08038 Barcelona, Spain
[2] Departament of Biology (Botany), Faculty of Sciences, Research Centre on Biodiversity and Global Change (CIBC-UAM), 28049 Madrid, Spain
[3] Key Laboratory of Plant Resources Conservation and Sustainable Utilization, South China Botanical Garden, Chinese Academy of Sciences, Guangzhou 510650, China
[4] Kochi Prefectural Makino Botanical Garden, 4200-6, Godaisan, Kochi 781-8125, Japan
[5] Systematics and Evolution of Vascular Plants (UAB) – Associated Unit to CSIC, Departament de Biologia Animal, Biologia Vegetal i Ecologia, Facultat de Biociències, Universitat Autònoma de Barcelona, 08193 Bellaterra, Spain
[6] Key Laboratory for Bio-Resources and Eco-Environment, College of Life Sciences, Sichuan University, Chengdu, China
[7] Department of Biological Sciences, Center for Biodiversity, University of Memphis, Memphis, Tennessee 38152, U.S.A.
[8] Department of Biology, Science and Research Branch, Islamic Azad University, Tehran, Iran
[9] Botanical Museum, Finnish Museum of Natural History, P.O. Box 7, 00014 University of Helsinki, Finland; and Herbarium, Komarov Botanical Institute of Russian Academy of Sciences, Prof. Popov str. 2, 197376 St. Petersburg, Russia

*Correspondence to be sent to: Botanic Institute of Barcelona (IBB, CSIC-Ajuntament de Barcelona), Pg. del Migdia, s.n., 08038 Barcelona, Spain; E-mail: sonia.herrando@gmail.com

In preparation for Systematic Biology

Abstract

The Irano-Turanian region is one of the world's richest floristic regions and the centre of diversity for numerous xerophytic plant lineages. However, we still have limited knowledge on the timing of evolution and biogeographic history of its flora, and potential drivers of diversification remain underexplored. To fill this knowledge gap, we focus on the Eurasian genus *Jurinea* (ca. 200 species), one of the largest plant radiations that diversified in the region. We applied a macroevolutionary integrative approach to explicitly test diversification hypotheses and investigate the relative roles of geography vs. ecology and niche conservatism vs. niche lability in speciation processes. To do so, we gathered a sample comprising 77% of total genus richness and obtained data about (1) its phylogenetic history, recovering 502 nuclear loci sequences; (2) growth forms; (3) ecological niche, compiling data of 21 variables for more than 2500 occurrences; and (4) paleoclimatic conditions, to estimate climatic stability. Our results revealed that climate was a key factor in the evolutionary dynamics of *Jurinea*. The main diversification and biogeographic events that occurred during past climate changes (that led to colder and drier conditions) are the following: (1) the origin of the genus (10.7 Ma); (2) long-distance dispersals from the Iranian Plateau to adjacent regions (~ 7–4 Ma); and (3) the diversification shift during Pliocene-Pleistocene Transition (ca. 3 Ma), when net diversification rate almost doubled. Our results supported the pre-adaptation hypothesis, i.e., the evolutionary success of *Jurinea* was linked to the retention of the ancestral niche adapted to aridity. Interestingly, the paleoclimatic analyses revealed that in the Iranian Plateau long-term climatic stability favored old-lineage persistence, resulting in current high species richness of semi-arid and cold adapted clades; in contrast, moderate climate oscillations stimulated allopatric diversification in the lineages distributed in the Circumboreal region and in mountains of Central Asia.

Keywords

Adaptive radiation
Asteraceae
Diversification
Evolution
Hyb-seq
Irano-Turanian region
Next-generation sequencing
Phylogeny

Index

1. Introduction

The study of evolutionary radiations is a hot topic in plant science research (Hughes et al. 2015), with the number of published articles on plant radiations increasing dramatically since 2000s (Stroud and Losos 2016). Most studies to date have focused on lineages with outstanding species richness in plant diversity hotspots such as the Cape Floristic Region, the Mediterranean Basin, oceanic islands and archipelagos such as Macaronesia, or mountain ranges such as the Andes or the Himalayas (Hughes and Atchison 2015). These types of studies have allowed researchers to identify the main evolutionary and environmental factors shaping biodiversity patterns in these hotspots and assess their relative contributions. However, there are still species-rich regions where large-continental radiations have taken place that remain understudied from an evolutionary and biogeographic perspective. One such knowledge gaps concerns the Irano-Turanian floristic region (hereafter IT; Takhtajan 1986), whose biodiversity origins and biogeographic relationships with other regions remain poorly studied despite its relevance in terms of plant diversity (32,000 species; Takhtajan 1986; Sales and Hedge 2013) and endemism (25–40%; Zohary 1981; Takhtajan 1986). Indeed, this region harbours two global biodiversity hotspots: the Irano-Anatolian region and the Mountains of Central Asia (Mittermeier et al. 2011).

The IT region is located in Western-Central Asia, encircled by the Circumboreal floristic region in the north, the eastern Mediterranean in the west, Arabia in the south, and East Asia in the east (see Manafzadeh et al. 2017 for a review of its circumscription). One of most remarkable characteristics of IT is its continental climate, with cold winters, warm-dry summers, and a high precipitation seasonality (Djamali et al. 2012). Topographically, IT presents high landscape heterogeneity and wide altitudinal gradients, with large mountains ranges (Taurus, Zagros, Alborz, Kopet-Dagh, Pamir, Tian Shan, Western Himalayas) and broad plateaus (Anatolian, Iranian, and Qinghai-Tibetan plateaus). The geo-climatic configuration of IT has likely favoured the diversification of xerophytes, which are adapted to arid or semi-arid environments under low water availability conditions, such as steppe elements or C4 metabolism plants (e.g., mega-diverse genera like *Acantholimon*, *Acanthophyllum*, *Astragalus*, *Cousinia*, *Eremurus*, and *Eremostachys*; Manafzadeh et al. 2017). Diversification is especially noticeable in mountain habitats: for instance, endemic richness of Compositae in Iran reaches maximum values at mid-elevations in mountain ranges (1400–2100 m a.s.l.; Noroozi et al. 2019).

However, limited information is available about the origins and drivers of diversification of the IT flora (see a synthesis in Manafzadeh et al. 2017). Regarding the timing of the origin of IT elements, several hypotheses have been proposed. The oldest scenario points to a Cretaceous origin (145–66 Ma; Takhtajan, 1986). Other studies suggest the early Eocene as a period of initial divergence of lineages (55 Ma; Manafzadeh et al. 2014), or during the drying of the Tethys Sea in the Neogene (34 Ma; Zohary, 1973). Nevertheless, recent time-calibrated phylogenies of diverse groups suggest that IT elements originated much more recently, from the Miocene onwards (Wu et al. 2015; Lauterbach et al. al. 2019), with special emphasis on the Pliocene-Pleistocene (Moharrek et al. 2019; Mahmoudi Shamsabad et al. 2020). Regarding potential drivers of diversification, the orogenic activity during the Miocene deeply shaped the IT landscape (Manafzadeh et al. 2017) and likely impacted the evolution of its flora: the second collision of the Afro-Arabian plate against the Eurasian plate originated the uplift of the Iranian Plateau ca. 13 Ma, and subsequent uplift and deformations occurred from 15 to 5 Ma along the Zagros, Alborz, Kopeth Dagh, and Caucasus mountains (Mouthereau 2011). Such orogenic activity may have been an important driver of species diversification increasing habitat heterogeneity and thus ecological opportunities and favouring allopatric speciation through vicariance. On the other hand, the distribution and diversity of organisms have also likely been impacted by past climatic changes such as: (1) the aridification of central Asia during mid-late Miocene (17–5 Ma; Miao et al. 2012); (2) the global climate cooling between ~3.2–2.7 Ma, a period known as the Pliocene-Pleistocene transition (PPT; Lisiecki and Raymo 2005); and (3) the drastic climatic oscillations during the Quaternary glacial-interglacial cycles (Zachos et al. 2001).

Recent methodological advances in comparative phylogenetic analyses (Borges et al. 2019), biogeographic history inference (Matzke 2013), and models of diversification (e.g. Etienne et al. 2012; Morlon et al. 2016), together with ecological niche modelling and high-resolution paleoclimatic datasets (*PaleoClim*; Brown et al. 2018), make it now possible to investigate the relative importance of abiotic and biotic factors on the diversification of IT taxa, and allow the testing of hypotheses about the macroevolutionary and biogeographic dynamics of the IT flora proposed by previous authors (reviewed in Manafzadeh et al. 2017). Two main scenarios, not mutually exclusive, may explain IT diversification: (1) vicariance through uplifts of mountain ranges (geographic speciation); and (2) ecological speciation. In the first scenario, species diverged by allopatry driven by geologic or climatic processes, and the ecological niche has been preserved among sister taxa (niche conservatism; Wiens et al. 2010). In the second, lineages diverged by adaptation to new ecological zones (niche divergence). Concerning the biogeographic interactions of IT with other regions, previous studies have suggested that IT constituted an important source of taxa to adjacent regions, especially contributing to the assemblage of the Mediterranean flora (Jabbour and Renner 2011; Salvo et al. 2011; Manafzadeh et al. 2014; Lauterbach et al. 2019; Moharrek et al. 2019; Peterson et al. 2019; Mahmoudi Shamsabad et al. 2020). Two main hypotheses have been proposed regarding the temporal framework of dispersals from IT to the Mediterranean region: (1) during the early to late Miocene (Zohary 1973; Manafzadeh et al. 2014); and (2) during Late

Quaternary (Magyari et al. 2008). On the other hand, it is unknown whether IT also constituted an important source of xerophytic lineages for other hotspots of the rest of Eurasia.

Integrative studies on mega-genera based on well-resolved phylogenies are needed to significantly advance in our understanding of the relative roles of geological, climatic, biogeographic and ecological factors in the diversification of the IT flora (Manafzadeh et al. 2017). Here, we studied one of the largest plant radiations of the IT: the genus *Jurinea* Cass. (tribe Cardueae, Compositae), which includes ca. 200 species, of which ca. 70% occur in the IT (Susanna and Garcia-Jacas 2007). *Jurinea* spans throughout Eurasia from latitude 20ºN to 60ºN, and from longitude 10ºW to 120ºE, and it is found in species-rich areas of the IT, especially in Iran, Afghanistan, and Central Asia Mountains (Rechinger and Wagenitz 1979; Szukala et al. 2019). The genus is also present along the Mediterranean Basin, central-eastern Europe, European Russia, Himalayas, and East Asia. Different growth forms are found within the genus, e.g., annual and biennial plants, rosette acaulescent perennials, rosette scapose perennials, cushion acaulescent perennials, prostrate woody-stemmed plants, and subshrubs (caulescent perennial forms with leafy stems). Ecologically, *Jurinea* species mainly grow on mountain and rocky slopes, grasslands, steppes, pine-forest stands, forest margins, lake and riverbanks, and subalpine-alpine meadows (Iljin 1962; Danin and Davis 1975; Rechinger and Wagenitz 1979). The predominant altitudinal range is around 1000–2000 m, although some species grow near the sea-level (like *J. kilaea* Azn.; Danin and Davis 1975) and others up to the alpine belt (4500 m like *J. gilesii* (Hemsl.) N.Garcia, Herrando & Susanna; Rechinger and Wagenitz 1979).

The complex taxonomy of the genus has contributed to its understudied status. Historically, up to nine satellite genera have been recognized within it, though a recent phylogeny of subtribe Saussureinae based on Hyb-Seq data has shown that all of them are more properly accommodated as part of the large genus *Jurinea* (Herrando-Moraira et al. 2020). The most complete phylogeny of *Jurinea* published to date includes only 81 spp., which represents one third of the genus diversity (Szukala et al. 2019). Although it provides a substantial contribution on the phylogenetic relationships, it is still notably incomplete in terms of taxon and biogeographic sampling, as it lacks many representatives from Central Asia, and it does not include a diversification analysis with explicit temporal and spatial frameworks. Therefore, the evolutionary history of the group is still to be unraveled.

In sum, our study aims to decipher the diversification history of *Jurinea* with the central goal to fill the knowledge gap concerning the evolution of the IT flora, investigating the three classical key issues addressed in radiation studies: tempo, mode, and diversification drivers. We did so by integrating phylogenetic data with biogeographic history inference, diversification modelling, ancestral state reconstruction, and ecological niche modelling. Specifically, we first obtained a highly resolved time-calibrated tree based on high throughput molecular methods yielding 502 family-specific loci and including a wide taxon sampling (77%). Then, we conducted an analysis of biogeographic history inference to elucidate the relative contributions of *in situ* speciation versus vicariance to the IT flora, and to explore the potential impact of range expansions and contractions in the diversification dynamics of *Jurinea*. Next, we explicitly tested whether species diversification varied with time, climate or species-diversity; and whether the major climatic shifts that occurred during the PPT significantly impacted diversification dynamics. Furthermore, we investigated the relative importance of niche conservatism vs. niche divergence in speciation by measuring the phylogenetic signal of the environmental niche. Last, we identified stable climatic areas since the Pliocene to the present to assess how regional climatic fluctuations may have impacted *Jurinea*.

2. Materials and methods

Methods are described in greater detail with full references in the Supplementary Information.

Sampling strategy

To reconstruct the evolutionary history of *Jurinea*, we sampled 187 species representing ca. 77% of the genus diversity covering all distribution areas (see Appendix 1 for a complete list of species). Plant material was extracted from dried leaves of herbarium collections or field expeditions. Based on previous phylogenetic studies (Susanna et al. 2006; Barres et al. 2013; Herrando-Moraira et al. 2019a), we added 58 outgroup species to estimate the time divergence of *Jurinea* lineages. In total, 245 species were analyzed, from which 131 have been sequenced for the first time in this study and 114 taken from previous ones (see Appendix 1).

Laboratory workflow

We followed the library prep protocol and sequence capture workflow detailed in Herrando-Moraira et al. (2018, 2019a, and 2020) based on Compositae COS 1061 loci kit (Mandel et al. 2014) except for the pooling step: here, groups up to eight libraries were arranged (500 ng in total per pool). The final enriched libraries were sequenced using 100 bp paired-end reads in the DNA Sequencing Core CGRC/ICBR of the University of Florida using an Illumina HiSeq 3000 (Illumina, USA) or in Macrogen Co. (Seoul, South Korea) in an Illumina HiSeq 4000 (Illumina, USA).

Target sequence extraction

A first read quality assessment was conducted with FastQC v.0.10.1(https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Trimmomatic v.0.36 (Bolger et al. 2014) was used to remove the adapters and clean the reads, and HybPiper v.1.3.1 (Johnson et al. 2016) to extract the target loci. We then removed in downstream analyses all loci flagged as potentially paralogous by HybPiper. Finally, MAFFT v.7.266 (Katoh and Standley 2013), trimAl v.14 (Capella-Gutiérrez et al. 2009), and FASconCAT-G v.1.02 (Kück and Longo 2014) were used sequentially to obtain the loci-separated alignments and the concatenated supermatrix. In order to reduce phylogenetic noise, we excluded positions with high substitution rates, following Fragoso-Martínez et al. (2017). However, after removing a total of 6109 positions from the initial supermatrix (159,512 bp), the same unsupported nodes remained (see comparison in Supplementary Fig. S1). Because of that, we used the initial

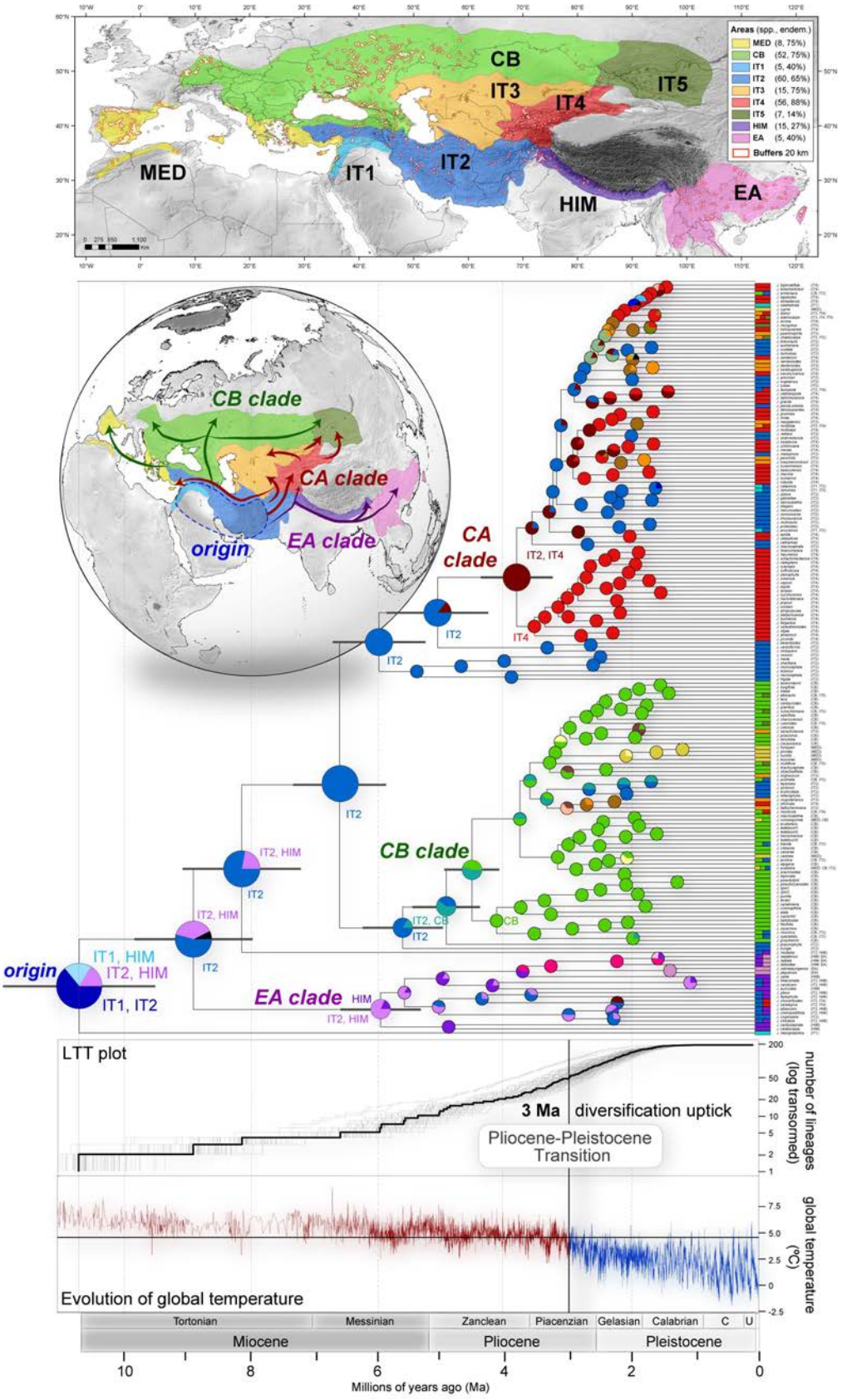


Figure 1. Distribution of *Jurinea* species diversity across the nine biogeographic areas (justification in supplementary text and descriptions in Supplementary Table S1) and buffer zones, which represent areas around 20 km of each occurrence record (2691 points; Appendix 2). Phylogenetic reconstruction of *Jurinea* diversification inferred from the supermatrix of 502 COS loci by maximum likelihood analysis (in RaxML), time-calibrated (in treePL), and reconstructed ancestral biogeographic areas (in RASP, BioGeoBEARS). Colors of tree nodes indicate the most likely estimation of ancestral areas. Tips are also colored according to the current geographic distribution of each species. Abbreviations of areas for inferred ancestral states are only specified for the main tree backbone nodes, for information about the rest of nodes and all color correspondence see Supplementary Table S6 and Supplementary Figs. S10 and S11. Gray bars on nodes show the 95% of confidence intervals (CI). The graphs below phylogeny show: the upper one, a lineage-through-time plot (LTT) that indicates the number of lineages emerged (log-transformed) in a temporal line. The black line represents the maximum clade credibility (MCC) tree chronogram and the grey ones correspond to 100 time-calibrated trees. The lower graphic shows the evolution of global mean temperature (ºC) since Miocene to present. Abbreviations of areas: MED = Mediterranean; CB = Circumboreal; IT = Irano-Turanian; HIM = Himalayas; EA = Eastern Asia. Abbreviations of time stages: C = Chibanian; U = ~~U~~pper".

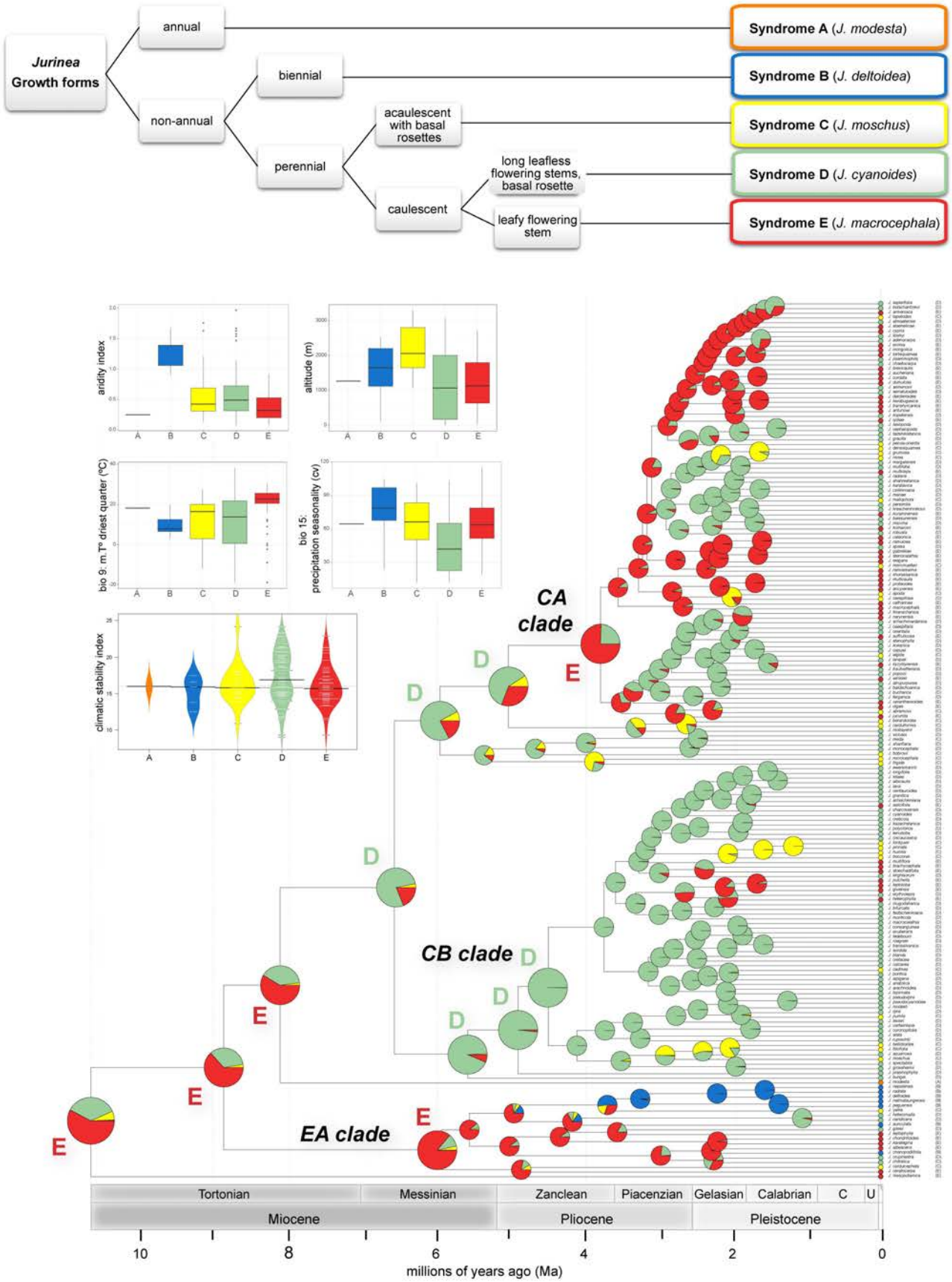


Figure 2. Results of analyses on growth form evolution. Dichotomic scheme of morphological classification of *Jurinea* according to five main morphological syndromes (coded as A, B, C, D, and E), followed by an example species of each category. For each morphotype, boxplots representing variance of several climate variables based on their current distribution areas (aridity index, bio 9, and bio 15), altitude, and beanplot showing the magnitude of paleoclimatic variation (climatic stability index). Phylogenetic reconstruction of *Jurinea* inferred from the supermatrix of 502 COS loci by maximum likelihood analysis (with RaxML), time-calibrated (with treePL), and reconstructed ancestral morphological syndromes (with *diversitree*). Colors of tree nodes indicate the most likely estimation of ancestral morphologies. Tips are also colored according to the current morphotype of each species. Abbreviations of growth forms for inferred ancestral states are only specified for the main tree backbone nodes. Abbreviations of time stages: C = Chibanian; U = ~~U~~pper".

unfiltered dataset in downstream analyses.

Phylogenetic inference and dating analyses

Phylogenetic inference of *Jurinea* was conducted twice, using concatenation and coalescence methods. For the first method, we performed a maximum likelihood (ML) analysis based on the supermatrix using RAxML-NG (Kozlov et al. 2019) from the CIPRES Science Gateway v.3.1 (Miller et al. 2010). Specifically, we run a search for the best-scoring ML tree and a slow bootstrap (BS) analysis, with ten randomized and parsimony starting trees and setting each locus as a different partition under the GTR+G nucleotide model. For the BS analysis, we applied the auto-stopping criterion "autoMRE". Branches with BS values > 70% were considered as statistically supported (Hillis and Bull 1993). The resulting tree was visualized with FigTree v.1.4.3 (http://tree.bio.ed.ac.uk/software/figtree/). The coalescence inference was performed with ASTRAL-III v.5.5.3 (Zhang et al. 2018) based on the individual gene trees for each locus previously obtained with a RAxML analysis applying the same settings as for the supermatrix, but with 200 resampling replicates. The local posterior probability (LPP) was used as a branch support metric, considering strongly supported branches with a LPP > 0.95 (Sayyari and Mirarab 2016).

Due to the large-size of the dataset, we used the penalized-likelihood approach implemented in treePL (Smith and O'Meara 2012) to time-calibrate the best-scoring ML tree obtained with the concatenation approach. We applied the same procedure and calibration points as in Herrando-Moraira et al. (2019a). As a brief description, we obtained the confidence intervals (95% CI) in the estimated node ages running 100 independent treePL analyses with calibration points constrained to a set of random values generated under normal and lognormal distributions, depending on each point (see Herrando-Moraira et al. 2019a). The resultant time-calibrated 100 trees were introduced in TreeAnnotator v.1.7.5 (Drummond et al. 2012) to obtain the maximum clade credibility (MCC) tree chronogram with median node heights and corresponding CIs. This final dated MCC tree is provided in newick format in Supplementary Material.

Ancestral area reconstructions

To infer the most probable ancestral ranges of *Jurinea* lineages, we performed a biogeographical analysis using the software RASP v.4.02 (Yu et al. 2015), which implements the R package *BioGeoBEARS* (Matzke 2013). As input we used the chronogram tree generated from treePL in which all species other than *Jurinea* were pruned. We defined nine biogeographic areas (see Fig. 1 and Supplementary Table S1) according to data on floristic regions (Takhtajan 1986), biogeographic syntheses of the IT (White and Léonard 1991; Djamali et al. 2012; Manafzadeh et al. 2017), and global climate delineations (Köppen and Geiger 1936; Beck et al. 2018).

A first model comparison step was run for the following models: Dispersal–Extinction–Cladogenesis (DEC; Ree et al. 2005), and the likelihood versions of Dispersal–Vicariance (DIVA-like), and BayArea (BAYAREALIKE). The addition of the parameter jump dispersal or founder event speciation "J" (Matzke 2013) to the models was discarded due to the recently reported possible statistical and conceptual problems derived from its inclusion (Ree and Sanmartín 2018). The DEC model resulted as the best-fit one according to AIC values (Supplementary Table S2), and thus it was used to compute the final analysis.

Ancestral state reconstruction of growth form

Reconstruction of ancestral growth form was performed on the *Jurinea* chronogram. We classified *Jurinea* species into five growth form categories based on our own observations and descriptions from local floras: (A) annuals; (B) biennials; (C) acaulescent perennials; (D) caulescent perennials formed by basal rosettes with leafless flowering stems; and (E) caulescent perennials with leafy flowering stems (see dichotomic diagram in Figure 2). Inferences about growth form evolution was done using Markov models (Mk) in the R package *diversitree* (FitzJohn 2012). Four possible discrete Mk models were fitted and compared using AIC values, a null model with different transition rates (ARD) plus three constrained models: a symmetrical model in which transition rate between any two states do not differ (SYM), a symmetrical model with equal rates for the three types of perennial lifeforms (SYM-ERp), and a model with all rates equal (ER).

Diversification analyses

As a first exploratory analysis, we outlined the accumulation of lineages over time with a lineage-through-time plot (LTT; Nee et al. 1992) with the R package *ape* and function *mltt.plot* (Paradis and Schliep 2019) setting in log scale axis "y" (number of lineages). As input, we used the maximum clade credibility tree chronogram and the 100 independent dated trees extracted from treePL, which were used to account for the uncertainty of the node age estimations (see details in Herrando-Moraira et al. 2019a).

We investigated diversification dynamics of *Jurinea* following a hypothesis-driven approach to avoid identifiability problems (Louca and Pennell 2020; Morlon 2020), i.e., we defined a set of diversification models to be fitted and compared in order to test explicit evolutionary hypotheses. Specifically, we tested whether species diversification rates varied (linearly or exponentially) with time, climate, or species-diversity, and included as null models constant-rate ones. We also tested for the effect of the Pliocene-Pleistocene Transition (PPT) by fitting diversity-dependent models in which diversification parameters are allowed to shift at a fixed time (in our case, 3 Ma). Finally, to account for the fact that *Jurinea* is composed of three main subclades with different ages and geographical contexts (and thus each subclade may be at a different stage of diversification), we fitted diversity-dependent models in which we allow one of the subclades to undergo its own diversification dynamics by enabling a decoupling of diversification rates and carrying capacities (Etienne et al. 2012). We only fitted models allowing variations in speciation rate to avoid potential flaws as indicated by Burin et al. (2019) on models assuming variable extinction rates. To account for incomplete taxon sampling, we applied in each model the analytical correction corresponding to the sampling fraction. The best fitting-model was selected based on AICc. All analyses were run with R packages *RPANDA* (Morlon et al. 2016) and *DDD* (Etienne et al. 2012).

Given that the best-fitting model according to AICc was a diversity-dependent model in which the diversification dynamics of the Eastern-Asian subclade were decoupled from the main clade, we performed a second round of diversification analyses: we fitted the set of models to the phylogenetic tree obtained after dropping the EA subclade plus *J. mesopotamica* (hereafter we refer to this subset tree as CB+CA), with the main aim to further investigate the drivers of diversification of the two main subclades of *Jurinea* (Circumboreal and Central-Asian ones).

Environmental niche: data assembly and analyses

To explore the impact of environmental factors in the evolution of main *Jurinea* lineages, we conducted comparative niche analyses. In total, we gathered 2691 records for 158 species, representing 84.5% of our taxon sampling (see Appendix 2 for the complete list of occurrences and sources). The gathered localities were extracted from several sources: GBIF portal, research papers, online floras and atlases, herbarium collections, and virtual herbaria. As environmental predictors, we collected an initial set of 45 variables (at 2.5 arc min or coarser resolutions) related with climate, topography, vegetation, soil properties, habitat heterogeneity, and solar radiation (Supplementary Table S3). We extracted the values of variables for each locality with ArcGIS v.10.2.2 (Esri, Redlands, California, USA 2014). We reduced the initial 45 variables to a set of 21 uncorrelated ones based on the results of a Pearson correlation analysis (Supplementary Fig. S2) and a principal component analysis (PCA) (Supplementary Table S4) to avoid model overfitting. See details of variable selection in Supplementary Information. Two datasets were tested, one with all occurrences and values, and another with the mean values for each species. Given that similar results were obtained (Supplementary Fig. S3C and S3D), we used the mean values as a summary of the species average potential niche.

We identified the variables responsible for niche variation of the main phylogenetic lineages of *Jurinea* through a standard PCA analysis conducted in R v.3.6.3 (R Development Core Team, 2019) using *FactoMineR*, *factoextra* (for PCAs) and *ggplot2*. To test the hypothesis of niche conservatism, expansion, and divergence among lineages, we also conducted the PCA-env analysis designed by Broennimann et al. (2012) with the R script reported in Herrando-Moraira et al. (2019b).

We reconstructed the most probable ancestral niche states with the R function *fastAnc* from *phytools* package (Revell 2012). As input variables for each species, we used the PC1 and PC2 scores from the standard PCA analysis, and the most variable environmental factors at ancestral nodes or at interspecific level. The changes in environmental values were visualized with the *contMap* function (also from *phytools*).

Phylogenetic signal in phenotype and environmental niche

We measured the phylogenetic signal of the categorical variable ‘growth form” with δ statistic (Borges et al. 2019) computed in R, and of the quantitative variables ‘altitude” and ‘environmental niche” (PC1 and PC2 values) with Blomberg‘s K (Blomberg et al. 2003) computed in RASP. These analyses were performed to discern between non-adaptive radiation (high phylogenetic signal expected, i.e. closely related species present similar phenotype and niche traits), and adaptive radiation (low phylogenetic signal expected, i.e. close relatives exhibit notable differences in trait values).

Climatic stability inference

We implemented an approach to assess the impact of climatic fluctuations on *Jurinea* diversification and to identify stable refugia or unstable climatic areas where the species occur (see schematic overview in Supplementary Fig. S4). We used the recently published database *PaleoClim* (Brown et al. 2018) at a spatial resolution of 2.5 arc-minutes. It includes surface temperature and precipitation estimates of 19 bioclimatic variables for 12 time periods (see Supplementary Fig. S4A), from the oldest dated at the Pliocene (3.3 Ma) to the present time (1979–2013) extracted from CHELSA project (Karger et al. 2017). Two spatial scales were initially tested, the first corresponding to the nine biogeographic areas and the second to buffer areas defined 20 km around each occurrence record (Fig. 1). Buffers were created with the function *gBuffer* of R package *rgeos* (Bivand et al. 2017). Then, buffers were cut by areas with the function *gIntersection*, considering that areas were previously independently digitalized in ArcGIS in shapefile format. As both scales, areas and buffers, showed similar results (Supplementary Figs. S5 and S6), we retained the first one to present the main results. See Supplementary Information for additional details of climate data extraction.

We followed the approach described in Tang et al. (2018) to identify stable refuge areas, i.e. regions where temperature and precipitation conditions may have remained relatively constant from the Pliocene to present. The standard deviation (SD) was calculated for each pixel in the 12 *PaleoClim* time slice rasters, independently for a subset of temperature (bio1, bio4, bio9) and precipitation variables (bio12, bio15, bio17). Calculations were performed with the ArcGIS tool “Cell Statistics” ignoring NoData values. Then, for each bioclimatic variable, the output SD values were normalized (scaled from 0 to 10) to get a sum of all standardized SDs, a measure that we call “climatic stability index”. The procedure is schematized in Supplementary Fig. S4C. For each species occurrence record, we extracted its corresponding climatic stability index score with the ArcGIS tool “Extract Multi Values to Points”. Then, the average was calculated for the species with more than one occurrence to obtain a final dataset with a unique index value per species. Climatic stability data plots were generated with *boxplot* from *graphics* R package and bean plots from portal http://shiny.chemgrid.org/boxplotr/, a web-based R application. The R scripts used for the paleoclimatic analyses are included in Appendix 3.

3. Results

Phylogenetic analyses

From the total 1061 target loci, 502 were finally retained after discarding potential paralogs. The supermatrix resulted in an alignment of 161,485 bp. Phylogenetic reconstruction using concatenation and coalescent-based approaches returned similar

topologies at deep tree nodes with strong branch supports (Supplementary Figs. S7 and S8), except for the position of *J. mesopotamica* Hand.-Mazz. Three main clades were recovered, which mostly corresponded to the three main geographic groups (Fig. 1). Hereafter, we refer to these clusters as "Eastern Asia clade" (EA clade), "Circumboreal clade" (CB clade), and "Central Asia clade" (CA clade).

Divergence times of Jurinea and biogeographic history inference

All numerical results from divergence time and ancestral area estimates are compiled in Supplementary Material: node age values (Supplementary Table S5), corresponding node identifiers (hereafter IDs; Supplementary Fig. S9), probabilities of ancestral areas for each node (Supplementary Table S6 and Supplementary Fig. S10) and inferred areas appended to nodes on the phylogenetic tree (Supplementary Fig. S10).

The estimated stem and crown ages for *Jurinea* were 12.8 Ma (11.5–14.5 95% CI) and 10.7 Ma (9.5–12.0), respectively. The most probable ancestral area for the root node of *Jurinea* was IT1–IT2 region (Mesopotamia-Iranian Plateau; Fig. 1). The first main divergence event occurred at 8.9 Ma (7.8–10.0 95% CI) when the EA segregated. In the Messinian period, the EA clade started to diverge at 5.9 Ma (5.2–6.5 95% CI) from an IT2 ancestor, giving rise mainly to Himalayan and East Asian lineages. Subsequent tree backbone nodes showed that the most likely ancestral area was IT2, pointing out that Iranian Plateau was the focal zone of subsequent dispersals to adjacent regions (Fig. 1). In the Miocene-Pliocene transition a second main dispersal occurred to the CB region, which started to diverge around 4.5 Ma (5.1–4.0 95% CI) in a IT2–CB transition, and spread later towards the following regions: IT2, IT3, IT4, IT5, CB, and MED (Fig. 1). Similarly, in the Miocene-Pliocene transition the other main range expansion took place, dated to around 3.8 Ma (3.4–4.6 95% CI) with an ancestral area in IT2–IT4, triggering the origin of the CA clade, which contains lineages that colonized all IT regions, CB, and MED (Fig. 1).

Growth form evolution

According to AIC values, the best fitting model of growth form evolution was SYM, followed by SYM-Erp (ΔAIC = 0.47; Supplementary Table S7). According to SYM, transitions were most frequent between different types of perennials (q_{54} = 0.104, q_{43} = 0.047, q_{53} = 0.037) than between perennials and biennials/annuals (q values <0.02). Congruently, according to the second best model, transition between perennial growth forms were also the most frequent ones ($q_{perennial}$ = 0.073). Indeed, most of *Jurinea* species show perennial biotypes. The most heterogeneous group in growth forms is the EA clade, which harbours species with four of the five distinct growth forms (all except annuals) (Fig. 2). Conversely, the CB clade is the most uniform, with type "D" (perennials with leaf rosettes and leafless flowering stems) predominating and with few transitions to types "C" (acaulescent perennials with leaf rosettes) and "E" (subshrubs with leafy flowering stems). The CA clade shows a wide heterogeneity of growth forms (syndromes "C", "D", and "E", being the most abundant the last two), besides exhibiting large disparities among closely related species.

Diversification analyses

The maximum slope in LTT plot occurred around 3–1.5 Ma (Fig. 1). We compared the fit of 30 models (Supplementary Table S8). According to AICc, the best-fitting model was the diversity-dependent model in which the diversification dynamics of the subclade EA was decoupled from those of the main clade. This model indicated that the speciation rate of EA was nearly half of the main clade (λ_{EA} = 0.85; $\lambda_{main\ clade}$= 1.84) whereas the extinction rate was one order of magnitude higher in EA (μ_{EA}=0.00023; $\mu_{main\ clade}$= 0.00001). Following this model, EA had reached its carrying capacity (K_{EA} = 25, which is indeed the estimated current species diversity), whereas the main clade had not yet ($K_{main\ clade}$ = 342, while *Jurinea* diversity estimations are of ca. 200 species).

Concerning the diversification models fitted for the CB+CA subset (Supplementary Table S9), the best-fitting model according to AICc was the diversity-dependent model with a major shift of diversification dynamics at the Pliocene-Pleistocene transition (PPT, ca. 3 Ma) and null extinction. This model indicated that the speciation rate nearly doubled after the shift (λ_1 = 0.93, λ_2 =1.77). The estimated values of carrying capacity (K) before and after the shift (K_1 = 6163 and K_2 = 331, respectively) suggested that diversity-dependence developed mainly after PPT. The ΔAIC between the best and second-best model was above the threshold of 2, that is typically chosen to discern models with confidence for both datasets (all *Jurinea* and CB+CA subset).

Environmental niche evolution and phylogenetic signal

Environmental variables that mainly contributed to the separation of the phylogenetic clusters were aridity (12.8% of contribution to PC1 axis of standard PCA) and altitude (19.2% in PC2 axis; Figs. 3A and 3B, Supplementary Table S10). In the present study case, altitude was highly correlated (Pearson's r > 0.4) with soil quality and other topographic variables (Supplementary Fig. S2). Standard PCA (Fig. 3B) and PCA-env (Supplementary Fig. S3B) analyses revealed that environmental niches of the three main clades (EA, CB, and CA) overlap considerably (D = ca. 30%; Supplementary Table S11). The clade EA showed the highest niche breadth and turned out to be equivalent to niches found in CB and CA (*p-value* < 0.05; Supplementary Table S11). Comparing the three main clades, the CB showed high niche expansion values with respect the other two clades (33% with respect to EA and 41% with respect to CA;

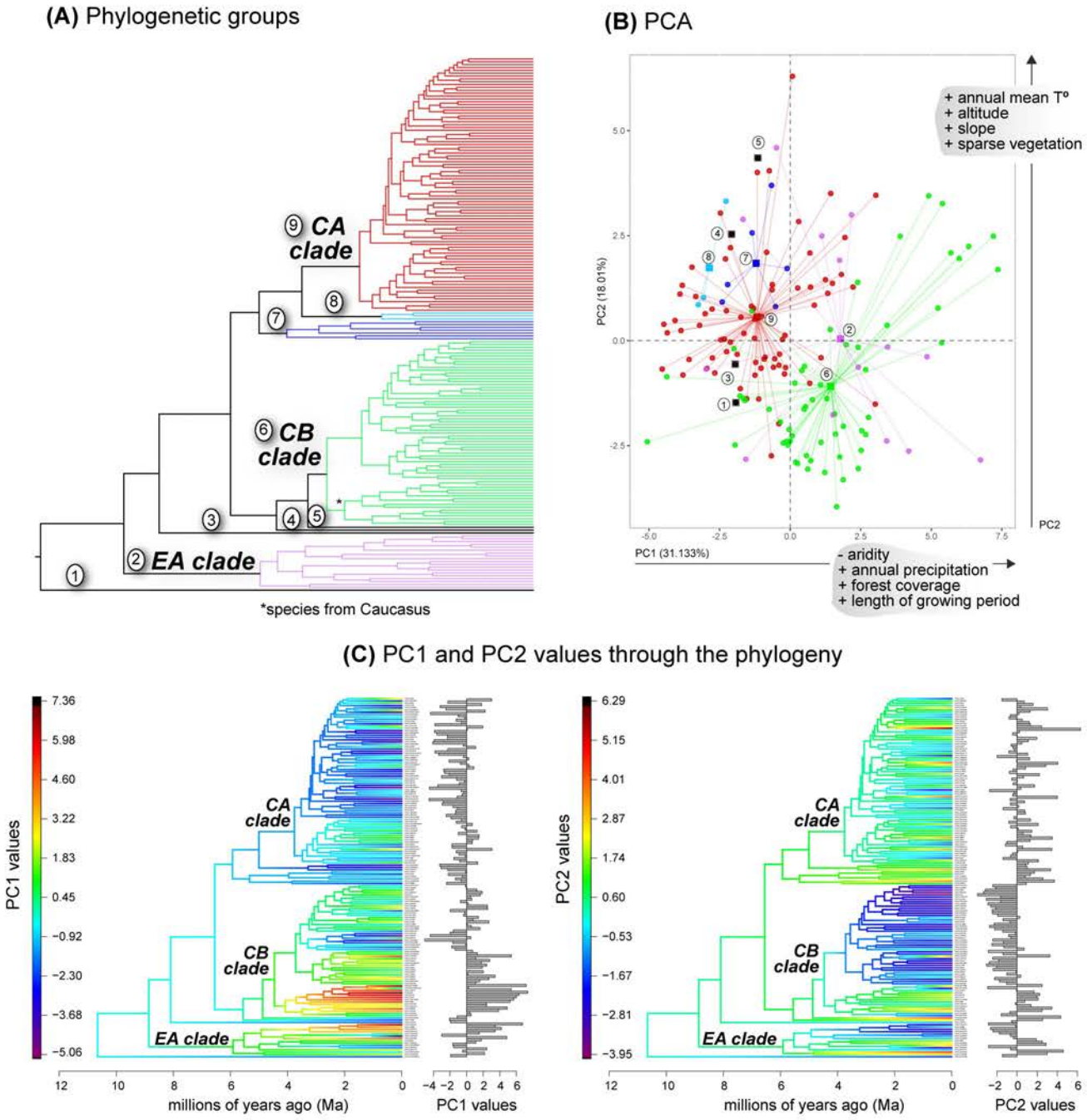


Figure 3. (A) Classification of phylogenetic groups (see Fig. 1). (B) Graphical output of Principal Component Analysis (PCA) analysis of the values of 21 climatic variables for each species, representing the global environmental space and the distribution of each phylogenetic group. (C) Ancestral state reconstructions (obtained with the *fasAnc* function of R package *phytools*) of PC1 and PC2 scores for each species, representing the ecological niche evolution along the phylogenetic history of *Jurinea*. Loading values for each species of both PCs are represented in vertical barplots besides the time-calibrated phylogeny.

Supplementary Table S11).

Ancestral reconstruction of the first two PC axes (Fig. 3C) showed that CB clade was the one including species with most diverging niches. Species from Caucasus showed preference for more humid habitats (Fig. 4, PC1 axis), and the rest of CB species showed preference for lower altitudes (Fig. 4, PC2 axis). We recovered a common trend in CB clade towards, on the one hand, higher values of humidity and length of growing period, and on the other hand, lower values of diurnal range of temperatures, temperatures on driest quarter, and precipitation seasonality, and minor preference for a sparse vegetation coverage (Fig. 4).

In general, the EA and CA clades included species with preferences for a wide variety of environmental conditions, and closely related species showed considerable disparity of values (Figs. 3 and 4). The easternmost distributed species from the EA clade, which are biennials (syndrome ~~–B~~"; Fig. 2), showed a high degree of niche differentiation from other clades of EA towards a

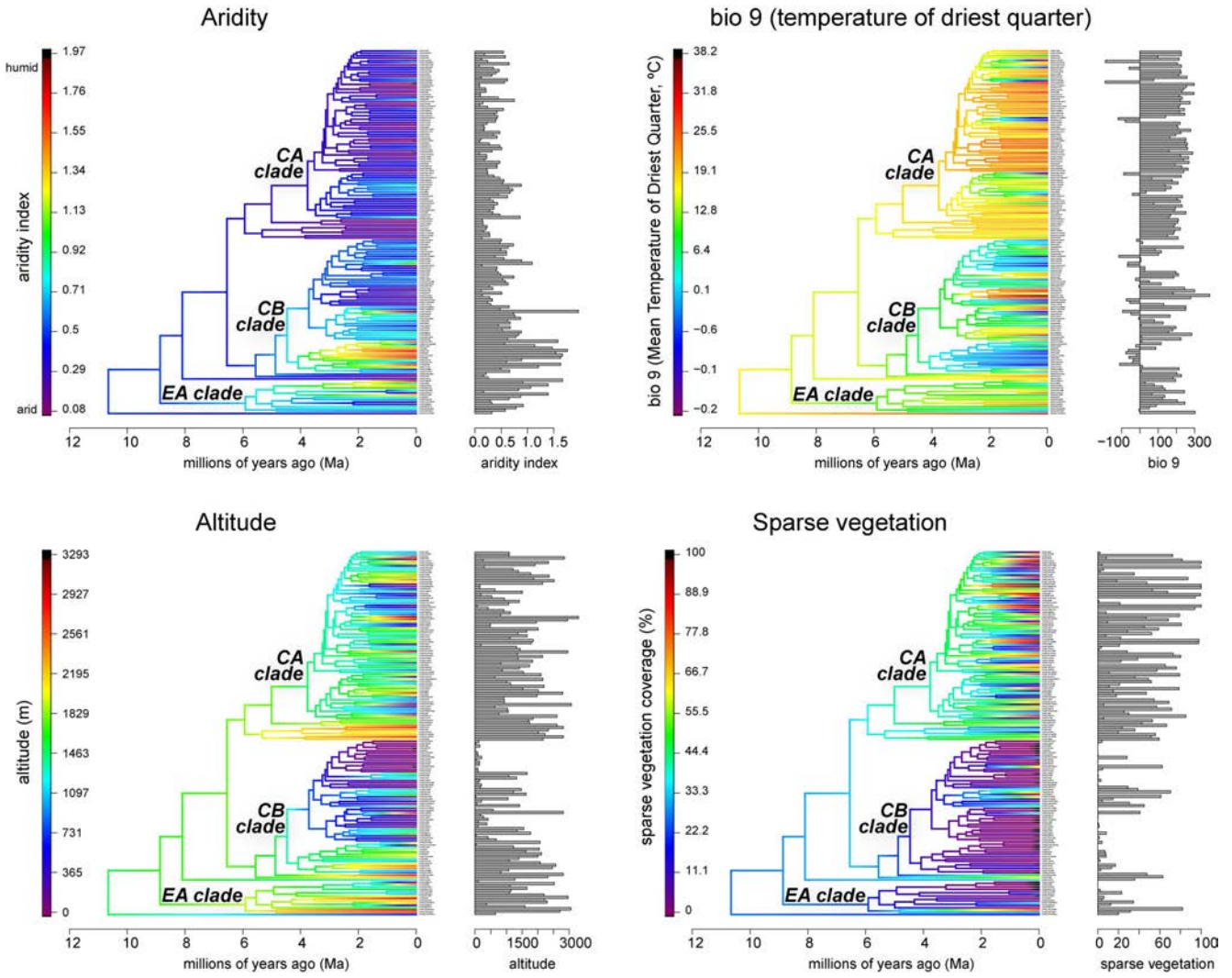


Figure 4. Estimation of ancestral states reconstructed with the *fasAnc* function of R package *phytools* for the most variable environmental traits across the *Jurinea* phylogeny. Loading values for each species are represented in vertical barplots besides the time-calibrated phylogeny.

preference for more humid conditions. For the species of the CA clade, we detected a predominant niche preference for arid habitats (Fig. 4).

Statistically significant levels of phylogenetic signal (Blomberg's $K > 1$; *p-value* = 0.001) were recovered only in CB clade for environmental niche and altitude. The δ-value related to growth form conservatism was also three-fold higher in CB than for the other clades (Table 1).

Paleoclimatic oscillations and long-term stable regions

Paleoclimatic data analyses identified both climatic stable refugia and highly variable regions along IT and surrounding territories (Figs. 5 and 6). We detected local stable climatic conditions since middle-late Pliocene to present in the Mediterranean region, Anatolia, Mesopotamia, Iranian Plateau, central-eastern Himalayas, and south-eastern Asia. As a general pattern, climatic fluctuations affected the northernmost areas (especially for temperature oscillations) and East Asia (especially for precipitation; Fig. 6). Accordingly, the CB clade, which has more northerly distributed species, showed higher median and range values (Q1–Q3) of climatic stability index (Fig. 7A and Table 1), i.e. species occur in regions that had more unstable climate. The top species rich-areas (IT2, IT4, CB) showed medium-high values of climatic oscillations (Fig. 7B).

4. Discussion

The present study offers new evidence on the tempo, mode, and drivers of evolution of the IT flora, which is one of the less studied Eurasian floras despite its high diversity and rate of endemism. The integrative approach applied here, combining data from multiple sources (phylogeny, biogeography, geology, ecology, morphology, and paleoclimate), revealed the importance of climate shifts and topographic heterogeneity as triggers or modulators of species radiations. Importantly, the paleoclimatic analyses highlighted that either long-term climatic stability or moderate climate oscillations could explain high species richness in two biogeographic regions depending on their general topography. In the Iranian Plateau, the high climatic stability detected would explain its high richness of xerophytic elements and its role as a museum and a cradle for the IT flora. Contrarily, regional climatic fluctuations in the Circumboreal

Table 1. Summary of main results of five key questions in *Jurinea* radiation. Species formation and growth form transition processes were calculated according to retention (same inferred area or habit) or shift in internal nodes and tips respect their respective MRCA node. Abbreviations: BS = bootstrap; E-C = expansion-contraction model; IT = Irano-Turanian; LPP = local posterior probability; MRCA = most recent common ancestor; p% = probability in percentage; Phy-signal = phylogenetic signal; spp = species; SS = strong significant phylogenetic signal.

Specific attributes or measurable traits	*Jurinea* (whole tree)	Eastern Asia (EA clade)	Circumboreal (CB clade)	Central Asia (CA clade)
Number of species	187	19	65	89
Estimated sampling fraction	77%	68%	74%	67%
Clade support (BS/LPP)	100/1	100/1	100/1	100/1
Divergence time (crown age)	Late Miocene (10.7 Ma)	Late Miocene (5.9 Ma)	Pliocene (4.5 Ma)	Pliocene (3.8 Ma)
Ancestral areas (p%)	IT1–IT2 (64%)	IT2–HIM (81%)	IT2–CB (55%)	IT2–IT2 (100%)
Species formation process	*in situ* (60%)	colonization (76%)	*in situ* (70%)	*in situ* (57%)
Geographic scale (spp/km^2)	5.9×10^{-6}	1.9×10^{-6}	3.3×10^{-6}	11.1×10^{-6}
Growth form conservation or transition	conservation (81%)	conservation (64%)	conservation (90%)	conservation (78%)
Phy-signal growth form: δ	6.21	1.33	18.89 SS	4.15
Phy-signal niche PC1: K	0.58	0.78	1.29 SS	0.80
Phy-signal niche PC2: K	0.52	0.96	1.19 SS	0.74
Phy-signal altitude: K	0.53	0.62	1.22 SS	0.75
Climatic stability index: median (Q1-Q3)	16.05 (14.70–18.68)	16.20 (15.60–17.35)	17.8 (14.63–20.70)	15.9 (14.8–18.0)
Radiation type proposed	Mixed model	Adaptive	Non-adaptive	Adaptive
Diversification drivers and history	- Climate changes (through cold-arid conditions) - Pre-adaptation hypothesis to aridity - Geographic speciation (longitudinal, latitudinal, altitudinal)	- Niche partitioning, adaptive radiation on an ecologically constrained adaptive landscape - Depauperate lineage without enough evolvability potential to diversify in mesophilous habitats	- Range and niche expansion to humid conditions - Niche and morphology conservatism among sister species - Past climatic oscillations stimulated speciation by allopatric mode (E-C)	- Climate fluctuations and active orogeny triggered speciation in Pamir-Tian Shan mountains - High morphological and niche lability among sister lineages in response to adaptation to local conditions

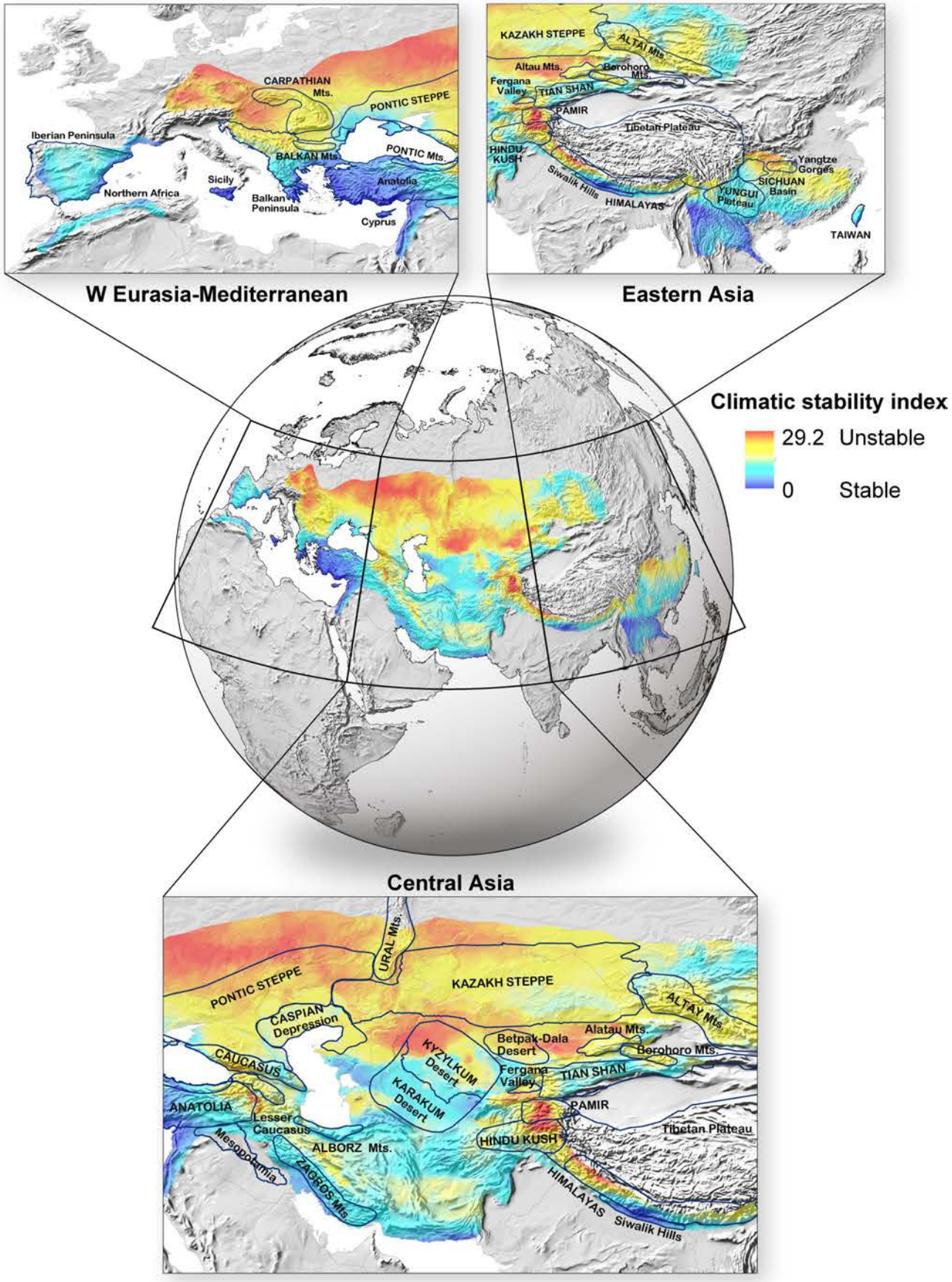


Figure 5. Mapped values of climatic stability index for the study area from Pliocene (3.3 Ma) to present, at a resolution of 5 × 5 km. Colors range from blue for low standard deviation (SD) values, which represent areas with low climatic fluctuations from Pliocene to present, to red for high SD values, which show areas where high climatic fluctuations probably took place. Main geographic locations (e.g. mountains, deserts) are indicated in the zoomed maps. This index is based on data from six variables (bio1 = annual mean temperature; bio4 = temperature seasonality; bio9 = mean temperature of driest quarter; bio12 = annual precipitation; bio15 = precipitation seasonality; and bio17 = precipitation of driest quarter).

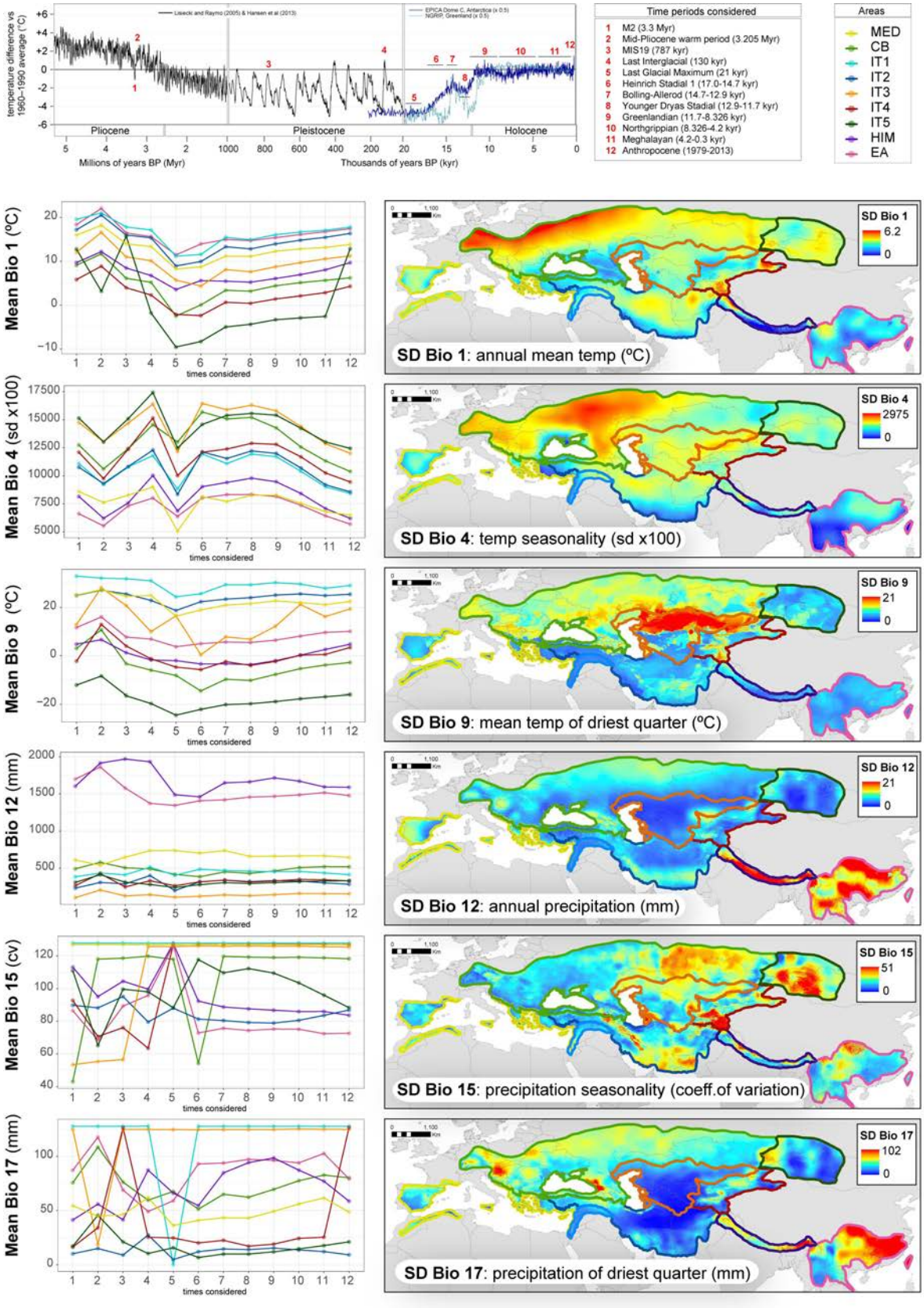


Figure 6. Representation of climatic variable oscillations through the 12 time intervals (see Supplementary Fig. S4) of *PaleoClim* dataset (Brown et al. 2018). On the left, we show values for temperature variables (bio1, bio4, bio9) and precipitation variables (bio12, bio15, bio17) for each of the nine biogeographic areas defined. On the right, for each variable and biogeographic area, the standard deviation (SD) along the 12 time periods considered is mapped.

region and mountains of Central Asia fostered allopatric diversification, especially in lineages adapted to semi-arid and cold areas.

Cold-arid climatic conditions as triggers of *Jurinea* radiation

The main evolutionary events of *Jurinea* were probably substantially influenced by climate-related turnovers, especially those that led to colder and drier conditions: (1) the origin of the genus (10.7 Ma; Fig. 1) occurred during a major global climate change, turning from warm-tropical and subtropical to cool-temperate at the onset of global cooling and marked seasonality during mid-late Miocene (Potter and Szatmari 2009; Pound et al. 2012; Herbert et al. 2016); (2) long-distance dispersals from IT to other floristic regions (~ 7–4 Ma; Fig. 1) occurred around the Miocene-Pliocene boundary, a period characterized by global biotic turnovers (LaJeunesse 2005), such as the expansion of C4 grasslands (Shen et al. 2018) in response to a major cooling trend between 7.0 and 5.5 Ma (Holbourn et al. 2018) and the aridification of Central Asia (Miao et al. 2012); and (3) diversification dynamics of the main lineages of *Jurinea* (i.e., the CB and CA clades) suffered a significant diversification shift during the Pliocene-Pleistocene Transition (ca. 3 Ma; Fig. 1), in which the net diversification rate almost doubled, from $\lambda = 0.93$ to $\lambda = 1.77$. During this later period, climate turned colder and drier at the onset of the Northern Hemisphere glaciations when large continental ice sheets appeared, resulting in important environmental changes (Willeit et al. 2015). Lineages that could not adapt to novel environments either dispersed to suitable habitats or faced extinction. *Jurinea* seems to have taken evolutionary advantage of these climate shifts towards dry-cold regimes. The diversification increase detected for the last 3 million years could be related to geographic range modifications or to adaptation to these new ecological opportunities and niches deserted by other plant groups unable to adapt to plummeting temperature and precipitation. Indeed, during this period *Jurinea* lineages also showed an increasing trend of growth form lability (Fig. 2) and ecological niche shifts (Figs. 3 and 4), probably as an adaptation to novel habitats and new geographic areas. Overall, *Jurinea* is an excellent example of the extensive impact of gradual global cooling trend over the past 15 Ma on diversification history and expansion of the temperate biota of the Northern Hemisphere (Folk et al. 2019, 2020 and references therein).

Insights into evolutionary and biogeographic hypotheses about the Irano-Turanian flora

In relation to the temporal origin of typical xerophytic species within the IT, we found that *Jurinea* originated in the middle-late Miocene, ca. 10.7 Ma (9.5–12.0; Fig. 1), a more recent origin than that suggested for the IT elements (Manafzadeh et al. 2017). Spatially, *Jurinea* started to diverge in Western Asia around Mesopotamia and the Iranian Plateau (region IT1–IT2 here; Fig. 1). Within IT, Western Asia is a centre of origin for numerous plant radiations that also occurred during middle-late Miocene (Barres et al. 2013; Karl and Koch, 2013; Lauterbach et al. 2019; Peterson et al. 2019). The definitive closure of the Mesopotamian Seaway in the middle Miocene at 13.8 Ma (Bialik et al. 2019) could have promoted this deep-lineage divergence, as new lands to be colonized were available (Potter and Szatmari, 2009; Herbert et al. 2016).

The colonization of adjacent regions by IT elements could be the result of independent dispersion waves. In *Jurinea*, the first colonized region out of IT was East Asia (late Miocene), the second the Circumboreal region (Pliocene), and the last one the Mediterranean region (Pleistocene), which was colonized repeatedly by different clades (Fig. 1). Among these regions, the genus actively diversified only in the Circumboreal region, whereas in East Asia the net diversification rate was significantly lower than in other regions. Internal movements and range expansions between IT areas (IT1-IT5) occurred during the Pliocene-Pleistocene boundary.

The Iranian Plateau as both a museum and a cradle for plant diversity

Our results suggest a double role of the Iranian Plateau as a ~~"m~~museum" (old-lineage persistence) and a ~~"c~~radle" center (*in situ* speciation; cf. Moreau and Bell 2013) for *Jurinea*, which has also been documented for other IT radiations such as *Acantholimon* (Moharrek et al. 2019). For instance, most of the deepest nodes of *Jurinea* were distributed in this area, according to our biogeographic estimations; in contrast, some recent nodes gave rise to several endemics (Fig. 1). The dual role of this region as a museum and a cradle is likely linked to one of our main findings: here we demonstrated for the first time that the Iranian Plateau is a region characterized by long-term high climatic stability (Fig. 5 and 7B), one of the main causes determining high species richness and endemism of world's biodiversity hotspots (Harrison and Noss 2017). Specifically, our analysis of paleoclimatic data revealed a low temporal variation in IT2 for annual precipitation and, particularly, precipitation of driest quarter (bio 12 and bio17, respectively; Fig. 6 and Supplementary Fig. S6); thus, the precipitation regime remained rather constant since the Pliocene, without abrupt regional changes even during the periods of strongest climatic changes such as the LGM (time ~~"5~~" in Fig. 6 and Supplementary Fig. S5). This factor probably favoured the persistence of *Jurinea* species as well as other xerophytic IT elements.

Takhtajan's (1986) old hypothesis is once again verified here: the IT flora originated along the Iranian Plateau and served as biodiversity source for the adjacent regions. For some millions of years during late Miocene (ca. 10.7–6 Ma) *Jurinea* probably remained within the Iranian Plateau (Fig. 1). Its diversification was constant during the so-called period of the ~~"P~~ikermian chronobiome", when arid-seasonal climates and open environments dominated in Eurasia (Hurka et al. 2019). The main transregional dispersal events took place around 6–4 Ma during the Miocene-Pliocene transition, when the genus successfully colonized three geographic areas: East Asia, the Circumboreal region, and Central Asia. This period is coincident with an intense phase of uplift and deformation along the Iranian mountain ranges: Zagros, Alborz, and Kopeh Dagh (Mouthereau 2011), which provided isolated geographic areas that promoted allopatric speciation, as well as new habitats that created opportunities for ecological speciation. It is also coincident with the onset of the Central Asia aridification (7–5.3 Ma; Sun et al. 2015) and a phase of intense cooling in Northern Hemisphere (the

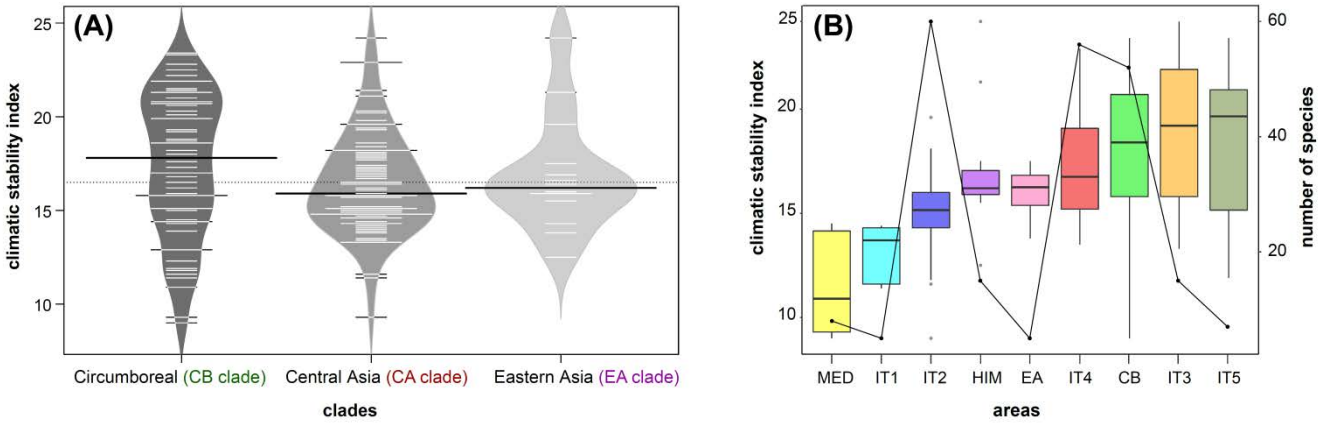


Figure 7. Plots showing range of climatic stability index values for (A) the three main clades and (B) the nine biogeographic areas (see Fig. 1) based on mean index values for each species. The graphical representation type is a beanplot in (A), in which white lines represent species values, and a boxplot in (B) with a superposed line showing the species richness for each region.

–Late Miocene Cooling"; 7–5.4 Ma; Zachos et al. 2001; Herbert et al. 2016). These climatic changes likely led to a decrease in native competitors and herbivores as shown by the regional extinction events of mammalian biochronological indicators (Ge et al. 2012; Casanovas-Vilar et al. 2010 and 2011). Consequently, new lineages such as *Jurinea*, other xerophytic plants and C4 grasses (Huang et al. 2007; Arakaki et al. 2011) could have spread as a result of antagonists decrease (Yoder et al. 2010) and to their pre-adapted niches to arid habitats. Here, we confirmed the pre-adaptation hypothesis based on the phylogenetically conserved values obtained for aridity (Fig. 4), which may have resulted in higher rates of successful establisment in new available areas by *Jurinea* lineages (Peterson 2003), as documented in alpine radiations by cold-adapted lowland clades (Ye et al. 2019; Uribe-Convers and Tank 2015).

Biogeographic history of *Jurinea*: Successful dispersal pathways and geographic dead ends

The first dispersal of *Jurinea* out of IT occurred at 5.9 Ma and led to the emergence of EA clade, extending the genus from the Iranian Plateau towards the Himalayas (probably across the mountain ranges connecting IT2 and HIM), Central Asia Mountains (IT4), and East Asia (subtropical China, Myanmar, Vietnam, and Taiwan). A similar dispersal corridor has been postulated for other IT elements (e.g., *Gagea* Salisb., Peterson et al. 2019) but, as in *Jurinea*, this dispersal event originated a species-poor clade. Below, we discuss hypotheses that could explain why some IT elements found a dead end along this East Asia corridor.

During the Miocene-Pliocene transition, early diverging clades that coexisted in the Iranian Plateau gave rise to CB (at 4.9 Ma) and CA clades (at 3.8 Ma), both of which colonized large parts of Eurasia (Fig. 1). The dispersal direction of the CB clade was inferred to occur from Central Anatolia and Transcaucasia to northern latitudes through landmasses between Black and Caspian seas, although one dispersal pathway remained exclusively along the Caucasus mountains. This is a currently known radiation focus with ca. 20 endemic species (Szukala et al. 2019). Another lineage within CB, with an inferred ancestral distribution in IT2–CB, was mainly dispersed through vast CB territories including N Anatolia, Balkan Peninsula, the river basins of Dnieper, Don, and Volga rivers, the Caspian Depression, or SC Russian uplands. Secondary routes also reached the rest of IT regions: the eastern Aralo-Caspian area (IT3); the Tian Shan (IT4); and SW Siberia, N Xinjiang, and W Mongolia (IT5). However, immigrated lineages from CB back to IT regions originating some species as well. Two main causes may account for such poor diversity: (1) founder effects associated with long distance dispersal, which implies low potential genetic plasticity to adapt to a new environmental space; or (2) high climatic instability of most of the newly colonized areas (especially in IT3 and IT5 regions; Figs. 5 and 6) that would have increased local extinction rates. These areas were later recolonized by recently-diverged *Jurinea* lineages, whose recent origin could explain the absence of large *in situ* radiations.

Jurinea also colonized the Mediterranean Basin several times during the Pleistocene, coincident with the oldest xeric period of the region and the emergence of Mediterranean-type vegetation (ca. 2.3 Ma, Suc 1984). Such transregional dispersal is responsible for the so-called –Kiermack disjunctions" between western Mediterranean and Central Asia (Ribera and Blasco-Zumeta 1998). Contrary to other IT elements that actively diversified in the Mediterranean (Banasiak et al. 2013; Manafzadeh et al. 2014; Peterson et al. 2019), this region did not become a secondary speciation center for *Jurinea* (only ca. 6 endemic species are known), as it would be expected for a long-term climatically stable area (Harrison and Noss, 2017). Reasons are probably two-fold: on the one hand, species of *Jurinea* reached the Mediterranean only in the Pleistocene, a too short time for a radiation to occur. On the other hand, Mediterranean vegetation would have been well-established at the arrival of the *Jurinea* ancestral taxa and, thus, with a low availability of empty niches.

The route towards Central Asia (CA clade) was established during the Pliocene (at ca. 3.8 Ma), probably in the transition zone of IT2–IT4, which is the intersection of various topographically complex mountain ranges (Hindu Kush, Karakoram, Pamir, and Altai). This singular geographic position would offer the opportunity of a rapid expansion along the main

mountain systems of Central Asia, and also of back-colonization of Iranian mountain belts, probably following a "stepping stone" dispersal mode throughout suitable patches at mid-altitude habitats. Members of this clade reached all IT regions, and even colonized Mongolia (IT5), Mesopotamia (IT1), and Cyprus Island (MED). More than half of *Jurinea* sampled species originated within this successful lineage. Unfortunately, a fine-scale discussion about the biogeographic history of this lineage is hindered by low support values recovered within CA clade (Supplementary Figs. S7 and S8). Further exploration of another type of high-throughput molecular markers may be necessary to unravel the CA radiation. COS target loci, which are conserved orthology loci identified from expressed sequence tags (Mandel et al. 2014), do not seem the most adequate markers for disentangling deep-hard species radiations when incomplete lineage sorting and contradictory signals among coding sequences occur, as is probably the case for the CA clade. Such pattern of conflicting gene information has also been suggested for radiation of Andean *Lupinus* L. by Nevado et al. (2016), who found that over 40% of genes had undergone adaptive evolution of coding sequences.

Parallel and independent diversification patterns of three *Jurinea* lineages

Given that the three main clades (EA, CB, and CA) followed independent evolutionary histories and were probably driven by distinct extrinsic forces, we discuss separately their diversification dynamics in the next sections.

The Circumboreal clade: a non-adaptive radiation triggered by climatic oscillations

Contrary to our initial expectations of a clear dominance of IT-distributed species along the phylogenetic tree of *Jurinea*, we recovered a high species rich clade dominated by CB taxa (65 species, representing 35% of the total explored richness). The CB clade originated with the expansion of *Jurinea* from IT2 to CB region at 4.5 Ma, coinciding with a climate reversal from cool to warm-humid conditions known as the "Pliocene climatic optimum" (Sniderman et al. 2016; Jiménez-Moreno et al. 2019). Such reversal probably facilitated northward expansions. In addition, we observe that ancestral CB lineages experienced a niche shift towards more humid and less continental-seasonal climate and less preference for open habitats (Figs. 3 and 4). Accordingly, paleoreconstructions of Pliocene vegetation showed that CB territories were generally dominated by deciduous forests, in contrast to grasslands in IT region (Salzmann et al. 2008).

Disruptive selection could have operated on ancestral populations at the onset of the CB clade radiation, favouring those with evolutionary potential to become different species as they became reproductively isolated as a result to adaptation to different climatic conditions (Stroud and Losos 2016, and references therein). This hypothesis fits well with the wide range of ecological niche diversity detected at the tree backbone in the CB clade (Figs. 3 and 4), where highly ecologically specialized groups emerged (Seehausen 2015; Sundue et al. 2015). However, closely related species conserved their ecological preferences (Wiens et al. 2010), as shown by the high phylogenetic signal recovered for niche, growth form, and altitude of CB clade in contrast to the others (Table 1). This may indicate a diversity-independent evolution, i.e. niche-neutral diversification and a high niche conservatism pattern, pointing for a non-adaptive radiation for the CB clade at more recent times. But then, what could have promoted its successful diversification?

Interestingly, we found that CB clade harbours a high number of species that currently grow in regions that have been climatically unstable since Pliocene to present (Figs. 5 and 7A). Living in territories with highly oscillating climatic conditions could have direct evolutionary implications (Dynesius and Jansson 2000). The geographic distributions of ancestral populations were probably strongly modified at each warm-to-cold or cold-to-warm reversals. Furthermore, migrations would have been successful more often in CB than in other regions due to the lack of great geographic barriers to dispersal (e.g. mountain ranges); most of the region is composed by great plains and river basins and the few mountain ranges rarely exceed 1000–1500 m. Such "range expansion-contraction" process could start with dispersion of propagules into multiple directions; some of them succumbed to extinction via competitive exclusion or drastic local climate changes, whereas others fell into unsuitable habitats. The survivors remained spatially isolated, hindering regular gene exchange with the original population. Over the process, niche preferences and phenotypes were conserved. Accordingly, it seems reasonable to assume that allopatry was the main speciation mechanism within this clade. The effect of past climatic fluctuations as a trigger of speciation was already proposed in other studies (Kadereit et al. 2004; Aguilée et al. 2011; Roquet et al. 2013; Sun et al. 2020), but the present one is the first to explicitly quantify long-term climatic stability to test its impact on a radiation burst.

The Central Asia clade: The classical "species pump" in mountain systems

The CA clade, the most diverse phylogenetic group of *Jurinea* (for which we sampled 89 species, ca. 50% of total richness of the genus), probably originated around 3.8 Ma. We identified two possible triggering factors for the first diverging events within CA. The first one is associated to the cold-arid period in the Asian interior dated to 3.6 Ma (Rea et al. 1998; An et al. 2001; Zheng et al. 2004) and to the MIS M2 glaciation at global scale (3.3 Ma). In response, plant communities experienced marked turnovers to drought tolerant taxa at these time frames (An et al. 2005; Koutsodendris et al. 2019). Congruently, we observed niche changes towards adaptation to higher aridity and open habitats in the origin of CA (Fig. 4), and a growth form reversion to the most ancestral habit (syndrome "E", caulescent perennials with leafy flowering stems; Fig. 2), likely providing adaptation to arid climates in the Mesopotamian-Iranian region (IT1–IT2; Fig. 1). Therefore, aridity could have acted as a speciation stimulus to CA radiation (Stebbins 1952).

The second possible trigger is associated with the Pamir-Tian Shan convergence, which began at ca. 3–5 Ma (Fu et al. 2010; Thompson et al. 2015) when Pamir advanced about 300 km northwards (Burtman and Molnar 1993; Burtman 2000). Additionally, parallel and subsequent pulses of deformation and rapid uplifts of Pamir-Tian Shan mountain belts occurred (Zhang et al. 2013). The uplift-driven diversification hypothesis (Xing and Ree 2017) is discarded here, since the onset of CA mountains

orogeny was long before (Buslov et al. 2008; Amidon and Hynek 2010). However, new ecological opportunities generated by later deformation phases could have triggered the diversification of CA lineage, a time delay that has also been observed in other alpine radiations (Hughes and Atchison 2015; Hauenschild et al. 2017).

We found poor phylogenetic resolution and retention of ancestral niches in the tree backbone of CA clade (Figs. 3 and 4, Supplementary Figs. S7 and S8), suggesting a first rapid diversification event likely by allopatry (similar examples in Gillespie et al. 2020). We hypothesize that, after this initial divergence, the ancestral lineages occupied multiple habitats provided by mountain ranges, partitioning their niche along altitudinal gradients differentiated by a series of variables: topography, microclimate, predator pressure, pollinators, soils, or vegetation structure (Karl and Koch 2013; Sundue et al. 2015; Seehausen 2015). This hypothesis agrees with the lack of phylogenetic signal detected for the global niche and the altitude variable, indicating niche divergence between sister species (Figs. 3 and 4, Table 1). Furthermore, we propose an additional factor that may have fostered speciation in this clade: the climatic oscillations detected in Pamir/S Tian Shan area (Figs. 5 and 6). Such climatic instability could have forced lineages to altitudinal and/or latitudinal migrations leading to reproductive isolation of populations. Niche partitioning among close lineages could have been possible thanks to the high lifeform flexibility (Onstein 2020) that we detected through ancestral growth form inference (Fig. 2) and poor phylogenetic signal in this trait (Table 1), which means a high ability to repeatedly reverse or change habit to adapt rapidly to new environments. Overall, the diversification of CA shows signs of adaptive radiation such as high morphological disparity and weak niche conservatism (Czekanski-Moir and Rudell 2019). In a similar way to the model proposed for adaptive radiations by Seehausen (2015), we suggest that the two main sources of eco-space variability (altitude and time) stimulated several successive cycles of adaptation, speciation, and niche shift.

The depauperate lineage: Himalayas and East Asia clade

In the evolutionary history of *Jurinea*, the Himalayas acted as a bridge between two climatic and floristic regions: the continental, summer-dry Irano-Turanian and the summer-monsoon East Asian regimes (Takhtajan 1986; Djamali et al. 2012). However, *Jurinea* did not explosively radiate in either the Himalayan corridor or in East Asia (Fig. 1 and Table 1), remaining there as a relatively depauperate lineage (Donoghue and Sanderson 2015) despite it shows wide ecological niche breadth (Figs. 3 and 4) and high morphological disparity (Fig. 2 and Table 1). Some species conserved part of their ancestral niche, morphology, and range (IT2), while others diverged by adapting to the East Asia mesic habitats and changing their lifeform to biennials (syndrome "B"; Fig. 2). Bringing all results together, an adaptive process arises as a plausible explanation for the EA clade diversification.

The low species richness of this clade could be due to a greater extinction rate (μ), associated with the fact that it is the most ancient diverging lineage (5.9 Ma), i.e., extinction may have eroded the early burst signal of radiation. However, the speciation rate (λ) for EA was about half of those for CB and CA clades. Thus, we suggest an alternative hypothesis: its low richness could be due to constraints on speciation. As speciation constraint, we suggest that the suitable adaptive landscape space was already almost filled by competitors when *Jurinea* arrived; examples of successful "native" competitors would be the closely-related, highly-speciose genus *Saussurea* DC., and other groups that successfully radiated in the region such as *Delphinium* L. and *Pedicularis* L. (Wen et al. 2014).

Given the xerophytic ancestral background of *Jurinea*, we hypothesize that the evolution from xerophytic towards mesophilous niches is probably complicated for *Jurinea.* In other words, *Jurinea* may not present enough evolutionary potential to successfully diversify in humid environments. Analogous cases have been found in two sister subtribes of Cardueae, Arctiinae and Onopordinae (Herrando-Moraira et al. 2019a). In the Arctiinae, the explosive radiation of the mostly xerophytic *Cousinia* Cass. (ca. 600 spp.) in the IT region contrasts with the depauperate genus *Arctium* L. (27 spp.), which is mainly adapted to mesophilous habitats (López-Vinyallonga et al. 2009; Susanna and Garcia-Jacas 2009). At a smaller scale, the subtribe Onopordinae shows a lineage with greater success in xeric habitats (the genus *Onopordum* L. with 60 spp.) in contrast to seven less-successful small genera in the mesic habitats of CA (*Alfredia* Cass., *Ancathia* DC., *Lamyropappus* Knorring & Tamamsch., *Olgaea* Iljin, *Syreitschikovia* Pavlov, *Synurus* Iljin, and *Xanthopappus* C.Winkl.).

5. Funding

Financial support from the Ministerio de Ciencia e Innovación (Project CGL2015-66703-P MINECO/FEDER, UE and Ph.D. grant to Sonia Herrando-Moraira) and the Catalan government ("Ajuts a grups consolidats" 2017-SGR1116) are also greatly acknowledged. This project has been carried out under the Ph.D. program "Plant Biology and Biotechnology" of the Autonomous University of Barcelona (UAB).

6. Acknowledgements

Authors thank Maria Luisa Gutiérrez, Fernando Castro, and Laia Barres for providing technical support during the laboratory process. We also thank the herbaria that provided material for the study: BC, DUSH, E, FRU, GDA, H, LE, MBK, MEM, MHA, MO, MU, OS, SANT, TAD, TARI, TI, TK, US, W, and WU. We gratefully acknowledge the University of Memphis computing facilities to allow us the utilization of its high-performance cluster (HPC).

7. References

Aguilée R., Lambert A., Claessen D. 2011. Ecological speciation in dynamic landscapes. J. Evol. Biol. 24:2663–2677.

Amidon W.H., Hynek S.A. 2010. Exhumational history of the north central Pamir. Tectonics 29:TC5017.

An Z., Huang Y., Liu W., Guo Z., Clemens S., Li L., Prell W., Ning Y., Cai Y., Zhou W., Lin B., Zhang Q., Cao Y., Qiang X., Chang H., Wu Z. 2005. Multiple expansions of C4 plant biomass in East Asia since 7 Ma coupled with strengthened

monsoon circulation. Geology 33:705.

An Z., Kutzbach J.E., Prell W.L., Porter S.C. 2001. Evolution of Asian monsoons and phased uplift of the Himalaya–Tibetan plateau since Late Miocene times. Nature 411:62–66.

Arakaki M., Christin P.A., Nyffeler R., Lendel A., Eggli U., Ogburn R.M., Spriggs E., Moore M.J., Edwards E.J. 2011. Contemporaneous and recent radiations of the world's major succulent plant lineages. Proc. Natl. Acad. Sci. USA 108:8379–8384.

Banasiak Ł., Piwczyński M., Uliński T., Downie S.R., Watson M.F., Shakya B., Spalik K. 2013. Dispersal patterns in space and time: a case study of Apiaceae subfamily Apioideae. J. Biogeogr. 40:1324–1335.

Barres L., Sanmartín I., Anderson C.L., Susanna A., Buerki S., Galbany-Casals M., Vilatersana R. 2013. Reconstructing the evolution and biogeographic history of tribe Cardueae (Compositae). Am. J. Bot. 100:867–882.

Beck H.E., Zimmermann N.E., McVicar T.R., Vergopolan N., Berg A., Wood E.F. 2018. Present and future Köppen-Geiger climate classification maps at 1-km resolution. Sci. Data 5:180214.

Bialik O.M., Frank M., Betzler C., Zammit R., Waldmann N.D. 2019. Two-step closure of the Miocene Indian Ocean Gateway to the Mediterranean. Sci. Rep. 9:1–10.

Bivand R., Rundel C., Pebesma E., Stuetz R., Hufthammer K.O., Bivand, M.R. 2017. Package ‗rgeos'. The Comprehensive R Archive Network (CRAN).

Blomberg S.P., Garland Jr. T., Ives A.R. 2003. Testing for phylogenetic signal in comparative data: behavioral traits are more labile. Evolution 57:717–745.

Bolger A.M., Lohse M., Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120.

Borges R., Machado J.P., Gomes C., Rocha A.P., Antunes A. 2019. Measuring phylogenetic signal between categorical traits and phylogenies. Bioinformatics 35:1862–1869.

Broennimann O., Fitzpatrick M.C., Pearman P.B., Petitpierre B., Pellissier L., Yoccoz N.G., Thuiller W., Fortin M.J., Randin C., Zimmermann N.E., Graham C.H., Guisan A. 2012. Measuring ecological niche overlap from occurrence and spatial environmental data. Glob. Ecol. Biogeogr. 21:481–497.

Brown J.L., Hill D.J., Dolan A.M., Carnaval A.C., Haywood A.M. 2018. PaleoClim, high spatial resolution paleoclimate surfaces for global land areas. Sci. Data 5:1–9.

Burin G., Alencar L.R., Chang J., Alfaro M.E., Quental T.B. 2019. How well can we estimate diversity dynamics for clades in diversity decline?. Syst. Biol. 68:47–62.

Burtman V.S. 2000. Cenozoic crustal shortening between the Pamir and Tien Shan and a reconstruction of the Pamir-Tien Shan transition zone for the Cretaceous and Palaeogene. Tectonophysics 319:69–92.

Burtman V.S., Molnar P.H. 1993. Geological and geophysical evidence for deep subduction of continental crust beneath the Pamir. Geol. S. Am. S. 281:1–76.

Buslov M.M., Kokh D.A., de Grave J. 2008. Mesozoic-Cenozoic tectonics and geodynamics of Altai, Tien Shan, and Northern Kazakhstan, from apatite fissiontrack data. Russ. Geol. Geophys. 49:648–654.

Capella-Gutiérrez S., Silla-Martínez J.M., Gabaldón T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25:1972–1973.

Casanovas-Vilar I., Alba D.M., Garcés M., Robles J.M., Moyà-Solà S. 2011. Updated chronology for the Miocene hominoid radiation in Western Eurasia. Proc. Natl. Acad. Sci. USA 108:5554–5559.

Casanovas-Vilar I., García-Paredes I., Alba D.M., van den Hoek Ostende L.W., Moyà-Solà S. 2010. The European Far West: Miocene mammal isolation, diversity and turnover in the Iberian Peninsula. J. Biogeogr. 37:1079–1093.

Czekanski-Moir J.E., Rundell R.J. 2019. The ecology of nonecological speciation and nonadaptive radiations. Trends Ecol. Evol. 34:400–415.

Danin A., Davis P.H. 1975. *Jurinea* Cass. In: Davis P.H., ed. Flora of Turkey and the East Aegean Islands, vol. 5. Edinburgh: Edinburgh Univ. Press, 439–450.

Djamali M., Brewer S., Breckle S.W., Jackson S.T. 2012. Climatic determinism in phytogeographic regionalization: a test from the Irano-Turanian region, SW and Central Asia. Flora 207:237–249.

Donoghue M.J., Sanderson M.J. 2015. Confluence, synnovation, and depauperons in plant diversification. New Phytol. 207:260–274.

Drummond A.J., Suchard M.A., Xie D., Rambaut A. 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. Mol. Biol. Evol. 29:1969–1973.

Dynesius M., Jansson R. 2000. Evolutionary consequences of changes in species' geographical distributions driven by Milankovitch climate oscillations. Proc. Natl. Acad. Sci. USA 97:9115–9120.

Etienne R.S., Haegeman B., Stadler T., Aze T., Pearson P.N., Purvis A., Phillimore A.B. 2012. Diversity-dependence brings molecular phylogenies closer to agreement with the fossil record. Proc. R. Soc. B Biol. Sci. 279:1300–1309.

FitzJohn R.G. 2012. Diversitree: comparative phylogenetic analyses of diversification in R. Meth. Ecol. Evol. 3: 1084–1032.

Folk R.A., Stubbs R.L., Mort M.E., Cellinese N., Allen J.M., Soltis P.S., Soltis D.E., Guralnick R.P. 2019. Rates of niche and phenotype evolution lag behind diversification in a temperate radiation. Proc. Natl. Acad. Sci. USA 116:10874–10882.

Folk R.A., Siniscalchi C.M., Soltis, D.E. 2020. Angiosperms at the edge: Extremity, diversity, and phylogeny. Plant Cell Environ. 43:2871–2893.

Fragoso-Martínez I., Salazar G.A., Martínez-Gordillo M., Magallón S., Sánchez-Reyes L., Lemmon E.M., Lemmon A.R., Sazatornil F., Mendoza C.G. 2017. A pilot study applying the plant Anchored Hybrid Enrichment method to New World sages (*Salvia* subgenus *Calosphace*; Lamiaceae). Mol. Phylogenet. Evol. 117:124–134.

Fu B., Ninomiya Y., Guo J. 2010. Slip partitioning in the northeast Pamir–Tian Shan convergence zone. Tectonophysics 483:344–364.

Ge D., Zhang Z., Xia L., Zhang Q., Ma Y., Yang Q. 2012. Did the expansion of C4 plants drive extinction and massive range contraction of micromammals? Inferences from food preference and historical biogeography of pikas. Palaeogeogr. Palaeoclimatol. Palaeoecol. 326:160–171.

Gillespie R.G., Bennett G.M., De Meester L., Feder J.L., Fleischer R.C., Harmon L.J., Hendry A.P., Knope M.L., Mallet J., Martin C., Parent C.E., Patton A.H., Pfennig K.S.,

Rubinoff D., Schluter D., Seehausen O., Shaw K.L., Stacy E., Stervander M., Stroud J.T., Wagner C., Wogan G.O.U. 2020. Comparing adaptive radiations across space, time, and taxa. J. Hered. 111:1–20.

Harrison S., Noss, R. 2017. Endemism hotspots are linked to stable climatic refugia. Ann. Bot. 119: 207–214.

Hauenschild F., Favre A., Schnitzler J., Michalak I., Freiberg M., Muellner-Riehl A.N. 2017. Spatio-temporal evolution of *Allium* L. in the Qinghai–Tibet-Plateau region: Immigration and in situ radiation. Plant Divers. 39:167–179.

Herbert T.D., Lawrence K.T., Tzanova A., Peterson L.C., Caballero-Gill R., Kelly C.S. 2016. Late Miocene global cooling and the rise of modern ecosystems. Nat. Geosci. 9:843–847.

Herrando-Moraira S., Nualart N., Herrando-Moraira A., Chung M.Y., Chung M.G., López-Pujol J. 2019b. Climatic niche characteristics of native and invasive *Lilium lancifolium*. Sci. Rep. 9:1–16.

Herrando-Moraira S., The Cardueae Radiations Group (Calleja J.A., Carnicero-Campmany P., Fujikawa K., Galbany-Casals M., Garcia-Jacas N., Im H.T., Kim S.C., Liu J.Q., López-Alvarado J., López-Pujol J., Mandel J.R., Massó S., Mehregan I., Montes-Moreno N., Pyak E., Roquet C., Sáez L., Sennikov A., Susanna A., Vilatersana R.). 2018. Exploring data processing strategies in NGS target enrichment to disentangle radiations in the tribe Cardueae (Compositae). Mol. Phylogenet. Evol. 128:69–87.

Herrando-Moraira S., The Cardueae Radiations Group (Calleja J.A., Galbany-Casals M., Garcia-Jacas N., Liu J.Q., López-Alvarado J., López-Pujol J., Mandel J.R., Massó S., Montes-Moreno N., Roquet C., Sáez L., Sennikov A., Susanna A., Vilatersana R.). 2019a. Nuclear and plastid DNA phylogeny of tribe Cardueae (Compositae) with Hyb-Seq data: A new subtribal classification and a temporal diversification framework. Mol. Phylogenet. Evol. 137:313–332.

Herrando-Moraira S., The Cardueae Radiations Group (Calleja J.A., Chen Y.S., Fujikawa K., Galbany-Casals M., Garcia-Jacas N., Kim S.C., Liu J.Q., López-Alvarado J., López-Pujol J., Mandel J.R., Mehregan I., Roquet C., Sennikov A., Susanna A., Vilatersana R., Xu L.S.). 2020. Generic boundaries in subtribe Saussureinae (Compositae: Cardueae): Insights from Hyb-Seq data. Taxon 69:694–714.

Hillis D.M., Bull J.J. 1993. An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. Syst. Biol. 42:182–192.

Holbourn A.E., Kuhnt W., Clemens S.C., Kochhann K.G., Jöhnck J., Lübbers J., Andersen N. 2018. Late Miocene climate cooling and intensification of southeast Asian winter monsoon. Nat. Commun. 9:1–13.

Huang Y., Clemens S.C., Liu W., Wang Y., Prell W.L. 2007. Large-scale hydrological change drove the late Miocene C4 plant expansion in the Himalayan foreland and Arabian Peninsula. Geology 35:531–534.

Hughes C.E., Atchison G.W. 2015. The ubiquity of alpine plant radiations: from the Andes to the Hengduan Mountains. New Phytol. 207:275–282.

Hughes C.E., Nyffeler R., Linder P.H. 2015. Evolutionary plant radiations: where, when, why and how? New Phytol. 207:249–253.

Hurka H., Friesen N., Bernhardt K.-G., Neuffer B., Smirnov S.V., Shmakov A.I., Blattner F.R. 2019. The Eurasian steppe belt: Status quo, origin and evolutionary history. Turczaninowia 22:5–71.

Iljin M.M. 1962. *Jurinea* Cass. In: Shiskin B.K., Bobrov E.G., eds. Flora of the USSR, vol. 27. Moscow: Izdatel'stvo Akademii Nauk SSSR, 538–704.

Jabbour F., Renner S.S. 2011. *Consolida* and *Aconitella* are an annual clade of *Delphinium* (Ranunculaceae) that diversified in the Mediterranean basin and the Irano-Turanian region. Taxon 60:1029–1040.

Jiménez-Moreno G., Pérez-Asensio J.N., Larrasoaña J.C., Sierro F.J., Garcia-Castellanos D., Salazar A., Salvany J.M., Ledesma S., Mata M.P., Mediavilla C. 2019. Early Pliocene climatic optimum, cooling and early glaciation deduced by terrestrial and marine environmental changes in SW Spain. Glob. Planet. Change 180:89–99.

Johnson M.G., Gardner E.M., Liu Y., Medina R., Goffinet B., Shaw A.J., Zerega N.J.C., Wickett N.J. 2016. HybPiper: Extracting coding sequence and introns for phylogenetics from high-throughput sequencing reads using target enrichment. App. Plant Sci. 4:1600016.

Kadereit J.W., Griebeler E.M., Comes H.P. 2004. Quaternary diversification in European alpine plants: pattern and process. Philos. Trans. R. Soc. Lond., B, Biol. Sci. 359:265–274.

Karger D.N., Conrad O., Böhner J., Kawohl T., Kreft H., Soria-Auza R.W., Zimmermann N.E., Linder H.P., Kessler M. 2017. Climatologies at high resolution for the earth's land surface areas. Sci. Data 4:170122.

Karl R., Koch M.A. 2013. A world-wide perspective on crucifer speciation and evolution: phylogenetics, biogeography and trait evolution in tribe Arabideae. Ann. Bot. 112:983–1001.

Katoh K., Standley D.M. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. 30:772–780.

Köppen W., Geiger R. 1936. Handbuch der Klimatologie. Berlin: Gebruder Borntraeger.

Koutsodendris A., Allstädt F.J., Kern O.A., Kousis I., Schwarz F., Vannacci M., Woutersen A., Appeld E., Berke M.A., Fang X., Friedrich O., Hoorn C., Salzmann U., Pross J. 2019. Late Pliocene vegetation turnover on the NE Tibetan Plateau (Central Asia) triggered by early Northern Hemisphere glaciation. Glob. Planet. Change 180:117–125.

Kozlov A.M., Darriba D., Flouri T., Morel B., Stamatakis A. 2019. RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. Bioinformatics 35:4453–4455.

Kück P., Longo G.C. 2014. FASconCAT-G: extensive functions for multiple sequence alignment preparations concerning phylogenetic studies. Front. Zool. 11:81.

LaJeuness T.C. 2005. ~~–~~Species" radiations of symbiotic dinoflagellates in the Atlantic and Indo-Pacific since the Miocene-Pliocene transition. Mol. Biol. Evol. 22:570–581.

Lauterbach M., Veranso-Libalah M.C., Sukhorukov A.P., Kadereit G. 2019. Biogeography of the xerophytic genus *Anabasis* L. (Chenopodiaceae). Ecol. Evol. 9:3539–3552.

Lisiecki L.E., Raymo M.E. 2005. A Pliocene-Pleistocene stack of 57 globally distributed benthic δ18O records. Paleoceanography 20:PA1003.

López-Vinyallonga S., Mehregan I., Garcia-Jacas N., Tscherneva O., Susanna A., Kadereit J.W. 2009. Phylogeny and evolution of the *Arctium-Cousinia* complex (Compositae, Cardueae-Carduinae). Taxon 58:153–171.

Louca S., Pennell M.W. 2020. Extant timetrees are consistent with a myriad of diversification histories. Nature 580:502–505.

Magyari E.K., Chapman J.C., Gaydarska B., Marinova E., Deli T., Huntley J.P., Allen J.R.M., Huntley B. 2008. The ‗oriental' component of the Balkan flora: evidence of presence on the Thracian Plain during the Weichselian late-glacial. J. Biogeogr. 35:865–883.

Mahmoudi Shamsabad M., Moharrek F., Assadi M., Nieto Feliner G. 2020. Biogeographic history and diversification patterns in the Irano-Turanian genus *Acanthophyllum* s.l. (Caryophyllaceae). Plant Biosyst. https://doi.org/10.1080/11263504.2020.1756974

Manafzadeh S., Salvo G., Conti E. 2014. A tale of migrations from east to west: the Irano-Turanian floristic region as a source of Mediterranean xerophytes. J. Biogeogr. 41:366–379.

Manafzadeh S., Staedler Y.M., Conti E. 2017. Visions of the past and dreams of the future in the Orient: the Irano-Turanian region from classical botany to evolutionary studies. Biol. Rev. 92:1365–1388.

Mandel J.R., Dikow R.B., Funk V.A., Masalia R.R., Staton S.E., Kozik A., Michelmore R.W., Rieseberg L.H., Burke J.M. 2014. A target enrichment method for gathering phylogenetic information from hundreds of loci: an example from the Compositae. App. Plant Sci. 2:1300085.

Matzke N.J. 2013. BioGeoBEARS: BioGeography with Bayesian (and likelihood) evolutionary analysis in R Scripts: CRAN: The Comprehensive R Archive Network. [WWW document] URL Available from: https://cran.r-project.org/.

Miao Y., Herrmann M., Wu F., Yan X., Yang S. 2012. What controlled Mid–Late Miocene long-term aridification in Central Asia?—Global cooling or Tibetan Plateau uplift: A review. Earth Sci. Rev. 112:155–172.

Miller M.A., Pfeiffer W., Schwartz T. 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. Proceedings of the Gateway Computing Environments Workshop (GCE). New Orleans, LA: IEEE, 1–8.

Mittermeier R.A., Turner W.R., Larsen F.W., Brooks T.M., Gascon C. 2011. Global biodiversity conservation: the critical role of hotspots. In: Zachos F.E., Habel J.C., eds. Biodiversity hotspots distribution and protection of conservation priority areas. Heidelberg: Springer, 3–22.

Moharrek F., Sanmartín I., Kazempour-Osaloo S., Nieto Feliner G. 2019. Morphological innovations and vast extensions of mountain habitats triggered rapid diversification within the species-rich Irano-Turanian genus *Acantholimon* (Plumbaginaceae). Front. Genet. 9:698.

Moreau C.S., Bell C.D. 2013. Testing the museum versus cradle tropical biological diversity hypothesis: phylogeny, diversification, and ancestral biogeographic range evolution of the ants. Evolution 67:2240–2257.

Morlon H. 2020. Prior hypotheses or regularization allow inference of diversification histories from extant timetrees. bioRxiv, p. 2020.07.03.185074.

Morlon H., Lewitus E., Condamine F.L., Manceau M., Clavel J., Drury J. 2016. RPANDA: an R package for macroevolutionary analyses on phylogenetic trees. Meth. Ecol. Evol. 7: 589–597.

Mouthereau F. 2011. Timing of uplift in the Zagros belt/Iranian plateau and accommodation of late Cenozoic Arabia–Eurasia convergence. Geol. Mag. 148:726–738.

Nee S., Mooers A.O., Harvey P.H. 1992. Tempo and mode of evolution revealed from molecular phylogenies. Proc. Natl. Acad. Sci. USA 89:8322–8326.

Nevado B., Atchison G.W., Hughes C.E., Filatov D.A. 2016. Widespread adaptive evolution during repeated evolutionary radiations in New World lupins. Nat. Commun. 7:1–9.

Noroozi J., Naqinezhad A., Talebi A., Doostmohammadi M., Plutzar C., Rumpf S.B., Asgarpour Z., Schneeweiss G.M. 2019. Hotspots of vascular plant endemism in a global biodiversity hotspot in Southwest Asia suffer from significant conservation gaps. Biol. Conserv. 237:299–307.

Onstein R.E. 2020. Darwin's second ‗abominable mystery': trait flexibility as the innovation leading to angiosperm diversity. New Phytol. 228:1741–1747. https://doi.org/10.1111/nph.16294

Paradis E., Schliep K. 2019. ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics 35:526–528.

Peterson A., Harpke D., Peterson J., Harpke A., Peruzzi L. 2019. A pre-Miocene Irano-Turanian cradle: Origin and diversification of the species-rich monocot genus *Gagea* (Liliaceae). Ecol. Evol. 9:5870–5890.

Peterson A.T. 2003. Predicting the geography of species' invasions via ecological niche modeling. Q. Rev. Biol. 78:419–433.

Potter P.E., Szatmari P. 2009. Global Miocene tectonics and the modern world. Earth-Sci. Rev. 96:279–295.

Pound M.J., Haywood A.M., Salzmann U., Riding J.B. 2012. Global vegetation dynamics and latitudinal temperature gradients during the Mid to Late Miocene (15.97–5.33 Ma). Earth Sci Rev 112:1–22.

Rea D.K., Snoeckx H., Joseph L.H. 1998. Late Cenozoic eolian deposition in theNorth Pacific: Asian drying, Tibetan uplift, and cooling of the north hemisphere. Paleoceanography 13:215–224.

Rechinger K.H., Wagenitz G. 1979. *Jurinea*. In: Rechinger K.H., ed. Flora Iranica, vol. 139a. Graz: Akademische Druck- u. Verlagsanstalt, 180–209.

Ree R.H., Moore B.R., Webb C.O., Donoghue M.J. 2005. A likelihood framework for inferring the evolution of geographic range on phylogenetic trees. Evolution 59:2299–2311.

Ree R.H., Sanmartín I. 2018. Conceptual and statistical problems with the DEC+ J model of founder-event speciation and its comparison with DEC via model selection. J. Biogeogr. 45:741–749.

Revell L.J. 2012. phytools: an R package for phylogenetic comparative biology (and other things). Methods Ecol. Evol. 3:217–223.

Ribera I., Blasco-Zumeta J. 1998. Biogeographical links between steppe insects in the Monegros region (Aragón, NE Spain), the eastern Mediterranean, and central Asia. J. Biogeogr. 25:969–986.

Roquet C., Boucher F.C., Thuiller W., Lavergne S. 2013. Replicated radiations of the alpine genus *Androsace* (Primulaceae) driven by range expansion and convergent key innovations. J. Biogeogr. 40:1874–1886.

Sales F., Hedge I.C. 2013. Genneric endemism in South-West Asia: an overview. Rostaniha 14:22–35.

Salvo G., Manafzadeh S., Ghahremaninejad F., Tojibaev K.,

Zeltner L., Conti E. 2011. Phylogeny, morphology, and biogeography of *Haplophyllum* (Rutaceae), a species-rich genus of the Irano-Turanian floristic region. Taxon 60:513–527.

Salzmann U., Haywood A.M., Lunt D.J., Valdes P.J., Hill D.J. 2008. A new global biome reconstruction and data-model comparison for the middle Pliocene. Glob. Ecol. Biogeogr. 17:432–447.

Sayyari E., Mirarab S. 2016. Fast coalescent-based computation of local branch support from quartet frequencies. Mol. Biol. Evol. 33:1654–1668.

Seehausen O. 2015. Process and pattern in cichlid radiations–inferences for understanding unusually high rates of evolutionary diversification. New Phytol. 207:304–312.

Shen X., Wan S., Colin C., Tada R., Shi X., Pei W., Tan Y., Jiang X., Li A. 2018. Increased seasonality and aridity drove the C4 plant expansion in Central Asia since the Miocene–Pliocene boundary. Earth Planet. Sci. Lett. 502:74–83.

Smith S.A., O'Meara B.C. 2012. treePL: divergence time estimation using penalized likelihood for large phylogenies. Bioinformatics 28:2689–2690.

Sniderman J.K., Woodhead J.D., Hellstrom J., Jordan G.J., Drysdale R.N., Tyler J.J., Porch N. 2016. Pliocene reversal of late Neogene aridification. Proc. Natl. Acad. Sci. USA 113:1999–2004.

Stebbins G.L. 1952. Aridity as a stimulus to plant evolution. Am. Nat. 86:33–44.

Stroud J.T., Losos J.B. 2016. Ecological opportunity and adaptive radiation. Annu. Rev. Ecol. Evol. Syst. 47:507–532.

Suc J.P. 1984. Origin and evolution of the Mediterranean vegetation and climate in Europe. Nature 307:429–432.

Sun J., Gong Z., Tian Z., Jia Y., Windley B. 2015. Late Miocene stepwise aridification in the Asian interior and the interplay between tectonics and climate. Palaeogeogr. Palaeoclimatol. Palaeoecol. 421:48–59.

Sun M., Folk R.A., Gitzendanner M.A., Soltis P.S., Chen Z., Soltis D.E., Guralnick R.P. 2020. Recent accelerated diversification in rosids occurred outside the tropics. Nat. Commun. 11:1–12.

Sundue M.A., Testo W.L., Ranker T.A. 2015. Morphological innovation, ecological opportunity, and the radiation of a major vascular epiphyte lineage. Evolution 69:2482–2495.

Susanna, A., Garcia-Jacas, N. 2007. Tribe Cardueae Cass. (1819). In: Kadereit J.W., Jeffrey C., eds. The families and genera of vascular plants, vol. 8. Berlin & Heidelberg: Springer, 123–146.

Susanna A., Garcia-Jacas N. 2009. Cardueae (Carduoideae). In: Funk V.A., Susanna A., Stuessy T.F., Bayer R.J., eds. Systematics, evolution, and biogeography of Compositae. Vienna: IAPT, 293–313.

Susanna A., Garcia-Jacas N., Hidalgo O., Vilatersana R., Garnatje T. 2006. The Cardueae (Compositae) revisited: insights from ITS, *trnL-trnF*, and *matK* nuclear and chloroplast DNA analysis. Ann. Missouri Bot. Gard. 93:150–171.

Szukala A., Korotkova N., Gruenstaeudl M., Sennikov A.N., Lazkov G.A., Litvinskaya S.A., Gabrielian E., Borsch T., Raab-Straube E. von. 2019. Phylogeny of the Eurasian genus *Jurinea* (Asteraceae: Cardueae): Support for a monophyletic genus concept and a first hypothesis on overall species relationships. Taxon 68:112–131.

Takhtajan A. 1986. Floristic Regions of the World. Berkeley: University of California Press.

Tang C.Q., Matsui T., Ohashi H., Dong Y.F., Momohara A., Herrando-Moraira S., Qian S., Yang Y., Ohsawa M., Luu H.T., Grote P.J., Krestov P.V., LePage B., Werger M., Robertson K., Hobohm C., Wang C.Y., Peng M.C., Chen X., Wang H.C., Su W.H., Zhou R., Li S., He L.H., Yan K., Zhu M.Y., Hu J., Yang R.H., Li W.J., Tomita M.,Wu Z.L., Yan H.Z., Zhang G.F., He H., Yi S.R., Gong H., Song K., Song D., Li X.S., Zhang Z.Y., Han P., Shen L.Q., Huang D.S., Luo K., López-Pujol J. 2018. Identifying long-term stable refugia for relict plant species in East Asia. Nat. Commun. 9:1–14.

Thompson J.A., Burbank D.W., Li T., Chen J., Bookhagen B. 2015. Late Miocene northward propagation of the northeast Pamir thrust system, northwest China. Tectonics 34:510–534.

Uribe-Convers S., Tank D.C. 2015. Shifts in diversification rates linked to biogeographic movement into new areas: An example of a recent radiation in the Andes. Am. J. Bot. 102:1854–1869.

Wen J., Zhang J., Nie Z.L., Zhong Y., Sun H. 2014. Evolutionary diversifications of plants on the Qinghai-Tibetan Plateau. Front. Genet. 5:4.

White F., Léonard J. 1991. Phytogeographic links between Africa and Southwest Asia. Flora Veg. Mundi 9:229–246.

Wiens J.J., Ackerly D.D., Allen A.P., Anacker B.L., Buckley L.B., Cornell H.V., Damschen E.I., Davies T.J., Grytnes J.A., Harrison S.P., Hawkins B.A., Holt R.D., McCain C.M., Stephens P.R. 2010. Niche conservatism as an emerging principle in ecology and conservation biology. Ecol. Lett. 13:1310–1324.

Willeit M., Ganopolski A., Calov R., Robinson A., Maslin M. 2015. The role of CO_2 decline for the onset of Northern Hemisphere glaciation. Quat. Sci. Rev. 119:22–34.

Wu S.D., Lin L., Li H.L., Yu S.X., Zhang L.J., Wang W. 2015. Evolution of Asian interior arid-zone biota: evidence from the diversification of Asian *Zygophyllum* (Zygophyllaceae). PloS ONE 10:e0138697.

Xing Y., Ree R.H. 2017. Uplift-driven diversification in the Hengduan Mountains, a temperate biodiversity hotspot. Proc. Natl. Acad. Sci. USA 114:E3444–E3451.

Ye X.Y., Ma P.F., Yang G.Q., Guo C., Zhang Y.X., Chen Y.M., Guo Z.H., Li D.Z. 2019. Rapid diversification of alpine bamboos associated with the uplift of the Hengduan Mountains. J. Biogeogr. 46:2678–2689.

Yoder J.B., Clancey E., Des Roches S., Eastman J.M., Gentry L., Godsoe W., Harmon L.J. 2010. Ecological opportunity and the origin of adaptive radiations. J. Evol. Biol. 23:1581–1596.

Yu Y., Harris A.J., Blair C., He X. 2015. RASP (Reconstruct Ancestral State in Phylogenies): a tool for historical biogeography. Mol. Phylogenet. Evol. 87:46–49.

Zachos J.C., Pagani M., Sloan L., Thomas E., Billups K. 2001. Trends, rhythms, and aberrations in global climate 65 Ma to present. Science 292:686–693.

Zhang C., Rabiee M., Sayyari E., Mirarab S. 2018. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. BMC Bioinformatics 19:153.

Zhang Z., Han W., Fang X., Song C., Li X. 2013. Late Miocene–Pleistocene aridification of Asian inland revealed by geochemical records of lacustrine-fan delta sediments from the western Tarim Basin, NW China. Palaeogeogr. Palaeoclimatol. Palaeoecol. 377:52–61.

Zheng H., Powell C.M., Rea D.K., Wang J., Wang P. 2004. Late Miocene and mid-Pliocene enhancement of the East Asian monsoon as viewed from the land and sea. Glob. Planet. Change 41:147–155.

Zohary M. 1973. Geobotanical Foundations of the Middle East (Volume I and II). Amsterdam: Gustav Fischer Verlag, and Stuttgart: Swets & Zeitlinger.

Zohary M. 1981. On the flora and vegetation of the Middle East structure and evolution. In: Uerpmann H.P., Frey W., eds. Contributions to the Environmental History of Southwest Asia, vol. 8. Wiesbaden: Reichert, 1–17.

8. Supplementary material

Supplementary Materials and Methods.
Supplementary Tables (S1–S11) and Figures (S1–S11).

Supplementary alignments Supermatrix and loci partitions.
Supplementary trees from concatenated approach and unfiltered dataset, from coalescence approach and unfiltered dataset, dated tree in Phylip format.

Materials and methods (extended version)

DNA extraction, library preparation, sequence capture, and sequencing

Dried leaf tissue was weighed (10–30 mg per sample) and homogenized using Mixer Mill MM 301 (Retsch®, Haan, Germany). The plant genomic DNA was extracted with the E.N.Z.A SP Plant DNA Mini Kit (Omega Bio-Tek Inc., Norcross, GA, USA) according manufacturer's protocol. After the extraction, we measured the DNA quantity with Qubit™ 3.0 Fluorometer (Thermo Scientific, Waltham, MA, USA) to standardize the samples at 1 µg of DNA in 70 µl dH2O. With a target fragment size of 400–500 bp, the DNA was shared using a Q800R2 Sonicator® machine (QSonica, Newtown, CT, USA). The sonicator was set to 20% of amplitude and was firstly run for 3 min (with 10 s pulse on, and 10 s pulse off). Later, we evaluated the fragment profile on a 1.2% (w/v) agarose gel. When the sonication was not enough to achieve most of the fragments at the target size, the sonication step was repeated. The barcoded libraries were constructed with the NEBNext Ultra II DNA Library Prep kit for Illumina (New England Biolabs, Ipswich, MA, USA). The libraries were pooled in groups up to eight samples, but most of them in four samples aiming for a DNA quantity of 500 ng. For the target enrichment step (sequence capture), we used the MyBaits COS 1Kv1 (MYcroarray, Ann Arbor, MI, USA) following manufacturer's specifications with slight modifications detailed in Herrando-Moraira et al. (2018) and final library spiking as Herrando-Moraira et al. (2019). The final enriched libraries were sequenced using 100 bp paired-end reads in the DNA Sequencing Core CGRC/ICBR of the University of Florida using an Illumina HiSeq 3000 (Illumina, USA) or in Macrogen Co. (Seoul, South Korea) in an Illumina HiSeq 4000 (Illumina, USA).

Target sequence extraction

A general inspection of raw reads quality was performed with FastQC v.0.10.1 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). The software Trimmomatic v.0.36 (Bolger et al. 2014) was used to remove the adapters and clean the reads according the quality threshold of 5:20 of sliding-window, short reads (< 36 bp length), and lack of both forward and reverse pair. To extract the COS loci, we used the workflow pipeline HybPiper v.1.3.1 (Johnson et al. 2016) using the BWA mapper (Li and Durbin 2009) and the assembler SPAdes (Bankevich et al. 2012). To prevent the inclusion of potential paralog loci, we removed all loci flagged by HybPiper for downstream analyses. Later, we used MAFFT v.7.266 (Katoh and Standley 2013) with the auto mode to align the sequences in each multi-fasta file of COS loci. The alignments were trimmed to clean the ambiguously aligned regions with trimAl v.14 (Capella-Gutiérrez et al. 2009) applying the automated1 flag. The individual trimmed alignments were concatenated into a supermatrix with FASconCAT-G v.1.02 (Kück and Longo 2014).

In order to reduce the phylogenetic noise, we followed the filtering strategy of position exclusion with high substitution rates described by Fragoso-Martínez et al. (2017). This filtering method has been proved useful to improve support values under the concatenation approach for rapidly evolving lineages (Wanke et al. 2017; Herrando-Moraira et al. 2018; Tomasello et al. 2020). The best Maximum Likelihood (ML) tree (see methods of phylogenetic inference below) was made ultrametric with R function ~~–~~chronos" from ape v.5.3 package (Paradis and Schliep 2019), setting lambda to 1.0 and age maximum to 1. For the following steps we proceed as the original Fragoso-Martínez et al. (2017) workflow. We applied a cut-off threshold of substitution rate higher than 1, according the results obtained by Herrando-Moraira et al. (2018) on a small representation of subtribe Arctiinae and Saussureinae. A total of 6109 positions were removed from the initial supermatrix (159,512 bp). However, the same unsupported nodes were generally maintained in the obtained trees, and even these resulted in lower support branch values than initial unfiltered dataset (see comparison in Supplementary Fig. S1). Similar results, but for only 28 species of Jurinea, were reported in Herrando-Moraira et al. (2018). In views of the lack of an

improvement with this filtering strategy, we used the unfiltered dataset for downstream analyses.

Definition of areas for historical biogeographic analysis

The distribution range of Jurinea was divided into nine main areas according to the genus distribution patterns and endemism centers, floristic regions [at both global level such as Takhtajan's classification (Takhtajan 1986) and also al regional level (see below)], and climatic regions following the widely-recognized Köppen-Geiger classification (Köppen and Geiger 1936; Beck et al. 2018). The defined areas were: Mediterranean (MED); Circumboreal (CB); Irano-Turanian (IT), in turn subdivided in five subareas (IT1, IT2, IT3, IT4, and IT5); Himalayas (HIM); and Eastern Asia (EA). For the MED, CB, IT, and EA we generally followed the delimitation of Takhtajan's floristic regions, with some exceptions (see below). The definitions of the nine areas (countries, regions, and/or mountain ranges) are summarized in Supplementary Table S1.

Takhtajan's (1986) classification, based on floristic affinities, includes the Western Himalayas within the IT region. However, based on climatic data, the Western Himalayas, as also occurs with the Eastern ones, show a temperate/mesothermal climate with monsoon influence according the Köppen-Geiger classification. In contrast, in the IT region dry (desert and semi-arid) and continental/microthermal climates predominate. Therefore, in order to take into account these climatic differences and to test for possible relation with diversification patterns, the Himalayas were treated as a unique, separate area that included both Western and Eastern ranges.

The subdivision of the IT region has been long discussed during the last decades (see revision in Manafzadeh et al. 2017). Although its limits differ among treatments, there is a wide consensus for a longitudinal division of the IT region, from west to east based on geological barriers and floristic similarities (Zohary 1973; Takhtajan 1986; Manafzadeh et al. 2017). This delimitation splits by half the main mountain ranges of Central Asia (Pamir Mts. and Tian Shan Mts.) and puts part of each into W-IT and E-IT regions (see Fig. 2 in Manafzadeh et al. 2017). We could not strictly follow this subdivision since the lack of accurate information on species distribution ranges specified on the floras examined (Flora of the USSR and Flora Iranica); e.g. for the vast majority of species distributed in the Tian Shan there is no indication whether it is located in western or in eastern Tian Shan (see Appendix 1). Certainly, in spite of the proposed W-IT and E-IT division by Manafzadeh et al. (2017), the authors warned that additional studies are required to assess if Pamir and eastern Tian Shan Mountains can be treated as an independent floristic region, which was already proposed by Grubov (1963). In line with this, White and Léonard (1991; based on floristic studies of Léonard 1981–1989), proposed the division of IT in four regions, taking into account phytogeographic regions but also regional climatic differences (see revision in Djamali et al. 2012). In our opinion, this proposal fits better the current distribution and endemism patterns of Jurinea. Thus, we followed this delimitation with slight modifications. We partitioned, however, the White and Léonard –IT4" into two regions (IT4: Pamir and Tian Shan; and IT5: Altai and western Mongolia), since the Altai range can be regarded as a transition zone between CB and IT climates, and different Jurinea lineages converged there.

Selection of environmental variables

The variable selection process was conducted performing a principal component analysis (PCA) with the 45 variables, extracting the variable contribution to new PCs, and further representing their relationships according to their correlation values in a dendrogram (Supplementary Fig. S2). Based on these results, we aimed to retain only one variable within each group of highly correlated variables ($r < 0.7$). The final variable retained per group was selected according to the highest percentage contribution value for the first two PCA axes. A total of 21 environmental variables were used for further analyses (see those ones selected on Supplementary Table S4). Three variables (coefficient of variation of habitat heterogenity, grass/scrub/woodland coverage, and topographic position index), although not highly correlated (Supplementary Fig. S2), were rejected due to their poor contribution to the environmental variation (Supplementary Table S4).

Climatic data extraction for areas and buffers

A schematic overview of the method used for further analyses is detailed in Supplementary Fig. S4. The climatic layers were clipped by each nine areas and nine buffer polygons with the function –mask" of R package raster (Hijmans et al. 2015). Subsequently, the clipped rasters were processed with the extension SDM Toolbox v2.2d (Brown 2014) in ArcGIS to reduce decimal positions with the tool –Raster Calculator- Float to Integer", and later to define the same extend with –Extract by Mask" from –10–140ºE to 10–60ºN. To estimate the average changes in each area and buffer along the temporal series of PaleoClim, the transformed rasters were input in R to calculate the mean values with the function –cellStats" from raster R package. Resultant calculations were plotted in R with ggplot2 package (Wickham et al. 2016).

References of supplementary text

Bankevich A., Nurk S., Antipov D., Gurevich A.A., Dvorkin M., Kulikov A.S., Lesin V.M., Nikolenko S.I., Pham S., Prjibelski A.D., Pyshkin A.V., Sirotkin A.V., Vyahhi N., Tesler G., Alekseyev M.A., Pyshkin A.V. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J. Comput. Biol. 19:455–477.

Beck H.E., Zimmermann N.E., McVicar T.R., Vergopolan N., Berg A., Wood E.F. 2018. Present and future Köppen-Geiger climate

classification maps at 1-km resolution. Sci. Data 5:180214.

Bolger A.M., Lohse M., Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120.

Brown J.L. 2014. SDM toolbox: a python-based GIS toolkit for landscape genetic, biogeographic and species distribution model analyses. Methods Ecol. Evol. 5: 694–700.

Capella-Gutiérrez S., Silla-Martínez J.M., Gabaldón T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25:1972–1973.

Djamali M., Brewer S., Breckle S.W., Jackson S.T. 2012. Climatic determinism in phytogeographic regionalization: a test from the Irano-Turanian region, SW and Central Asia. Flora 207:237–249.

Fragoso-Martínez I., Salazar G.A., Martínez-Gordillo M., Magallón S., Sánchez-Reyes L., Lemmon E.M., Lemmon A.R., Sazatornil F., Mendoza C.G. 2017. A pilot study applying the plant Anchored Hybrid Enrichment method to New World sages (Salvia subgenus Calosphace; Lamiaceae). Mol. Phylogenet. Evol. 117:124–134.

Herrando-Moraira S., The Cardueae Radiations Group (Calleja J.A., Carnicero-Campmany P., Fujikawa K., Galbany-Casals M., Garcia-Jacas N., Im H.T., Kim S.C., Liu J.Q., López-Alvarado J., López-Pujol J., Mandel J.R., Massó S., Mehregan I., Montes-Moreno N., Pyak E., Roquet C., Sáez L., Sennikov A., Susanna A., Vilatersana R.). 2018. Exploring data processing strategies in NGS target enrichment to disentangle radiations in the tribe Cardueae (Compositae). Mol. Phylogenet. Evol. 128:69–87.

Herrando-Moraira S., The Cardueae Radiations Group (Calleja J.A., Galbany-Casals M., Garcia-Jacas N., Liu J.Q., López-Alvarado J., López-Pujol J., Mandel J.R., Massó S., Montes-Moreno N., Roquet C., Sáez L., Sennikov A., Susanna A., Vilatersana R.). 2019. Nuclear and plastid DNA phylogeny of tribe Cardueae (Compositae) with Hyb-Seq data: A new subtribal classification and a temporal diversification framework. Mol. Phylogenet. Evol. 137:313–332.

Hijmans R.J., et al. (2015). Package ‗raster'. R package, 734.

Johnson M.G., Gardner E.M., Liu Y., Medina R., Goffinet B., Shaw A.J., Zerega N.J.C., Wickett N.J. 2016. HybPiper: Extracting coding sequence and introns for phylogenetics from high-throughput sequencing reads using target enrichment. App. Plant Sci. 4:1600016.

Katoh K., Standley D.M. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. 30:772–780.

Köppen W., Geiger R. 1936. Handbuch der Klimatologie. Berlin, Germany: Gebruder Borntraeger.

Kück P., Longo G.C. 2014. FASconCAT-G: extensive functions for multiple sequence alignment preparations concerning phylogenetic studies. Front. Zool. 11:81.

Li H., Durbin R. 2009. Fast and accurate short read alignment with burrowswheeler transform. Bioinformatics 25:1754–1760.

Manafzadeh S., Staedler Y.M., Conti E. 2017. Visions of the past and dreams of the future in the Orient: the Irano-Turanian region from classical botany to evolutionary studies. Biol. Rev. 92:1365–1388.

Paradis E., Schliep K. 2019. ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics 35:526–528.

Takhtajan A. 1986. Floristic Regions of the World. Berkeley, USA: University of California Press.

Tomasello S., Karbstein K., Hodač L., Paetzold C., Hörandl, E. 2020. Phylogenomics unravels Quaternary vicariance and allopatric speciation patterns in temperate-montane plant species: a case study on the Ranunculus auricomus species complex. Mol. Ecol. 26:2031–2049.

Wanke S., Mendoza C.G., Müller S., Guillén A.P., Neinhuis C., Lemmon A.R., Lemmon E.M., Samain M.S. 2017. Recalcitrant deep and shallow nodes in Aristolochia (Aristolochiaceae) illuminated using anchored hybrid enrichment. Mol. Phylogenet. Evol. 117:111–123.

White F., Léonard J. 1991. Phytogeographic links between Africa and Southwest Asia. Flora Veg. Mundi 9:229–246.

Wickham H., Chang W., Wickham M.H. 2016. Package ‗ggplot2'. Create Elegant Data Visualisations Using the Grammar of Graphics. Version 2:1–189.

Zohary M. 1973. Geobotanical Foundations of the Middle East (Volume I and II). Gustav Fischer Verlag, Stuttgart and Swets & Zeitlinger, Amsterdam.

Supplementary tables

Supplementary Table S1. Delimitations of the distribution areas for *Jurinea*.

Code	Geographic area	Main areas included (countries, regions, and/or mountain ranges)
MED	Mediterranean	Morocco, N Algeria, Portugal, Spain, S France (Mediterranean coast), Italy (Sicily), Albania, Greece (including Aegean Islands), Cyprus, and SW Turkey (i.e. the area parallel to the Mediterranean coastline)
CB	Circumboreal	Germany, Austria, Czech Republic, Slovenia, Croatia, Bosnia and Herzegovina, Hungary, Serbia, Montenegro, Macedonia, Albania, Slovakia, Poland, Bulgaria, Romania, Moldova, Belarus, Ukraine, SE European Russia and SW Siberia, Georgia, Armenia and Azerbaijan (except its SW part), Turkey (N range parallel to Black Sea coastline), and N Kazakhstan
IT1	Irano-Turanian 1	Turkey (the southernmost part of Anatolia), inland parts of Syria, Lebanon, and Israel, western Jordan, and N Iraq
IT2	Irano-Turanian 2	C Turkey (most of Anatolia), SW Azerbaijan ranges, Iran, S mountain range of Turkmenistan (Koppeh Dagh), Afghanistan (including Hindu Kush Mts.), and W Pakistan
IT3	Irano-Turanian 3	Eastern coast of the Caspian sea, Ustyurt Plateau, Karakum Desert, and Kyzylkum Desert (C-S Kazakhstan, most of Uzbekistan and Turkmenistan), and Zhetysu Region (E Kazakhstan)
IT4	Irano-Turanian 4	Pamir-Alay Mts. (Tajikistan, S Kyrgyzstan, narrow NE border in Afghanistan, W Xinjiang in China) and Tian Shan Mts. (SE Kazakhstan, Kyrgyzstan, E Uzbekistan, NW Xinjiang in China)
IT5	Irano-Turanian 5	Altai, Sayan, and Khangai Mts. (range intersection among Altai and Tuva republics in Russia, W Mongolia, E Kazakhstan, and N Xinjiang in China)
HIM	Western and Eastern Himalayas	Karakoram range and Himalayas [NE Pakistan through N India (Jammu and Kashmir, Ladakh and Himachal Pradesh), Nepal, Bhutan, Arunachal Pradesh of India, S Tibet of China]
EA	Eastern Asia	S China (including Taiwan), E Myanmar, N Laos, N Thailand, N Vietnam

Supplementary Table S2. Summary of data likelihoods under each model, and results of statistical model choice from RASP software using BioGeoBEARS package. The model selected was the one with the highest AICc_wt value.

	LnL	numparams	d	e	AICc	AICc wt
DEC	-360.7	2	0.013	0.0026	**725.5**	**0.98**
DIVALIKE	-364.4	2	0.015	5.0e-09	732.9	0.023
BAYAREALIKE	-373.7	2	0.0076	0.13	751.5	2.2e-06

Supplementary Table S3. Initial set of 45 environmental variables tested.

Variable set (layers)	Description	Resolution	Source	Link
Annual Mean UVB1 (1)	Annual mean UV-B (UVB1) derived from the monthly mean covering the period of 2004-2013.	0.25º (15 arc min)	Ref[7]	https://www.ufz.de/gluv/index.php?en=32435
Aridity (1)	Represents average yearly precipitation divided by average yearly potential evapotranspiration. Lower values represent hyperarid areas and higher values humid areas.	0.16º (10 arc min)	FAO GeoNetwork	"Global map of aridity - 10 arc minutes" http://www.fao.org/geonetwork/srv/en/main.home
Bioclimatic variables (19: bio1–bio19)	Bioclimatic variables represent annual trends of temperature, and precipitation values.	0.0416º (2.5 arc min)	Ref[5], Ref[6]	https://chelsa-climate.org/downloads/
Evapotransporation (1)	Map of annual average evapotranspiration for the 1961–1990 time period.	0.5º (30 arc min)	FAO GeoNetwork	"Global map of monthly reference evapotranspiration - 30 arc minutes" http://www.fao.org/geonetwork/srv/en/main.home
Habitat heterogeneity (5: variance, standard deviation, dissimilarity, coefficient of variation, contrast)	This layers quantify the spatial heterogeneity of global habitat based on the textural features of Enhanced Vegetation Index (EVI) imagery acquired by the Moderate Resolution Imaging Spectroradiometer (MODIS).	0.0416º (2.5 arc min)	Ref[2]	https://www.earthenv.org/texture
Length of Growing Period (LGP) (1)	The period during the year (in total days) when both moisture availability and temperature are conducive to crop growth. Thus, LGP refers to the number of days within the period of temperatures above 5°C when moisture conditions are considered adequate.	0.5º (30 arc min)	FAO GeoNetwork	"Global length of growing periods" http://www.fao.org/geonetwork/srv/en/main.home
Soil pH (1)	Acidity measure of the soil divided in 5 classesof pH: <4.5, >=4.5–5.5, >5.5–7.2, >7.2–8.5, >8.5	0.083º (5 arc min)	FAO GeoNetwork	"PH - TOPSOIL" http://www.fao.org/geonetwork/srv/en/main.home
Soil depth (1)	Soil depth is indicated by 5 classes : <10cm, 10–50cm, 50–100cm, 100–150cm and 150-300cm. Estimates for the % occurrence of each depth class is derived from algorithms and the WORLD764.xls.	0.083º (5 arc min)	FAO GeoNetwork	"Effective soil depth (cm)" http://www.fao.org/geonetwork/srv/en/main.home
Soil quality (7: workability, toxicity, rooting conditions, oxygen availability to roots, nutrient retention capacity, nutrient availability, excess salts)	Based on soil parameters seven key soil qualities are extracted (see Ref[1] for details).	0.083º (5 arc min)	Ref[1]	http://www.fao.org/soils-portal/soil-survey/soil-maps-and-databases/harmonized-world-soil-database-v12/en/ http://webarchive.iiasa.ac.at/Research/LUC/External-World-soil-database/HTML/SoilQuality.html?sb=10
Topographic (5: altitude, slope, roughness, profile/tangential curvature, topographic position index)	Terrain features based on the digital elevation model products of global 250 m GMTED2010 and near-global 90 m SRTM4.1dev	0.0416º (2.5 arc min)	Ref[3]	https://www.earthenv.org/topography
Vegetation coverage (3: forest land, grass/scrub/woodland, barren/very sparsely vegetated land)	Land cover layers represented by the percentage of cover of three vegetation types.	0.083º (5 arc min)	Ref[4]	http://webarchive.iiasa.ac.at/Research/LUC/External-World-soil-database/HTML/LandUseShares.html?sb=9

Ref[1] Beckmann M., Václavík T., Manceur A.M., Šprtová L., von Wehrden H., Welk E., Cord A.F. 2014 glUV: A global UV-B radiation dataset for macroecological studies. Methods Ecol. Evol. 5:372–383.

Ref[2] Karger D.N., Conrad O., Böhner J., Kawohl T., Kreft H., Soria-Auza R.W., Zimmermann N.E., Linder, H.P., Kessler M. 2017. Climatologies at high resolution for the earth's land surface areas. Sci. Data 4:170122.

Ref[3] Karger D.N., Conrad O., Böhner J., Kawohl T., Kreft H., Soria-Auza R.W., Zimmermann N.E., Linder H.P., Kessler M. 2017. Data from: Climatologies at high resolution for the earth's land surface areas. Dryad Digital Repository.

Ref[4] Tuanmu M.N., Jetz W. 2015. A global, remote sensing-based characterization of terrestrial habitat heterogeneity for biodiversity and ecosystem modeling. Glob. Ecol. Biogeogr. 24: 1329–1339.

Ref[5] Fischer G., Nachtergaele F., Prieler S., van Velthuizen H.T., Verelst L., Wiberg D. 2008. *Global Agro-ecological Zones Assessment for Agriculture (GAEZ 2008)*. IIASA, Laxenburg, Austria and FAO, Rome, Italy.

Ref[6] Amatulli G., Domisch S., Tuanmu M.N., Parmentier B., Ranipeta A., Malczyk J., Jetz W. 2018. A suite of global, cross-scale topographic variables for environmental and biodiversity modeling. Sci. Data 5: 180040.

Ref[7] Fischer G., Nachtergaele F., Prieler S., van Velthuizen H.T., Verelst L., Wiberg D. 2008. *Global Agro-ecological Zones Assessment for Agriculture (GAEZ 2008)*. IIASA, Laxenburg, Austria and FAO, Rome, Italy.

Supplementary Table S4. Percentage of variables contribution to the initial PCA perform to select variables from the initial set of 45 potential predictors. Final selected variables used for environmental niche analyses are marked in bold.

variable	PC1	PC2	PC3	PC4	PC5
aridity	7.030	0.057	0.029	0.372	0.476
bio_1	2.041	9.035	0.572	1.010	1.785
bio_12	5.639	2.252	1.042	0.337	0.459
bio_15	1.728	0.210	8.026	1.746	0.002
bio_19	1.363	1.777	1.869	0.049	0.242
bio_2	3.580	1.512	3.047	1.444	0.301
bio_5	5.316	3.083	0.052	0.779	1.328
bio_7	3.913	3.915	0.431	0.001	0.035
bio_8	0.001	2.152	2.173	4.239	0.027
bio_9	2.723	3.210	2.171	0.193	2.186
dissim_h	3.969	0.121	1.269	0.299	11.547
altitude	0.511	3.354	6.549	4.360	0.039
forest	3.929	0.754	0.056	0.001	1.709
lgp	4.110	4.757	0.070	0.650	0.193
nutrient_a	0.395	2.138	2.483	10.119	0.137
ph	1.407	0.013	0.001	0.141	0.104
soil_depth	0.346	3.062	1.124	0.637	0.000
sparsely_veg	3.292	1.927	1.020	0.622	0.067
toxicity_	0.187	2.337	3.085	10.824	0.257
workabilit	1.411	3.410	4.835	1.083	0.360
slope	1.993	0.180	5.190	2.680	1.026
cv_h	0.237	0.681	0.079	0.510	18.581
grass_scrub	0.600	0.385	0.116	4.857	0.304
tpi	0.022	0.000	0.005	0.077	0.547
bio_10	3.991	5.271	0.000	1.557	1.266
bio_11	0.427	10.964	1.636	0.384	1.870

variable	PC1	PC2	PC3	PC4	PC5
bio_13	3.107	3.353	2.949	0.000	0.872
bio_14	5.037	0.075	0.363	2.825	0.091
bio_16	3.208	3.301	2.909	0.001	0.873
bio_17	5.120	0.114	0.309	2.865	0.099
bio_18	3.470	1.352	0.348	0.940	2.047
bio_3	0.390	0.027	8.102	2.740	0.645
bio_4	2.295	3.144	3.381	0.526	0.293
bio_6	0.164	11.949	0.654	0.757	1.616
contrast_h	2.856	0.253	0.941	0.308	14.253
evapotrans	3.751	1.143	4.289	0.005	1.205
excess_sal	0.193	1.326	1.665	10.545	1.487
nutrient_r	0.060	1.769	3.080	11.679	0.088
oxygen	0.015	2.227	1.892	11.832	0.387
rooting	1.365	3.033	4.934	0.944	0.243
roughness	1.999	0.141	5.461	2.631	0.890
std_h	3.782	0.064	1.043	0.248	13.623
UVB1_annual	0.240	0.004	10.058	1.906	0.703
variance_h	2.787	0.171	0.692	0.278	15.736

Supplementary Table S5. Median estimated ages and 95% of confidence intervals (CI) for the best ML phylogenetic tree by penalized likelihood method implemented in treePL software (see text for details). Node ID numbers correspond to those represented in Supplementary Figure S9.

Node ID	Type of Node	Median age (Ma)	Lower 95% CI	Upper 95% CI
246	internal	69.31	63.12	74.82
247	internal	54.02	46.82	60.60
248	internal	48.96	40.62	57.47
249	internal	47.24	38.10	56.31
250	internal	38.21	32.50	44.55
251	internal	35.88	30.40	41.32
252	internal	31.05	27.11	35.44
253	internal	30.11	26.37	34.18
254	internal	25.25	22.25	28.46
255	internal	21.85	19.91	25.33
256	internal	20.46	18.03	23.12
257	internal	18.46	16.59	20.81
258	internal	17.42	15.59	19.38
259	internal	16.69	15.20	18.88
260	internal	14.86	13.63	16.93
261	internal	13.82	12.69	15.88
262	internal	12.77	11.45	14.50
263	internal	10.66	9.46	12.01

Node ID	Type of Node	Median age (Ma)	Lower 95% CI	Upper 95% CI
264	internal	8.86	7.93	9.98
265	internal	8.10	7.26	9.16
266	internal	6.55	5.87	7.55
267	internal	5.93	5.32	6.90
268	internal	5.00	4.48	5.91
269	internal	3.76	3.42	4.60
270	internal	3.52	3.22	4.37
271	internal	3.24	2.96	4.03
272	internal	3.19	2.90	3.95
273	internal	3.13	2.84	3.85
274	internal	3.07	2.76	3.74
275	internal	2.85	2.58	3.40
276	internal	2.76	2.49	3.23
277	internal	2.70	2.41	3.13
278	internal	2.59	2.31	2.95
279	internal	2.48	2.17	2.76
280	internal	2.36	2.03	2.57
281	internal	2.32	1.98	2.53
282	internal	2.25	1.89	2.45
283	internal	2.16	1.81	2.37
284	internal	2.06	1.72	2.28
285	internal	1.91	1.56	2.08
286	internal	1.82	1.47	1.99
287	internal	1.72	1.37	1.88
288	internal	1.65	1.31	1.80
289	internal	1.53	1.20	1.67
290	internal	1.40	1.09	1.53
291	internal	1.59	1.33	1.76
292	internal	1.93	1.62	2.10
293	internal	1.66	1.39	1.81
294	internal	2.25	1.99	2.55
295	internal	2.02	1.80	2.31
296	internal	1.63	1.46	1.88
297	internal	1.94	1.72	2.20
298	internal	1.98	1.76	2.26
299	internal	1.61	1.44	1.85
300	internal	1.94	1.72	2.33
301	internal	2.57	2.27	3.04
302	internal	2.29	2.03	2.72
303	internal	1.88	1.66	2.23
304	internal	1.38	1.22	1.64
305	internal	2.87	2.58	3.47
306	internal	2.81	2.48	3.34
307	internal	2.61	2.31	3.07
308	internal	2.46	2.18	2.87
309	internal	2.26	2.04	2.65

Node ID	Type of Node	Median age (Ma)	Lower 95% CI	Upper 95% CI
310	internal	2.13	1.92	2.49
311	internal	1.61	1.45	1.86
312	internal	1.87	1.68	2.18
313	internal	1.95	1.73	2.29
314	internal	2.42	2.14	2.86
315	internal	2.23	1.97	2.64
316	internal	2.00	1.77	2.37
317	internal	1.60	1.42	1.89
318	internal	2.12	1.88	2.50
319	internal	1.75	1.55	2.04
320	internal	2.70	2.43	3.28
321	internal	2.25	2.03	2.75
322	internal	1.67	1.51	2.04
323	internal	2.74	2.48	3.46
324	internal	2.28	2.00	2.86
325	internal	2.16	1.90	2.71
326	internal	1.63	1.43	2.04
327	internal	2.11	1.85	2.65
328	internal	1.57	1.38	1.99
329	internal	2.33	2.05	2.95
330	internal	2.79	2.43	3.57
331	internal	2.15	1.85	2.82
332	internal	1.66	1.43	2.21
333	internal	1.98	1.71	2.65
334	internal	2.64	2.34	3.29
335	internal	3.48	3.15	4.23
336	internal	3.32	2.99	4.01
337	internal	3.11	2.79	3.71
338	internal	2.96	2.61	3.47
339	internal	2.80	2.47	3.25
340	internal	2.67	2.40	3.11
341	internal	2.60	2.34	3.02
342	internal	2.37	2.13	2.73
343	internal	2.23	2.00	2.56
344	internal	1.84	1.63	2.08
345	internal	1.89	1.69	2.16
346	internal	2.00	1.80	2.34
347	internal	2.35	2.12	2.72
348	internal	2.14	1.93	2.47
349	internal	1.67	1.50	1.92
350	internal	1.49	1.34	1.73
351	internal	2.18	1.92	2.55
352	internal	2.95	2.64	3.52
353	internal	2.51	2.25	3.00
354	internal	2.13	1.88	2.50
355	internal	2.75	2.49	3.33

Node ID	Type of Node	Median age (Ma)	Lower 95% CI	Upper 95% CI
356	internal	2.24	2.03	2.71
357	internal	3.28	2.93	3.85
358	internal	2.60	2.33	3.05
359	internal	5.33	4.78	6.19
360	internal	4.64	4.16	5.35
361	internal	3.96	3.56	4.56
362	internal	2.55	2.30	2.92
363	internal	2.44	2.19	2.82
364	internal	3.84	3.45	4.48
365	internal	5.56	5.00	6.43
366	internal	4.87	4.39	5.63
367	internal	4.46	4.00	5.14
368	internal	3.71	3.28	4.17
369	internal	3.56	3.11	3.96
370	internal	3.24	2.78	3.54
371	internal	3.15	2.69	3.45
372	internal	3.07	2.61	3.36
373	internal	2.93	2.46	3.21
374	internal	2.67	2.22	2.95
375	internal	2.49	2.10	2.79
376	internal	2.36	1.98	2.64
377	internal	2.05	1.66	2.25
378	internal	1.83	1.47	2.01
379	internal	1.36	1.08	1.49
380	internal	1.50	1.19	1.64
381	internal	2.12	1.76	2.36
382	internal	1.79	1.45	1.96
383	internal	1.71	1.39	1.88
384	internal	2.67	2.23	2.92
385	internal	2.43	2.03	2.67
386	internal	1.89	1.58	2.09
387	internal	1.84	1.53	2.01
388	internal	2.03	1.71	2.23
389	internal	1.56	1.30	1.71
390	internal	1.15	0.96	1.26
391	internal	2.96	2.56	3.24
392	internal	2.35	2.04	2.59
393	internal	3.28	2.87	3.65
394	internal	2.62	2.30	2.93
395	internal	2.08	1.82	2.32
396	internal	1.64	1.44	1.83
397	internal	2.03	1.79	2.28
398	internal	2.97	2.59	3.29
399	internal	2.66	2.31	2.92
400	internal	2.22	1.92	2.42
401	internal	3.11	2.75	3.54

Node ID	Type of Node	Median age (Ma)	Lower 95% CI	Upper 95% CI
402	internal	2.95	2.61	3.35
403	internal	2.77	2.46	3.16
404	internal	2.48	2.20	2.80
405	internal	2.40	2.13	2.71
406	internal	1.84	1.63	2.07
407	internal	1.90	1.68	2.13
408	internal	2.04	1.80	2.29
409	internal	1.56	1.37	1.73
410	internal	2.45	2.18	2.81
411	internal	2.82	2.51	3.23
412	internal	2.52	2.24	2.88
413	internal	2.31	2.05	2.63
414	internal	2.02	1.79	2.30
415	internal	2.06	1.86	2.42
416	internal	4.08	3.67	4.73
417	internal	3.70	3.33	4.32
418	internal	3.32	2.93	3.88
419	internal	2.74	2.41	3.23
420	internal	2.44	2.15	2.89
421	internal	2.03	1.79	2.40
422	internal	1.23	1.08	1.44
423	internal	1.86	1.66	2.24
424	internal	1.73	1.55	2.09
425	internal	3.22	2.90	3.75
426	internal	3.48	3.13	4.03
427	internal	2.88	2.56	3.28
428	internal	2.38	2.11	2.67
429	internal	2.01	1.78	2.25
430	internal	1.93	1.70	2.28
431	internal	5.91	5.24	6.49
432	internal	5.53	4.88	6.08
433	internal	4.93	4.33	5.42
434	internal	3.67	3.21	4.03
435	internal	3.22	2.81	3.54
436	internal	2.18	1.91	2.40
437	internal	1.54	1.34	1.69
438	internal	1.35	1.17	1.48
439	internal	4.14	3.63	4.55
440	internal	1.03	0.90	1.13
441	internal	4.99	4.37	5.49
442	internal	2.94	2.56	3.24
443	internal	2.23	1.93	2.46
444	internal	2.26	1.98	2.48
445	internal	4.31	3.75	4.74
446	internal	2.18	1.88	2.40
447	internal	3.53	3.07	3.90

Node ID	Type of Node	Median age (Ma)	Lower 95% CI	Upper 95% CI
448	internal	4.83	4.31	5.37
449	internal	8.16	7.40	9.40
450	internal	6.01	5.52	7.02
451	internal	5.25	4.80	6.12
452	internal	4.25	3.88	4.95
453	internal	4.35	3.98	5.08
454	internal	4.21	3.87	4.96
455	internal	8.79	7.85	10.97
456	internal	6.87	6.12	8.63
457	internal	6.17	5.45	8.05
458	internal	7.99	6.84	9.22
459	internal	5.92	5.24	6.86
460	internal	5.25	4.65	6.04
461	internal	4.59	4.04	5.25
462	internal	5.22	4.56	5.69
463	internal	4.09	3.59	4.40
464	internal	12.57	11.36	14.07
465	internal	11.54	9.97	12.63
466	internal	9.93	8.62	11.12
467	internal	9.27	7.84	10.27
468	internal	8.78	7.40	9.81
469	internal	8.20	6.88	9.26
470	internal	7.65	6.35	8.73
471	internal	6.70	5.43	7.66
472	internal	6.00	4.85	6.82
473	internal	5.58	4.51	6.28
474	internal	3.14	2.47	3.54
475	internal	4.97	3.97	5.62
476	internal	5.97	4.96	6.98
477	internal	9.50	8.49	10.67
478	internal	15.75	13.39	17.72
479	internal	13.99	11.49	16.24
480	internal	9.75	8.15	11.16
481	internal	14.08	12.57	16.39
482	internal	6.33	5.79	7.34
483	internal	24.01	21.15	27.08
484	internal	21.13	18.51	24.10
485	internal	23.98	20.99	27.22
486	internal	29.70	25.96	33.97
487	internal	16.49	14.03	18.95
488	internal	13.84	12.09	15.76
489	internal	45.51	39.40	50.73

Supplementary Table S6. RASP distribution node area inference. Codes correspondence are: A= MED; B = CB; C = IT1; D = IT2; E = IT3; F = IT4; G = IT5; H = HIM; and I =EA.

Node ID	1st prob		2nd prob		3rd prob		4th prob		5th prob	
	area	%	area	%	area	%	area	%	area	%
188	F	100								
189	DF	38.58	BF	38.54	F	22.89				
190	F	100								
191	F	100								
192	F	58.18	CF	41.82						
193	F	45.83	AF	44.43	CF	9.75				
194	F	86.21	EF	13.79						
195	F	100								
196	F	100								
197	FG	67.35	F	32.65						
198	EF	100								
199	EF	100								
200	E	79.75	EF	20.25						
201	DE	81.62	E	18.38						
202	D	100								
203	D	100								
204	EF	49.25	E	20.84	DE	11.24	F	9.49	DF	9.18
205	DE	55.55	DF	25.41	D	19.04				
206	DE	82.99	DF	17.01						
207	E	100								
208	EF	79.58	E	20.42						
209	DE	82.64	DF	17.36						
210	D	100								
211	D	100								
212	D	100								
213	DF	55.03	F	44.97						
214	DF	52.22	F	47.78						
215	DF	53.49	F	46.51						
216	DF	57.35	D	42.65						
217	D	78.98	DF	21.02						
218	F	100								
219	F	100								
220	EF	100								
221	F	100								
222	F	100								
223	D	100								
224	DF	100								
225	F	100								
226	F	100								
227	DF	56.81	F	43.19						
228	E	100								
229	EF	100								

Node ID	1st prob		2nd prob		3rd prob		4th prob		5th prob	
	area	%	area	%	area	%	area	%	area	%
230	F	61.48	DF	25.83	EF	12.69				
231	DF	85.2	F	14.8						
232	F	100								
233	F	100								
234	F	100								
235	DF	84.42	F	15.58						
236	D	76.62	DF	23.38						
237	D	78.23	CD	21.77						
238	D	100								
239	D	100								
240	D	100								
241	D	100								
242	D	100								
243	D	100								
244	D	100								
245	D	100								
246	D	100								
247	D	100								
248	F	100								
249	DF	100								
250	DF	79.23	D	20.77						
251	D	100								
252	DF	77.42	D	22.58						
253	F	100								
254	F	100								
255	F	100								
256	F	100								
257	F	100								
258	F	100								
259	F	100								
260	F	100								
261	F	100								
262	F	100								
263	F	100								
264	F	100								
265	F	100								
266	F	100								
267	F	100								
268	F	100								
269	F	100								
270	F	100								
271	F	100								
272	F	100								
273	F	100								

Node ID	1st prob		2nd prob		3rd prob		4th prob		5th prob	
	area	%	area	%	area	%	area	%	area	%
274	F	100								
275	DF	100								
276	D	100								
277	D	100								
278	D	85.42	DF	14.58						
279	D	100								
280	D	100								
281	D	100								
282	D	100								
283	D	100								
284	D	100								
285	D	100								
286	B	100								
287	B	100								
288	B	100								
289	B	100								
290	B	100								
291	B	100								
292	B	100								
293	B	100								
294	B	100								
295	B	100								
296	BE	79.12	B	20.88						
297	B	100								
298	B	100								
299	B	100								
300	B	100								
301	A	100								
302	A	100								
303	A	72.69	AB	27.31						
304	B	52	AB	48						
305	B	100								
306	B	100								
307	B	57.44	BE	42.56						
308	B	100								
309	D	50.42	BD	49.58						
310	D	52.6	BD	47.4						
311	D	100								
312	D	51.8	BD	48.2						
313	EF	100								
314	EF	69.89	E	30.11						
315	BF	65.04	BE	34.96						
316	BD	62.46	B	37.54						
317	B	50.52	BD	49.48						

Node ID	1st prob		2nd prob		3rd prob		4th prob		5th prob	
	area	%	area	%	area	%	area	%	area	%
318	B	100								
319	B	100								
320	B	100								
321	B	100								
322	B	100								
323	B	100								
324	B	100								
325	B	100								
326	B	100								
327	AB	61.77	B	38.23						
328	B	100								
329	B	100								
330	B	100								
331	B	100								
332	B	100								
333	B	52.4	BD	47.6						
334	B	100								
335	B	100								
336	B	100								
337	B	100								
338	B	100								
339	B	100								
340	B	100								
341	B	100								
342	B	100								
343	B	100								
344	B	100								
345	B	100								
346	B	79.46	BD	20.54						
347	B	100								
348	B	100								
349	BD	55.06	B	44.94						
350	BD	59.99	D	40.01						
351	D	83.81	BD	16.19						
352	D	100								
353	D	77.87	DH	22.13						
354	HI	51.65	I	25.32	H	23.03				
355	HI	100								
356	I	100								
357	HI	100								
358	H	57.62	HI	42.38						
359	DH	77.05	H	22.95						
360	H	80.46	DH	19.54						
361	H	82.92	DH	17.08						

Node ID	1st prob		2nd prob		3rd prob		4th prob		5th prob	
	area	%	area	%	area	%	area	%	area	%
362	D	63.81	DH	36.19						
363	D	53.29	DF	46.71						
364	D	73.94	DH	26.06						
365	D	45.58	DH	44.19	H	10.23				
366	D	73.59	DH	26.41						
367	DH	51.28	D	48.72						
368	DH	56.12	D	43.88						
369	DH	83.66	H	16.34						
370	H	100								
371	DH	81.38	H	18.62						
372	D	54.31	DH	37.93	H	7.76				
373	CD	64.12	CH	19.49	DH	16.39				

Supplementary Table S7. Results of fitted models for ancestral character reconstruction. Model abbreviations correspond to: ARD, null model with different transition rates for all possible transitions (–all rates differ"); SYM, symmetrical model in which rates between any two states do not differ; SYM-ERp, symmetrical model with equal rates for the three types of perennial lifeforms; ER, model with all rates equal. The best-fitting model according to AIC is indicated in bold, together with the other model with a ΔAIC<2. Subscripted numbers in transition rates (q) correspond to the following growth form types: 1, annual; 2, biennial; 3, acaulescent perennials; 4, caulescent perennials formed by basal rosettes with leafless flowering stems; 5, caulescent perennials with leafy flowering stem.

Model	log-likelihood	Number of parameters	AIC	ΔAIC	Estimated transition rates
ER	-203.94	1	409.88	41.02	$q=0.032$
SYM	**-174.43**	**10**	**368.86**	**0**	**$q_{12}<0.001$, $q_{13}<0.001$, $q_{23}=0.014$, $q_{14}<0.001$, $q_{14}<0.001$, $q_{24}<0.001$, $q_{34}=0.047$, $q_{51}=0.006$, $q_{25}=0.016$, $q_{35}=0.037$, $q_{45}=0.104$**
SYM-ERp	-176.66	8	369.33	0.47	$q_{perennial}=0.073$, $q_{12}<0.001$, $q_{13}=0.013$, $q_{23}=0.042$, $q_{14}<0.001$ $q_{24}<0.001$, $q_{51}<0.001$, $q_{25}<0.001$
ARD	-166.35	20	372.70	3.84	

Supplementary Table S8. Fit of diversification models applied to the phylogenetic tree of *Jurinea.* The best-fitting model accoding to AICc is highlighted in bold. Abbreviations: NP, number of free parameters; logL, log-likelihood; ΔAICc, the difference in AIC score between the model with the lowest AIC (i.e. best-fit model) and the model being compared; λ, speciation rate; α, rate of variation of speciation through time (in time-dependent models) or according to temperature (in climate-dependent models); μ, extinction rate; K, carrying-capacity in diversity-dependent models; λ_{main}, speciation rate of the main clade; $\lambda_{subclade}$, speciation rate of the decoupled subclade; μ_{main}, extinction rate of the main clade; $\mu_{subclade}$, extinction rate of the subclade; K_{main}, carrying-capacity of the main clade; $Ks_{ubclade}$, carrying-capacity of the subclade; λ_1, speciation rate before the shift; λ_2, speciation rate after the shift; μ_1, extinction rate before the shift; μ_2, extinction rate after the shift; K_1, carrying-capacity before the shift; K_2, carrying-capacity after the shift.

MODEL TYPE	MODEL DESCRIPTION	MODEL ACRONYM	NP	logL	AICc	ΔAICc	λ	α	μ
Constant-rate	Constant speciation and no extinction	BCST	1	-364.2	730.4		0.430	–	–
	Constant speciation and constant extinction	BCSTDCST	2	-364.2	732.4		0.430	–	<0.00001
Time-dependent	Speciation exponentially correlated with time, no extinction	BTimeVar_EXPO	2	-347.0	698.1		0.280	0.197	–
	Speciation exponentially correlated with time, constant extinction	BTimeVarDCST_EXPO	3	-347.0	700.1		0.280	0.197	<0.00001
	Speciation linearly correlated with time, no extinction	BTimeVar_LIN	2	-306.8	617.7		0.080	0.279	–
	Speciation linearly correlated with time, constant extinction	BTimeVarDCST_LIN	3	-306.8	619.8		0.080	0.279	<0.00001
Climate-dependent	Speciation exponentially correlated with temperature, no extinction	BEnv.Var_EXPO	2	-330.3	664.7		0.130	0.437	–
	Speciation exponentially correlated with temperature, constant extinction	BEnv.VarDCST_EXPO	3	-330.3	666.8		0.130	0.437	<0.00001
	Speciation linearly correlated with temperature, no extinction	BEnv.Var_LIN	2	-293.6	591.3		0.480	0.360	–
	Speciation linearly correlated with temperature, constant extinction	BEnv.VarDCST_LIN	3	-293.6	593.3		0.470	0.357	<0.00001

MODEL TYPE	MODEL DESCRIPTION	MODEL ACRONYM	NP	logL	AICc	ΔAICc	λ	μ	K
Diversity dependent	Linear dependence in speciation rate, no extinction, K corresponds to carrying capacity where speciation = extinction	DDL	2	-271.6	547.3	68.2	1.630	–	319

	Linear dependence in speciation rate, with extinction, K corresponds to carrying capacity where speciation = extinction	DDL+E	3	-288.8	583.8	104.7	1.610		0.00119		243	
	Linear dependence in speciation rate, no extinction, K' corresponds to carrying capacity where speciation = 0	DDL_K'	2	-265.5	535.0	55.9	1.630		–		332	
	Linear dependence in speciation rate, with extinction, K' corresponds to carrying capacity where speciation = 0	DDL_K'+E	3	-272.6	551.4	72.3	1.630		0.00021		319	
	Exponential dependence in speciation rate, with extinction, K corresponds to carrying capacity where speciation = extinction	DDX+E	3	-354.7	715.5		4.754		0.57129		432	
MODEL TYPE	**MODEL DESCRIPTION**	**MODEL ACRONYM**	**NP**	**logL**	**AICc**	**ΔAICc**	**λ_{main}**	**$\lambda_{subclade}$**	**μ_{main}**	**$\mu_{subclade}$**	**K_{main}**	**$K_{subclade}$**
Diversity-dependent, diversification dynamics of CA decoupled	Linear dependence in speciation rate, decoupling of CA, no extinction, K' corresponds to carrying capacity where speciation = 0	DDL_K'_CA	5	-249.5	509.3	30.2	0.550	2.025	0.00000	0.00000	263	172
	Linear dependence in speciation rate, decoupling of CA, with extinction, K' corresponds to carrying capacity where speciation = 0	DDL_K' +E_CA	7	-251.6	517.9	38.8	0.575	2.103	0.00000	0.00000	260	170
	Linear dependence in speciation rate, decoupling of CA, no extinction	DDL_CA	5	-245.7	501.7	22.6	0.726	3.117	0.00000	0.00000	312	119
	Linear dependence in speciation rate, decoupling of CA, with extinction	DDL+E_ CA	7	-281.4	577.5	98.4	1.123	2.892	0.49162	0.14187	319	116
Diversity-dependent, diversification dynamics of CB decoupled	Linear dependence in speciation rate, decoupling of CB, no extinction, K' corresponds to carrying capacity where speciation = 0	DDL_K'_CB	5	-272.6	555.6	76.5	1.270	0.960	0.00000	0.00000	319	92
	Linear dependence in speciation rate, decoupling of CB, with extinction, K' corresponds to carrying capacity where speciation = 0	DDL_K' +E_CB	7	-296.6	607.8	128.7	1.235	0.512	0.00775	0.00430	256	190
	Linear dependence in speciation rate, decoupling of CB, no extinction	DDL_CB	5	-286.3	583.0	103.9	1.218	0.704	–	–	319	134
	Linear dependence in speciation rate, decoupling of CB, with extinction	DDL+E_ CB	7	-329.1	672.7	193.6	1.609	1.614	1.21	0.31	347	319

Diversity-dependent, diversification dynamics of EA decoupled	Linear dependence in speciation rate, decoupling of EA, no extinction, K' corresponds to carrying capacity where speciation = 0	DDL_K'_EA	5	-236.2	482.8	3.7	1.760	0.951	–	–	342	24
	Linear dependence in speciation rate, decoupling of EA, with extinction, K' corresponds to carrying capacity where speciation = 0	**DDL_K' +E_EA**	**7**	**-232.2**	**479.1**	**0.0**	**1.838**	**0.851**	0.00001	0.00023	**342**	**25**
	Linear dependence in speciation rate (model 1), decoupling of EA, no extinction	DDL_EA	5	-239.2	488.7	9.6	1.510	0.869	–	–	343	24
	Linear dependence in speciation rate (model 1), decoupling of EA, with extinction	DDL+E_ EA	7	-237.0	488.6	9.6	1.793	0.331	0.00681	<0.00001	347	41
MODEL TYPE	**MODEL DESCRIPTION**	**MODEL ACRONYM**	**NP**	**logL**	**AICc**	**ΔAICc**	**λ_1**	**λ_2**	**μ_1**	**μ_2**	**K_1**	**K_2**
Diversity-dependent, shifting parameters at 3 mya	Linear dependence in speciation rate with shifting parameters at time t=3.0, no extinction	DDL_SR	4	-238.1	484.4	5.3	1.036	1.919	0.00000	0.00000	335	353
	Linear dependence in speciation rate with shifting parameters at time t=3.0, with extinction	DDL+E_SR	6	-244.8	502.1	23.1	0.996	1.793	0.39171	0.00021	228	319

Supplementary Table S9. Fit of diversification models applied to the phylogenetic tree corresponding to the CB+CA subset. The best-fitting model accoding to AICc is highlighted in bold. Abbreviations: NP, number of free parameters; logL, log-likelihood; ΔAICc, the difference in AIC score between the model with the lowest AIC (i.e. best-fit model) and the model being compared; λ, speciation rate; α, rate of variation of speciation through time (in time-dependent models) or according to temperature (in climate-dependent models); μ, extinction rate; K, carrying-capacity in diversity-dependent models; λ_1, speciation rate before the shift; λ_2, speciation rate after the shift; μ_1, extinction rate before the shift; μ_2, extinction rate after the shift; K_1, carrying-capacity before the shift; K_2, carrying-capacity after the shift.

MODEL TYPE	MODEL DESCRIPTION	MODEL ACRONYM	NP	logL	AICc	ΔAICc	λ	α	μ
Constant-rate	Constant speciation and no extinction	BCST	1	-314.1	630.2	259.2	0.453	0.000	0.00000
	Constant speciation and constant extinction	BCSTDCST	2	-314.1	632.3	261.3	0.453	0.000	0
Time-dependent	Speciation exponentially correlated with time, no extinction	BTimeVar_EXPO	2	-281.4	566.8	195.8	0.227	0.364	–
	Speciation exponentially correlated with time, constant extinction	BTimeVarDCST_EXPO	3	-281.3	568.8	197.8	0.227	0.364	0
	Speciation linearly correlated with time, no extinction	BTimeVar_LIN	2	-237.6	479.3	108.3	0.164	0.384	–
	Speciation linearly correlated with time, constant extinction	BTimeVarDCST_LIN	3	-237.6	481.4	110.4	0.164	0.383	0
Temperature-dependent	Speciation exponentially correlated with temperature, no extinction	BEnvVar_EXPO	2	-266.8	537.6	166.6	0.089	0.608	–
	Speciation exponentially correlated with temperature, constant extinction	BEnvVarDCST_EXPO	3	-266.8	539.7	168.7	0.089	0.609	<0.00001
	Speciation linearly correlated with temperature, no extinction	BEnvVar_LIN	2	-230.7	465.4	94.4	0.649	0.465	–
	Speciation linearly correlated with temperature, constant extinction	BEnvVarDCST_LIN	3	-230.7	467.5	96.5	0.649	0.465	<0.00001

MODEL TYPE	MODEL DESCRIPTION	MODEL ACRONYM	NP	logL	AICc	ΔAICc	λ	μ	K
Diversity-dependent	Linear dependence in speciation rate, no extinction, K corresponds to carrying capacity where speciation = extinction	DDL	2	-215.6	435.2	64.2	1.997	–	373
	Linear dependence in speciation rate, with extinction, K corresponds to carrying capacity where speciation = extinction	DDL+E	3	-213.8	433.7	62.7	1.997	0	273
	Linear dependence in speciation rate, no extinction, K' corresponds to carrying capacity where speciation = 0	DDL_K'	2	-194.2	392.4	21.4	2.010	–	341
	Linear dependence in speciation rate, with extinction, K' corresponds to carrying capacity where speciation = 0	DDL_K'+E	3	-217.2	440.6	69.6	1.997	0	273
	Exponential dependence in speciation rate, with extinction, K corresponds to carrying capacity where speciation = extinction	DDX+E	3	-326.0	658.2	287.2	0.953	0.92000	642

MODEL TYPE	MODEL DESCRIPTION	MODEL ACRONYM	NP	logL	AICc	ΔAICc	λ1	λ2	μ1	μ2	K1	K2
Diversity-dependent, shifting parameters at 3 mya	**Linear dependence in speciation rate with shifting parameters at time t=3.0, no extinction**	**DDL_SR**	**4**	**-181.4**	**371.0**	**0.0**	**0.937**	**1.767**	**0.00000**	**0.00000**	**6163**	**331**
	Linear dependence in speciation rate with shifting parameters at time t=3.0, with extinction	DDL+E_SR	6	-182.2	376.8	5.8	0.694	1.841	0.05263	0.00000	6160	330

Supplementary Table S10. Percentage of variables contribution to each axis of the PCA analyses: (A) Mean and Unique values and (B) Multiple and Unique values.

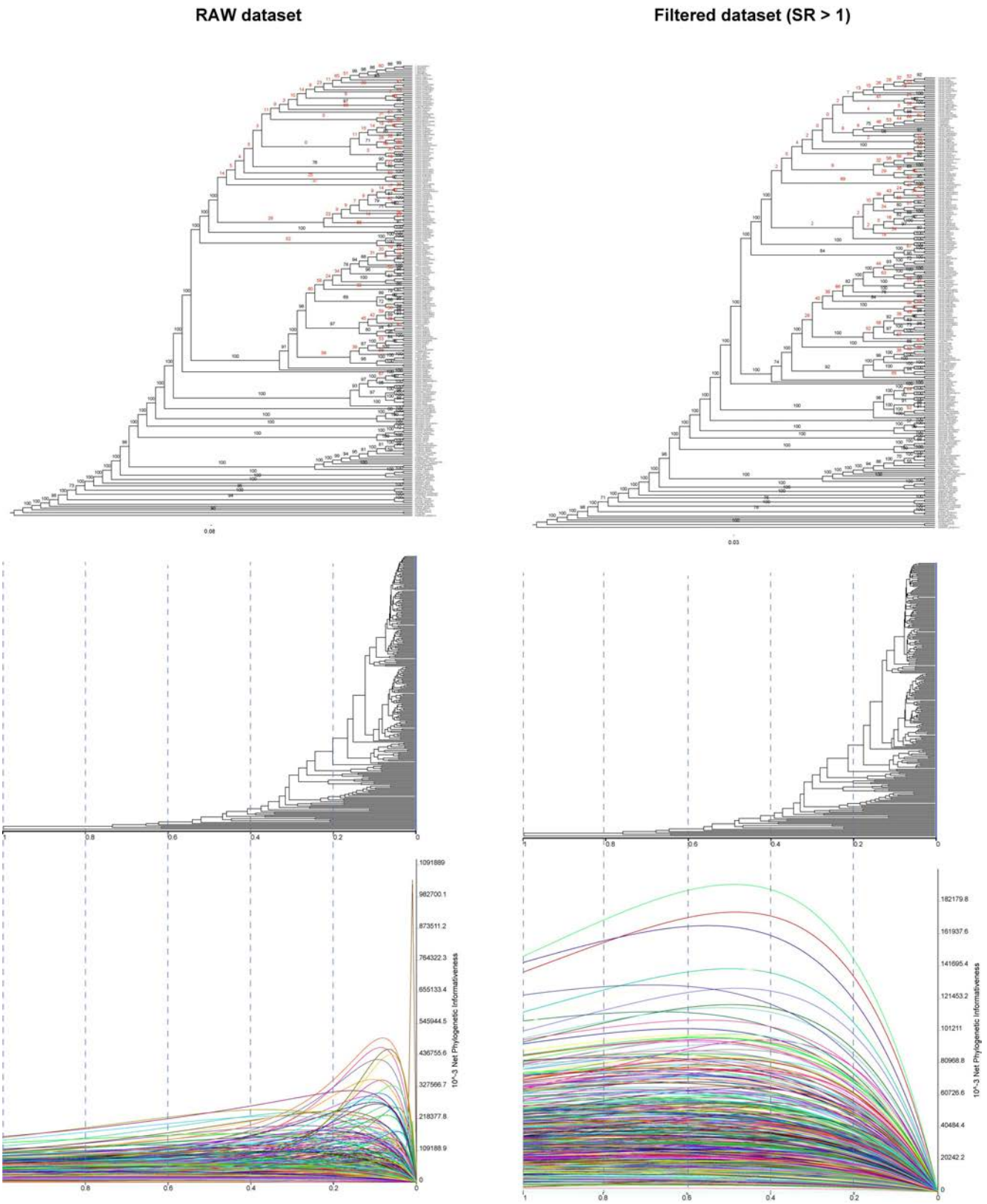
Variable	(A) PC1	(A) PC2	(B) PC1	(B) PC2
aridity	12.8	0.0	13.0	2.2
bio_1	2.9	7.3	0.3	0.6
bio_12	9.9	0.1	13.2	0.0
bio_15	3.3	3.9	0.1	9.6
bio_19	2.9	0.0	8.3	4.1
bio_2	7.6	6.3	4.1	11.5
bio_5	8.8	4.1	9.2	0.5
bio_7	8.3	1.2	10.5	0.0
bio_8	0.0	10.9	1.1	12.0
bio_9	4.1	0.3	0.0	12.6
habitat heterogenity	5.4	0.1	1.2	2.6
altitude	0.4	19.2	2.0	15.0
forest	8.5	0.0	8.0	0.4
LGP	8.6	3.2	9.3	2.9
nutrient availability	0.2	4.8	1.2	0.1
pH	2.7	0.0	6.0	0.5
slope	3.3	7.5	4.8	9.0
soil depth	0.3	10.5	0.3	1.4
sparsely vegetation	7.2	4.2	4.1	4.9
toxicity	0.9	4.1	0.4	1.4
workability	1.8	12.2	3.0	8.7

Supplementary Table S11. Environmental niche metric values (in percentage) of comparisons among phylogenetic groups of PCA-env analysis.

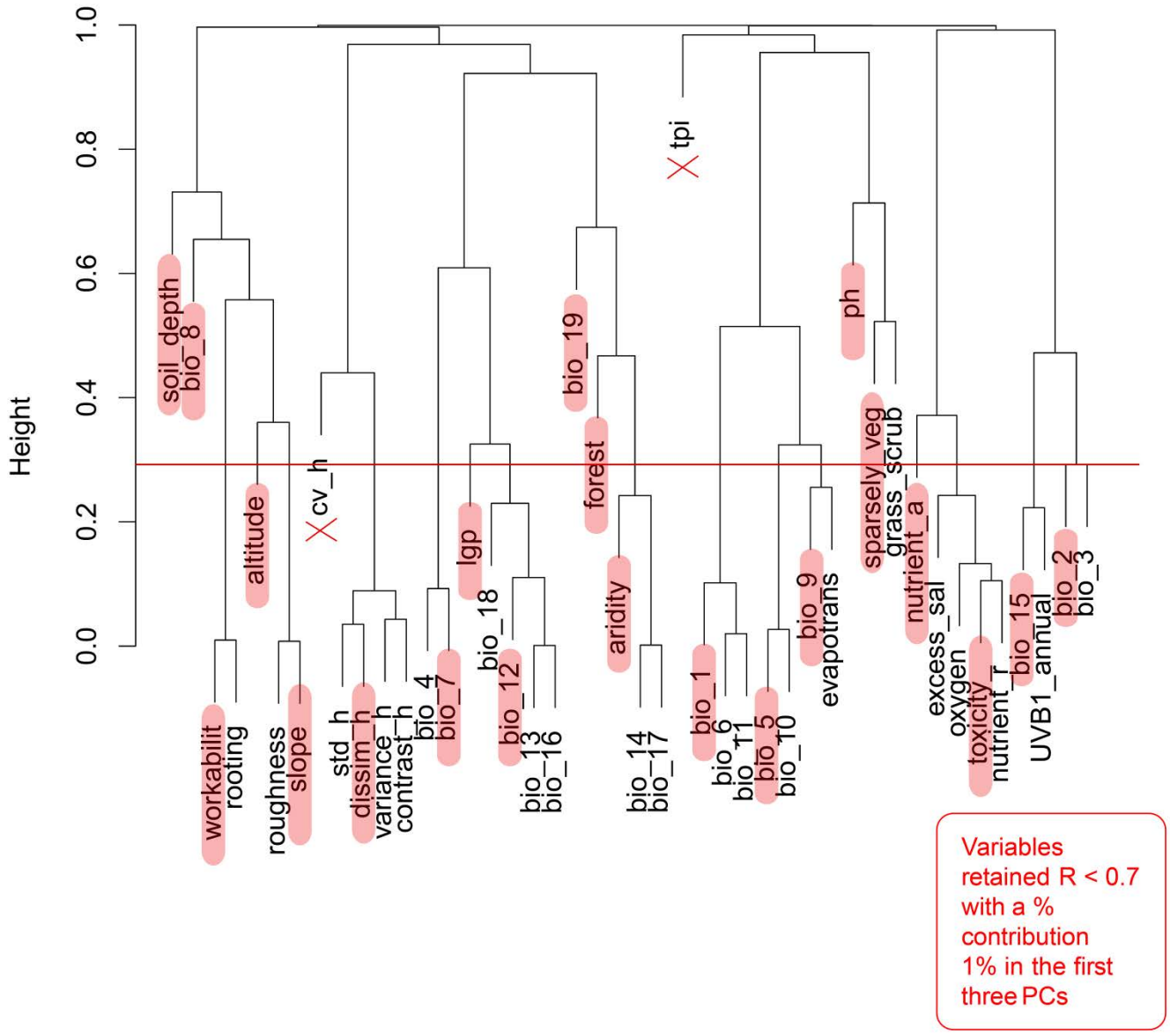
Niche overlap (*D*) values (%)					
Phylo groups	**EA**	**CB**	**5**	**6**	**CA**
EA	-	-	-	-	-
CB	0.33	-	-	-	-
5	0.09	0.04	-	-	-
6	0.001	0.002	0.12	-	-
CA	0.32	0.30	0.24	0.11	-
Equivalency test (*p*-value < 0.05 = equivalent niches)					
Phylo groups	**EA**	**CB**	**5**	**6**	**CA**
EA	-	**0.01**	0.99	NA	**0.01**
CB	**0.01**	-	1.00	NA	0.68
5	0.98	1.00	-	1.00	0.86
6	NA	NA	1.00	-	0.50
CA	**0.01**	0.73	0.82	0.48	-

Equivalency test (*p*-value < 0.05 = divergent niches)					
Phylo groups	**EA**	**CB**	**5**	**6**	**CA**
EA	-	1.00	**0.04**	NA	1.00
CB	1.00	-	**0.01**	NA	0.31
5	**0.04**	**0.01**	-	**0.01**	0.20
6	NA	NA	**0.01**	-	0.56
CA	1.00	0.24	0.20	0.56	-
Similarity test (*p*-value < 0.05 = similar niches)					
Phylo groups	**EA**	**CB**	**5**	**6**	**CA**
EA	-	**0.01**	**0.02**	0.25	**0.01**
CB	**0.02**	-	0.27	0.39	**0.04**
5	**0.03**	0.33	-	0.36	**0.01**
6	0.34	0.39	0.35	-	0.30
CA	**0.01**	**0.04**	**0.01**	0.39	-
Similarity test (*p*-value < 0.05 = non-similar niches)					
Phylo groups	**EA**	**CB**	**5**	**6**	**CA**
EA	-	0.98	0.98	0.70	1.00
CB	1.00	-	0.77	0.52	0.99
5	0.96	0.65	-	0.73	1.00
6	0.70	0.60	0.72	-	0.70
CA	0.99	1.00	1.00	0.64	-
Niche unfilling metric (%)					
Phylo groups	**EA**	**CB**	**5**	**6**	**CA**
EA	-	0.33	0.96	1.00	0.51
CB	0.03	-	0.91	0.99	0.09
5	0.64	0.54	-	0.47	0.00
6	0.87	0.78	0.33	-	0.00
CA	0.44	0.41	0.50	0.62	-
Niche expansion metric (%)					
Phylo groups	**EA**	**CB**	**5**	**6**	**CA**
EA	-	0.03	0.64	0.87	0.44
CB	0.33	-	0.54	0.78	0.41
5	0.96	0.91	-	0.33	0.50
6	1.00	1.00	0.47	-	0.62
CA	0.51	0.09	0.00	0.00	-
Niche stability metric (%)					
Phylo groups	**EA**	**CB**	**5**	**6**	**CA**
EA	-	0.97	0.36	0.13	0.56
CB	0.67	-	0.46	0.22	0.60
5	0.04	0.10	-	0.67	0.50
6	0.004	0.006	0.53	-	0.39
CA	0.50	0.91	1.00	1.00	-

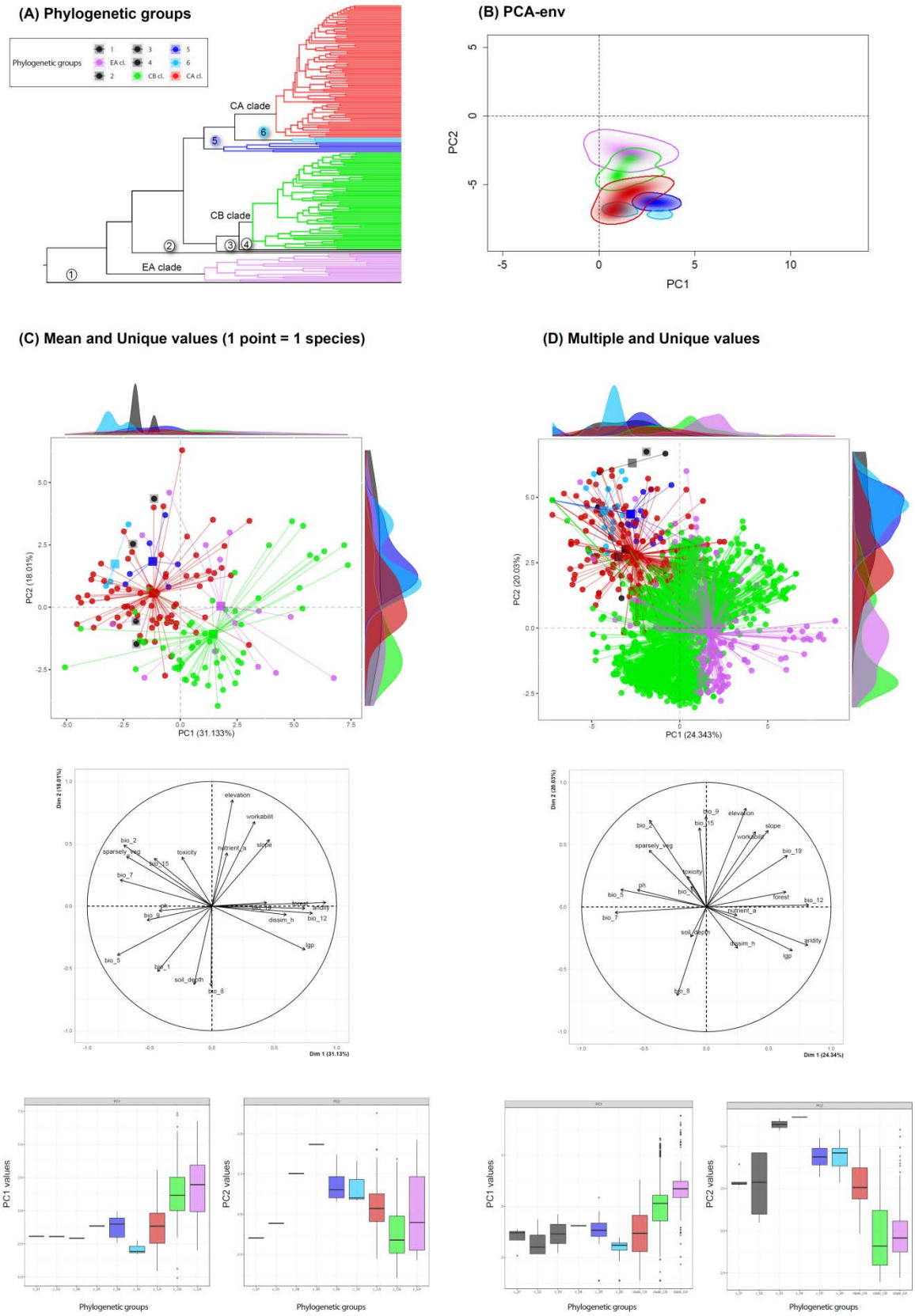
Supplementary figures



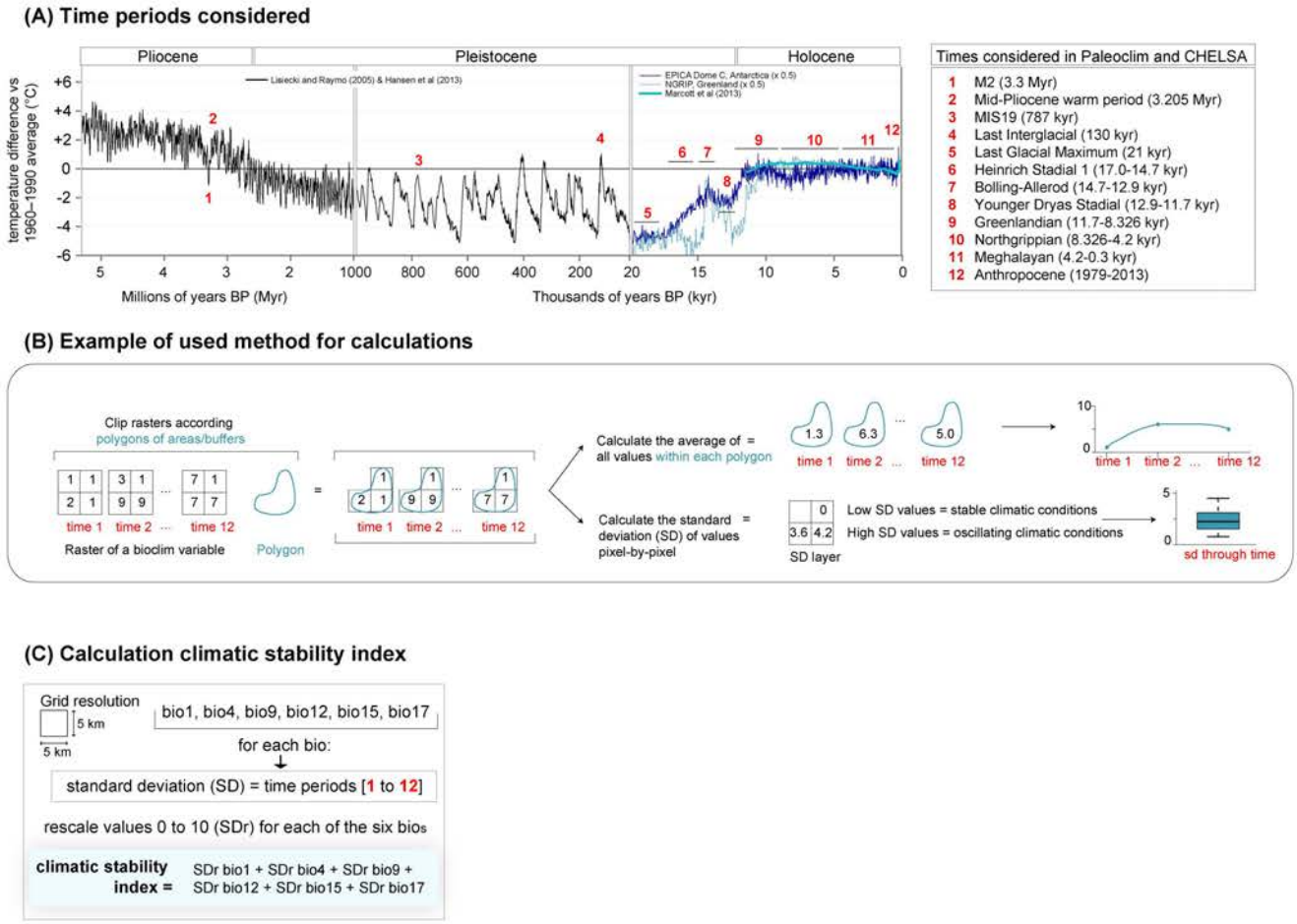
Supplementary Figure S1. Phylogenetic informativeness analyses comparing raw dataset and filtered dataset with a threshold value of substitution rate (SR) higher than 1. Trees were obtained using the concatenation approach with a maximum likelihood inference. In first figure row the trees show in red branches with low bootstrap (bs) support values (bs < 70). In the second row, the trees are transformed to ultrametric and scaled to an arbitrary scale of 1 (at the root) to 0 (at the tips). The last row shows net phylogenetic informativeness profiles displaying curves for each locus in different colors.



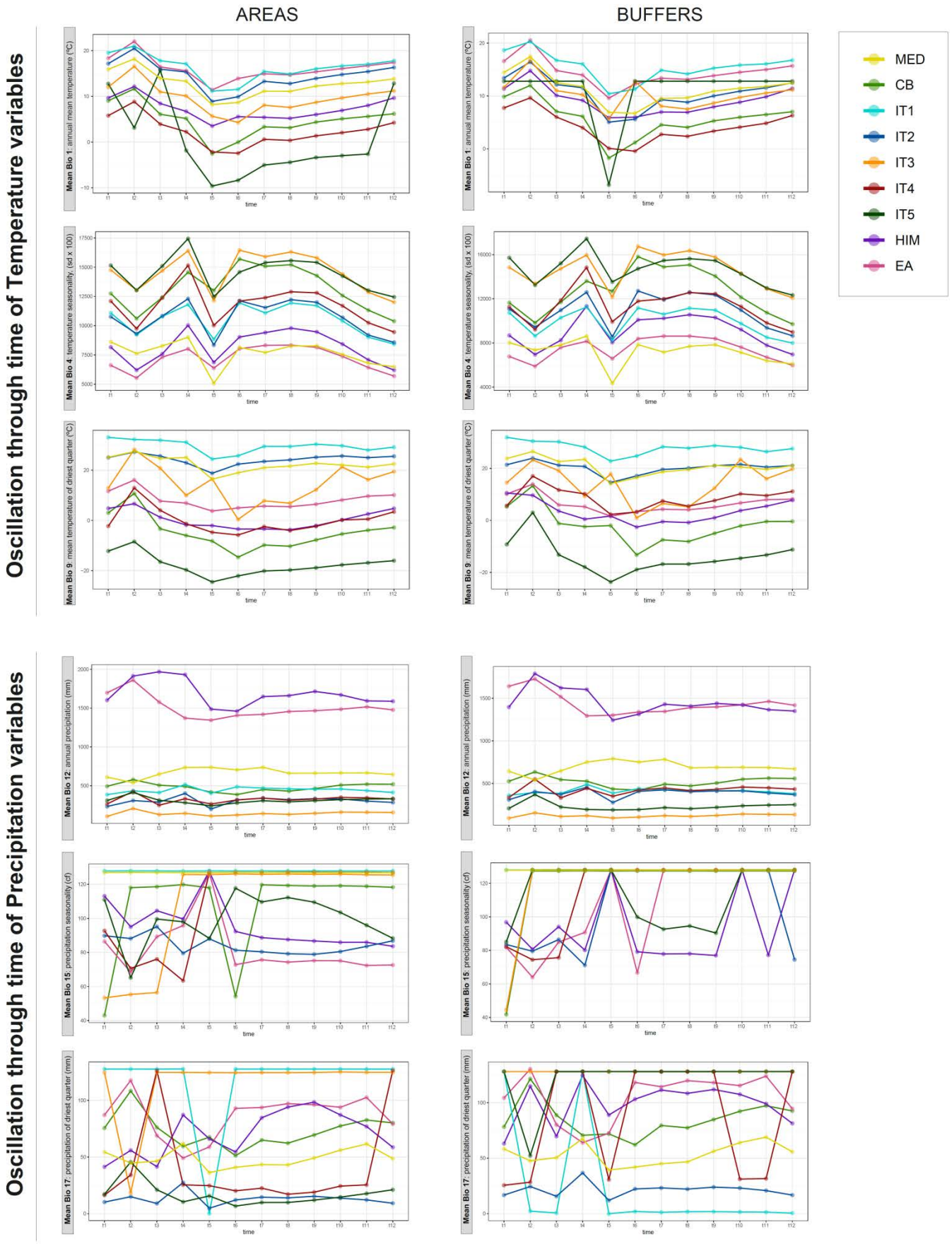
Supplementary Figure S2. Dendrogram of values from Pearson correlation analysis of initial set of 45 environmental variables. The final 21 variables retained are highlighted in red.



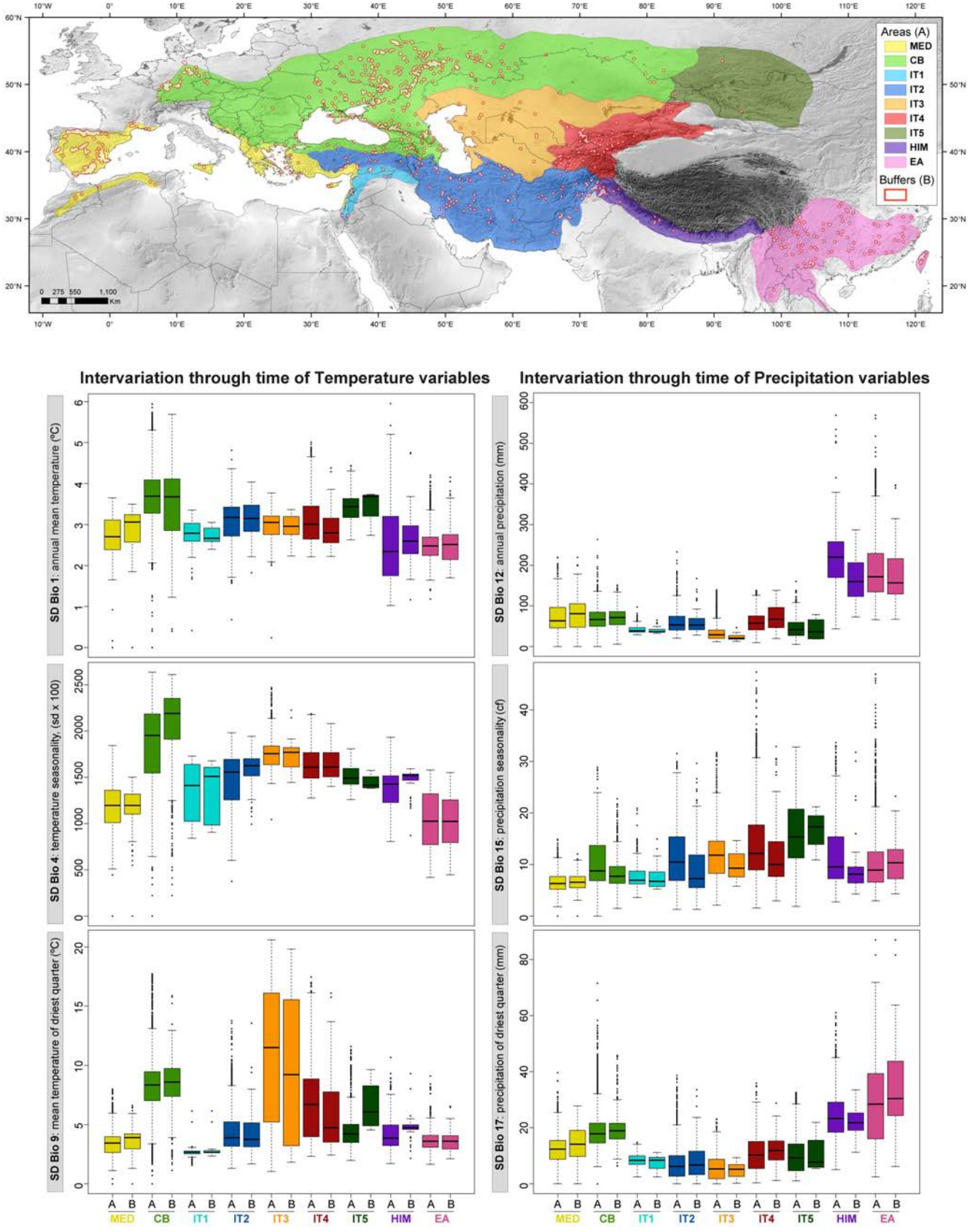
Supplementary Figure S3. Classification of phylogenetic groups (A) and Graphical output of Principal Component Analysis (PCA) of PCA-env (B) and standard PCA for the dataset with unique values (C) for each species (some calculated the average and the rest obtained by the single occurrence record found) and the complete (D) dataset (with all occurrence records). The PCA graphs represent the global environmental space those occupied by each species colored according their phylogenetic group. The direction of main contributing variables to the two-first components of the standard PCAs (C and D) are also outlined. Values range of each PCs of standard PCAs (C and D) are also drawn by each phylogenetic group and each dataset.



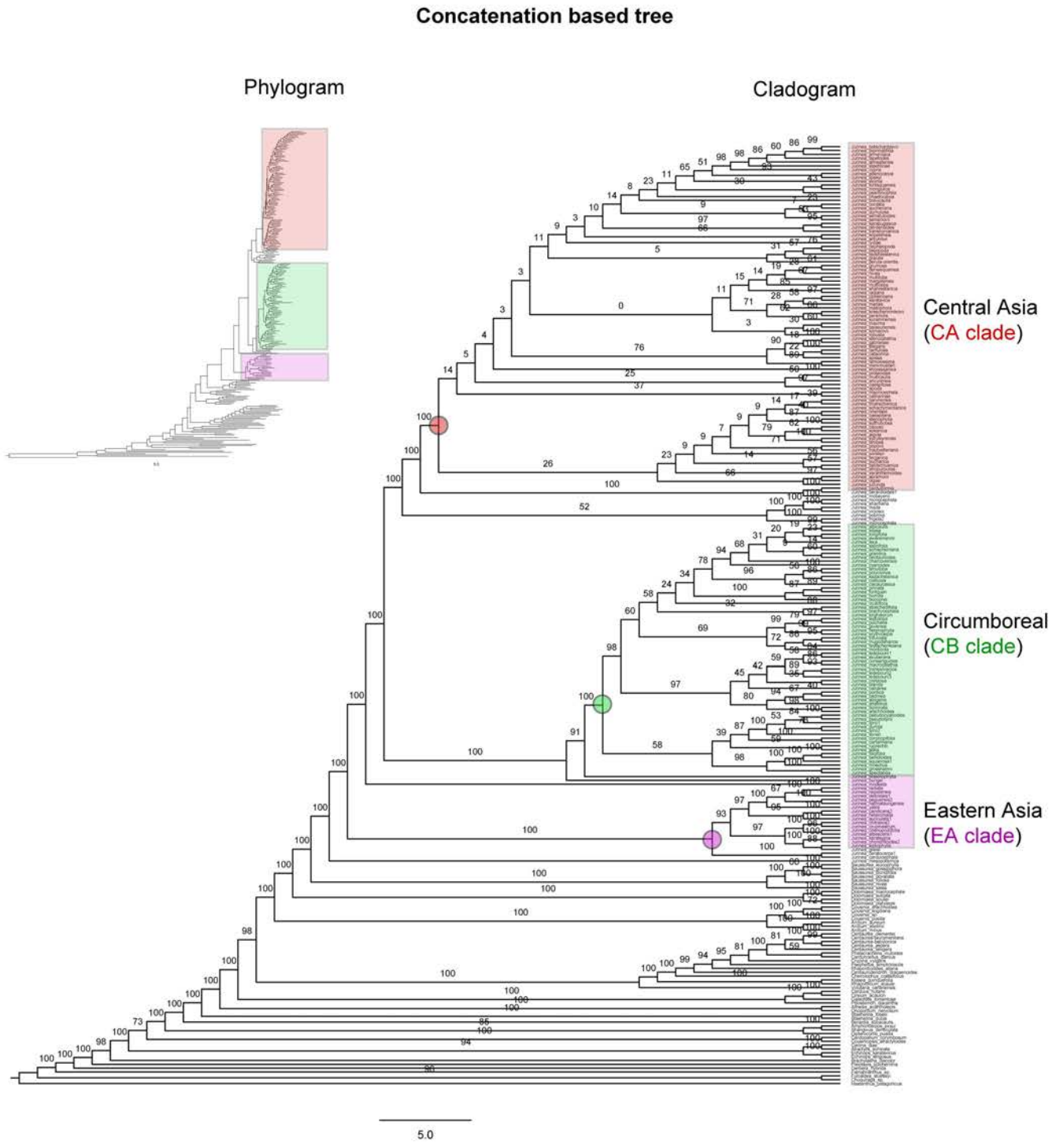
Supplementary Figure S4. (A) Time periods considered in paleoclimatic analysis from *PaleoClim* dataset (Brown et al. 2018, which include CHELSA data by Karger et al. 2017 for periods 5 and 12) outlined in a temperature through time graph modified from <https://en.wikipedia.org/wiki/Geologic_temperature_record>. (B) Graphical representation of methodology employed to extract the paleoclimatic oscillations data. (B) Graphical summary of the workflow employed to calculate the climatic stability index.



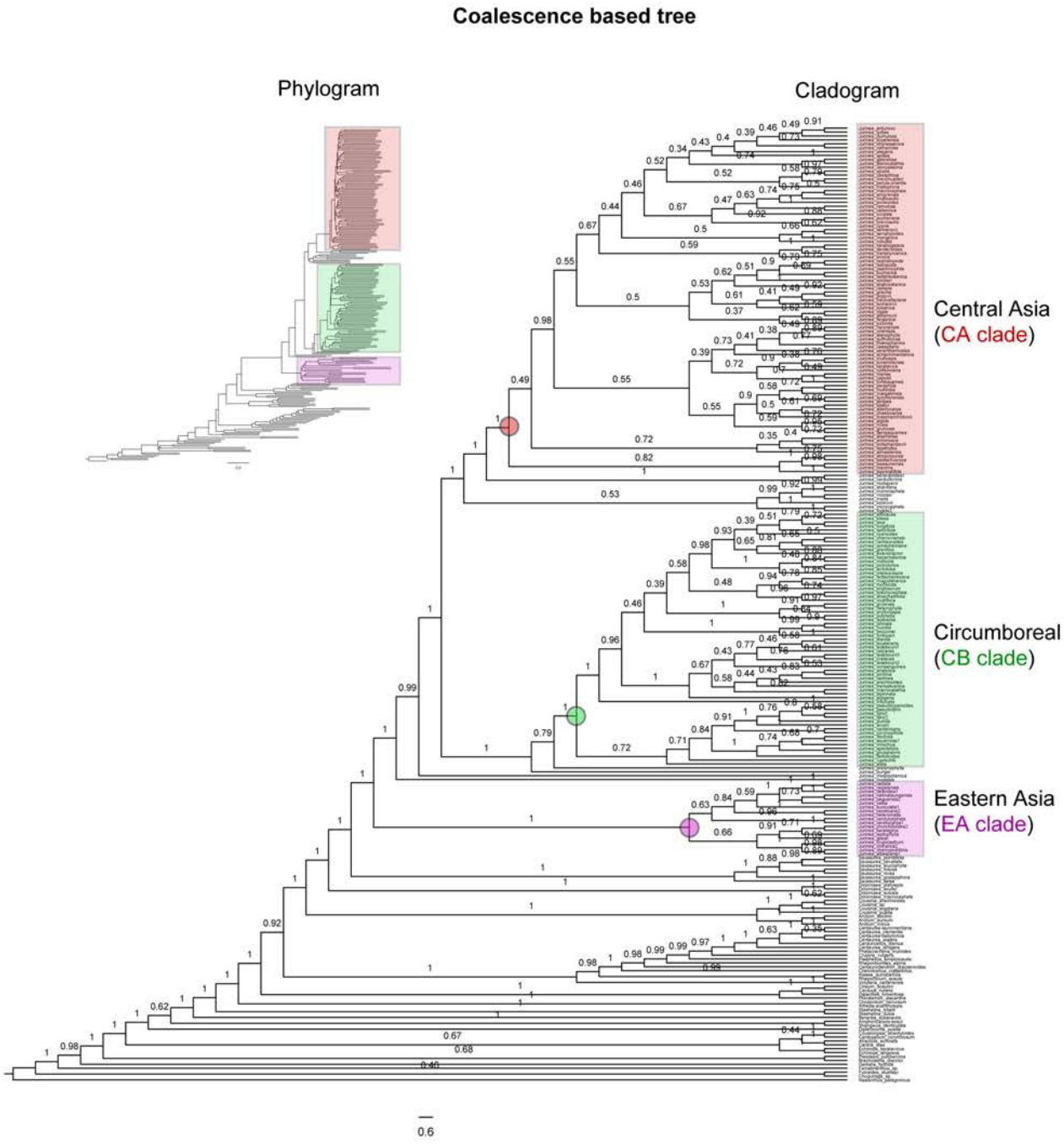
Supplementary Figure S5. Representation of oscillations through 12 time intervals (see Supplementary Fig. S4A) of Paleclimatic dataset (Brown et al. 2018) according temperature variables (bio1, bio4, bio9) and precipitation variables (bio12, bio15, bio17) for nine biogeographic areas and buffers around occurrence records (Fig. 2).



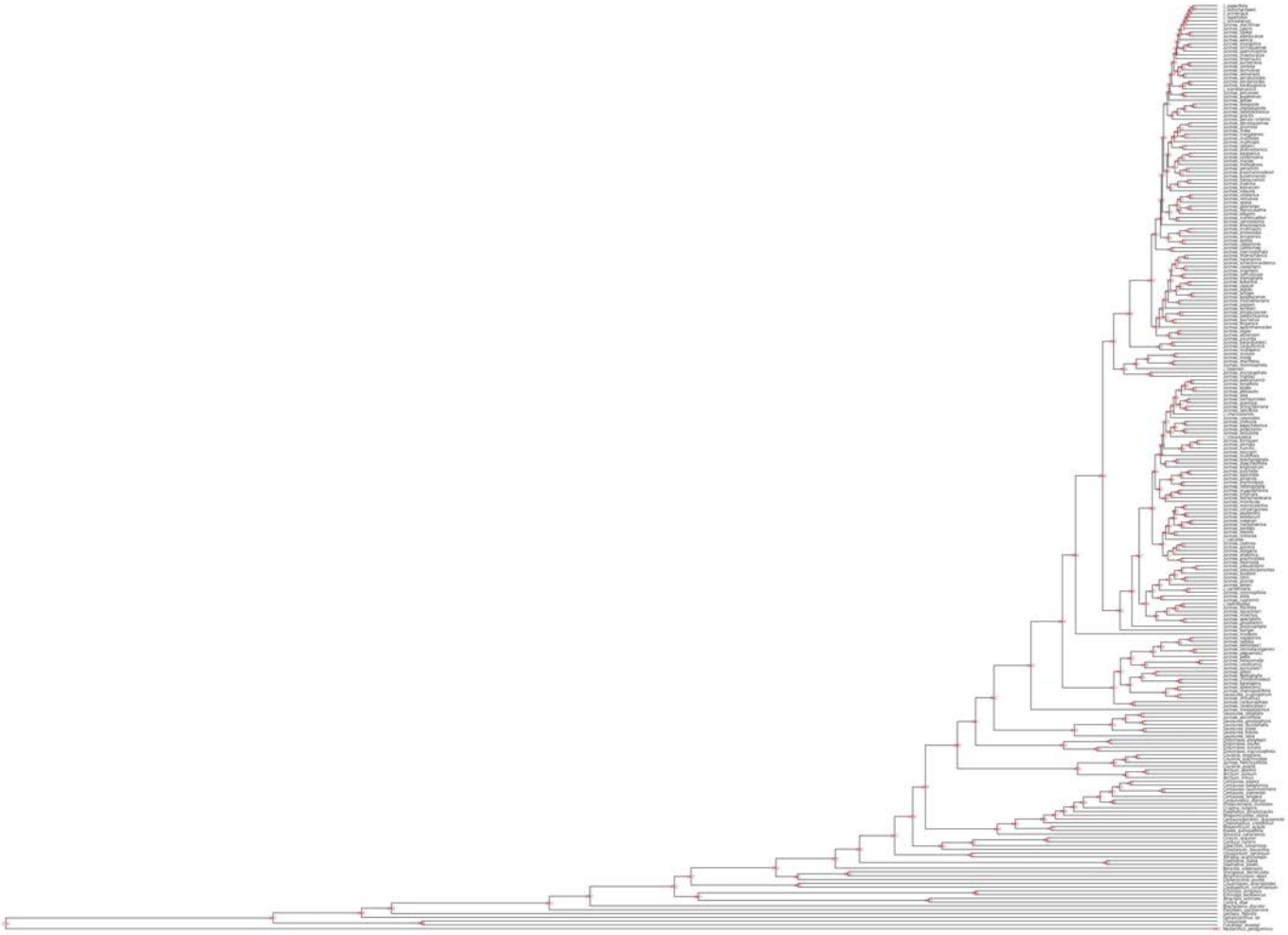
Supplementary Figure S6. Value range of the standard deviation (SD) for the 12 time periods considered (see Supplementary Fig. S4A) of each climatic variable (temperature: bio1, bio4, bio9; precipitation: bio12, bio15, bio17) in different categories nine biogeographic areas and buffers around occurrence records.



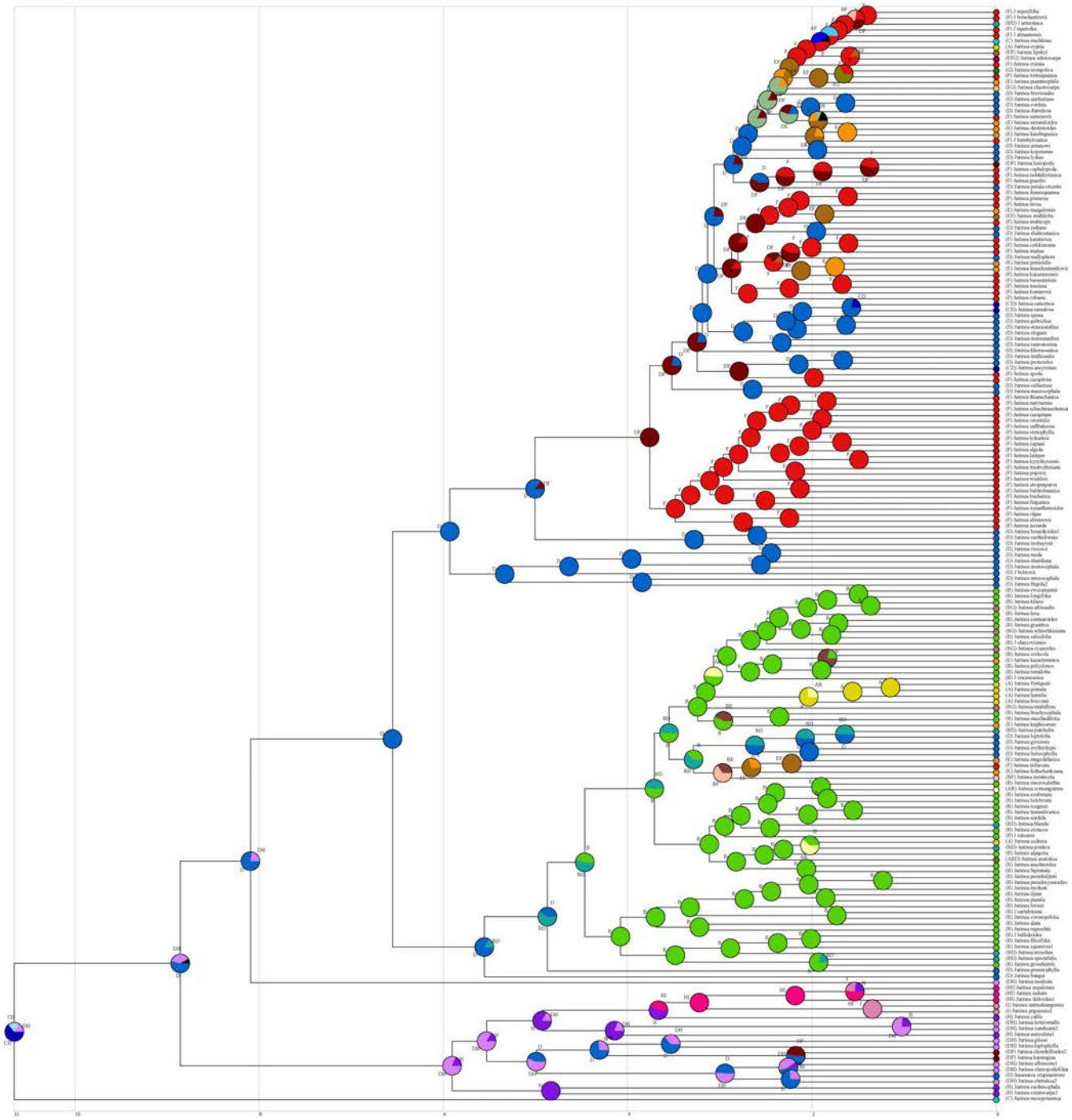
Supplementary Figure S7. Phylogenetic tree inferred from the concatenation approach with the unfiltered dataset. Bootstrap (BS) support values are shown over branches.



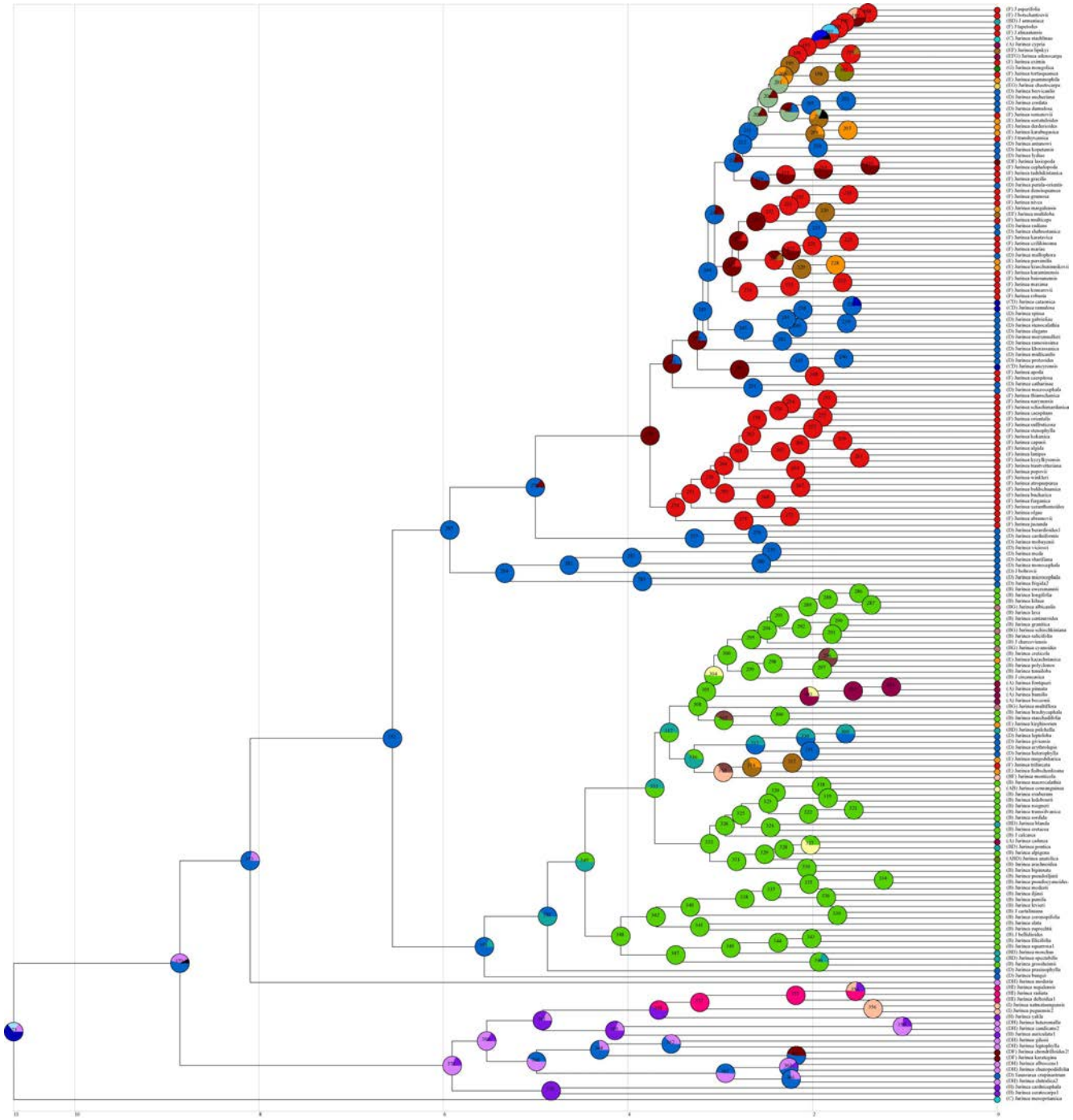
Supplementary Figure S8. Phylogenetic tree inferred from the coalescence approach with the unfiltered dataset. Local posterior probabilities (LPP) are shown over branches.



Supplementary Figure S9. Best Maximum Likelihood (ML) phylogenetic tree with node IDs corresponding to dating analysis with treePL. See Supplementary Table S5 for node IDs, median age (Ma), lower and upper 95% CI.



Supplementary Figure S10. Time-calibrated phylogeny and biogeographic inference with specified node IDs. Corresponding percentage values are in Supplementary Table S6.



Supplementary Figure S11. Time-calibrated phylogeny and biogeographic inference with specified on tree nodes the most probable ancestral areas. Corresponding percentage values are in Supplementary Table S6. Codes correspondence are: A= MED; B = CB; C = IT1; D = IT2; E = IT3; F = IT4; G = IT5; H = HIM; and I =EA.

Chapter 6

Testing evolutionary hypotheses for the world's hotspot of temperate alpine flora: the Tibetan-Himalaya-Hengduan region

Sonia Herrando-Moraira[1*] and The Cardueae Radiations Group
In alphabetical order:

Juan Antonio Calleja[2]
You-Sheng Chen[3]
Kazumi Fujikawa[4]
Mercè Galbany-Casals[5]
Núria Garcia-Jacas[1]
Seung-Chul Kim[6]
Jian-Quan Liu[7]
Javier López-Alvarado[5]
Jordi López-Pujol[1]
Jennifer R. Mandel[8]
Iraj Mehregan[9]
Cristina Roquet[5]
Llorenç Sáez[5]
Alexander Sennikov[10]
Alfonso Susanna[1]
Roser Vilatersana[1]
Lian-Sheng Xu[3]

1 Botanic Institute of Barcelona (IBB, CSIC-Ajuntament de Barcelona), Pg. del Migdia, s.n., 08038 Barcelona, Spain
2 Departament of Biology (Botany), Faculty of Sciences, Research Centre on Biodiversity and Global Change (CIBC-UAM), 28049 Madrid, Spain
3 Key Laboratory of Plant Resources Conservation and Sustainable Utilization, South China Botanical Garden, Chinese Academy of Sciences, Guangzhou 510650, China
4 Kochi Prefectural Makino Botanical Garden, 4200-6, Godaisan, Kochi 781-8125, Japan
5 Systematics and Evolution of Vascular Plants (UAB) – Associated Unit to CSIC, Departament de Biologia Animal, Biologia Vegetal i Ecologia, Facultat de Biociències, Universitat Autònoma de Barcelona, 08193 Bellaterra, Spain
6 Department of Biological Sciences, Sungkyunkwan University, Gyeonggi-do 440-746, Republic of Korea
7 Key Laboratory for Bio-Resources and Eco-Environment, College of Life Sciences, Sichuan University, Chengdu, China
8 Department of Biological Sciences, Center for Biodiversity, University of Memphis, Memphis, Tennessee 38152, U.S.A
9 Department of Biology, Science and Research Branch, Islamic Azad University, Tehran, Iran
10 Botanical Museum, Finnish Museum of Natural History, P.O. Box 7, 00014 University of Helsinki, Finland; and Herbarium, Komarov Botanical Institute of Russian Academy of Sciences, Prof. Popov str. 2, 197376 St. Petersburg, Russia

*Correspondence to be sent to: Botanic Institute of Barcelona (IBB, CSIC-Ajuntament de Barcelona), Pg. del Migdia, s.n., 08038 Barcelona, Spain; E-mail: sonia.herrando@gmail.com

In preparation for National Science Review

Abstract

The Tibetan-Himalaya-Hengduan massif is the mountain region with the highest concentration of temperate alpine plant species on Earth and the reasons for this richness are a fascinating mystery. In this paper, we tested the hypotheses proposed by constructing a nearly complete phylogeny of one of top explosive radiations of the region, genus *Saussurea* (400–500 species). Using a macroevolutionary integrative approach, we found that net diversification rate of the genus almost doubled around 6 Ma, corroborating the proposed synergic effect of mountain uplift, intensification of Asian monsoons, and colder conditions that occurred in Messinian (late Miocene). The enormous species accumulation, particularly in Hengduan Mountains is best explained by the mountain geo-diversity hypothesis, for which the steep ecological gradients, extremely rugged terrain, N-S mountain orientation, low-moderate climate oscillations, and historical connectivity of sky-island patches are confirmed as key factors.

Keywords

Climatic stability
Qinghai-Tibetan Plateau
Evolutionary radiations
Diversification
Mountain-geobiodiversity hypothesis
Saussurea

Index

1. Introduction

Mountain regions are among the top-ranked hotspots of biological richness on Earth (Rahbek et al., 2019). In the last decade, scientific efforts to disentangle diversification histories of alpine plant groups has increased exponentially (Hughes and Atchison, 2015). In floristic terms, the Andes in the Southern Hemisphere is probably the greatest biodiversity hotspot of tropical alpine taxa (Pérez-Escobar et al., 2017). In the Northern Hemisphere, the world's species richest area of temperate alpine flora is located in the Tibet-Himalaya-Hengduan mountain system (THH; Ding et al., 2020). However, little is known about the evolutionary history of the THH flora, and it is still one of the poorest explored regions from a botanical perspective (Yang et al., 2014). Unfortunately, the recent increase of anthropic pressures in the region in conjunction with global warming effects are critically endangering its montane-alpine biomes (Dullinger et al., 2012). Thus, we could be in front of the so-called "species extinction before we know them" effect (e.g. Lees and Pimm, 2015), highlighting the need and urgency to fill the knowledge gap concerning the diversity and evolutionary patterns in the THH region.

The THH is located in the Qinghai-Tibet Plateau (QTP) and harbours more than 20,000 seed plant species (Singh & Hajra, 1996). Two out of the 36 global biodiversity hotspots are recognized within the region (Marchese, 2015). The present-day species rich biotas of THH may be explained by the complex and active orogenic history of the region, which likely provided ecological opportunities to lineage radiations. Since the first emergence of the "proto Tibetan-Plateau" when India and Eurasian plates impacted around 50–40 Ma, the region has undergone many mountain uplift and deformation pulses during the Neogene period (Favre et al., 2015). Nowadays, tectonic activity is still present (e.g. Wenchuan earthquake Ms 8.0 in 2008).

Within the THH, the Hengduan Mountains have been of interest of evolutionary biologists particularly during the last decade (Sun et al., 2017; Xing and Ree, 2017). Higher in situ speciation rates of alpine flora have been reported for Hengduan Mountains in comparison with Himalayas or QTP *sensu stricto* (Xing and Ree, 2017). More than 8590 seed plant species are documented to occur within Hengduan Mountains (Zhang et al. 2009a), more than 25% of them endemic (Sun et al., 2017; López-Pujol et al., 2011; Zhao et al., 2016). Large plant radiations, particularly of temperate and alpine elements, are also recorded to have taken place within it: at least, 16 genera with over 100 species are recognized (Sun et al., 2017). All these exceptional biological estimates are in part due to the physiographic features of these mountains. Hengduan's range occupies ca. 500,000 km^2 (20% of total THH extension), with elevations averaging 2000–4500 m a.s.l. and including seven main parallel mountain valleys dissected by rivers (Sun et al., 2017). These mountains are located in the south-eastern edge and the geologically youngest area of the QTP (Favre et al., 2015). In fact, Hengduan Mountains are one of the most heterogeneous places on Earth in terms of topography and climate, in addition to harbour two outstanding features: an abrupt slope aspect and a north-south direction of large fault zones or gorges. For example, the Gongga Shan harbors an impressive altitudinal range from the valley, 1050 m a.s.l. to the mountain peak, 7556 m a.s.l. in less than 30 km of distance.

Biogeographic connections with neighbouring regions have also shaped the current assembly of THH flora. Many plant lineages of Northern Hemisphere diverged within the THH and then dispersed to the rest of Eurasia. Additionally, some of the immigrated lineages back-colonized the THH region. According to the "out-of-QTP hypothesis" (Wen et al., 2014) THH would have constituted an important biodiversity source to adjacent biogeographic areas and the cradle of many Eurasian plant lineages. However, no consensus has yet been reached concerning the timing of origin of the alpine flora in the THH and the colonization of adjacent regions. According to recent phylogenetic meta-analyses of 18 plant groups highly diversified within THH (Xing and Ree, 2017; Ding et al., 2020), in situ speciation rates would have peaked in two main periods. The first was inferred to occur during the late Oligocene to early Miocene, but with distinct time diversification shifts in Hengduan range (27–24 Ma) than in Himalayas (19–17 Ma). The second has been identified after the mid-Miocene climatic optimum, but again with remarkable differences regarding pulses (8–7 Ma in Hengduan Mountains, and 13 Ma and 10–5 Ma in the Himalayas). Geological and paleoclimatic events have been correlated to these diversification explosions of alpine temperate flora. In the first interval (late Oligocene to early Miocene) the transition from global warm climates to gradual cooling started, in conjunction with an active period of tectonic activity (Guo and Wilson, 2019; Shen et al., 2020). As for the second burst phase (late Miocene), three potential drivers, and accordingly three derived hypotheses, have been suggested to act synchronously in the Hengduan Mountains region (Ding et al., 2020): (1) pulses of intense tectonic activity ("uplift-driven diversification" hypothesis; Xing and Ree, 2017); (2) monsoon intensification ("monsoon-driven diversification"; Ding et al., 2020); and (3) cooler climates after warm periods ("warm-cold colonization"; Meng et al., 2017). Further studies based on broadly sampled phylogenies of plant radiations are needed to test the proposed hypotheses and deepen our knowledge on the biogeographic and evolutionary significance of THH at global scale.

For a complete understanding of the high present-day levels of montane-alpine species in THH region, another type of deduction has been recently formulated, but yet not explicitly tested: the "mountain-geobiodiversity hypothesis" (Mosbrugger et al., 2018). This theoretical explanation includes other key interacting factors that maximize the punctual effects of orogeny and climatic turnovers, since a time lag is observed between

Table 1. Hypotheses proposed to explain the alpine flora evolution in THH region (Tibetan-Himalaya-Hengduan). *Saussurea* is studied here as representative model to test the previously suggested hypotheses. References: [1]Wen et al. (2014), [2]Wang et al. (2015), [3]Ding et al. (2020), [4]Xing and Ree (2017), [5]Stokstad (2020), [6]Meng et al. (2017), [7]Mosbrugger et al. (2018), [8]Flantua and Hooghiemstra (2018), [9]Muellner-Riehl (2019).

Question addressed	Hypothesis tested
(1) Is the THH region a biogeographic source of lineages distributed in Eurasia and NorthAmerica?	(1.1.) **Out-of-Tibet or Out-of-QTP**[1,2] It postulates the role of QTP (or THH) as source (center of origin and source of dispersal) to other adjacent Northern Hemisphere regions. It is based on the assumption that ancestral lineages that migrated to other regions benefitted from the Late Miocene cooling onwards thanks to the pre-adaptation to cold-tolerant conditions.
(2) When the rate of species accumulation increased in THH?	(2.1.) **Late Oligocene to early Miocene**[3] Hengduan Mountains at 27–24 Ma Himalayas at 19–17 Ma (2.2.) **Late Miocene**[3] Hengduan Mountains at 8–7 Ma Himalayas at 13 Ma and 10–5 Ma
(3) Which processes or drivers contributed to diversification burst in THH?	(3.1.) **Uplift-driven diversification**[4] Orogeny and mountain-building activity (3.2.) **Monsoon-driven diversification**[3,5] Moonson rains that provide moisture and erosion that carved up the landscape (i.e. reshaping landscape[5]) (3.3.) **Warm-cold colonization**[6] Climatic cooling promoted divergence of some warm lowland adapted clades to cold highlands
(4) Which are the key factors to explain the present-day biodiversity in THH?	(4.1.) **Mountain-geodiversity**[7,] The effects on diversification of mountain surface uplift and paleoclimatic changes are enhanced by four additional factors: (a) steep zonation; (b) climate oscillations; (c) rugged terrain; and (d) high connectivity dynamics (flickering connectivity system[8,9]).

major mountains uplifts and major post mid-Miocene THH speciation events; certainly, the maximum elevations of mountains within THH were almost already reached during the Oligocene to mid-Miocene; Favre et al., 2015; Fang et al., 2020). One of the suggested factors is the steep zonation present in THH mountains, where thanks to its latitude position and temperate climate conditions, the altitudinal gradient is partitioned into distinct eco-zones: lowland, montane, alpine, and subnival. Another factor are the climate oscillations, which in a moderate degree would stimulate species divergence (turnover-pulse hypothesis; Simões et al., 2016). A further factor directly related with the others is the extremely rugged terrain, i.e. an extraordinary topographical complexity of THH with multiple combinations of slope, aspect, curvature, exposition, ruggedness, and soil types, which creates a huge diversity of habitats along each steep gradient. Finally, the last factor recently proposed by Muellner-Riehl (2019) as a complementary concept to the ~~"m~~ountain-geobiodiversity hypothesis", is the so-called ~~"fl~~ickering connectivity system" (Flantua and Hooghiemstra, 2018). It suggests that mountain areas that have been more biologically connected during the past exhibit higher species richness at present time. This premise was proposed for the Andes in the context of Pleistocene oscillations, when repeated cycles of mixing-isolation-mixing (MIM speciation model; He et al., 2018) induced by warm-cold transitions could have resulted in higher diversification rates that those expected from exclusively geographic isolation.

The Compositae genus *Saussurea* DC. is one of the most enigmatic cases of large plant radiations of the THH (ca. 460–493 species; Chen 2015; Raab-Straube 2017). The total diversity of the genus is still unknown, as in recent years many species have been described (e.g. 41 new species in China the last 5-year period; Chen and Xu, 2020). *Saussurea* has a broad and disjunct continental distribution in the Northern Hemisphere, and its major center of diversity corresponds to montane and alpine zones within THH (235 species, from which 63.4% are endemics; Chen, 2015). The largest diversity hotspot is located in the Hengduan Mountains with more than 119 species described, 53.8% being endemics (Zhang et al., 2009b). The genus is also present along the Central Asia Mountains (within the Irano-Turanian floristic region), Eastern Asia (C, N, and NE China, Korea, Japan; which approximately corresponds to the Sino-Japanese floristic region), and northern latitudes corresponding to Circumboreal floristic region (CE Europe, European Russia, Russian Far East, and North America). *Saussurea* can be found in an enormous variety of habitats, e.g. alpine meadows, scree slopes, moist forest, or steppe deserts, and in a wide range of elevational gradients, commonly from 1000–5000 m; indeed, it includes a species living at the highest elevations where vascular plants have been found (*S. gnaphalodes*, 6400 m; Dentant, 2018).

These intrinsic particularities of *Saussurea* have historically attracted the interest of scientists to explore its evolutionary history (Wang et al., 2009; Xing and Ree, 2017; Xu et al., 2019; Zhang et al., 2021). However, despite several efforts to unravel its origin and biogeographic-diversity patterns, *Saussurea*'s evolution is still to be unravelled given its complicated taxonomy and its wide global distribution area, including inaccessible alpine zones or socio-political conflict areas. In consequence, previous

studies had a limited taxon coverage (~10–30% of total species richness) and presented a considerable sampling gap of Central Asian species. Furthermore, their time-calibrated phylogenies were based on chloroplast data, which, due to its maternal inheritance, could result in a misleading signal of the evolutionary history, and its utility to test diversification hypotheses is still under debate (Gonçalves et al., 2019).

Our study aims to unravel the evolutionary history of *Saussurea* as a study-case to test hypotheses proposed by previous authors on the origins and persistence of the outstanding alpine flora found in the THH region (summarized in Table 1). To do so, we first build a nearly complete time-calibrated phylogeny for *Saussurea*—including 324 species (70% of total genus richness)—with Hyb-Seq data (1061 family-specific loci; Mandel et al. 2014). We then apply an integrative approach that combines biogeographic history inference, diversification modelling, ecological niche inference, and climatic stability estimation to address the following specific questions: (1) is the THH region, and particularly the Hengduan Mountains, the origin and dispersal center to adjacent regions of *Saussurea*, as for other alpine elements?; (2) is the diversification burst of *Saussurea* coincident in time and mode with that of other THH elements?; (3) which were the main triggers of the *Saussurea* radiation?; and (4) can diversity patterns of *Saussurea* be explained by the ~~"~~mountain-geodiversity hypothesis"?

2. Results and discussion

The role of THH as a biodiversity source to adjacent regions

The biogeographic analysis based on the phylogenomic data here recovered for *Saussurea* (Supplementary Figs. S1-S3) corroborates the relevant significance of THH as a diversification center and dispersal origin for northern temperate plants (Fig. 1). The Hengduan Mountains, located in the south-eastern edge of THH, are highlighted as the most important primary donor region, since floras of all adjacent regions have species emigrated from Hengduan Mountains lineages. However, we cannot conclude whether Hengduan Mountains are the origin area, since the oldest lineages were inferred to have an uncertain ancestral range (either around the Irano-Turanian region or Hengduan Mountains; Supplementary Figs. S1-S3, Supplementary Tables S1 and S2). Thus, the out-of-QTP hypothesis is not completely fulfilled here for *Saussurea*. The ancestral state reconstruction for the minimum temperatures of the coldest month revealed that both extinct and living species were well adapted to cool climates: below 0ºC up to -30ºC (Supplementary Fig. S4). It is possible then that pre-adaptation to cold environments played an important role in the diversification and evolutionary success of *Saussurea* to become a North Hemisphere wide distributed genus as postulated for other genera like *Gentiana* (Favre et al., 2015) and also for vertebrates (Wang et al., 2015).

Thanks to its high cold-tolerance, *Saussurea* could have taken an evolutionary advantage of the harsh climatic turnovers to coldness. Indeed, the genus divergence dated back to middle Miocene at ca. 15.3 Ma (11.4–17.5; Supplementary Fig. S5 and Table S3), during a global cooling period with a marked species extinction peak (Lewis et al., 2008). Later, while global temperatures were decreasing during the Tortonian stage in late Miocene (11.6–7.3 Ma), three clades emerged mainly through three floristic regions: (1) the Irano-Turanian; (2) the Circumboreal; and (3) the Hengduan Mountains (Supplementary Fig. S3). According to our results, a significant diversification burst with a two-fold increase in speciation rate occurred during the Messinian (7.3–5.3 Ma), one of the most global cold-arid stages (Holbourn et al., 2018). Additionally, we detected a dispersal from northern to southern latitudes, e.g. from Circumboreal to Sino-Japanese, suggesting a southwards colonization in response to cold and dry climate extremes.

Overall, although the phylogeny of *Saussurea* shows a strong biogeographic clustering, none of the regions harbours an exclusive, endemic clade. This result may highlight the evolutionary importance of biogeographic exchanges among floristic regions in multiple directions, which seems to have been very common during the evolution of *Saussurea*. But, which are the major routes of floristic exchanges? Here we found a complex network of dispersal pathways between THH and the Irano-Turanian region, supporting the ~~"~~Central Asiatic Highland Corridor" (Li et al., 2014). According to our ancestral niche reconstructions (Supplementary Figs. S6 and S7), lineage dispersion among these two regions could have been triggered by the lack of a profound niche shift (niche conservatism) or large ecophysiological changes to evolve in the alpine colonized areas (Xing and Ree, 2017). Conversely, fewer interchanges of Sino-Japanese and Circumboreal regions with other regions have been detected. We hypothesize that the establishment of *Saussurea* in these regions could have involved a niche shift. Ancestral lineages would have adapted into a novel evolutionary eco-space, from alpine to more humid and forest lowland habitats (Supplementary Figs. S8 and S9). Thus, further dispersals of niche-differentiated lineages among different geographic regions could have been more difficult and unlikely.

Late Miocene conditions triggered *Saussurea* diversification

The allopatric mode of speciation resulting from geographic isolation may be considered as the main lineage splitting mechanism, as we did not detect large macroclimatic niche differences among closely related species (Supplementary Figs. S6 and S7). However, climate changes (e.g. those occurred in late Miocene, Pliocene or Pleistocene glacial-interglacial cycles) could have contributed to local ecological speciation, since ancestral lineages could have been forced to disperse along the altitudinal gradient looking for their optimal climatic niche. Thus, in *Saussurea* as in other cases reported in the THH region, ecological divergence could have accelerated the species differentiation process and facilitated the lineage accumulation through time (Dong et al., 2020).

The diversification burst detected here for *Saussurea* coincides in timing with the late Miocene species radiation previously estimated for alpine THH elements (18 representative clades; Ding et al., 2020). *Saussurea* diversification dynamics was diversity-dependent according to the best-fitting diversification model (Supplementary Table S4), which estimated a double speciation rate increase (λ_1= 0.63 spp/Myr, λ_2=1.14 spp/Myr) at 6 Ma (Fig. 2). Similarly, Ding et al. (2020) recovered

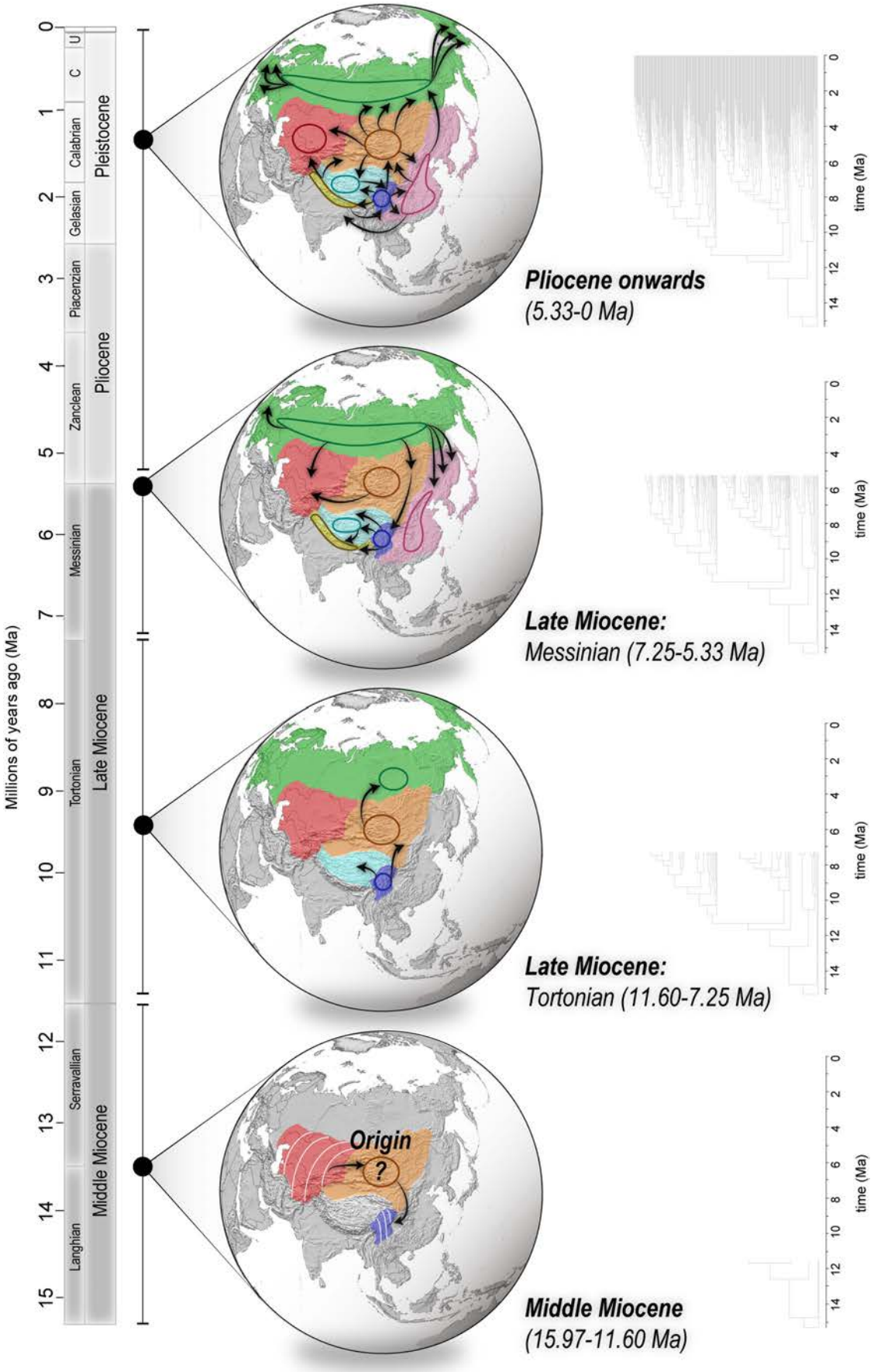


Figure 1. Biogeographic origin and major dispersal routes of *Saussurea* based on ancestral area estimates. Four geological time scales are displayed from old to recent order, representing main evolutionary movements: Middle Miocene epoch, Tortonian and Messinian ages from Late Miocene epoch, and Pliocene epoch to present. Lineage emergence and accumulation through time is also represented for each geological time scale. For detailed biogeographic results see Supplementary Figs. S2 and S3, and Supplementary Table S2.

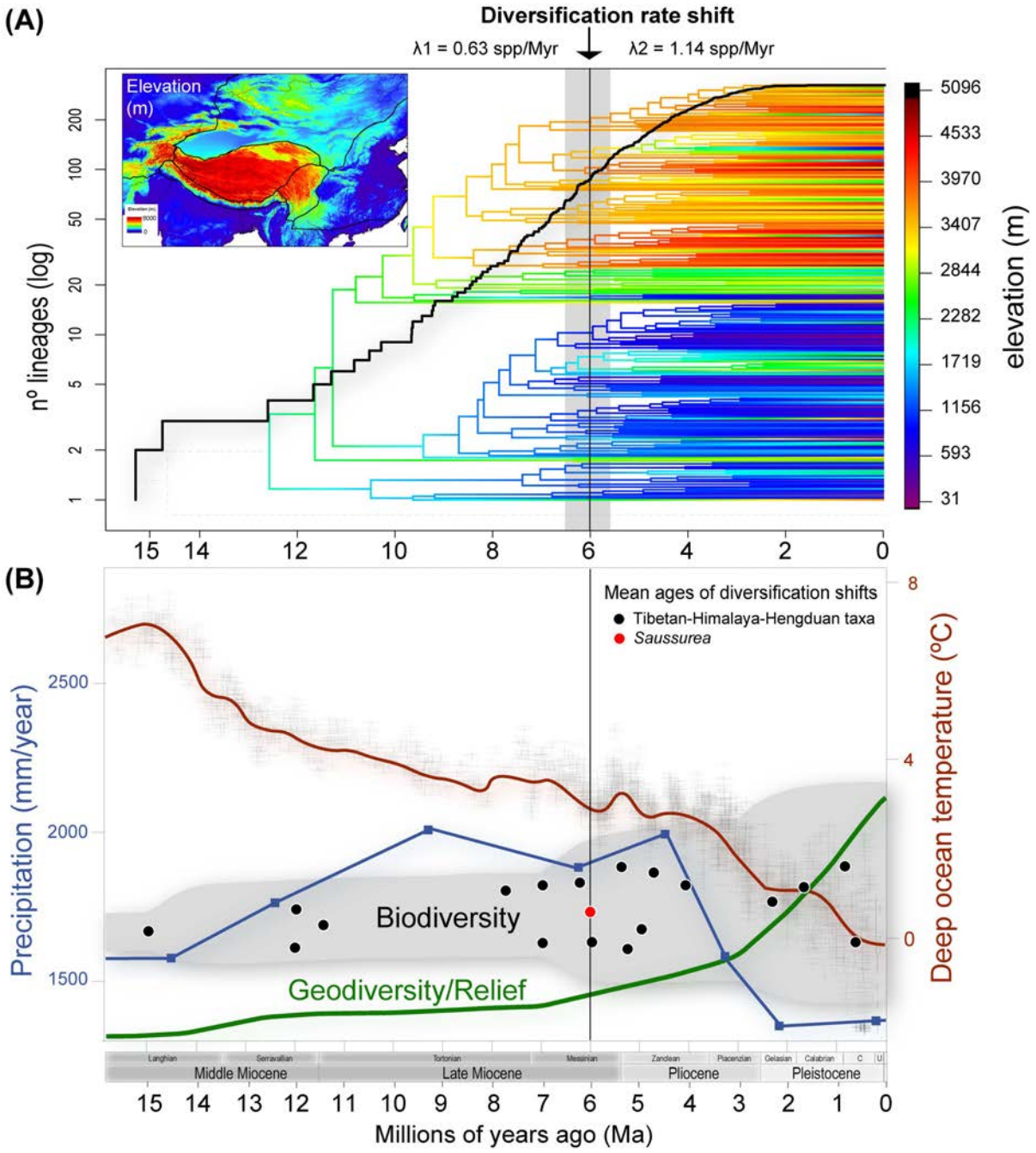


Figure 2. Diversification dynamics of *Saussurea* in relation with the alpine flora assembly in the Tibetan-Himalaya-Hengduan region (THH) and evolution of extrinsic factors postulate as main diversification drivers of alpine THH elements (orogeny and climate). (A) Estimation of ancestral states reconstructed with the *fasAnc* function of R package *phytools* for the elevation across the *Saussurea* time-calibrated phylogeny. The elevation gradient is displayed on a rainbow scale, showing low values in blue and high elevations in red. The species with explicit georeferenced data (ca. 70% of sampled species) are those represented in the phylogenetic tree. A lineage-through-time plot (LTT) is also outlined with a black line, which indicates the number of lineages emerged (log-transformed) in a temporal line. **(B)** Climate evolution displaying the global temperature, represented by deep-sea oxygen isotope records (grey slight lines; Zachos et al., 2008) and deep ocean temperature (red line; Hansen et al., 2013), and modelled mean annual precipitation for East Asia (Farnsworth et al., 2019), representing monsoon conditions (blue line) at idealized CO_2 concentrations (blue squares). In conjunction, estimates of geodiversity accumulation or increases of terrain ruggedness is showed (Mosbrugger et al., 2018). It is also represented the biodiversity accumulation (Mosbrugger et al., 2018) and estimates of diversification rate shifts of THH elements represented in black points (Muellner-Riehl et al., 2019). Age of diversification shift detected for *Saussurea* is marked as a red point. Note that estimates of geodiveristy and biodiversity are schematically represented and are not vertically scaled.

a notable increase in situ speciation rate of alpine THH flora at ~7 Ma, particularly in Hengduan Mountains. Other biotic turnovers have been already documented during late Miocene in the THH region, e.g. the expansion of C4 plants (Shen et al., 2018), extinction of mammalian faunas associated with changes of forest-type habitats (Li et al., 2020), and explosive diversification of multiple plant genera (see Muellner-Riehl et al., 2019).

Our results show that major diversification events and the species pump of *Saussurea* at 6 Ma are coincident with three geo-environmental factors that took place during Late Miocene in THH and surrounding areas (Fig. 2), as previously proposed (Meng et al., 2017; Ding et al., 2020). The first is a pulse in orogenic activity, uplifts and deformations, which resulted on an increase in the geodiversity or topographic ruggedness of mountain chains (Yang et al., 2019; Govin et al., 2020). In

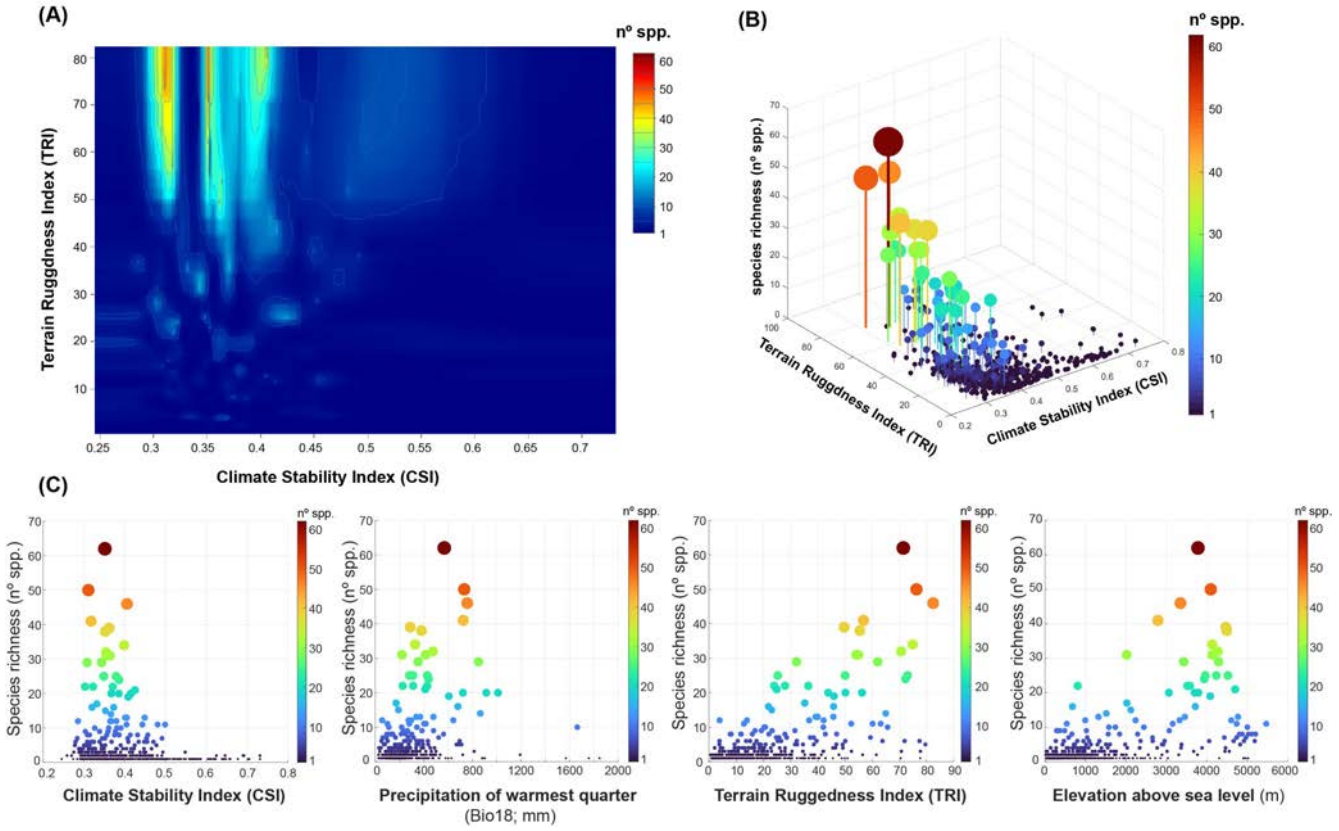


Figure 3. Graphs to test the combination of abiotic factors required to maximize the effects of mountain orogeny in *Saussurea* diversity within the Tibet-Himalaya-Hengduan region (THH) (~~"m~~ountain-geodiversity hypothesis"; Mosbrugger et al., 2018). Richness estimated and abiotic data used are derived from gird cells of 2 × 2 decimal degrees (see Methods for details). **(A)** Contour plot of the estimated *Saussurea* species richness across local climatic oscillations from Pliocene (3.3 Ma) to present (climatic stability index; Herrando-Moraira et al., in prep.) and terrain ruggedness index (TRI; Karger et al., 2017). Contour plot shows the predicted number of species based on interaction of climatic stability and terrain ruggedness. **(B)** 3D scatter plot of species richness (z-axis) and terrain ruggedness index (TRI; x-axis) and climate stability index (CSI; y-axis). In x-y plane are displayed lines that extend up to z-axis values. **(C)** Scatterplots combining species richness in y-axis with four abiotic factors related with the ~~"m~~ountain-geodiversity hypothesis". For (B) and (C) points are coloured according to the number of estimated species, representing blue colours a low number of species in cells of 2 × 2 decimal degrees, and yellow-red colours a high number of species in each square.

congruence, we found that around 6 Ma the ancestral lineages of *Saussurea* started to colonize the higher belts (< 3.500 m; Fig. 2). The second factor is the intensification of the East Asian monsoonal system, with rainfall increments greater than 25%, from ~1700– to 2100 mm/year (Farnsworth et al., 2019). And finally, the third factor is the sharp drop of temperatures at global scale during the ephemeral Northern Hemisphere glaciation, a late Miocene intense cooling event (~7–5.4 Ma; Chen et al., 2019). These three factors could have occurred synchronously and even synergically, deeply impacting diversification of *Saussurea* ancestors located in the most species richness center of the genus: the Hengduan Mountains. The southern range of Hengduan Mountains experienced a rapidly and significantly uplift around 8–6 Ma (Wang et al., 2008; 2018). Given this mountain uplift, the newly emerged elevation could have blocked the Indian monsoon (warm-wet moisture) arriving from the west (Li et al., 2020). Consequently, Hengduan Mountains turned somewhat cooler and drier which—in addition to produce the regional extinction of hominoids (Li et al., 2020)—could be advantageous conditions for the diversification of cold-tolerant alpine lineages like *Saussurea*. All of these factors could thus have promoted the formation of a great number of macro and microhabitats in a reduced geographical and temporal space (Ding et al., 2020). Overall, extrinsic factors are postulated as the most explanatory triggers of the THH plant radiations explained by three main hypotheses, the ~~"u~~plift-driven diversification", the ~~"m~~onsoon-driven diversification", and the ~~"w~~arm-cold colonization".

Evidences for the mountain-geodiversity hypothesis in THH and new key factors to be considered

Our biogeographic analysis revealed that the initial premise of the mountain-geodiversity hypothesis is here confirmed. *Saussurea* showed a time-dependent diversification shift (6 Ma) in a context of active orogeny and of a regime change to cooler-drier climates (Fig. 2). The impact of these events could have magnified the diversification burst in THH region by the combination of the three factors hypothesized by Mosbrugger et al. (2018): (1) the steep eco-climatic gradients along full altitudinal zonation in narrow distances (Supplementary Figs. S10E and S10F); (2) the low-moderate climate oscillations that

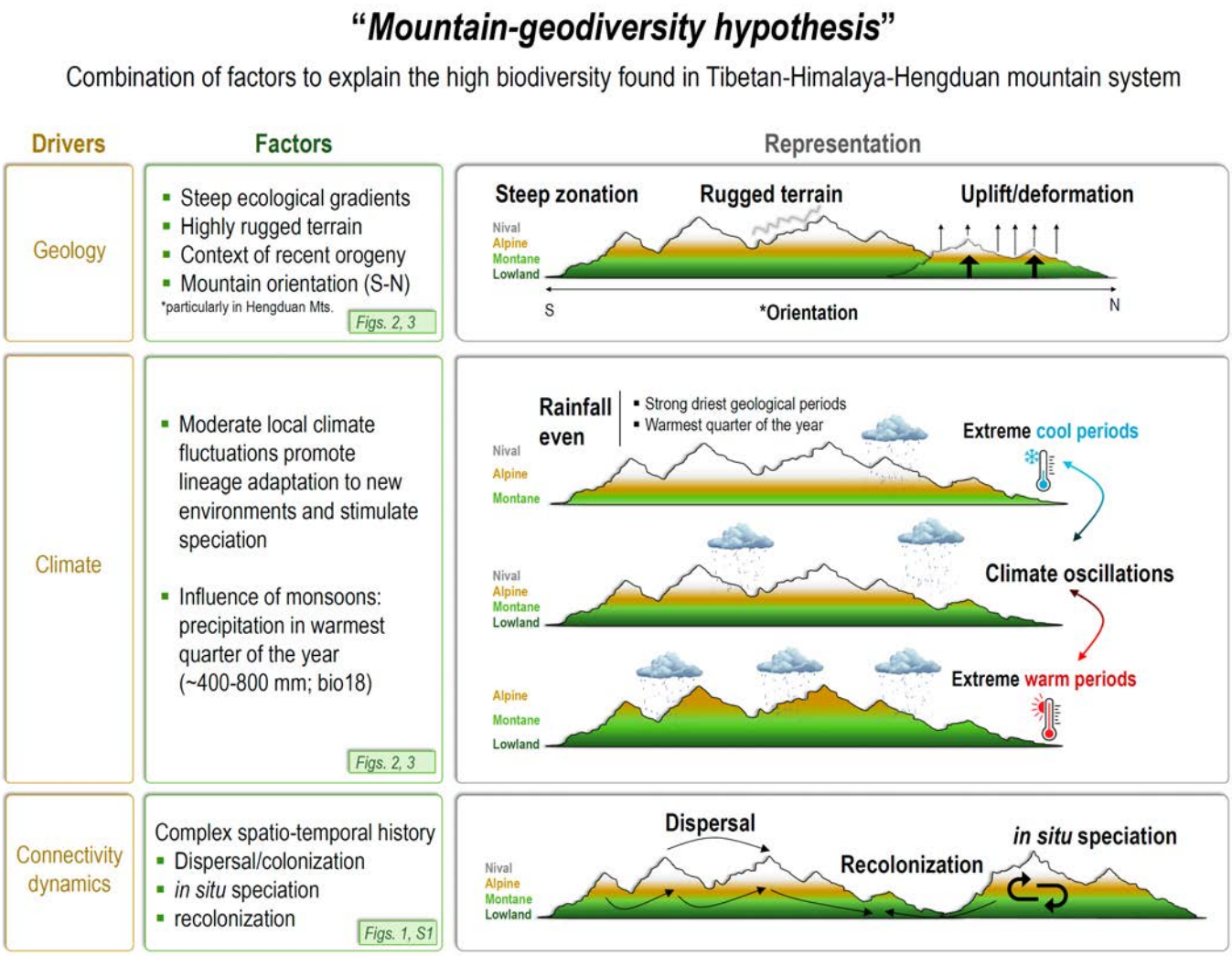


Figure 4. Schematic view of the "mountain-geodiversity hypothesis" (Mosbrugger et al. 2018; Muellner-Riehl, 2019). Illustration of the key factors postulated as biodiversity maximizers in the Tibet-Himalaya-Hengduan region (THH). Highlighted as "new" is showed a novel factor to be considered as the results reported here suggested (see further explanation in section "Results and Discussion"). Evidence found to support each proposed factor are referred to the related figures.

occurred in the region (Supplementary Fig. S10B), as showed by the climatic stability index values (CSI; Herrando-Moraira et al., in prep.); and (3) the highly rugged terrain, being the south-eastern margins of THH the greatest world's region with high rugged terrain values (Terrain Ruggedness Index; Supplementary Figs. S10D).

Our results shed light on one of the current biogeographic and evolutionary "enigmas" of THH: why Hengduan Mountains are a particular diversity hotspot within the region? This is the only area where species richness of *Saussurea* exceeds 36 species/2° × 2° grid cells (Supplementary Figs. S10A). We found that the combination of high terrain ruggedness (>50) with low-moderate climate variation (0.3–0.4) are the variable that correlates the most with maximum local diversity (Figs. 3A and 3B). Thus, it is confirmed here the importance of complex orography and moderate local variations of climate to understand the diversification of the temperate alpine flora in Hengduan mountains, two factors that are very often associated to biodiversity-rich regions (e.g. López-Pujol et al., 2011). Our biogeographic analysis highlighted noteworthy historical biotic exchanges between Hengduan Mountains and adjacent regions (Fig. 1 and Supplementary Fig. S2). The results obtained are in line with the last concept connected to the mountain-geodiversity hypothesis, which is the "flickering connectivity system" (Flantua and Hooghiemstra (2018); Muellner-Riehl (2019). Analogously to the Andes flora, the high species richness of Hengduan flora can be related with the complex spatio-temporal history of the region, in which multiple processes of in situ speciation, dispersal, and back-colonization events have taken place (Xing and Ree, 2017; Ding et al., 2020). Moreover, the north-south direction of mountain ranges (and thus, river valleys and gorges) within the Hengduan Mountains, in contrast with the west-east oriented Himalayas-QTP, is another key factor to take into account since, as previously reported, it has many direct biodiversity implications (Muellner-Riehl, 2019). On one hand, it hinders west-to-east dispersals among wavy mountain creases, constituting large geological barriers for biotic exchanges (Li and Sun, 2017; Sun et al., 2017), and on the other hand, during large climate changes, such as Quaternary glacial cycles, it allows species to persist through altitudinal migrations and it may foster allopatric speciation, in addition to ecological divergence to local conditions, linked to longitudinal dispersals across sky island areas.

Interestingly, our integration of abiotic with biotic data

revealed that at least one more variable should be added to the mountain-geodiversity hypothesis. We detected that distribution and diversity patterns of *Saussurea* are also related to the precipitation of the warmest quarter (bio18; Fig. 3C). Species richness increases at a range of 400–800 mm of rainfall, probably caused by the influence of monsoon system. This precipitation variable seems crucial to explain why in western Himalayan range, e.g. Pamir Mountains, species richness is not as high as in Hengduan Mountains, despite it fulfills the five above-mentioned factors (see maps in Supplementary Fig. S10). As an example, the endemic species of *Saussurea* in Pamir are located in areas with only 85 mm of precipitation during the warmest quarter of the year. Hence, the vegetation of Pamir Mountains is more constrained by water availability (lower moisture conditions and shorter length of growing period) especially during the harshest season of the year in terms of high temperatures. In consequence, its flora is not as rich as the Hengduan flora and plants are highly adapted to harsher conditions such as cold and arid/semi-arid environments.

We schematize the new updated framework of the mountain-geodiversity hypothesis in Fig. 4. In sum, the combination of all these factors proposed by historical hypotheses of the region (Table 1), could have been the responsible of a net diversification rate increase within THH, which is particularly accentuated in Hengduan Mountains. Diversification processes can be driven primarily by an increase on speciation rate alone, by a decreasing on extinction rate, or confluence of both processes when ―ultra-radiations" are described to occur (Donoghue and Sanderson, 2015). The last scenario seems plausible in the case of THH. The high local geodiversity could have allowed species to occupy areas of niche optimal preferences after short-distance dispersals during climate fluctuations, when range expansions-contractions through altitudinal or latitudinal gradient took place. In addition, microhabitats created could have acted as buffer zones or refugia during harsh climate changes, leading the species to avoid the extinction. We detected that *Saussurea* reached its diversity peak at mid-mountain elevations (around 3000–4500 m; Fig. 3C) and, as pointed out by Mullner-Riehl (2019), this position could have allowed the species dispersion either moving downslope or upslope mountains during cold or warm periods, respectively.

3. Materials and mehods

Please refer to the Supplementary data online.

4. Supplementary data

Supplementary Tables (S1–S7) and **Figures** (S1–S11).
Supplementary alignments Supermatrix and loci partitions.
Supplementary trees from concatenated approach and unfiltered dataset, from coalescence approach and unfiltered dataset, dated tree in Phylip format.

5. Ackowledgments

Authors thank Maria Luisa Gutiérrez, Fernando Castro, and Laia Barres for providing technical support during the laboratory process. We also thank the herbaria that provided material for the study: BC, DUSH, E, FRU, GDA, H, LE, MBK, MEM, MHA, MO, MU, OS, SANT, TAD, TARI, TI, TK, US, W, and WU. We gratefully acknowledge the University of Memphis computing facilities to allow us the utilization of its high-performance cluster (HPC).

6. Funding

Financial support from the Ministerio de Ciencia e Innovación (Project CGL2015- 66703-P MINECO/FEDER, UE and Ph.D. grant to S.H.M) and the Catalan government (―Ajuts a grups consolidats" 2017-SGR1116) is also greatly acknowledged. This study has been performed under the Ph.D. program ―Plant Biology and Biotechnology" of the Autonomous University of Barcelona (UAB).

7. References

Chen C, Bai Y and Fang X *et al.* A Late Miocene Terrestrial Temperature History for the Northeastern Tibetan Plateau's Period of Tectonic Expansion. *Geophys Res Lett* 2019; **46**: 8375–8386.

Chen YS and Xu LS. *Saussurea xinjiangensis* sp. nov. (Asteraceae, Cardueae), a new species from Xinjiang, China. *Nord J Bot* 2020; **38**: 9.

Chen YS. *Flora of Pan-Himalaya, vol. 48, Asteraceae II: Saussurea.* Cambridge: Cambridge University Press, 2015.

Dentant C. The highest vascular plants on Earth. *Alp Bot* 2018; **128**: 97–106.

Ding WN, Ree RH and Spicer RA *et al*. Ancient orogenic and monsoon-driven assembly of the world's richest temperate alpine flora. *Science* 2020; **369**: 578–581.

Dong F, Hung CM and Yang XJ. Secondary contact after allopatric divergence explains avian speciation and high species diversity in the Himalayan-Hengduan Mountains. *Mol Phylogenet Evol* 2020; **143**: 106671.

Donoghue MJ and Sanderson MJ. Confluence, synnovation, and depauperons in plant diversification. *New Phytol* 2015; **207**: 260–274.

Dullinger S, Gattringer A and Thuiller W *et al*. Extinction debt of high-mountain plants under twenty-first-century climate change. *Nat Clim Change* 2012; **2**: 619–622.

Fang X, Dupont-Nivet G and Wang C *et al*. Revised chronology of central Tibet uplift (Lunpola Basin). *Sci Adv* 2020; **6**: eaba7298.

Farnsworth A, Lunt DJ and Robinson SA *et al*. Past East Asian monsoon evolution controlled by paleogeography, not CO_2. *Sci Adv* 2019; **5**: eaax1697.

Favre A, Päckert M and Pauls SU *et al*. The role of the uplift of the Qinghai-Tibetan Plateau for the evolution of Tibetan biotas. *Biol Rev* 2015; **90**: 236–253.

Flantua S and Hooghiemstra H. Historical connectivity and mountain biodiversity. In: Hoorn C, Perrigo A and Antonelli A (eds.). *Mountains, Climate and Biodiversity*. Hoboken, NJ: Wiley-Blackwell, 2018, 17–36.

Gonçalves DJ, Simpson BB and Ortiz EM et al. Incongruence between gene trees and species trees and phylogenetic signal variation in plastid genes. *Mol Phylogenet Evol* 2019; **138**: 219–232.

Govin G, van der Beek P and Najman Y *et al*. Early onset and

late acceleration of rapid exhumation in the Namche Barwa syntaxis, eastern Himalaya. *Geology* 2020; **48**: 1139–1143.

Guo Z, Wilson M. Late Oligocene–early Miocene transformation of postcollisional magmatism in Tibet. *Geology* 2019; **47**: 776–780.

Hansen J, Sato M and Russell G *et al*. Climate sensitivity, sea level and atmospheric carbon dioxide. *Philos Trans A Math Phys Eng Sci* 2013; **371**: 20120294.

He Z, Li X and Yang M *et al*. Speciation with gene flow via cycles of isolation and migration: insights from multiple mangrove taxa. *Nat Sci Rev* 2018; **6**: 275–288.

Herrando-Moraira S, Nualart N and Galbany-Casals M *et al*. Climate Stability Index maps, a global high resolution cartography of climate stability from Pliocene to 2100. Submitted to *Sci Data* June 2021.

Holbourn AE, Kuhnt W and Clemens SC *et al*. Late Miocene climate cooling and intensification of southeast Asian winter monsoon. *Nat Commun* 2018; **9**: 1–13.

Hughes CE and Atchison GW. The ubiquity of alpine plant radiations: from the Andes to the Hengduan Mountains. *New Phytol* 2015; **207**: 275–282.

Karger DN, Conrad O and Böhner J *et al*. *Climatologies at high resolution for the earth's land surface areas*. Dryad Digital Repository: 2017.

Lees AC and Pimm SL. Species, extinct before we know them? *Curr Biol* 2015; **25**: R177–R180.

Lewis AR, Marchant DR and Ashworth AC *et al*. Mid-Miocene cooling and the extinction of tundra in continental Antarctica. *Proc Natl Acad Sci USA* 2008; **105**: 10676–10680.

Li GD, Kim C and Zha HG *et al*. Molecular phylogeny and biogeography of the arctic-alpine genus *Lagotis* (Plantaginaceae). *Taxon* 2014; **63**: 103–115.

Li S, Ji X and Harrison T *et al*. Uplift of the Hengduan Mountains on the southeastern margin of the Tibetan Plateau in the late Miocene and its paleoenvironmental impact on hominoid diversity. *Palaeogeogr Palaeoclimatol Palaeoecol* 2020; **553**: 109794.

Li X and Sun H. Phylogenetic pattern of alpine plants along latitude and longitude in Hengduan Mountains Region. *Plant Div* 2017; **39**: 37–43.

López-Pujol J, Zhang FM and Sun H *et al*. Centres of plant endemism in China: places for survival or for speciation? *J Biogeogr* 2011; **38**: 1267–1280.

Marchese C. Biodiversity hotspots: A shortcut for a more complicated concept. *Glob Ecol Conserv* 2015; **3**: 297–309.

Meng HH, Su T and Gao XY *et al*. Warm–cold colonization: response of oaks to uplift of the Himalaya–Hengduan Mountains. *Mol Ecol* 2017; **26**: 3276–3294.

Mosbrugger V, Favre A and Muellner-Riehl AN *et al*. Cenozoic evolution of geo-biodiversity in the Tibeto-Himalayan region. In: Hoorn C, Perrigo A and Antonelli A (eds.). *Mountains, Climate and Biodiversity*. Hoboken, NJ: Wiley-Blackwell, 2018, 429–448.

Muellner-Riehl AN, Schnitzler J and Kissling WD *et al*. Origins of global mountain plant biodiversity: Testing the 'mountain-geobiodiversity hypothesis'. *J Biogeogr* 2019; **46**: 2826–2838.

Muellner-Riehl AN. Mountains as evolutionary arenas: patterns, emerging approaches, paradigm shifts, and their implications for plant phylogeographic research in the Tibeto-Himalayan region. *Front Plant Sci* 2019; **10**: 195.

Pérez-Escobar OA, Chomicki G and Condamine FL *et al*. Recent origin and rapid speciation of Neotropical orchids in the world's richest plant biodiversity hotspot. *New Phytol* 2017; **215**: 891–905.

Raab-Straube E von. *Taxonomic revision of Saussurea subgenus Amphilaena (Compositae, Cardueae)*. Englera 34. Berlin: Botanic Garden and Botanical Museum Berlin, Freie Universität Berlin, 2017.

Rahbek C, Borregaard MK and Colwell RK *et al*. Humboldt's enigma: What causes global patterns of mountain biodiversity? *Science* 2019; **365**: 1108–1113.

Shen T, Wang G and Replumaz A *et al*. Miocene subsidence and surface uplift of southernmost Tibet induced by Indian subduction dynamics. *Geochemistry, Geophys. Geosystems* 2020; **21**: e2020GC009078.

Shen X, Wan S and Colin C *et al*. Increased seasonality and aridity drove the C_4 plant expansion in central Asia since the Miocene-Pliocene boundary. *Earth Planet Sci Lett* 2018; **502**: 74–83.

Simões M, Breitkreuz L and Alvarado M *et al*. The evolving theory of evolutionary radiations. *Trends Ecol Evol* 2016; **31**: 27–34.

Singh D and Hajra DK. Floristic diversity. In: Gujral GS and Sharma V (eds.). *Changing Perspective of Biodiversity Status in the Himalaya*. New Delhi, India: The British Council, 1996, 23–38.

Stokstad E. Mountains and monsoons created Tibetan biodiversity. *Science* 2020; **369**: 493.

Sun H, Zhang J and Deng T *et al*. Origins and evolution of plant diversity in the Hengduan Mountains, China. *Plant Div* 2017; **39**: 161–166.

Wang G, Wan J and Wang E *et al*. Late Cenozoic to recent transtensional deformation across the Southern part of the Gaoligong shear zone between the Indian plate and SE margin of the Tibetan plateau and its tectonic origin. *Tectonophysics* 2008; **460**: 1–20.

Wang X, Wang Y and Li Q *et al*. Cenozoic vertebrate evolution and paleoenvironment in Tibetan Plateau: Progress and prospects. *Gondwana Res* 2015; **27**: 1335–1354.

Wang Y, Zhang P and Schoenbohm LM *et al*. Two-phase exhumation along major shear zones in the SE Tibetan Plateau in the late Cenozoic. *Tectonics* 2018; **37**: 2675–2694.

Wang YJ, Susanna A and Raab-Straube E von *et al*. Island-like radiation of *Saussurea* (Asteraceae: Cardueae) triggered by uplifts of the Qinghai–Tibetan Plateau. *Biol J Linn Soc* 2009; **97**: 893–903.

Wen J, Zhang J and Nie ZL *et al*. Evolutionary diversifications of plants on the Qinghai-Tibetan Plateau. *Front Genet* 2014; **5**: 4.

Xing Y and Ree RH. Uplift-driven diversification in the Hengduan Mountains, a temperate biodiversity hotspot. *Proc Natl Acad Sci USA* 2017; **114**: E3444–E3451.

Xu LS, Herrando-Moraira S and Susanna A *et al*. Phylogeny, origin and dispersal of *Saussurea* (Asteraceae) based on chloroplast genome data. *Mol Phylogenet Evol* 2019; **141**: 106613.

Yang R, Yang Y and Fang X *et al*. Late miocene intensified tectonic uplift and climatic aridification on the Northeastern Tibetan Plateau: evidence from clay mineralogical and geochemical records in the Xining Basin. *Geochemistry,*

Geophys. Geosystems 2019; **20**: 829–851.
Yang W, Ma K and Kreft H. Environmental and socio-economic factors shaping the geography of floristic collections in China. *Glob Ecol Biogeogr* 2014; **23**: 1284–1292.
Zachos JC, Dickens GR and Zeebe RE. An early Cenozoic perspective on greenhouse warming and carbon-cycle dynamics. *Nature* 2008; **451**: 279–283.
Zhang DC, Boufford DE and Ree RH *et al*. The 29° N latitudinal line: an important division in the Hengduan Mountains, a biodiversity hotspot in southwest China. *Nord J Bot* 2009a; **27**: 405–412.
Zhang DC, Zhang YH and Boufford DE *et al*. Elevational patterns of species richness and endemism for some important taxa in the Hengduan Mountains, southwestern China. *Biodivers Conserv* 2009b; **18**: 699–716.
Zhang Y, Chen J and Sun H. Alpine speciation and morphological innovations: revelations from a species-rich genus in the northern hemisphere. *AoB Plants* 2021; **13**: plab018.
Zhao L, Li J and Liu H *et al*. Distribution, congruence and hotspots of higher plants in China. *Sci Rep* 2016; **6**: 19080.

10. Supplementary material

MATERIALS AND METHODS

Taxon sampling

We sampled a total of 324 species of *Saussurea*, which represents 70% of the genus diversity. The sampling strategy was designed to cover all distribution areas of the genus to extract solid conclusions about its biogeographical history. We also added 60 outgroup species (at subtribe, tribe and family level) to estimate time divergence of *Saussurea* clades. In this study, 254 species of a total of 384 species have been sequenced for the first time. See Appendix 1 for a complete list of sampled species with additional information.

Laboratory protocol

We implemented the Hyb-Seq technique based on target loci designed in the Compositae COS 1061 loci kit (Mandel et al., 2014) to obtain molecular data for phylogenetic inference. Plant material was extracted from dried leaves collected in field expeditions or from herbarium collections. For the DNA extraction, fragmentation, library preparation, and sequence capture, we followed the same laboratory workflow detailed in Herrando-Moraira et al. (2018, 2019, and 2020). The only difference here was the number of libraries grouped into pools, which were set up to eight libraries considering a total DNA per pool of 500 ng. Enriched libraries were sequenced in the University of Florida with an Illumina HiSeq 3000 (Illumina, USA) and in the Macrogen Co. (Seoul, South Korea) with an Illumina HiSeq 4000 (Illumina, USA).

Target sequence extraction

We checked the quality of the 100-bp paired-end reads obtained with FastQC v.0.10.1 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Adapters removal and read cleaning was conducted with Trimmomatic v.0.36 (Bolger et al., 2014) with a sliding-window set to 5:20. The target loci were extracted with the workflow implemented in HybPiper v.1.3.1 software (Johnson et al., 2016). The loci flagged as potentially paralogous by HybPiper were removed in subsequent analyses. We aligned selected target loci with MAFFT v.7.266 (Katoh and Standley, 2013) and cleaned alignments with trimAl v.14 (Capella-Gutiérrez et al., 2009). We constructed two datasets to apply two approaches of phylogenetic inference: (1) the concatenated supermatrix dataset, obtained with FASconCAT-G v.1.02 (Kück and Longo, 2014), to conduct concatenation-based analysis; and (2) the loci-separated alignments for coalescence-based analysis.

Phylogenetic reconstruction

The concatenation-based analysis was performed under a maximum likelihood (ML) inference as implemented RAxML-NG (Kozlov et al., 2019) from the CIPRES Science Gateway v.3.1 (Miller et al., 2010). We run a first search of the best-scoring ML tree, with ten randomized and parsimony starting trees. Then, we run five independent bootstrap (BS) analysis resampling 100 replicates each (500 bootstrap trees in total). With the –f b option, we drew bipartition information on the best ML tree based on 500 trees from bootstrap analysis. Each locus was set as a different partition unit under the GTR+G nucleotide model. For the BS analysis, we applied the auto-stopping option –autoMRE", in which the search stops when convergence is achieved. The coalescence-based analysis was conducted in ASTRAL-III v.5.5.3 (Zhang et al., 2018) with the individual gene trees previously obtained with RAxML specifying a bootstrap resampling of 200 replicates. The local posterior probability (LPP) was used as a branch support metric. Tree branches were considered as statistically supported with BS values > 70% (Hillis and Bull, 1993) and LPP > 0.95 (Sayyari and Mirarab, 2016). Resulting trees were exported with FigTree v.1.4.3 (http://tree.bio.ed.ac.uk/software/figtree/).

Time divergence analysis

As a molecular dating method, we used the penalized-likelihood approach implemented in treePL (Smith and O'Meara 2012), which is particularly suitable for dating large high throughput sequencing datasets. The best-scoring ML tree obtained with concatenation-based analysis was used as input tree. For calibration points and analysis procedure, we followed the same settings specified in Herrando-Moraira et al. (2019). The confidence intervals (95% CI) of node ages were obtained running 100 independent treePL analyses, each one with calibration points constrained to a set of random values generated under normal and lognormal distributions (see details in Herrando-Moraira et al. 2019). Finally, TreeAnnotator v.1.7.5 (Drummond et al. 2012) was used to obtain the maximum clade credibility (MCC) tree chronogram from the 100 time-calibrated trees. The MCC tree is provided in Supplementary Material.

Ancestral area reconstructions

A total of seven biogeographical ranges were defined to reconstruct the most probable ancestral areas of *Saussurea* (see Supplementary Fig. S2). For the area delimitation, we followed the definition of floristic regions by Takhtajan's (1986) with slight modifications specified below. The Irano-Turanian (IT) region was delimited according to the new limits proposed in the revision of Manafzadeh et al. (2017), which divided the IT into western (W-IT) and eastern (E-IT), but we separated the Tibetan Plateau from the rest of E-IT, and treated it as QTP *sensu stricto*. To account for the geographic distribution of the genus diversity, the Eastern Asiatic floristic region of Takhtajan (1986) was divided into Himalayas, Hengduan Mountains, and Sino-Japanese region. The last area considered was the Circumboreal floristic region that includes temperate areas of Northern Hemisphere.

The ancestral area reconstruction analysis was performed in RASP 4.02 (Yu et al., 2015), which implements the R package *BioGeoBEARS* (Matzke, 2013). We pruned the outgroup from the time-calibrated MCC tree, and this trimmed tree was then used as input chronogram. We compared the fit of three biogeographical models: (1) Dispersal–Extinction–Cladogenesis (DEC; Ree et al., 2005); (2) the likelihood versions of Dispersal–Vicariance (DIVA-like); and (3) BayArea (BAYAREALIKE). Following recent suggestions of Ree and Sanmartín (2018), we did not consider those models with the additional parameter jump dispersal or founder event speciation –J" (Matzke, 2013). We selected the best-fitting model based on AIC (DEC model) and applied it to estimate the most probable biogeographic history of *Saussurea*.

Diversification analyses

A lineages-through-time curve (LTT; Nee et al., 1992) was performed as a first exploratory analysis to visualize the tendency of lineages accumulation through time. We used the R package *ape* and function *mltt.plot* (Paradis and Schliep, 2019) setting in log scale the –y" axis (number of lineages). As input chronograms, we used the MCC tree and the 100 independent trees used to estimate for the node age uncertainty (see details in Herrando-Moraira et al., 2019).

The diversification dynamics of *Saussurea* was tested through a hypothesis-driven approach to avoid identifiability problems (Louca and Pennell, 2020; Morlon, 2020). We investigated whether species diversification rates varied linearly or exponentially with time, climate, or species-diversity. Null models (constant-rate ones) were also included. In models that assume variable extinction rates, we only tested models that allow variations in speciation rate to avoid potential errors reported by Burin et al. (2019). We accounted for incomplete taxon sampling by setting the sampling fraction for our *Saussurea* dataset to each model. Final model selection of best fitting-model was based on AICc values. Diversification analyses were run with R packages *RPANDA* (Morlon et al., 2016) and *DDD* (Etienne et al., 2012).

Ecological niche analyses

We gathered 10,877 georeferenced records for 226 species (ca. 70% of the total number of species sampled) to explore ecological niche evolution through time along the phylogeny of *Saussurea*. In Appendix 2 we list the geographic coordinates, the source, and precision of each occurrence record. As environmental variables, we used an initial set of 35 variables. Description of each variable, or set of variables, resolution, and source of extraction are specified in Supplementary Table S5. For each *Saussurea* occurrence, we extracted the values of 35 variables with ArcGIS v.10.2.2 (Esri, Redlands, California, USA 2014). Then, for species with more than one occurrence, mean values were calculated to get a database in which the average niche values of variables are provided for each species.

To avoid the inclusion of highly correlated variables in further analyses, we selected 19 uncorrelated ones (Supplementary Fig. S11) based on the results of a Pearson correlation analysis, retaining those above the threshold $r > |0.8|$ or only the best informative one from sets of highly correlated variables ($r > 0.8$). The variable relevance was assessed through a principal component analysis (PCA) and the weight of each to new principal component axes to explain the ecological species variance (Supplementary Table S6). The final variable selection and its weight to each PC axis is summarized in Supplementary Table S7. The analysis was conducted in R v.3.6.3 (R Development Core Team, 2019) using function *cor* for the Pearson correlation analysis, and *FactoMineR*, *factoextra* (for PCAs) and *ggplot2* for the PCA.

Finally, we used the R functions *fastAnc* and *contMap* (from *phytools* package; Revell 2012) to track the most probable ancestral niche states for the lineages of *Saussurea*. As input variables, we used the PC1 and PC2 scores obtained for each species from the PCA analysis and also the mean altitudinal value, bio6, aridity, and forest coverage.

Environmental data to test the mountain-geodiversity hypothesis

To explore the influence of the variables proposed to explain the high alpine flora diversity in the THH by Mosbrugger et al. (2018), we collected four map sets: (1) the climatic stability index (CSI; Herrando-Moraira in prep.), an estimate of climate oscillations for 12 time periods, from Pliocene (3.3 Ma) to the present time (1979–2013), covering the major part of variability during cold-warm cycles of middle Pliocene onwards; (2) the precipitation of warmest quarter (bio18) from Worldclim 2 (Fick and Hijmans, 2017). This variable was selected in order to test the possible effect of summer monsoons (i.e. values of East Asia vs. Central Asia); (3) the terrain ruggedness index (TRI) from Karger et al. (2017); and (4) the elevation above sea level (Karger et al., 2017).

To test whether *Saussurea* diversity is concentrated on a particular combination of the four variables cited above, we first extracted a species richness map based on the occurrence database. The tool *calculate species richness* from SDMToolbox (Brown, 2014) was used to obtain the species richness map. In line with recommendations of the *Saussurea* study conducted by Zhang et al. (2021), we selected a cell resolution of 2 × 2 decimal degrees, given that explicit geo-referenced records are missing in some areas where *Saussurea* taxa are distributed. After building the richness map, we extracted values for each cell of estimated species richness and the four environmental variables. We represented the combination of TRI, CSI, and species richness with a contour plot type with the R function *plot_ly* from *plotly* package (Sievert, 2020). In addition, five scatter plots were also visualized and exported with matlab, setting different combination of the species richness and four variables.

Bibliography (Materials and Methods)

Bolger AM, Lohse M and Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 2014; **30**: 2114–2120.

Brown JL. SDM toolbox: a python-based GIS toolkit for landscape genetic, biogeographic and species distribution model analyses. *Methods Ecol Evol* 2014; **5**: 694–700.

Burin G, Alencar LR and Chang J *et al*. How well can we estimate diversity dynamics for clades in diversity decline? *Syst Biol* 2019; 68: 47–62.

Capella-Gutiérrez S, Silla-Martínez JM and Gabaldón T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 2009; **25**: 1972–1973.

Drummond AJ, Suchard MA and Xie D *et al*. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol Biol Evol* 2012; **29**: 1969–1973.

Etienne RS, Haegeman B and Stadler T *et al*. Diversity-dependence brings molecular phylogenies closer to agreement with the fossil record. *Proc R Soc B Biol Sci* 2012; **279**: 1300–1309.

Favre A, Michalak I and Chen CH *et al*. Out-of-Tibet: The spatio-temporal evolution of *Gentiana* (Gentianaceae). *J Biogeogr* 2016: **43**; 1967–1978.

Fick SE and Hijmans RJ. WorldClim 2: new 1km spatial resolution climate surfaces for global land areas. *Int J Climatol* 2017; **37**: 4302–4315.

Herrando-Moraira S and The Cardueae Radiations Group. Exploring data processing strategies in NGS target enrichment to disentangle radiations in the tribe Cardueae (Compositae). *Mol Phylogenet Evol* 2018; **128**: 69–87.

Herrando-Moraira S and The Cardueae Radiations Group. Generic boundaries in subtribe Saussureinae (Compositae: Cardueae): Insights from Hyb-Seq data. *Taxon* 2020; **69**: 694–714.

Herrando-Moraira S and The Cardueae Radiations Group. Nuclear and plastid DNA phylogeny of tribe Cardueae (Compositae) with Hyb-Seq data: A new subtribal classification and a temporal diversification framework. *Mol Phylogenet Evol* 2019; **137**: 313–332.

Herrando-Moraira S, Nualart N and Galbany-Casals M *et al*. Climate Stability Index maps, a global high resolution cartography of climate stability from Pliocene to 2100. Submitted to *Sci Data* June 2021.

Hillis DM and Bull JJ. An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. *Syst Biol* 1993; **42**: 182–192.

Johnson MG, Gardner EM and Liu Y *et al*. HybPiper: Extracting coding sequence and introns for phylogenetics from high-throughput sequencing reads using target enrichment. *App Plant Sci* 2016; **4**: 1600016.

Karger DN, Conrad O and Böhner J *et al*. *Climatologies at high resolution for the earth's land surface areas*. Dryad Digital Repository: 2017.

Katoh K and Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 2013; **30**: 772–780.

Kozlov AM, Darriba D and Flouri T *et al*. RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics* 2019; **35**:4453–4455.

Kück P and Longo GC. FASconCAT-G: extensive functions for multiple sequence alignment preparations concerning phylogenetic studies. *Front Zool* 2014; **11**: 81.

Louca S and Pennell MW. Extant timetrees are consistent with a myriad of diversification histories. *Nature* 2020; **580**: 502–505.

Manafzadeh S, Staedler YM and Conti E. Visions of the past and dreams of the future in the Orient: the Irano-Turanian region from classical botany to evolutionary studies. *Biol Rev* 2017; **92**: 1365–1388.

Mandel JR, Dikow RB and Funk VA. A target enrichment method for gathering phylogenetic information from hundreds of loci: an example from the Compositae. *App Plant Sci* 2014; **2**: 1300085.

Matzke NJ. BioGeoBEARS: BioGeography with Bayesian (and likelihood) evolutionary analysis in R Scripts: CRAN: The

Comprehensive R Archive Network: 2013. Available from: https://cran.r-project.org/

Miller MA, Pfeiffer W and Schwartz T. *Creating the CIPRES Science Gateway for inference of large phylogenetic trees. Proceedings of the Gateway Computing Environments Workshop (GCE)*. New Orleans: IEEE, 2010, 1–8.

Morlon H, Lewitus E and Condamine FL *et al*. RPANDA: an R package for macroevolutionary analyses on phylogenetic trees. *Methods Ecol Evol* 2016; **7**: 589–597.

Morlon H. Prior hypotheses or regularization allow inference of diversification histories from extant timetrees. BioRxiv 2020: p. 2020.07.03.185074.

Mosbrugger V, Favre A and Muellner-Riehl AN *et al*. Cenozoic evolution of geo-biodiversity in the Tibeto-Himalayan region. In: Hoorn C, Perrigo A and Antonelli A (eds.). *Mountains, Climate and Biodiversity*. Hoboken, NJ: Wiley-Blackwell, 2018, 429–448.

Nee S, Mooers AO and Harvey PH. Tempo and mode of evolution revealed from molecular phylogenies. *Proc Natl Acad Sci USA* 1992; **89**: 8322–8326.

Paradis E and Schliep K. ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* 2019; **35**: 526–528.

Ree RH and Sanmartín I. Conceptual and statistical problems with the DEC+ J model of founder-event speciation and its comparison with DEC via model selection. *J Biogeogr* 2018; **45**: 741–749.

Ree RH, Moore BR and Webb CO *et al*. A likelihood framework for inferring the evolution of geographic range on phylogenetic trees. *Evolution* 2005: **59**: 2299–2311.

Revell LJ. phytools: an R package for phylogenetic comparative biology (and other things). *Methods Ecol Evol* 2012; **3**: 217–223.

Sayyari E and Mirarab S. Fast coalescent-based computation of local branch support from quartet frequencies. *Mol Biol Evol* 2016; **33**: 1654–1668.

Sievert C. *Interactive web-based data visualization with R, plotly, and shiny*. CRC Press, 2020.

Smith SA and O'Meara BC. treePL: divergence time estimation using penalized likelihood for large phylogenies. *Bioinformatics* 2012: **28**: 2689–2690.

Stokstad E. Mountains and monsoons created Tibetan biodiversity. *Science* 2020; **369**: 493.

Takhtajan A. *Floristic Regions of the World*. Berkeley: University of California Press, 1986.

Yu Y, Harris AJ and Blair C *et al*. RASP (Reconstruct Ancestral State in Phylogenies): a tool for historical biogeography. *Mol Phylogenet Evol* 2015; **87**: 46–49.

Zhang C, Rabiee M and Sayyari E *et al*. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics* 2018; **19**: 153.

Zhang Y, Chen J and Sun H. Alpine speciation and morphological innovations: revelations from a species-rich genus in the northern hemisphere. *AoB Plants* 2021; **13**: plab018.

Supplementary tables

Supplementary Table S1. Summary of data likelihoods under each model, and results of statistical model choice from RASP software using BioGeoBEARS package. The model selected was the one with the highest AICc_wt value.

	LnL	numparams	d	e	AICc	AICc wt
DEC	-655.8	2	0.011	2×10^{-8}	**1316**	**1**
DIVALIKE	-670.4	2	0.014	1×10^{-12}	1345	4.7×10^{-7}
BAYAREALIKE	-697.4	2	0.0049	0.1	1399	9×10^{-19}

Supplementary Table S2. RASP distribution node area inference. For nodes IDs see Supplementary Fig. S3. Codes correspondence are: A= Hengduan Mountains; B = Himalayas; C = QTP; D = Sino-Japanese; E = Western Irano-Turanian; F = Eastern Irano-Turanian; G = Circumboreal.

Node ID	1st prob		2nd prob		3rd prob		4th prob		5th prob	
	area	%	area	%	area	%	area	%	area	%
325	A	89.99	AB	4.59	AC	4.58	ABC	0.84	AF	0.00
326	A	100.00	AB	0.00	AC	0.00	AF	0.00	AE	0.00
327	A	98.67	AB	0.65	AC	0.63	ABC	0.04	AF	0.00
328	A	100.00	AB	0.00	AC	0.00	AF	0.00	AE	0.00

Node ID	1st prob		2nd prob		3rd prob		4th prob		5th prob	
	area	%	area	%	area	%	area	%	area	%
329	A	99.72	AB	0.15	AC	0.13	ABC	0.01	AF	0.00
330	A	100.00	AB	0.00	AC	0.00	AF	0.00	AE	0.00
331	A	100.00	AB	0.00	AC	0.00	AF	0.00	AE	0.00
332	A	99.95	AB	0.04	AC	0.01	ABC	0.00	AF	0.00
333	A	99.97	AB	0.03	AC	0.00	ABC	0.00	AF	0.00
334	AB	53.21	B	46.79	ABC	0.00	ABG	0.00	ABF	0.00
335	AB	50.91	B	46.20	BC	2.35	ABC	0.41	AC	0.12
336	AB	85.17	A	12.88	ABC	1.84	AC	0.11	B	0.00
337	AB	93.21	A	6.79	B	0.00	ABC	0.00	ABG	0.00
338	B	51.10	AB	48.90	A	0.00	ABC	0.00	ABF	0.00
339	AB	91.39	A	6.01	B	1.98	ABC	0.61	AC	0.01
340	AB	91.90	A	7.81	ABC	0.28	AC	0.01	B	0.00
341	B	100.00	AB	0.00	BC	0.00	BF	0.00	BE	0.00
342	B	100.00	AB	0.00	BC	0.00	BF	0.00	BE	0.00
343	B	100.00	AB	0.00	BC	0.00	BF	0.00	BE	0.00
344	B	100.00	AB	0.00	BC	0.00	BF	0.00	BG	0.00
345	AB	99.57	B	0.29	ABC	0.14	BC	0.00	A	0.00
346	A	81.28	AB	18.69	AC	0.02	ABC	0.02	AD	0.00
347	A	100.00	AB	0.00	AC	0.00	AF	0.00	AD	0.00
348	A	95.76	AB	2.09	AC	1.98	ABC	0.18	AD	0.00
349	A	95.73	AB	3.92	AC	0.33	ABC	0.02	B	0.00
350	A	96.62	AB	3.38	AC	0.00	AD	0.00	AF	0.00
351	A	98.36	AB	1.55	AC	0.09	ABC	0.00	B	0.00
352	A	100.00	AB	0.00	AC	0.00	AD	0.00	AF	0.00
353	A	100.00	AB	0.00	AC	0.00	AD	0.00	AF	0.00
354	A	99.80	AB	0.19	AC	0.01	ABC	0.00	AD	0.00
355	A	95.73	AC	4.27	AD	0.00	AB	0.00	AE	0.00
356	A	98.93	AC	1.07	AD	0.00	AB	0.00	AE	0.00
357	A	100.00	AD	0.00	AC	0.00	AB	0.00	AF	0.00
358	A	99.90	AC	0.10	AD	0.00	AB	0.00	AE	0.00
359	A	92.68	AC	7.32	AD	0.00	ACD	0.00	AB	0.00
360	A	89.51	AC	5.91	AB	3.85	ABC	0.72	C	0.01
361	A	97.55	AC	1.58	AB	0.81	ABC	0.06	C	0.00
362	D	100.00	AD	0.00	CD	0.00	BD	0.00	DF	0.00
363	AD	98.38	ACD	1.07	ABD	0.49	CD	0.05	BD	0.00
364	AC	100.00	A	0.00	C	0.00	ACD	0.00	AD	0.00
365	A	54.30	AD	34.54	AC	9.87	ACD	1.03	ABD	0.17
366	A	96.14	AD	2.92	AC	0.89	ACD	0.02	AB	0.02
367	A	87.44	AC	12.56	AB	0.00	AD	0.00	AF	0.00
368	A	99.38	AC	0.62	AB	0.00	AD	0.00	AF	0.00
369	A	99.37	AD	0.43	AC	0.19	AB	0.00	ACD	0.00
370	AC	81.35	A	18.65	ABC	0.00	ACD	0.00	ACF	0.00
371	AC	95.58	C	4.42	A	0.00	ABC	0.00	AB	0.00
372	A	80.45	AB	10.24	ABC	4.69	AC	4.63	AD	0.00
373	A	79.33	AC	8.83	AB	8.07	ABC	3.51	C	0.18
374	A	92.09	AC	7.91	AB	0.00	ABC	0.00	AD	0.00
375	A	83.35	AB	9.08	AC	5.15	ABC	2.41	C	0.01
376	B	100.00	AB	0.00	BC	0.00	BD	0.00	BF	0.00
377	AB	95.47	ABC	4.45	B	0.05	BC	0.02	A	0.00

Node ID	1st prob		2nd prob		3rd prob		4th prob		5th prob	
	area	%	area	%	area	%	area	%	area	%
378	A	83.13	AB	15.79	AC	0.62	ABC	0.36	AD	0.09
379	A	98.64	AB	1.27	AC	0.07	ABC	0.01	AD	0.01
380	A	100.00	AB	0.00	AC	0.00	AF	0.00	AE	0.00
381	A	100.00	AB	0.00	AC	0.00	AF	0.00	AE	0.00
382	A	100.00	AB	0.00	AC	0.00	AF	0.00	AE	0.00
383	A	100.00	AB	0.00	AC	0.00	AF	0.00	AE	0.00
384	A	100.00	AB	0.00	AC	0.00	AF	0.00	AE	0.00
385	A	100.00	AB	0.00	AC	0.00	AF	0.00	AE	0.00
386	A	100.00	AB	0.00	AC	0.00	AF	0.00	AE	0.00
387	A	100.00	AB	0.00	AC	0.00	AF	0.00	AE	0.00
388	A	99.69	AB	0.30	AC	0.01	AD	0.00	ABC	0.00
389	A	85.20	AB	10.68	AC	3.01	ABC	0.95	B	0.16
390	A	98.49	AB	1.16	AC	0.32	ABC	0.03	AD	0.00
391	B	100.00	AB	0.00	BC	0.00	BF	0.00	BE	0.00
392	B	100.00	AB	0.00	BC	0.00	BF	0.00	BG	0.00
393	AB	86.28	ABC	13.39	BC	0.33	A	0.00	AC	0.00
394	A	100.00	AB	0.00	AC	0.00	AG	0.00	AD	0.00
395	A	77.98	AB	18.34	AC	2.03	ABC	1.66	AF	0.00
396	A	96.23	AB	3.77	AC	0.00	AF	0.00	AE	0.00
397	A	96.97	AB	2.50	AC	0.44	ABC	0.08	B	0.01
398	A	96.31	AB	1.84	AC	1.68	ABC	0.11	C	0.05
399	A	100.00	AC	0.00	AB	0.00	AF	0.00	AE	0.00
400	AC	100.00	A	0.00	AB	0.00	AF	0.00	AE	0.00
401	A	83.14	AC	16.44	AB	0.37	ABC	0.04	C	0.00
402	A	100.00	AC	0.00	AB	0.00	AF	0.00	AE	0.00
403	A	92.48	AC	7.33	AB	0.17	ABC	0.02	AF	0.00
404	A	94.64	AB	5.36	AC	0.00	AF	0.00	AE	0.00
405	A	99.19	AC	0.53	AB	0.28	ABC	0.00	B	0.00
406	A	100.00	AC	0.00	AB	0.00	AF	0.00	AE	0.00
407	A	75.99	AB	10.59	AC	10.59	ABC	2.52	B	0.13
408	A	96.35	AB	1.75	AC	1.75	ABC	0.15	AF	0.00
409	A	99.60	AC	0.21	AB	0.18	ABC	0.00	B	0.00
410	A	99.55	AB	0.31	AC	0.14	ABC	0.00	AD	0.00
411	A	100.00	AB	0.00	AC	0.00	AF	0.00	AE	0.00
412	A	94.50	AB	5.50	AC	0.00	AF	0.00	AE	0.00
413	A	97.09	AB	2.91	AC	0.00	AF	0.00	AE	0.00
414	A	69.30	AB	14.04	ABC	10.03	AC	6.62	AF	0.00
415	A	75.68	AB	13.51	ABC	6.67	AC	3.52	B	0.56
416	ABC	63.53	BC	16.50	AC	10.20	C	9.77	AB	0.00
417	ABC	43.02	AB	32.29	AC	8.44	B	7.27	BC	5.08
418	B	100.00	BF	0.00	BC	0.00	AB	0.00	BE	0.00
419	BF	100.00	B	0.00	BCF	0.00	ABF	0.00	BC	0.00
420	B	72.77	BF	27.23	BC	0.00	AB	0.00	BCF	0.00
421	B	94.97	BF	5.03	BC	0.00	AB	0.00	BE	0.00
422	B	40.85	ABC	31.23	AB	14.63	BC	12.81	BF	0.21
423	B	44.43	BC	24.83	ABC	22.08	AB	4.03	C	1.63
424	AB	100.00	B	0.00	ABC	0.00	ABE	0.00	ABF	0.00
425	B	100.00	AB	0.00	BC	0.00	BE	0.00	BF	0.00
426	B	91.67	AB	8.33	BC	0.00	BE	0.00	BF	0.00

Node ID	1st prob		2nd prob		3rd prob		4th prob		5th prob	
	area	%	area	%	area	%	area	%	area	%
427	B	85.77	AB	6.72	BC	5.74	ABC	1.76	A	0.01
428	B	100.00	BC	0.00	AB	0.00	BE	0.00	BF	0.00
429	B	91.57	BC	4.46	AB	2.74	ABC	1.24	BE	0.00
430	B	44.90	BC	27.78	ABC	20.29	AB	3.97	C	1.59
431	BC	47.70	C	22.07	ABC	13.42	AC	7.91	B	6.86
432	C	51.86	BC	25.21	AC	14.31	ABC	8.54	CE	0.04
433	A	82.29	AC	17.71	C	0.00	AB	0.00	ABC	0.00
434	AC	39.56	A	29.62	ABC	17.70	C	6.27	AB	5.81
435	A	46.32	AC	44.26	AB	4.23	ABC	2.97	C	2.17
436	AC	61.98	A	16.79	AB	8.34	ABC	7.90	C	3.09
437	AC	37.13	AB	22.70	ABC	19.18	A	7.75	C	6.87
438	AC	39.94	AB	29.02	B	10.64	ABC	9.82	BC	5.97
439	A	79.96	AB	20.04	B	0.00	AC	0.00	ABC	0.00
440	A	44.63	AB	29.37	AC	12.51	B	6.61	ABC	6.12
441	AB	100.00	B	0.00	ABF	0.00	ABE	0.00	A	0.00
442	AB	67.06	B	32.94	A	0.00	ABF	0.00	ABE	0.00
443	AB	58.90	A	28.77	B	3.96	ABC	3.91	AC	3.75
444	AF	21.84	AE	21.76	ABF	20.79	ABE	20.69	AEF	4.51
445	A	64.87	AC	12.65	AB	9.34	AF	3.84	AE	3.70
446	A	95.52	AC	1.57	AB	1.35	AF	0.75	AE	0.54
447	F	100.00	AF	0.00	EF	0.00	DF	0.00	FG	0.00
448	F	95.29	EF	4.71	AF	0.00	DF	0.00	FG	0.00
449	F	99.08	EF	0.92	AF	0.00	AEF	0.00	DF	0.00
450	F	100.00	EF	0.00	AF	0.00	DF	0.00	FG	0.00
451	EF	100.00	F	0.00	AEF	0.00	AF	0.00	DEF	0.00
452	F	81.57	EF	18.20	E	0.24	AF	0.00	AEF	0.00
453	EF	100.00	F	0.00	AEF	0.00	AF	0.00	E	0.00
454	F	70.37	EF	24.50	E	5.13	AF	0.00	AEF	0.00
455	A	86.01	AD	13.99	ADF	0.00	AF	0.00	AE	0.00
456	A	85.62	AD	14.35	D	0.03	AF	0.00	ADF	0.00
457	AF	76.58	AE	13.16	AEF	7.29	ADF	2.49	ADE	0.41
458	A	100.00	AF	0.00	AE	0.00	AD	0.00	AG	0.00
459	A	82.99	AF	13.22	AE	2.41	AEF	0.77	AD	0.31
460	A	91.70	AF	6.09	AE	0.84	AEF	0.39	AC	0.37
461	A	94.84	AD	5.16	AF	0.00	AE	0.00	AG	0.00
462	A	98.51	AD	1.49	AF	0.00	AE	0.00	AG	0.00
463	D	100.00	AD	0.00	DF	0.00	DE	0.00	DG	0.00
464	AD	96.85	D	3.15	A	0.00	ADF	0.00	ADE	0.00
465	A	91.62	AD	8.37	D	0.01	AF	0.00	AE	0.00
466	A	98.57	AD	1.43	AF	0.00	AE	0.00	AG	0.00
467	A	94.22	AF	4.38	AD	0.57	AE	0.32	AEF	0.28
468	A	94.79	AF	4.17	AD	0.44	AE	0.25	AEF	0.24
469	D	100.00	DG	0.00	DF	0.00	AD	0.00	BD	0.00
470	D	100.00	DG	0.00	DF	0.00	AD	0.00	BD	0.00
471	D	100.00	DG	0.00	DF	0.00	AD	0.00	BD	0.00
472	D	100.00	DG	0.00	DF	0.00	AD	0.00	BD	0.00
473	D	100.00	DG	0.00	DF	0.00	AD	0.00	BD	0.00
474	D	100.00	DG	0.00	DF	0.00	AD	0.00	BD	0.00
475	D	100.00	DG	0.00	DF	0.00	AD	0.00	BD	0.00

Node ID	1st prob		2nd prob		3rd prob		4th prob		5th prob	
	area	%	area	%	area	%	area	%	area	%
476	D	100.00	DG	0.00	DF	0.00	AD	0.00	BD	0.00
477	D	100.00	DG	0.00	DF	0.00	AD	0.00	BD	0.00
478	B	100.00	BD	0.00	BF	0.00	BG	0.00	AB	0.00
479	BF	100.00	BD	0.00	B	0.00	BDF	0.00	DF	0.00
480	D	100.00	DF	0.00	BD	0.00	DG	0.00	AD	0.00
481	BDF	37.35	DF	31.35	BD	31.30	D	0.00	BDG	0.00
482	D	57.88	DF	16.39	BD	16.30	BDF	9.42	DG	0.00
483	DG	100.00	D	0.00	G	0.00	DFG	0.00	BDG	0.00
484	DG	93.84	D	3.93	DFG	0.60	BDG	0.53	DF	0.48
485	DG	83.87	G	15.00	DFG	0.62	BDG	0.47	FG	0.02
486	DG	83.03	G	16.16	DFG	0.46	BDG	0.32	FG	0.02
487	DG	97.02	D	2.56	DFG	0.28	BDG	0.13	DF	0.01
488	G	100.00	DG	0.00	FG	0.00	AG	0.00	BG	0.00
489	DG	94.09	G	5.60	DFG	0.23	BDG	0.08	FG	0.00
490	D	100.00	DG	0.00	DF	0.00	AD	0.00	BD	0.00
491	D	100.00	DG	0.00	DF	0.00	AD	0.00	BD	0.00
492	D	100.00	DG	0.00	DF	0.00	AD	0.00	BD	0.00
493	D	98.05	DF	1.95	DG	0.00	AD	0.00	BD	0.00
494	D	92.79	DG	6.95	DF	0.23	DFG	0.03	FG	0.00
495	DG	91.08	D	8.47	DFG	0.26	G	0.10	BDG	0.05
496	D	100.00	DG	0.00	DF	0.00	AD	0.00	BD	0.00
497	D	100.00	DG	0.00	DF	0.00	AD	0.00	BD	0.00
498	D	100.00	DG	0.00	DF	0.00	AD	0.00	BD	0.00
499	D	100.00	DG	0.00	DF	0.00	AD	0.00	BD	0.00
500	DG	87.55	D	12.19	DFG	0.21	DF	0.04	BDG	0.01
501	AD	100.00	A	0.00	D	0.00	ADF	0.00	ADG	0.00
502	AD	87.85	A	12.15	ADF	0.00	ADG	0.00	ABD	0.00
503	AD	86.11	A	13.89	ADF	0.00	ADG	0.00	ABD	0.00
504	AD	99.75	D	0.25	A	0.00	ADF	0.00	ADG	0.00
505	A	91.83	AD	8.17	D	0.00	AF	0.00	AG	0.00
506	A	100.00	AD	0.00	AF	0.00	AG	0.00	AB	0.00
507	AD	100.00	A	0.00	AF	0.00	AG	0.00	D	0.00
508	A	85.21	AD	14.77	D	0.02	AF	0.00	AG	0.00
509	A	100.00	AF	0.00	AG	0.00	AD	0.00	AB	0.00
510	A	89.24	AD	10.76	AF	0.00	AG	0.00	ADF	0.00
511	D	100.00	AD	0.00	DF	0.00	DG	0.00	BD	0.00
512	D	100.00	AD	0.00	DF	0.00	DG	0.00	BD	0.00
513	D	100.00	AD	0.00	DF	0.00	DG	0.00	BD	0.00
514	D	96.08	DF	3.92	AD	0.00	ADF	0.00	DG	0.00
515	AD	48.87	ADF	25.78	ADG	21.88	DF	1.63	DG	1.34
516	AD	64.43	ADF	17.11	ADG	13.58	A	3.48	AF	1.04
517	AD	73.45	ADF	13.72	ADG	9.98	D	1.28	AF	1.11
518	A	95.16	AD	4.84	AF	0.00	AG	0.00	AB	0.00
519	A	98.62	AD	1.38	AF	0.00	AG	0.00	AB	0.00
520	A	58.46	AD	20.11	AF	6.60	ADF	6.20	AG	4.31
521	A	68.63	AF	9.30	AD	8.15	AG	6.30	D	3.34
522	FG	100.00	AFG	0.00	F	0.00	G	0.00	DFG	0.00
523	AFG	63.51	FG	11.31	DFG	7.41	AF	5.14	AG	3.11
524	FG	49.84	AFG	29.44	DFG	7.41	F	6.96	DF	3.55

Node ID	1st prob		2nd prob		3rd prob		4th prob		5th prob	
	area	%	area	%	area	%	area	%	area	%
525	G	44.49	DG	17.75	FG	16.60	DFG	10.43	AFG	4.46
526	G	100.00	DG	0.00	FG	0.00	AG	0.00	EG	0.00
527	G	100.00	DG	0.00	FG	0.00	AG	0.00	EG	0.00
528	G	100.00	DG	0.00	FG	0.00	AG	0.00	EG	0.00
529	G	100.00	DG	0.00	FG	0.00	AG	0.00	EG	0.00
530	G	100.00	DG	0.00	FG	0.00	AG	0.00	EG	0.00
531	G	100.00	DG	0.00	FG	0.00	AG	0.00	EG	0.00
532	G	100.00	DG	0.00	FG	0.00	AG	0.00	EG	0.00
533	G	100.00	DG	0.00	FG	0.00	AG	0.00	EG	0.00
534	G	100.00	DG	0.00	FG	0.00	AG	0.00	EG	0.00
535	G	100.00	DG	0.00	FG	0.00	AG	0.00	EG	0.00
536	G	100.00	DG	0.00	FG	0.00	AG	0.00	EG	0.00
537	G	100.00	DG	0.00	FG	0.00	AG	0.00	EG	0.00
538	G	100.00	DG	0.00	FG	0.00	AG	0.00	EG	0.00
539	G	100.00	DG	0.00	FG	0.00	AG	0.00	EG	0.00
540	G	92.20	FG	3.14	DG	2.82	AG	0.87	DFG	0.50
541	G	100.00	FG	0.00	AG	0.00	DG	0.00	EG	0.00
542	G	100.00	FG	0.00	AG	0.00	DG	0.00	EG	0.00
543	G	100.00	FG	0.00	AG	0.00	DG	0.00	EG	0.00
544	G	100.00	FG	0.00	AG	0.00	DG	0.00	EG	0.00
545	G	100.00	FG	0.00	AG	0.00	DG	0.00	EG	0.00
546	G	100.00	FG	0.00	AG	0.00	DG	0.00	EG	0.00
547	G	100.00	FG	0.00	AG	0.00	DG	0.00	EG	0.00
548	G	97.30	FG	1.26	DG	0.69	AG	0.30	AFG	0.27
549	EG	100.00	G	0.00	DEG	0.00	DG	0.00	FG	0.00
550	G	85.52	EG	14.48	DG	0.00	DEG	0.00	FG	0.00
551	G	94.22	EG	5.78	DG	0.00	DEG	0.00	FG	0.00
552	DG	95.30	DEG	4.68	DE	0.01	G	0.00	EG	0.00
553	DG	84.23	G	13.30	DEG	2.38	EG	0.09	DFG	0.00
554	DG	82.40	G	16.77	DEG	0.80	EG	0.03	DFG	0.00
555	DG	98.57	D	1.07	DEG	0.36	DE	0.00	G	0.00
556	DG	55.21	G	44.60	DEG	0.15	EG	0.04	D	0.00
557	D	100.00	DG	0.00	DE	0.00	DF	0.00	AD	0.00
558	DG	65.99	D	34.01	G	0.00	DEG	0.00	DFG	0.00
559	G	60.32	DG	39.33	D	0.32	DEG	0.02	EG	0.00
560	G	91.04	DG	8.96	EG	0.00	DEG	0.00	FG	0.00
561	G	97.80	DG	2.20	EG	0.00	DEG	0.00	FG	0.00
562	G	100.00	FG	0.00	DG	0.00	AG	0.00	EG	0.00
563	G	100.00	FG	0.00	DG	0.00	AG	0.00	EG	0.00
564	G	99.45	DG	0.55	EG	0.00	DEG	0.00	FG	0.00
565	G	99.84	DG	0.16	EG	0.00	DEG	0.00	FG	0.00
566	G	98.57	FG	0.79	AFG	0.23	DG	0.15	AG	0.15
567	D	100.00	DG	0.00	DF	0.00	AD	0.00	DE	0.00
568	DG	100.00	D	0.00	DFG	0.00	ADG	0.00	DEG	0.00
569	D	80.14	DG	19.86	AD	0.00	DF	0.00	DE	0.00
570	D	100.00	DG	0.00	AD	0.00	DF	0.00	DE	0.00
571	D	88.33	DG	11.67	G	0.00	AD	0.00	DF	0.00
572	D	100.00	DG	0.00	AD	0.00	DF	0.00	DE	0.00
573	D	100.00	DG	0.00	AD	0.00	DF	0.00	DE	0.00

Node ID	1st prob		2nd prob		3rd prob		4th prob		5th prob	
	area	%	area	%	area	%	area	%	area	%
574	G	100.00	DG	0.00	AG	0.00	FG	0.00	EG	0.00
575	DG	100.00	G	0.00	D	0.00	ADG	0.00	DFG	0.00
576	DG	92.55	G	6.49	D	0.96	ADG	0.00	DFG	0.00
577	EG	100.00	G	0.00	DEG	0.00	EFG	0.00	AEG	0.00
578	G	100.00	EG	0.00	DG	0.00	FG	0.00	AG	0.00
579	G	89.37	EG	10.63	DG	0.00	FG	0.00	AG	0.00
580	G	81.05	DG	17.88	EG	0.95	DEG	0.12	DE	0.00
581	G	93.02	DG	5.55	FG	0.77	AFG	0.23	EG	0.16
582	F	82.22	FG	17.78	AFG	0.00	G	0.00	AF	0.00
583	F	63.95	FG	28.61	G	7.44	AFG	0.00	DFG	0.00
584	F	65.31	FG	34.13	G	0.55	AFG	0.00	DFG	0.00
585	F	100.00	FG	0.00	AF	0.00	DF	0.00	EF	0.00
586	F	66.85	FG	33.15	AFG	0.00	AF	0.00	DFG	0.00
587	G	100.00	FG	0.00	AG	0.00	DG	0.00	EG	0.00
588	FG	99.53	G	0.47	F	0.00	AFG	0.00	DFG	0.00
589	FG	62.50	F	37.50	G	0.00	AFG	0.00	DFG	0.00
590	F	100.00	FG	0.00	AF	0.00	DF	0.00	EF	0.00
591	FG	100.00	F	0.00	G	0.00	AFG	0.00	DFG	0.00
592	FG	83.96	G	10.09	F	5.95	AFG	0.00	DFG	0.00
593	FG	82.49	G	12.09	DG	3.43	DFG	0.85	F	0.46
594	DF	100.00	F	0.00	ADF	0.00	D	0.00	AF	0.00
595	F	35.09	FG	30.44	DF	19.19	DFG	9.84	DG	3.62
596	AF	55.23	ADF	18.13	AFG	12.26	AD	6.81	AG	4.44
597	A	100.00	AF	0.00	AD	0.00	AE	0.00	AG	0.00
598	A	100.00	AF	0.00	AD	0.00	AE	0.00	AG	0.00
599	A	100.00	AF	0.00	AD	0.00	AE	0.00	AG	0.00
600	AF	51.67	ADF	16.62	A	11.52	AFG	8.87	AD	6.69
601	E	100.00	EF	0.00	EG	0.00	DE	0.00	AE	0.00
602	EF	100.00	E	0.00	F	0.00	DEF	0.00	EFG	0.00
603	EF	86.61	E	13.39	DEF	0.00	EFG	0.00	AEF	0.00
604	EF	80.02	E	19.98	F	0.00	DEF	0.00	EFG	0.00
605	F	100.00	EF	0.00	FG	0.00	DF	0.00	AF	0.00
606	FG	100.00	F	0.00	EFG	0.00	EF	0.00	DFG	0.00
607	F	85.45	FG	14.55	EF	0.00	EFG	0.00	DF	0.00
608	F	84.82	FG	15.13	G	0.05	EF	0.00	EFG	0.00
609	EF	68.24	F	30.15	EFG	1.31	FG	0.20	EG	0.09
610	E	100.00	EF	0.00	EG	0.00	DE	0.00	AE	0.00
611	E	100.00	EF	0.00	EG	0.00	DE	0.00	AE	0.00
612	E	63.29	EF	36.71	DEF	0.00	AEF	0.00	EFG	0.00
613	E	68.17	EF	31.83	DEF	0.00	AEF	0.00	EFG	0.00
614	EF	70.17	F	28.22	EFG	0.80	E	0.72	EG	0.06
615	F	100.00	DF	0.00	AF	0.00	EF	0.00	FG	0.00
616	DF	62.13	ADF	36.35	AF	1.51	D	0.00	F	0.00
617	DF	74.98	ADF	13.92	D	10.57	AD	0.54	F	0.00
618	F	82.31	DF	13.26	FG	3.80	DFG	0.55	DG	0.07
619	F	45.51	DF	44.46	ADF	6.28	AF	2.32	FG	0.67
620	F	77.00	EF	17.93	DF	2.85	DEF	0.73	AF	0.54
621	F	58.27	FG	14.46	EFG	14.14	EF	13.14	DF	0.00
622	F	83.34	FG	7.73	DF	4.24	EF	1.86	DFG	1.43

Node ID	1st prob		2nd prob		3rd prob		4th prob		5th prob	
	area	%	area	%	area	%	area	%	area	%
623	F	78.64	EF	12.46	DF	3.46	FG	2.68	EFG	0.73
624	FG	100.00	F	0.00	EFG	0.00	G	0.00	EF	0.00
625	EF	75.13	F	11.10	EFG	7.68	FG	3.99	EG	2.10
626	E	100.00	EF	0.00	EG	0.00	CE	0.00	AE	0.00
627	EF	52.12	E	45.19	EFG	2.26	EG	0.43	F	0.00
628	E	81.89	EF	18.11	F	0.00	CE	0.00	EG	0.00
629	EF	60.81	E	37.72	EFG	1.29	EG	0.13	F	0.05
630	EF	61.00	E	38.29	EFG	0.63	EG	0.08	F	0.00
631	FG	100.00	F	0.00	EFG	0.00	CFG	0.00	AFG	0.00
632	F	88.57	FG	11.43	EF	0.00	EFG	0.00	CF	0.00
633	F	100.00	EF	0.00	CF	0.00	AF	0.00	FG	0.00
634	F	100.00	EF	0.00	CF	0.00	AF	0.00	FG	0.00
635	F	97.07	FG	2.93	EF	0.00	EFG	0.00	CF	0.00
636	EF	56.05	F	42.56	EFG	1.14	FG	0.25	EG	0.01
637	F	76.67	EF	22.67	EFG	0.41	FG	0.26	CF	0.00
638	F	84.83	EF	8.59	DF	6.32	DEF	0.09	FG	0.09
639	A	100.00	AF	0.00	AC	0.00	AE	0.00	AD	0.00
640	AC	100.00	ACF	0.00	C	0.00	CF	0.00	ACE	0.00
641	C	54.04	AC	45.96	CF	0.00	ACF	0.00	ACE	0.00
642	CF	56.62	AF	19.15	ACF	10.25	CEF	5.14	CDF	3.87
643	F	44.35	CF	18.35	AF	14.66	ACF	5.15	EF	4.12
644	AF	26.20	F	20.86	ADF	13.79	A	6.20	DF	5.45
645	E	100.00	EF	0.00	AE	0.00	DE	0.00	EG	0.00
646	EF	37.34	AE	20.05	DEF	11.90	ADE	10.28	AEF	8.56
647	DEF	51.33	ADE	32.64	DE	4.40	CDE	3.44	ADF	2.82

Supplementary Table S3. Median estimated ages and 95% of confidence intervals (CI) for the best ML phylogenetic tree by penalized likelihood method implemented in treePL software (see text for details). Node ID numbers correspond to those represented in Supplementary Fig. S5.

Node code	Tyoe of node	Median age (Myr)	Lower 95% CI	Upper 95% CI
385	internal	70.02	63.93	74.50
386	internal	54.50	48.68	60.19
387	internal	49.46	42.29	56.11
388	internal	47.60	40.39	55.59
389	internal	39.82	34.08	44.58
390	internal	37.53	31.71	42.17
391	internal	33.02	28.40	37.58
392	internal	32.24	27.72	36.59
393	internal	27.67	22.41	31.44
394	internal	24.62	20.08	28.86
395	internal	23.10	18.93	27.17
396	internal	20.65	16.34	23.72
397	internal	19.39	15.42	22.17
398	internal	18.77	14.98	21.46
399	internal	17.62	13.74	20.15

Node code	Tyoe of node	Median age (Myr)	Lower 95% CI	Upper 95% CI
400	internal	16.73	12.79	19.15
401	internal	16.09	12.21	18.41
402	internal	15.28	11.38	17.49
403	internal	14.73	10.82	16.87
404	internal	12.57	8.16	14.48
405	internal	11.64	6.95	13.46
406	internal	11.27	6.54	13.05
407	internal	10.80	6.19	12.51
408	internal	10.25	5.82	11.88
409	internal	9.62	5.47	11.15
410	internal	9.21	5.30	10.67
411	internal	8.53	4.90	9.89
412	internal	8.01	4.49	9.30
413	internal	7.73	4.31	8.97
414	internal	7.46	4.16	8.65
415	internal	6.82	3.81	7.92
416	internal	5.99	3.41	6.95
417	internal	5.58	3.22	6.47
418	internal	5.18	3.05	5.99
419	internal	4.81	2.90	5.55
420	internal	4.56	2.86	5.30
421	internal	3.99	2.45	4.59
422	internal	3.78	2.39	4.40
423	internal	3.26	2.07	3.79
424	internal	2.86	1.82	3.30
425	internal	2.43	1.55	2.80
426	internal	2.26	1.45	2.60
427	internal	2.78	1.71	3.19
428	internal	3.41	2.16	3.96
429	internal	2.67	1.69	3.08
430	internal	4.25	2.71	4.88
431	internal	3.75	2.46	4.25
432	internal	3.45	2.22	3.83
433	internal	2.79	1.88	3.12
434	internal	3.18	2.08	3.65
435	internal	2.33	1.54	2.64
436	internal	4.33	2.60	5.00
437	internal	3.81	2.36	4.44
438	internal	3.00	1.87	3.45
439	internal	3.57	2.11	4.13
440	internal	5.08	2.91	5.89
441	internal	4.40	2.50	5.09
442	internal	5.03	2.82	5.84
443	internal	6.20	3.43	7.21
444	internal	4.94	2.69	5.74
445	internal	6.76	3.76	7.84
446	internal	6.39	3.57	7.42
447	internal	5.93	3.35	6.88
448	internal	4.95	2.96	5.77

Node code	Tyoe of node	Median age (Myr)	Lower 95% CI	Upper 95% CI
449	internal	4.67	2.82	5.41
450	internal	4.26	2.62	4.90
451	internal	3.92	2.40	4.52
452	internal	2.97	1.81	3.47
453	internal	2.46	1.50	2.85
454	internal	2.39	1.47	2.76
455	internal	3.54	2.16	3.97
456	internal	2.58	1.65	2.88
457	internal	3.84	2.33	4.46
458	internal	3.96	2.36	4.63
459	internal	5.60	3.09	6.51
460	internal	2.96	1.60	3.44
461	internal	6.10	3.40	7.09
462	internal	5.64	3.15	6.54
463	internal	5.40	3.03	6.26
464	internal	4.45	2.53	5.15
465	internal	3.45	2.05	3.99
466	internal	4.54	2.54	5.26
467	internal	3.84	2.13	4.46
468	internal	4.80	2.63	5.57
469	internal	7.34	4.08	8.51
470	internal	6.62	3.72	7.67
471	internal	5.60	3.19	6.47
472	internal	4.98	2.95	5.81
473	internal	4.32	2.57	5.05
474	internal	3.47	2.10	4.00
475	internal	5.44	2.83	6.35
476	internal	4.29	2.20	5.01
477	internal	6.79	3.71	7.90
478	internal	8.21	4.76	9.51
479	internal	7.63	4.53	8.81
480	internal	6.25	4.08	7.20
481	internal	5.95	3.95	6.81
482	internal	5.65	3.81	6.42
483	internal	5.30	3.64	5.98
484	internal	4.65	3.28	5.30
485	internal	4.13	2.96	4.70
486	internal	3.60	2.62	4.06
487	internal	2.72	2.03	3.07
488	internal	4.01	2.86	4.57
489	internal	4.33	3.03	4.89
490	internal	5.49	3.66	6.27
491	internal	4.86	3.26	5.54
492	internal	5.08	3.18	5.83
493	internal	5.62	3.30	6.50
494	internal	7.32	4.03	8.50
495	internal	6.22	3.39	7.23
496	internal	5.07	2.65	5.94
497	internal	8.39	4.98	9.69

Node code	Tyoe of node	Median age (Myr)	Lower 95% CI	Upper 95% CI
498	internal	7.65	4.70	8.80
499	internal	6.92	4.40	7.93
500	internal	6.47	4.32	7.47
501	internal	5.90	4.05	6.74
502	internal	5.41	3.78	6.19
503	internal	4.75	3.39	5.41
504	internal	4.32	3.13	4.93
505	internal	3.99	2.92	4.55
506	internal	3.48	2.59	3.93
507	internal	2.97	2.25	3.35
508	internal	2.43	1.86	2.72
509	internal	3.05	2.22	3.51
510	internal	3.59	2.60	4.10
511	internal	3.69	2.67	4.20
512	internal	2.85	2.09	3.24
513	internal	2.40	1.78	2.71
514	internal	1.53	1.14	1.72
515	internal	5.09	3.59	5.80
516	internal	4.60	3.29	5.24
517	internal	4.36	3.14	4.96
518	internal	3.29	2.41	3.74
519	internal	3.46	2.52	3.94
520	internal	3.78	2.70	4.30
521	internal	6.87	4.27	7.89
522	internal	3.81	2.40	4.37
523	internal	5.87	3.68	6.74
524	internal	6.96	3.84	8.09
525	internal	6.24	3.36	7.27
526	internal	5.60	2.97	6.52
527	internal	4.98	2.63	5.80
528	internal	3.97	2.08	4.63
529	internal	4.49	2.37	5.23
530	internal	5.81	3.12	6.77
531	internal	4.48	2.40	5.22
532	internal	9.15	4.97	10.62
533	internal	8.43	4.51	9.79
534	internal	7.38	3.96	8.55
535	internal	6.51	3.55	7.52
536	internal	5.93	3.18	6.86
537	internal	4.71	2.46	5.47
538	internal	4.49	2.41	5.20
539	internal	4.70	2.69	5.25
540	internal	3.69	2.23	4.11
541	internal	5.51	2.80	6.41
542	internal	5.34	2.66	6.23
543	internal	3.21	1.58	3.73
544	internal	5.71	2.80	6.78
545	internal	9.61	5.35	11.14
546	internal	8.79	4.83	10.21

Node code	Tyoe of node	Median age (Myr)	Lower 95% CI	Upper 95% CI
547	internal	6.00	3.22	6.98
548	internal	4.91	2.65	5.71
549	internal	7.43	4.04	8.62
550	internal	5.92	3.14	6.89
551	internal	9.42	4.89	10.95
552	internal	8.69	4.43	10.11
553	internal	8.30	4.25	9.65
554	internal	8.18	4.20	9.50
555	internal	7.87	4.02	9.14
556	internal	7.64	3.92	8.88
557	internal	7.12	3.74	8.25
558	internal	6.66	3.57	7.72
559	internal	5.95	3.42	6.94
560	internal	5.66	3.32	6.52
561	internal	5.37	3.22	6.13
562	internal	4.86	2.96	5.41
563	internal	4.65	2.96	5.17
564	internal	4.17	2.77	4.63
565	internal	3.71	2.56	4.12
566	internal	3.54	2.47	3.93
567	internal	3.28	2.32	3.63
568	internal	2.46	1.81	2.73
569	internal	2.50	1.80	2.77
570	internal	3.43	2.28	3.81
571	internal	4.91	2.97	5.57
572	internal	4.24	2.59	4.81
573	internal	3.76	2.26	4.19
574	internal	3.08	1.86	3.43
575	internal	2.19	1.36	2.46
576	internal	3.13	1.89	3.49
577	internal	4.08	2.48	4.63
578	internal	3.51	2.09	3.92
579	internal	3.12	1.86	3.47
580	internal	4.93	2.84	5.74
581	internal	5.62	3.23	6.54
582	internal	5.06	2.94	5.82
583	internal	4.33	2.55	4.92
584	internal	3.88	2.25	4.33
585	internal	3.13	1.90	3.49
586	internal	6.30	3.28	7.30
587	internal	5.70	2.84	6.62
588	internal	4.60	2.20	5.36
589	internal	5.37	2.80	6.22
590	internal	6.80	3.61	7.88
591	internal	6.50	3.49	7.52
592	internal	6.25	3.39	7.23
593	internal	5.72	3.09	6.62
594	internal	5.21	2.82	6.03
595	internal	4.61	2.51	5.32

Node code	Tyoe of node	Median age (Myr)	Lower 95% CI	Upper 95% CI
596	internal	4.10	2.26	4.73
597	internal	3.72	2.12	4.31
598	internal	3.16	1.81	3.64
599	internal	4.70	2.53	5.43
600	internal	3.99	2.15	4.62
601	internal	5.08	2.71	5.88
602	internal	4.12	2.18	4.78
603	internal	5.13	3.03	5.93
604	internal	4.99	2.97	5.73
605	internal	4.65	2.82	5.32
606	internal	3.64	2.35	4.06
607	internal	3.10	2.06	3.42
608	internal	2.79	1.90	3.10
609	internal	2.44	1.70	2.74
610	internal	4.36	2.61	5.00
611	internal	3.84	2.31	4.41
612	internal	3.94	2.24	4.54
613	internal	3.87	2.03	4.49
614	internal	7.26	3.56	8.44
615	internal	6.72	3.21	7.82
616	internal	6.24	2.92	7.26
617	internal	5.82	2.71	6.77
618	internal	4.81	2.23	5.60
619	internal	3.57	1.65	4.15
620	internal	4.56	2.11	5.30
621	internal	4.53	2.08	5.28
622	internal	3.68	1.74	4.27
623	internal	6.33	3.01	7.35
624	internal	5.57	2.62	6.47
625	internal	4.79	2.25	5.56
626	internal	4.86	2.28	5.65
627	internal	5.66	2.67	6.59
628	internal	7.02	3.01	8.20
629	internal	6.05	2.45	7.12
630	internal	5.47	2.20	6.42
631	internal	4.62	1.86	5.40
632	internal	4.93	1.91	5.92
633	internal	6.29	2.64	7.36
634	internal	4.15	1.71	4.86
635	internal	6.81	3.77	7.86
636	internal	6.25	3.60	7.19
637	internal	5.87	3.57	6.76
638	internal	5.61	3.49	6.40
639	internal	5.41	3.41	6.11
640	internal	4.77	3.18	5.30
641	internal	4.52	3.14	5.14
642	internal	4.16	2.93	4.60
643	internal	3.54	2.64	3.90
644	internal	3.13	2.43	3.50

Node code	Tyoe of node	Median age (Myr)	Lower 95% CI	Upper 95% CI
645	internal	2.93	2.31	3.29
646	internal	2.66	2.14	2.98
647	internal	2.22	1.77	2.43
648	internal	4.38	2.74	5.01
649	internal	3.56	2.23	4.06
650	internal	5.27	2.92	6.06
651	internal	4.32	2.35	4.97
652	internal	7.60	3.68	8.85
653	internal	6.89	3.30	8.01
654	internal	6.23	3.00	7.25
655	internal	5.50	2.63	6.39
656	internal	4.33	2.08	5.03
657	internal	3.52	1.71	4.09
658	internal	5.31	2.57	6.16
659	internal	6.09	2.87	7.09
660	internal	5.40	2.53	6.29
661	internal	4.00	1.87	4.66
662	internal	4.44	2.06	5.20
663	internal	6.49	2.97	7.62
664	internal	5.53	2.51	6.49
665	internal	4.82	2.16	5.89
666	internal	8.07	3.85	9.40
667	internal	7.14	3.32	8.36
668	internal	6.61	3.05	7.74
669	internal	6.27	2.89	7.32
670	internal	5.55	2.58	6.46
671	internal	4.67	2.17	5.43
672	internal	4.13	1.93	4.80
673	internal	5.49	2.51	6.45
674	internal	5.22	2.38	6.13
675	internal	6.50	2.94	7.72
676	internal	5.50	2.49	6.71
677	internal	6.51	2.99	8.36
678	internal	6.33	3.79	7.92
679	internal	4.55	2.70	5.88
680	internal	4.50	2.69	5.72
681	internal	10.49	6.79	12.22
682	internal	7.39	5.15	9.26
683	internal	7.02	4.94	8.85
684	internal	6.51	4.63	8.25
685	internal	6.17	4.42	7.82
686	internal	4.40	3.13	5.76
687	internal	4.05	2.89	5.31
688	internal	3.13	2.21	4.22
689	internal	3.33	2.39	4.29
690	internal	5.30	3.87	6.66
691	internal	3.49	2.66	4.29
692	internal	2.91	2.22	3.60
693	internal	3.16	2.42	3.87

Node code	Tyoe of node	Median age (Myr)	Lower 95% CI	Upper 95% CI
694	internal	4.82	3.49	6.19
695	internal	3.90	2.88	4.86
696	internal	2.78	2.09	3.33
697	internal	3.22	2.23	4.53
698	internal	4.76	3.36	6.19
699	internal	4.21	2.99	5.48
700	internal	3.08	2.19	4.04
701	internal	1.56	1.13	1.95
702	internal	3.90	2.74	5.16
703	internal	6.23	4.32	7.99
704	internal	2.72	1.90	3.48
705	internal	9.63	6.10	11.36
706	internal	9.19	5.78	10.86
707	internal	7.45	4.61	8.89
708	internal	6.78	4.20	8.02
709	internal	6.37	4.00	7.42
710	internal	5.96	3.80	6.93
711	internal	5.67	3.66	6.58
712	internal	4.38	3.03	4.98
713	internal	3.47	2.47	3.91
714	internal	4.25	2.73	4.94
715	internal	4.30	2.61	5.16
716	internal	5.73	3.36	6.86
717	internal	4.91	2.86	5.86
718	internal	4.04	2.34	4.81
719	internal	4.96	2.91	6.07
720	internal	3.66	2.11	4.49
721	internal	5.96	3.82	7.35
722	internal	3.45	2.20	4.30
723	internal	1.94	1.24	2.35
724	internal	2.93	2.39	3.31
725	internal	9.81	8.04	11.57
726	internal	5.88	5.03	6.62
727	internal	2.80	2.49	3.14
728	internal	2.05	1.83	2.28
729	internal	3.66	3.21	4.16
730	internal	6.58	5.20	7.79
731	internal	5.36	4.22	6.35
732	internal	11.76	8.03	14.23
733	internal	7.79	4.98	9.86
734	internal	9.42	6.36	11.42
735	internal	8.07	6.84	9.63
736	internal	5.49	4.80	6.58
737	internal	4.22	3.68	5.05
738	internal	5.91	5.11	7.05
739	internal	4.88	4.26	5.93
740	internal	13.53	11.56	15.29
741	internal	12.32	10.80	13.83
742	internal	10.46	9.39	11.81

Node code	Tyoe of node	Median age (Myr)	Lower 95% CI	Upper 95% CI
743	internal	9.69	8.53	10.80
744	internal	9.02	7.84	10.10
745	internal	8.31	7.16	9.41
746	internal	7.28	6.24	8.48
747	internal	6.70	5.65	7.52
748	internal	6.11	4.99	6.90
749	internal	5.67	4.72	6.43
750	internal	4.99	4.17	5.66
751	internal	3.01	2.49	3.41
752	internal	7.88	6.84	8.81
753	internal	10.21	8.92	11.51
754	internal	16.53	13.90	18.57
755	internal	13.94	11.25	16.23
756	internal	9.73	7.91	11.31
757	internal	16.08	12.17	18.59
758	internal	6.74	5.33	7.74
759	internal	26.38	21.61	30.34
760	internal	23.31	18.69	27.04
761	internal	10.11	6.00	13.21
762	internal	2.73	1.46	3.88
763	internal	30.84	26.62	34.88
764	internal	16.31	14.60	18.22
765	internal	14.30	12.24	16.36
766	internal	27.07	22.82	30.65
767	internal	42.87	38.55	47.61

Supplementary Table S4. Fit of diversification models applied to the phylogenetic tree of *Saussurea*. The best-fitting model accoding to AICc is highlighted in bold. Abbreviations: NP, number of free parameters; logL, log-likelihood; λ, speciation rate; α, rate of variation of speciation through time (in time-dependent models) or according to temperature (in climate-dependent models); μ, extinction rate; K, carrying-capacity in diversity-dependent models; λ_1, speciation rate before the shift; λ_2, speciation rate after the shift; μ_1, extinction rate before the shift; μ_2, extinction rate after the shift; K_1, carrying-capacity before the shift; K_2, carrying-capacity after the shift; t, time estimated of diversification shift in diversity-dependent diversification models.

MODEL TYPE	MODEL DESCRIPTION	MODEL ACRONYM	NP	logL	AICc	λ	α	μ
Constant-rate	Constant speciation and no extinction	BCST	1	-825.8	1653.6	0.242	–	–
	Constant speciation and constant extinction	BCSTDCST	2	-825.8	1655.7	0.243	–	0.00000
Time-dependent	Speciation exponentially correlated with time, no extinction	BTimeVar_EXPO	2	-825.8	1655.7	0.243	0.00000	–
	Speciation linearly correlated with time, no extinction	BTimeVar_LIN	2	-668.9	1341.8	0.905	0.00000	–
Climate-dependent	Speciation exponentially correlated with temperature, no extinction	BEnv.Var_EXPO	2	-772.7	1549.4	0.155	0.0245	–

	Speciation exponentially correlated with temperature, constant extinction	BEnv.VarDCST_EXPO	3	-772.7	1551.5	0.155	0.0245	<0.00001					
	Speciation linearly correlated with temperature, no extinction	BEnv.Var_LIN	2	-678.2	1360.4	0.018	0.018	–					
	Speciation linearly correlated with temperature, constant extinction	BEnv.VarDCST_LIN	3	-678.2	1362.4	0.018	0.018	<0.00001					
MODEL TYPE	**MODEL DESCRIPTION**	**MODEL ACRONYM**	**NP**	**logL**	**AICc**	**λ**		**μ**		**K**			
Diversity-dependent	Linear dependence in speciation rate, no extinction, K corresponds to carrying capacity where speciation = extinction	DDL	2	-669.7	1343.5	0.966		–		500			
	Linear dependence in speciation rate, with extinction, K corresponds to carrying capacity where speciation = extinction	DDL+E	3	-599.0	1204.1	0.961		0.00245		662.2			
	Linear dependence in speciation rate, no extinction, K' corresponds to carrying capacity where speciation = 0	DDL_K'	2	-607.5	1219.2	0.969		–		662.2			
	Linear dependence in speciation rate, with extinction, K' corresponds to carrying capacity where speciation = 0	DDL_K'+E	3	-595.6	1197.29	0.969		0.00016		662.2			
MODEL TYPE	**MODEL DESCRIPTION**	**MODEL ACRONYM**	**NP**	**logL**	**AICc**	**λ_1**	**λ_2**	**μ_1**	**μ_2**	**K_1**	**K_2**	**t**	
Diversity-dependent, shifting parameters at 12 mya	Linear dependence in speciation rate with shifting parameters at time t=12, no extinction	DDL_SR_12	4	-635.2	1278.6	2.428	0.984	0.000	0.000	3	602		
Diversity-dependent, shifting parameters at 8 mya	Linear dependence in speciation rate with shifting parameters at time t=8, no extinction	DDL_SR_8	4	-588.4	1184.9	0.466	1.039	0.000	0.000	437	664		
Diversity-dependent, shifting parameters at 7 mya	Linear dependence in speciation rate with shifting parameters at time t=7, no extinction	DDL_SR_7	4	-596.2	1200.6	0.496	1.085	0.000	0.000	1509	631		
Diversity-dependent, shifting parameters at 6 mya	Linear dependence in speciation rate with shifting parameters at time t=6, no extinction	DDL_SR_6	4	-585.2	1178.5	0.627	1.139	0.000	0.000	1335	667		
Diversity-dependent,	Linear dependence in speciation rate with shifting parameters at time t (t to be optimized), no extinction	DDL_SR_OPTIM	5	-589.0	1188.3	0.626	1.142	0.000	0.000	1572	647	6	
Diversity-dependent,	Linear dependence in speciation rate with shifting parameters at time t (t to be optimized), with extinction	DDL_SR_OPTIM_EXT	7	-645.6	1305.4	0.697	2.445	0.000	0.000	753	500	2.78	

Supplementary Table S5. Initial set of environmental variables tested.

Variable set (layers)	Description	Resolution	Source	Link
Annual Mean UVB1 (1)	Annual mean UV-B (UVB1) derived from the monthly mean covering the period of 2004-2013.	0.25º (15 arc min)	Ref[1]	https://www.ufz.de/gluv/index.php?en=32435
Aridity (1)	Represents average yearly precipitation divided by average yearly potential evapotranspiration. Lower values represent hyperarid areas and higher values humid areas.	0.16º (10 arc min)	FAO GeoNetwork	~~–G~~lobal map of aridity - 10 arc minutes" http://www.fao.org/geonetwork/srv/en/main.home
Bioclimatic variables (19: bio1–bio19)	Bioclimatic variables represent annual trends of temperature, and precipitation values.	0.0416º (2.5 arc min)	Ref[2], Ref[3]	https://chelsa-climate.org/downloads/
Habitat heterogeneity (5: variance, standard deviation, dissimilarity, coefficient of variation, contrast)	This layers quantify the spatial heterogeneity of global habitat based on the textural features of Enhanced Vegetation Index (EVI) imagery acquired by the Moderate Resolution Imaging Spectroradiometer (MODIS).	0.0416º (2.5 arc min)	Ref[4]	https://www.earthenv.org/texture
Length of Growing Period (LGP) (1)	The period during the year (in total days) when both moisture availability and temperature are conducive to crop growth. Thus, LGP refers to the number of days within the period of temperatures above 5°C when moisture conditions are considered adequate.	0.5º (30 arc min)	FAO GeoNetwork	~~–G~~lobal length of growing periods" http://www.fao.org/geonetwork/srv/en/main.home
Topographic (5: altitude, slope, roughness, profile/tangential curvature, topographic position index)	Terrain features based on the digital elevation model products of global 250 m GMTED2010 and near-global 90 m SRTM4.1dev	0.0416º (2.5 arc min)	Ref[5]	https://www.earthenv.org/topography
Vegetation coverage (3: forest land, grass/scrub/woodland, barren/very sparsely vegetated land)	Land cover layers represented by the percentage of cover of three vegetation types.	0.083º (5 arc min)	Ref[6]	http://webarchive.iiasa.ac.at/Research/LUC/External-World-soil-database/HTML/LandUseShares.html?sb=9

Ref[1] Beckmann M., Václavík T., Manceur A.M., Šprtová L., von Wehrden H., Welk E., Cord A.F. 2014 glUV: A global UV-B radiation dataset for macroecological studies. Methods Ecol. Evol. 5:372–383.

Ref[2] Fischer G., Nachtergaele F., Prieler S., van Velthuizen H.T., Verelst L., Wiberg D. 2008. *Global Agro-ecological Zones Assessment for Agriculture (GAEZ 2008)*. IIASA, Laxenburg, Austria and FAO, Rome, Italy.

Ref[3] Amatulli G., Domisch S., Tuanmu M.N., Parmentier B., Ranipeta A., Malczyk J., Jetz W. 2018. A suite of global, cross-scale topographic variables for environmental and biodiversity modeling. Sci. Data 5: 180040.

Ref[4] Karger D.N., Conrad O., Böhner J., Kawohl T., Kreft H., Soria-Auza R.W., Zimmermann N.E., Linder, H.P., Kessler M. 2017. Climatologies at high resolution for the earth's land surface areas. Sci. Data 4:170122.

Ref[5] Karger D.N., Conrad O., Böhner J., Kawohl T., Kreft H., Soria-Auza R.W., Zimmermann N.E., Linder H.P., Kessler M. 2017. Data from: Climatologies at high resolution for the earth's land surface areas. Dryad Digital Repository.

Ref[6] Tuanmu M.N., Jetz W. 2015. A global, remote sensing-based characterization of terrestrial habitat heterogeneity for biodiversity and ecosystem modeling. Glob. Ecol. Biogeogr. 24: 1329–1339.

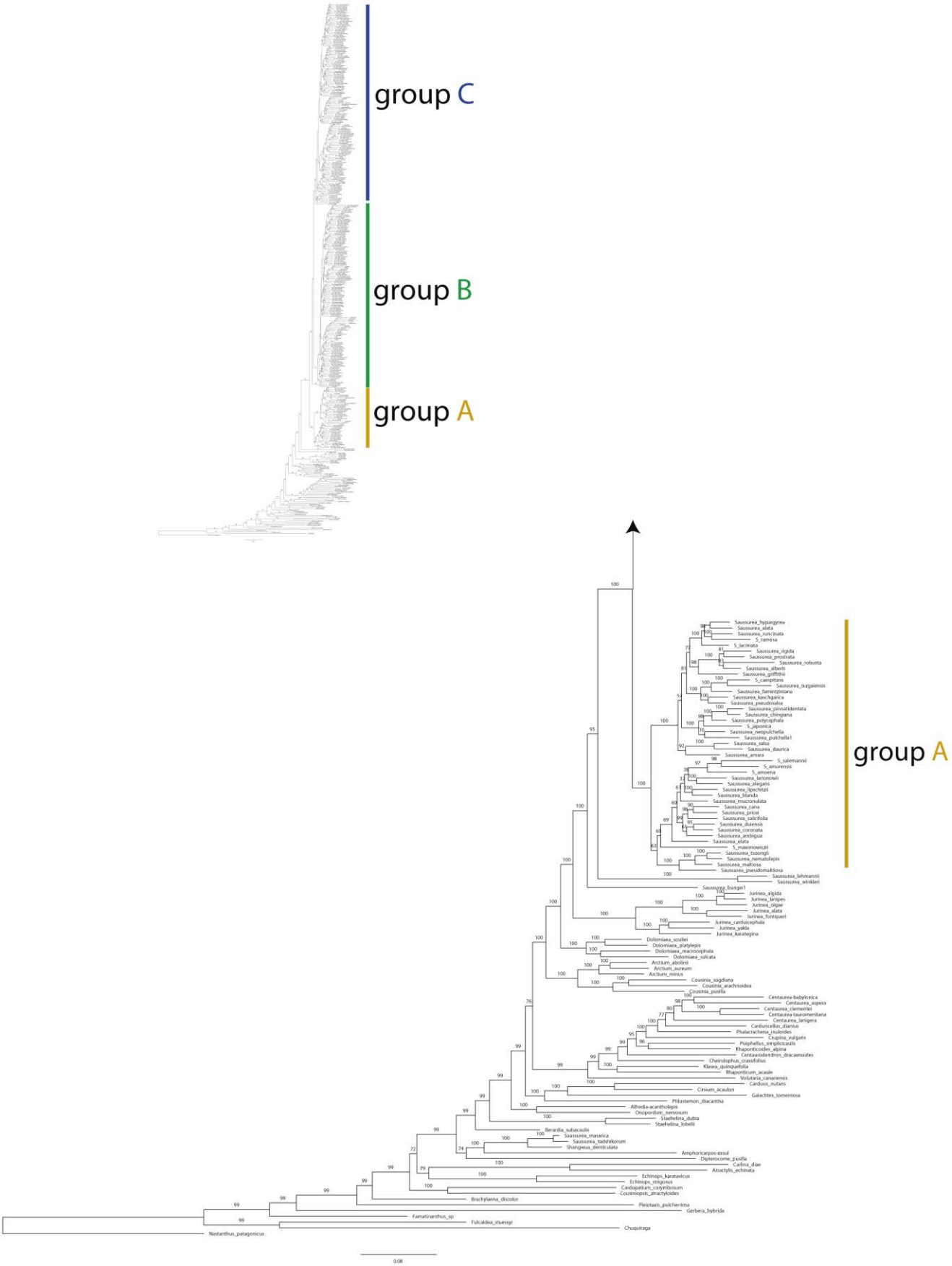
Supplementary Table S6. Percentage of variables contribution to the initial PCA perform to select variables from the initial set of 45 potential predictors. Final selected variables used for environmental niche analyses are marked in bold.

Variable	PC1	PC2	PC3	PC4	PC5
bio_12	7.14	0.48	1.49	0.81	1.30
lgp	6.80	0.09	5.0×10^{-2}	0.37	0.33
aridity	6.03	0.75	0.56	3.21	2.2×10^{-4}
bio_16	5.41	1.24	1.21	8.0×10^{-2}	7.57
bio_13	5.40	1.13	1.17	0.02	7.89
bio_6	5.33	2.62	1.27	2.32	3.31
bio_1	5.29	0.18	2.60	7.01	1.23
bio_17	4.84	0.74	1.34	5.01	4.64
bio_14	4.62	0.88	1.25	5.02	5.03
bio_2	4.58	0.03	5.7×10^{-4}	2.50	0.45
bio_18	4.43	1.26	1.21	0.04	10.16
bio_11	4.38	3.54	1.27	3.53	2.92
bio_9	4.33	2.55	0.92	1.83	7.94
bio_19	3.92	0.44	1.41	5.85	6.16
forest	3.91	1.43	8.0×10^{-3}	1.05	2.69
bio_10	3.59	3.32	2.33	5.68	1.4×10^{-2}
sparsely_v	3.45	0.01	1.05	0.36	5.61
bio_8	2.68	2.29	2.48	8.49	1.70
bio_5	2.17	4.76	2.51	5.79	0.29
bio_7	1.91	8.95	1.7×10^{-2}	1.2×10^{-4}	2.19
grass	1.82	4.69	0.18	1.9×10^{-4}	1.24
cv_h	1.56	1.39	4.39	1.11	2.13
elevation	1.43	9.53	0.89	0.24	3.1×10^{-3}
bio_15	1.40	1.48	4.3×10^{-2}	7.58	10.19
bio_4	0.89	10.33	1.5×10^{-2}	0.15	3.15
tpi	0.69	1.97	0.29	0.03	2.5×10^{-4}
tcurv	0.63	0.47	0.79	0.01	0.40
roughness	0.40	5.07	3.59	2.86	1.00
slope	0.34	4.84	3.60	2.99	0.98
dissim_h	0.17	0.35	10.51	5.41	2.20
UVB1	0.12	11.63	9.3×10^{-4}	0.77	2.0×10^{-3}
bio_3	0.10	10.15	7.6×10^{-2}	1.70	1.72
std_h	0.04	0.68	10.60	5.47	0.89
contrast_h	0.03	0.16	10.57	6.18	1.02
variance_h	0.003	0.40	10.52	6.46	0.46

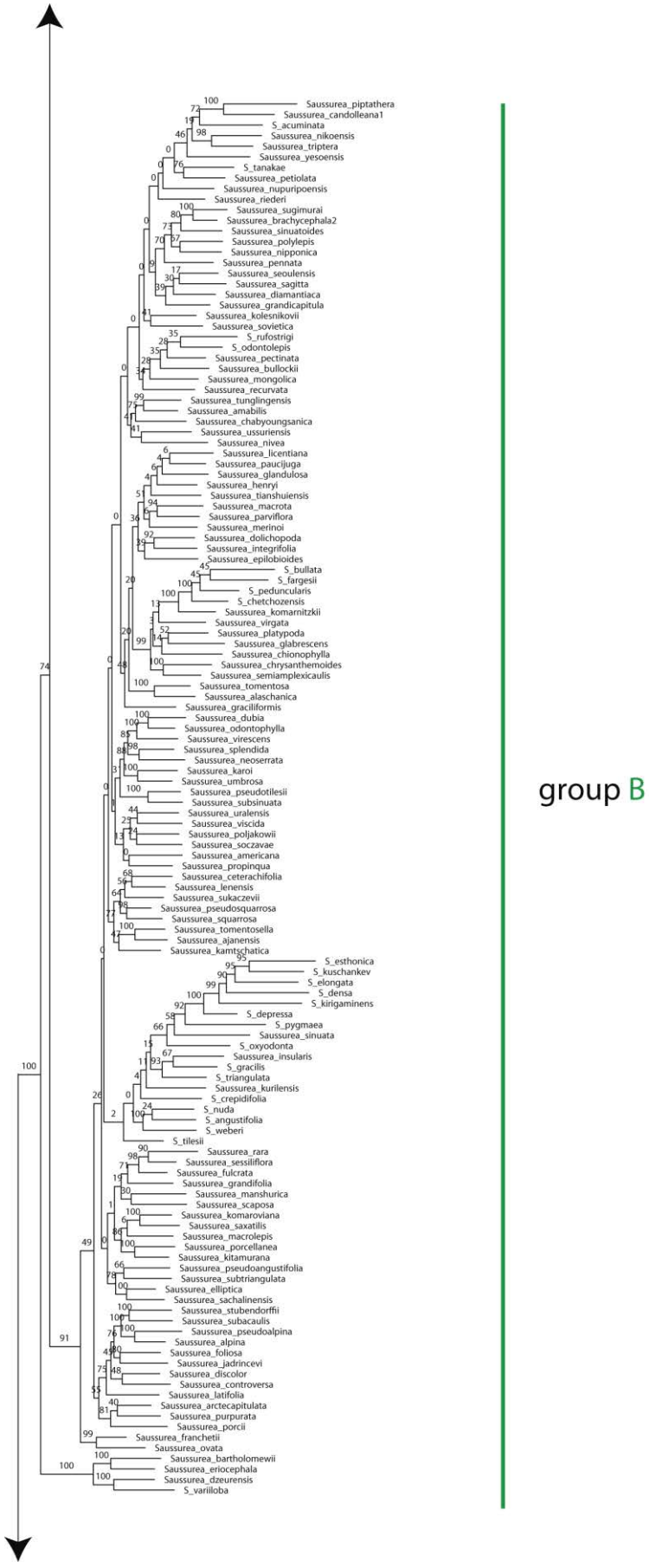
Supplementary Table S7. Percentage of variables contribution to each axis of the PCA analysis with the selected set of variables.

Variable	PC1	PC2	PC3	PC4	PC5
aridity	11.66	0.23	3.17	0.161	0.46
forest	9.52	0.07	4.45E-04	4.8×10^{-2}	0.72
bio_17	9.21	0.08	1.21E+01	1.16	0.02
lgp	8.77	4.24	4.01	0.18	0.12
bio_2	8.43	2.00	2.55E-01	1.66	0.40
bio_12	7.86	5.91	2.01	0.26	0.08
bio_10	6.72	2.30	1.73E+01	5.5×10^{-4}	1.09
elevation	6.50	11.21	3.46	1.16	0.26
grass	5.68	5.42	1.66E-01	2.33	2.80
sparsely_v	4.98	2.27	1.45E+01	1.35	1.42
bio_15	4.68	0.51	1.39E+01	10.34	0.01
cv_h	4.15	0.23	6.07E-01	10.59	28.26
bio_6	3.66	10.52	6.92	0.10	4.04
UVB1	3.36	17.31	4.97E-01	0.98	1.04
tpi	2.45	2.47	2.61	10.43	25.43
tcurv	1.85	0.27	2.62	20.68	16.61
bio_7	0.27	21.91	1.88E-02	4.7×10^{-2}	1.73
dissim_h	0.17	1.03	3.65	10.99	13.14
roughness	0.003623	11.93	1.21E+01	0.25	2.28

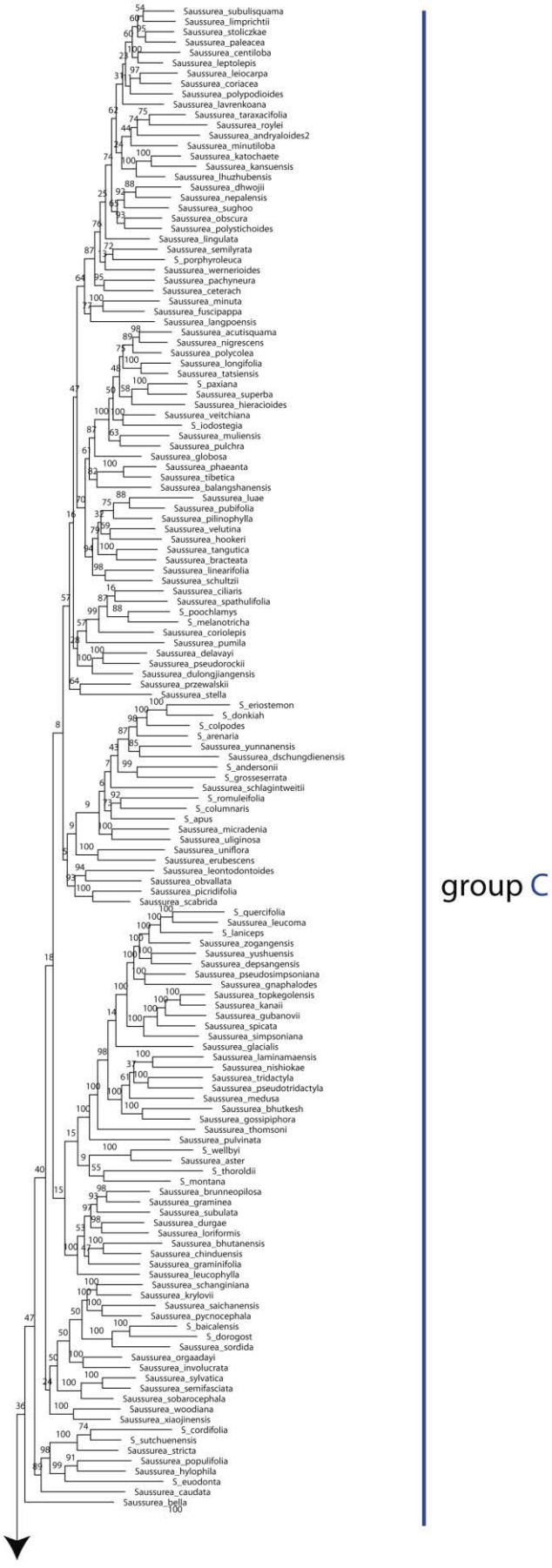
Supplementary figures



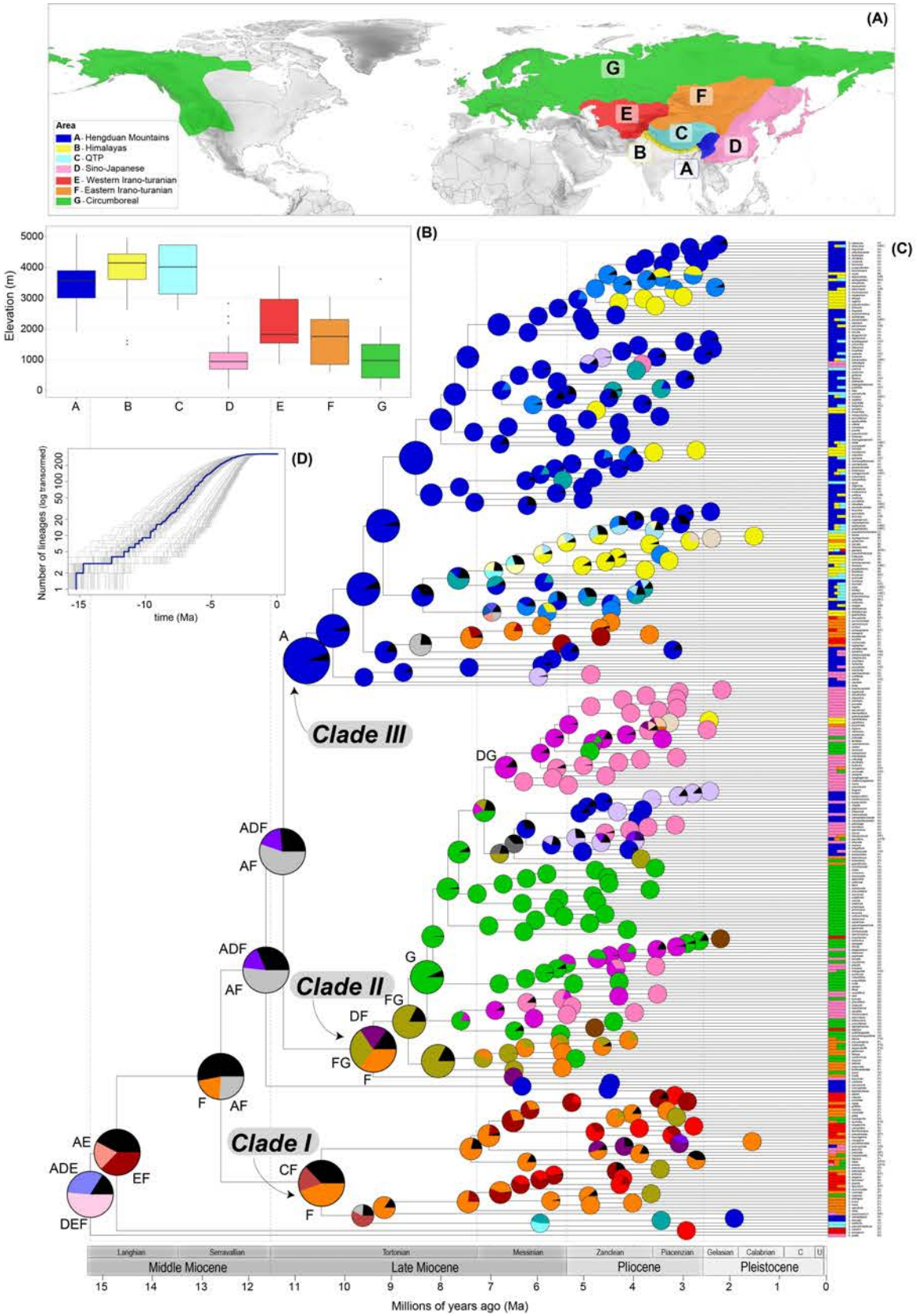
Supplementary Figure S1. Phylogenetic tree inferred from the concatenation approach with the unfiltered dataset. Bootstrap (BS) support values are shown over branches.



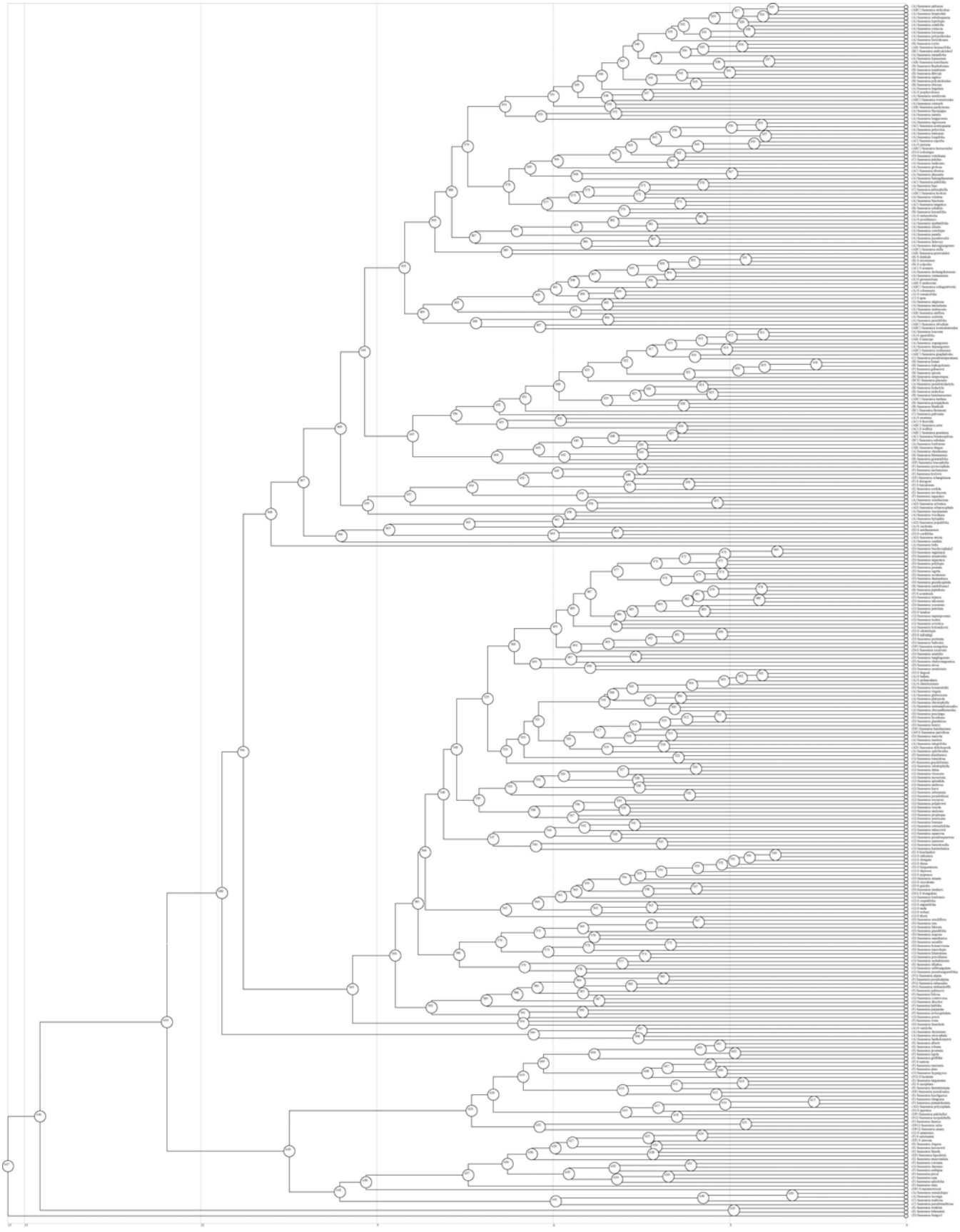
Cont. Supplementary Figure S1. Phylogenetic tree inferred from the concatenation approach with the unfiltered dataset. Bootstrap (BS) support values are shown over branches.



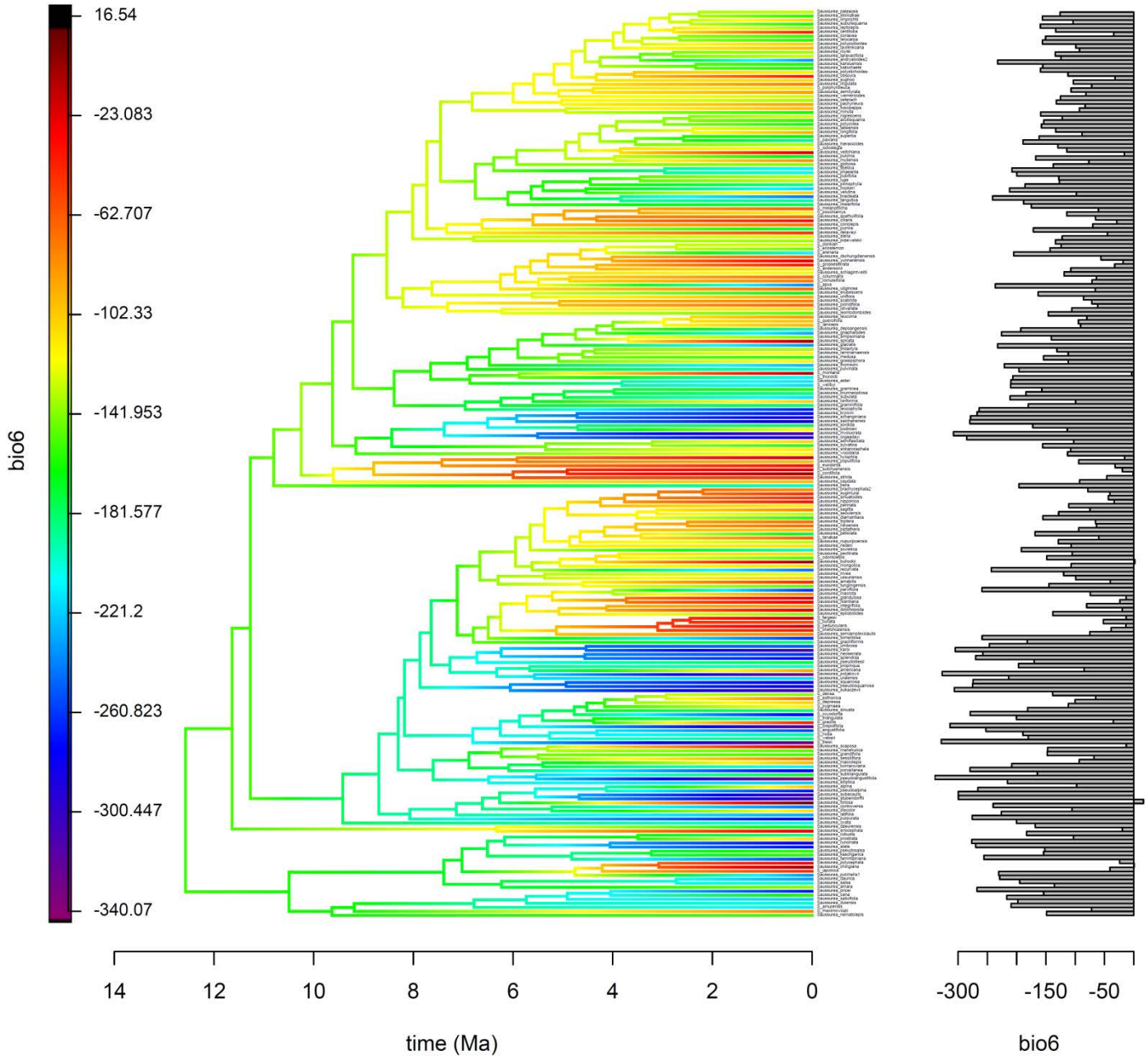
Cont. Supplementary Figure S1. Phylogenetic tree inferred from the concatenation approach with the unfiltered dataset. Bootstrap (BS) support values are shown over branches.



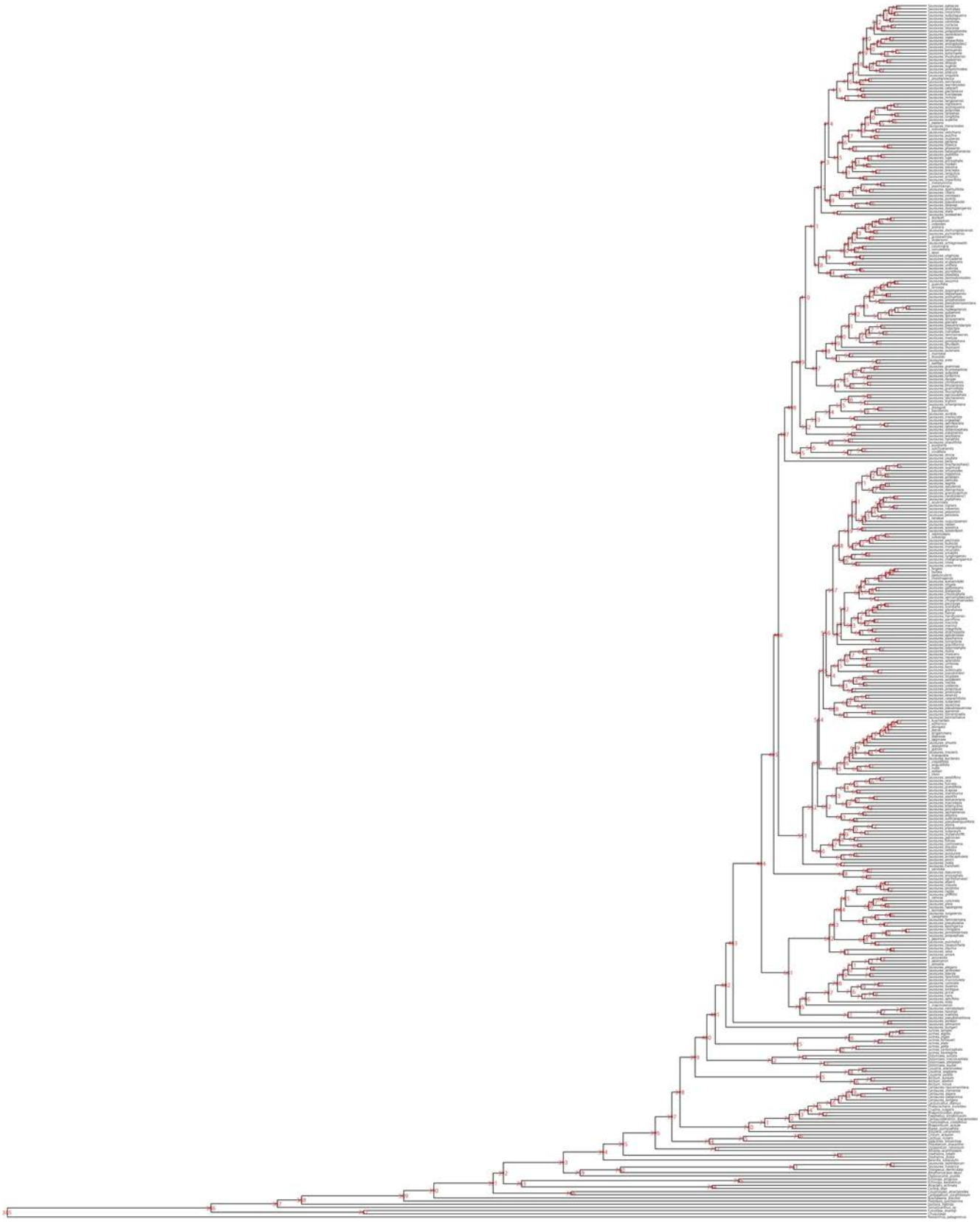
Supplementary Figure S2. (A) Distribution of *Saussurea* species across the seven biogeographic areas. **(B)** Graphical representation by boxplots of endemic species on the areas along altitudinal gradient. **(C)** Phylogenetic reconstruction of *Saussurea* diversification inferred from the supermatrix of 489 COS loci (total of 158,106 bp) by maximum likelihood analysis (in RaxML), time-calibrated (in treePL), and reconstructed ancestral biogeographic areas (in RASP implementing BioGeoBEARS). Colors of tree nodes indicated the most likely estimation of ancestral areas. Tips are also colored according their assigned regions. Abbreviations of areas for inferred ancestral states are only specified for the main tree backbone nodes, for information about the rest of nodes and all color correspondence see Supplementary Table S2 and Supplementary Fig. S4. **(D)** Lineage-through-time plot (LTT) that indicates the number of lineages emerged (log-transformed) in a temporal line. Blue line represents the maximum clade credibility (MCC) tree chronogram and the grey ones correspond to 100 time-calibrated trees.



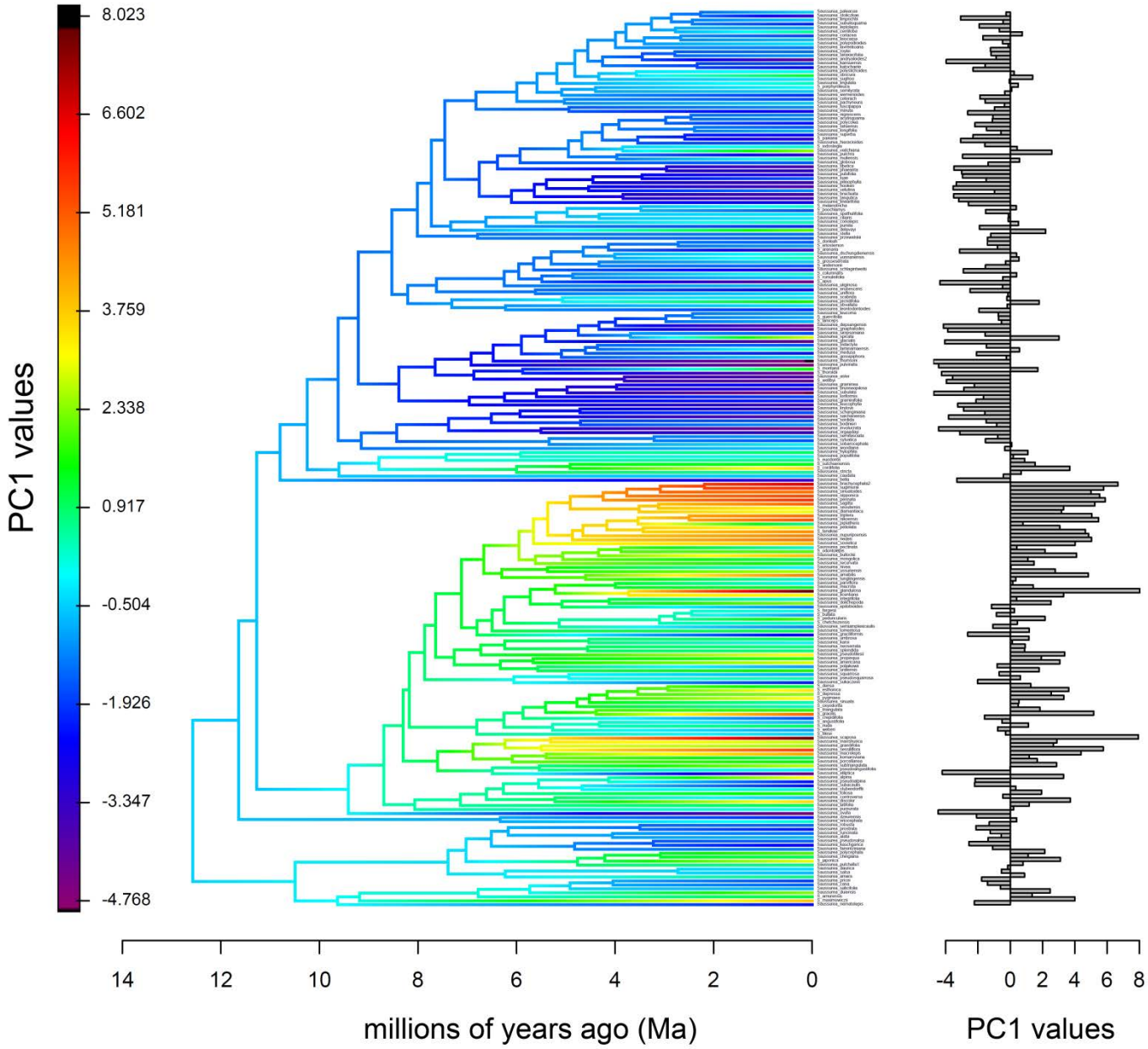
Supplementary Figure S3. Time-calibrated phylogeny and biogeographic inference with node IDs corresponding to biogeographic analysis with BioGeoBEARS implemented in RASP. Percentage values of each estimate area for each node are in Supplementary Table S2.



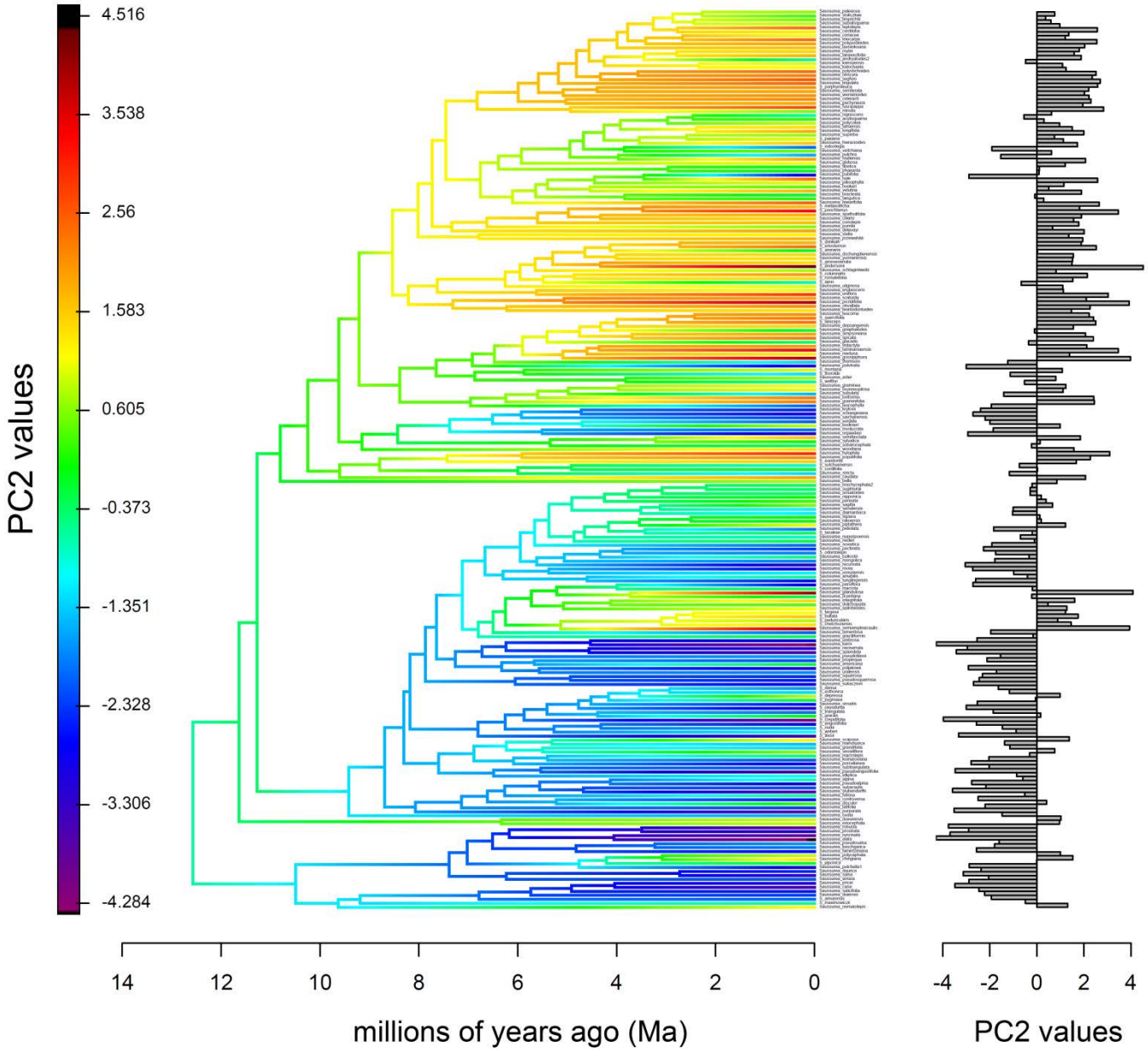
Supplementary Figure S4. Ancestral state reconstructions (obtained by *fasAnc* function of R package *phytools*) of the bio6 values (minimum temperature of the coldest month; ºC) for each species. Loading values for each species of bio6 are represented in vertical barplots besides time-calibrated phylogeny.



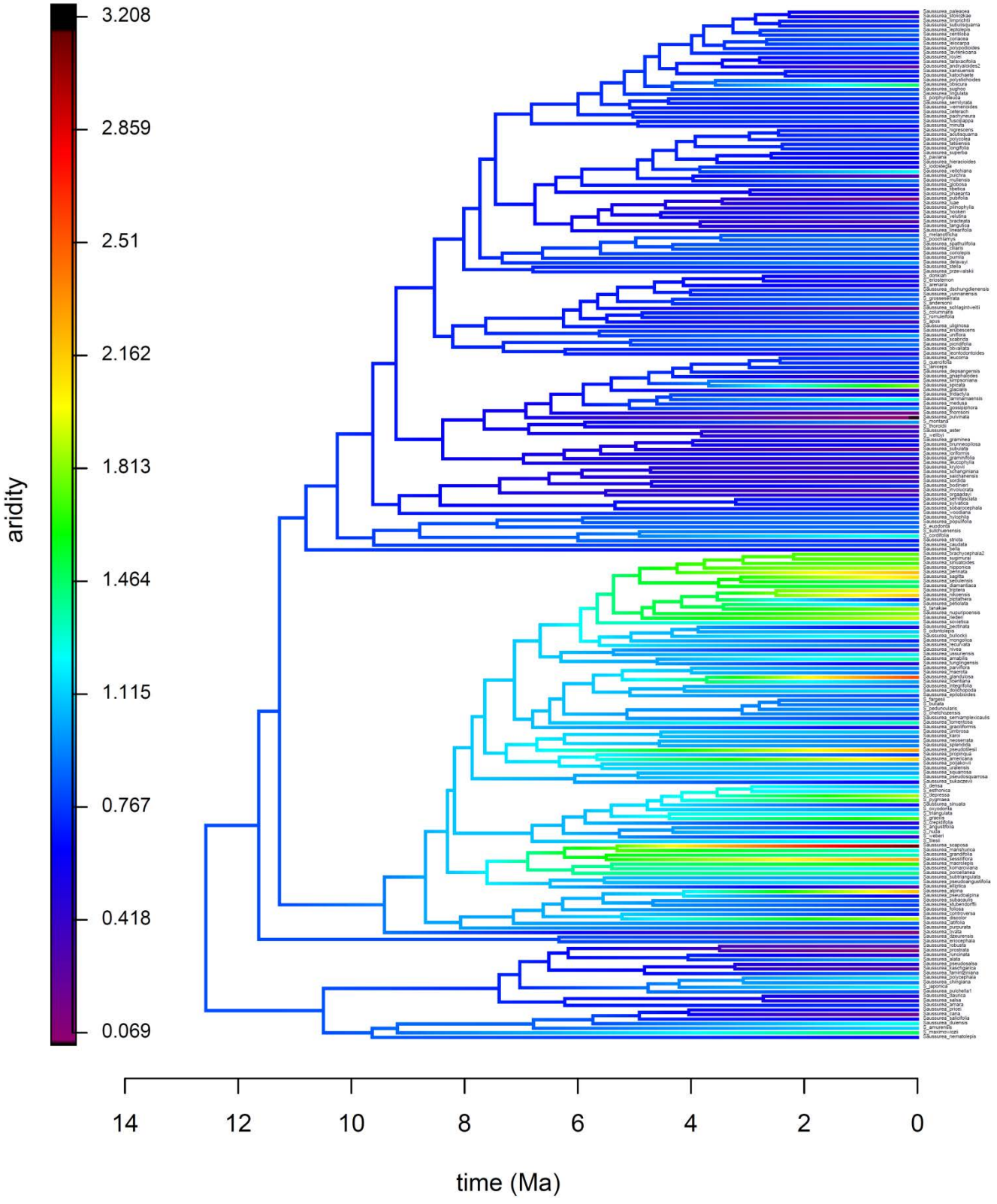
Supplementary Figure S5. Best Maximum Likelihood (ML) phylogenetic tree with node IDs corresponding to dating analysis with treePL. See Supplementary Table S3 for node IDs, median age (Ma), lower and upper 95% CI.



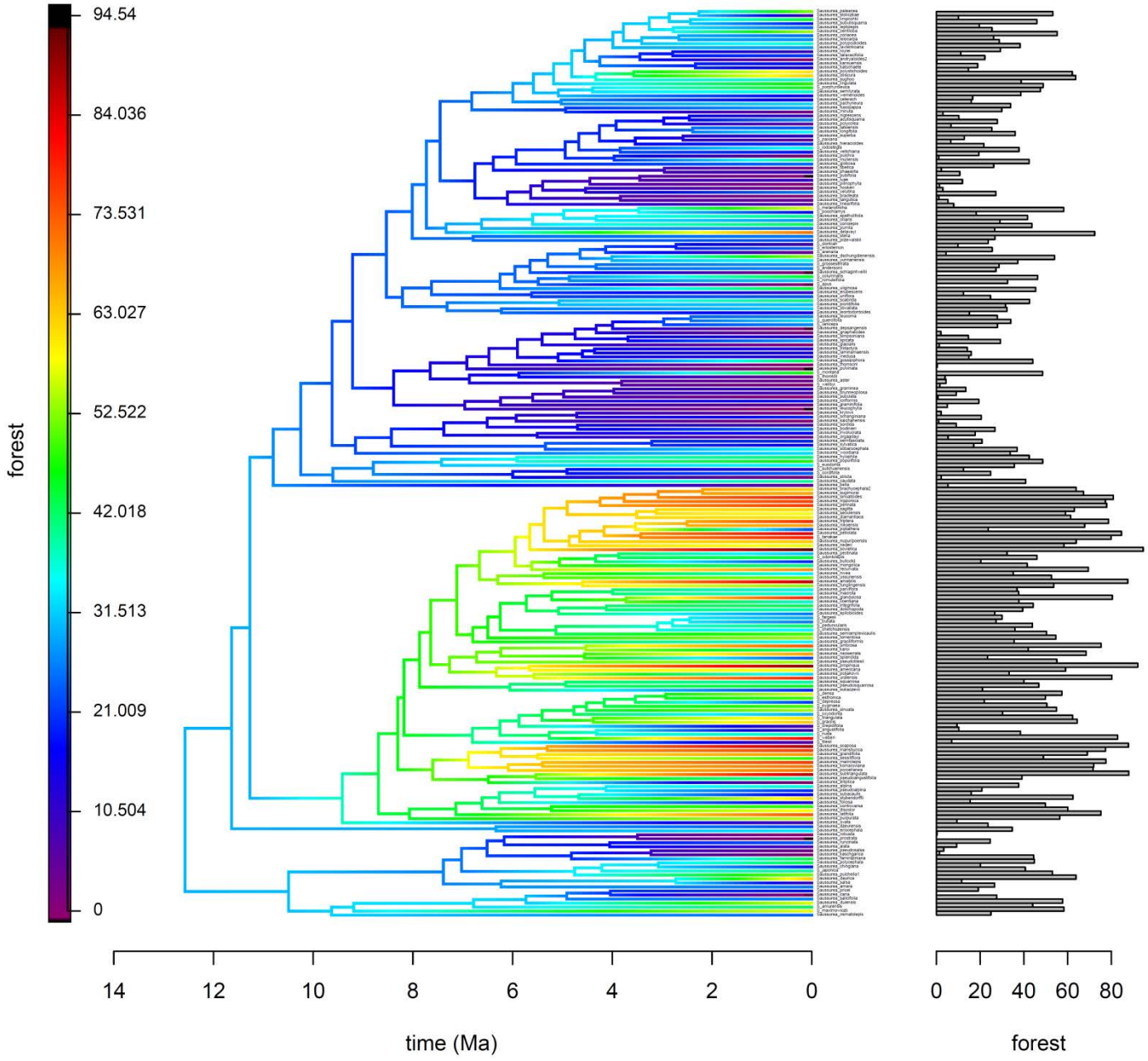
Supplementary Figure S6. Ancestral state reconstructions (obtained by *fasAnc* function of R package *phytools*) of PC1 scores for each species, representing the ecological niche evolution along the phylogenetic history of *Saussurea*. Loading values for each species of PC1 are represented in vertical barplots besides time-calibrated phylogeny.



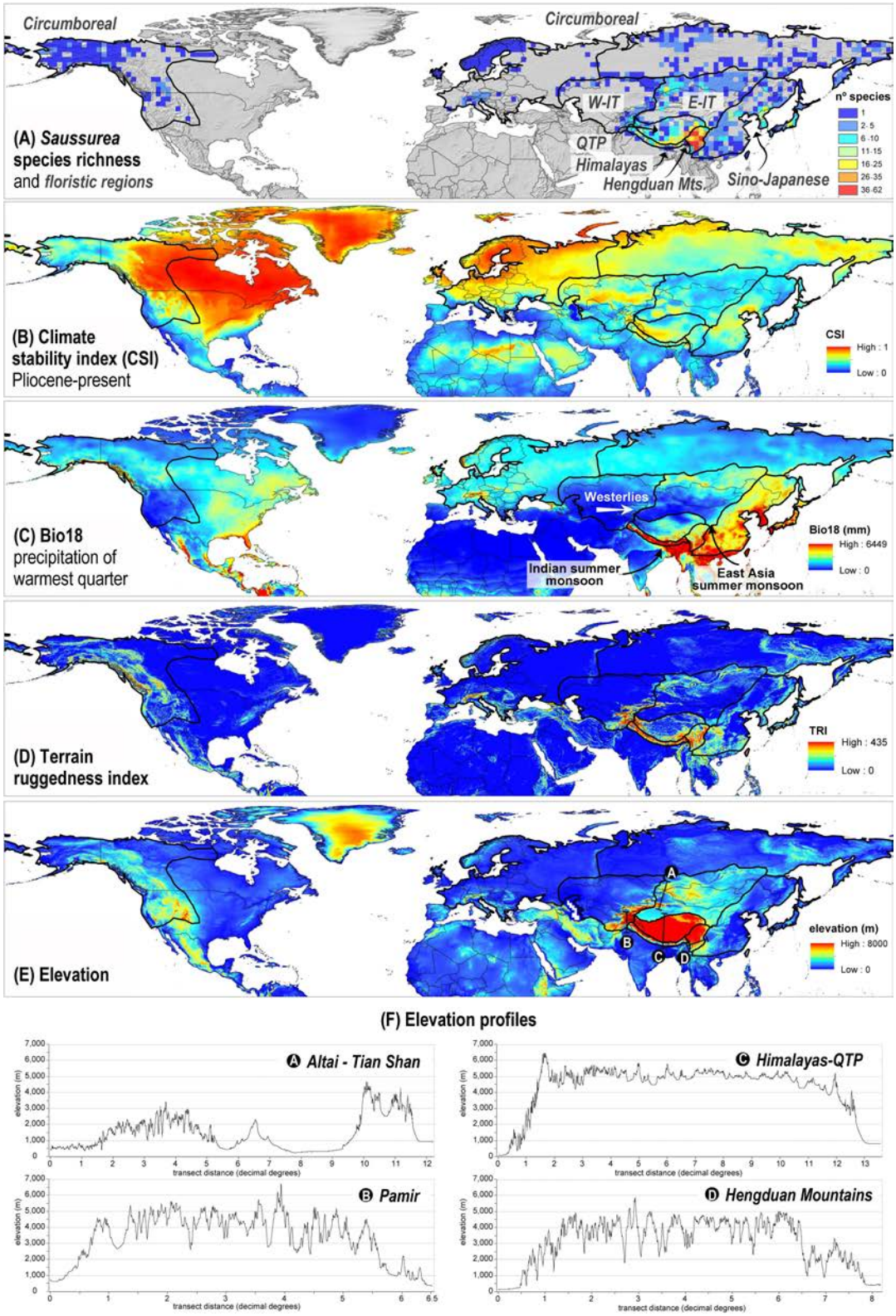
Supplementary Figure S7. Ancestral state reconstructions (obtained by *fasAnc* function of R package *phytools*) of PC2 scores for each species, representing the ecological niche evolution along the phylogenetic history of *Saussurea*. Loading values for each species of PC2 are represented in vertical barplots besides time-calibrated phylogeny.



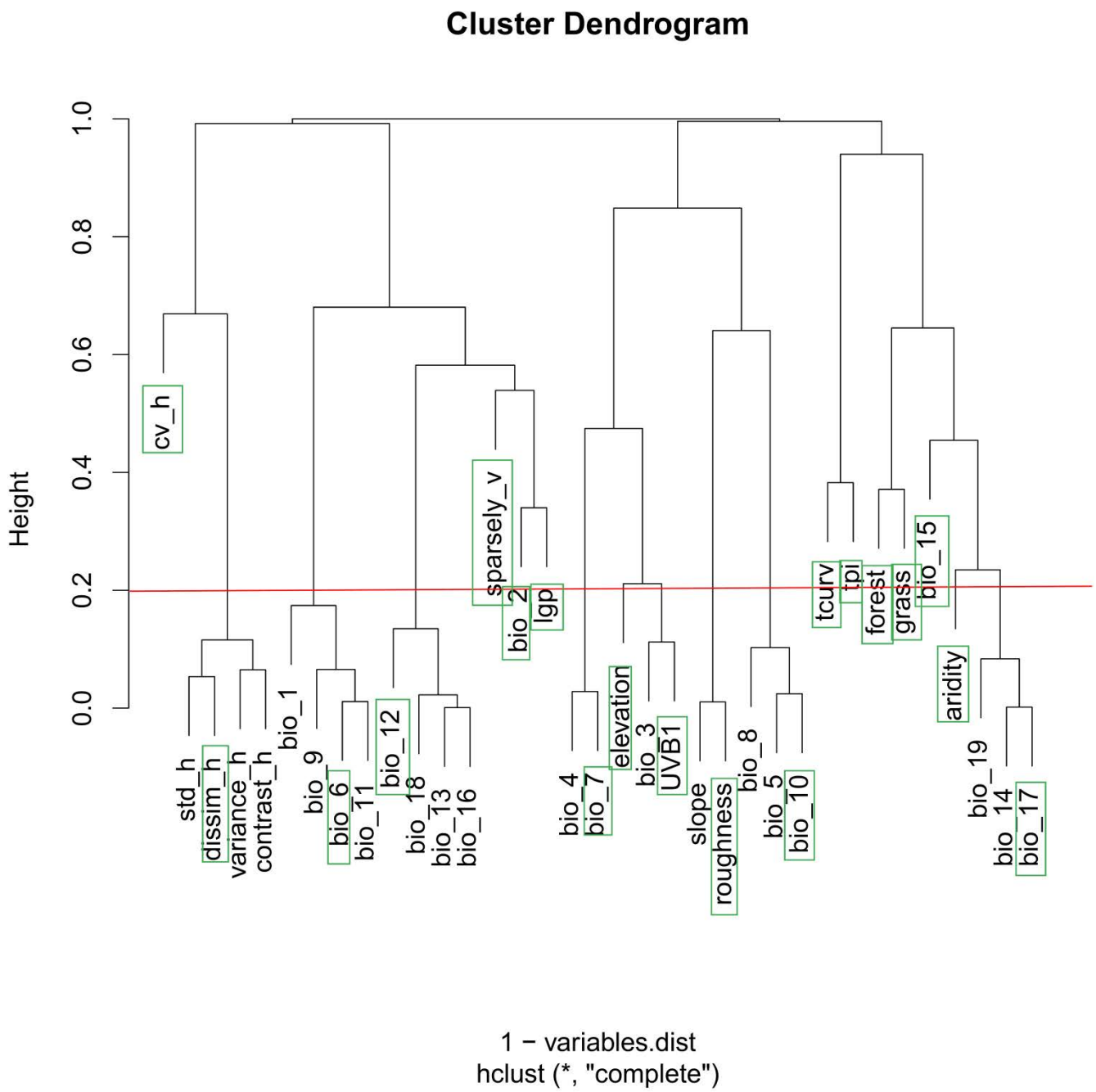
Supplementary Figure S8. Ancestral state reconstructions (obtained by *fasAnc* function of R package *phytools*) of the aridity index values for each species. Loading values for each species of aridity are represented in vertical barplots besides time-calibrated phylogeny.



Supplementary Figure S9. Ancestral state reconstructions (obtained by *fasAnc* function of R package *phytools*) of the forest coverage (in percentatge) for each species. Loading values for each species of forest coverage are represented in vertical barplots besides time-calibrated phylogeny.



Supplementary Figure S10. (A) Distribution of estimated *Saussurea* species richness in squares of 2 × 2 decimal degrees. The seven floristic regions where *Saussurea* is present are showed by black contoured polygons in each of presented maps in this figure. **(B)** Mapped values of climatic stability index (CSI; Herrando-Moraira et al., in prep.) for the study area from Pliocene (3.3 Ma) to present. Colors range from blue for low standard deviation (SD) values, which represent areas with low climatic fluctuations from Pliocene to present, to red for high SD values, which show areas where high climatic fluctuations probably took place. **(C)** Mapped values of rainfall during the warmest quarter of the year (bio18 from *Worldclim 2* database; Fick and Hijmans, 2017). Type and direction of monsoons are also indicated in map. **(D)** Mapped values of terrain ruggedness index (TRI; Karger et al., 2017). **(E)** Mapped values of elevation above sea level (Karger et al., 2017). In black lines are marked the selected transects to outline the elevation profiles. **(F)** Elevation gradients showing an overview of altitudinal variations through the defined transects. Profile plots were extracted from the Global Multi-resolution Terrain Elevation Data 2010 (GMTED2010) at 7.5 arc-seconds resolution obtained with the *Create Profile Graph* tool from *3D Analyst toolbar* of ArcGIS v.10.2.2 (Esri, Redlands, California, USA 2014).



Supplementary Figure S11. Dendrogram of values from Pearson correlation analysis of initial set of 35 environmental variables. The final 19 variables retained are marked in a green square.

Discusión general

La investigación llevada a cabo en la presente tesis doctoral ha resultado en avances científicos que contribuyen al conocimiento de la sistemática y evolución de plantas en varios niveles:

- metodológico;
- taxonómico;
- evolutivo; y
- biogeográfico.

A continuación se discuten aquellos resultados más relevantes que han sido obtenidos a lo largo del estudio.

Nivel metodológico

Desempeño del panel COS loci para reconstrucción filogenética

¿Qué se ha hecho?

Se ha confirmado la adecuación de la técnica de secuenciación de alto rendimiento *Hyb-Seq* usando el panel diseñado para la familia *Compositae* (1061 COS loci; Mandel et al., 2014) al estudio de la tribu *Cardueae*, y se han comparado varios métodos de extracción de datos (**CAPÍTULO 1)**. A continuación, se ha aplicado esta aproximación a varias escalas taxonómicas: tribu y subtribu (**CAPÍTULO 2**), relaciones entre géneros (**CAPÍTULO 3)** y entre especies cercanas (**CAPÍTULOS 1, 5 y 6**).

¿Cuáles son las nuevas aportaciones?

Se confirma que el panel COS loci *Compositae* proporciona información filogenética para un amplio rango taxonómico dentro de las Compuestas, una de las familias más ricas en especies de todas las Angiospermas. Su aplicación ha servido para esclarecer clasificaciones anteriores problemáticas (ver nivel 2 taxonómico) y, en cuanto a análisis, para explorar qué métodos proporcionan resultados filogenéticos más robustos. Con la inspección de varios enfoques bioinformáticos, se ha establecido un protocolo óptimo para obtener topologías más estables y valores de apoyo de rama más altos.

En el **CAPÍTULO 1** se probó por primera vez el método de extracción de secuencias HybPiper (Johnson et al., 2016), el cual permitió recuperar un 99% de los loci en vez de cerca del 60% que se conseguía con el software PHYLUCE (Mandel et al., 2014, 2015; Faircloth, 2015). A partir de nuestro estudio, los siguientes trabajos de la presente tesis doctoral (**CAPÍTULOS 2, 3, 5 y 6**), así como otros estudios realizados en Compuestas (p.ej. Jones et al., 2019; Watson et al., 2020) utilizaron HybPiper. Se concluye que, para la extracción de secuencias *target* o regiones específicas derivadas de *Hyb-Seq*, la estrategia que minimiza los datos faltantes y los potenciales efectos del ruido filogenético consiste utilizar métodos que primero mapeen las lecturas frente a las regiones *target* y luego realicen el ensamblaje de estas lecturas previamente mapeadas, como hacen HybPiper u otros (aTRAM, Allen et al., 2015; HybPhyloMarker, Fér and Schmickl, 2018).

Aunque los COS loci fueron diseñados para recuperar genes de baja copia y aparentemente ortólogos, se ha visto en los **CAPÍTULOS 1, 2, 3, 5 y 6** que éstos pueden contener loci con potenciales parálogos, cuyo número varía según los táxones muestreados en cada estudio. Como patrón general, cuantas más especies se han incluído y éstas eran más distantes filogenéticamente, mayor número de loci se han identificado como potencialmente afectados por paralogía. En los **CAPÍTULOS 3, 5 y 6** se optó por la eliminación manual de los loci potencialmente afectados por paralogía señalados por HybPiper, aunque por omisión este software permita incluirlos. En futuros estudios, este procedimiento debería automatizarse para optimizar el tiempo empleado en este paso bioinformático.

También se ha comprobado, en el grupo de estudio utilizado, que la aproximación bajo coalescencia es mucho más robusta para la reconstrucción filogenética que la de concatenación (**CAPÍTULO 1**), en congruencia con evidencias previas reportadas por otros autores (Kubatko & Degnan, 2007; Liu et al., 2015). De hecho, ya hay estudios basados en *Hyb-Seq* que únicamente están empleando la aproximación bajo coalescencia, como Siniscalchi et al. (2021).

Con los resultados presentados en esta tesis doctoral se ha podido comprobar hasta qué punto la técnica *Hyb-Seq*, y específicamente los COS loci, son útiles para el estudio de la evolución de géneros altamente diversificados, como son *Saussurea* (400–500 especies) y *Jurinea* (200 especies). En general, resulta una técnica útil para establecer cuáles son los grandes linajes que se pueden reconocer sobre la base de altos valores estadísticos de apoyo (**CAPÍTULOS 3, 5 y 6**). Sin embargo, para clados más "someros", es decir, relaciones entre especies filogenéticamente cercanas, la precisión es variable. Se han recuperado clados bien apoyados, pero, por el contrario, también hay otros con valores de apoyo muy bajos y con una topología incongruente entre las aproximaciones de concatenación y coalescencia. Este último hecho es especialmente reseñable en el caso de *Jurinea*. En el **CAPÍTULO 1**, la falta de apoyo estadístico para las relaciones entre especies de este género se atribuyó al posible efecto de la falta de muestreo, pero en el **CAPÍTULO 5** se comprobó que, aun con la inclusión de gran parte de su diversidad, el grupo que forma la radiación en las montañas de Asia Central sigue quedando no resuelto.

Desafortunadamente, el pre-tratamiento o filtrado de las secuencias basado en la eliminación de posiciones de rápida evolución (Fragoso-Martínez et al., 2017), que pusimos a prueba en el **CAPÍTULO 1**, tampoco ha ayudado a la resolución de las ramas no apoyadas (**CAPÍTULO 5**). Este es el motivo por el cual, a falta de futuros estudios, se desestima como técnica útil para resolver radiaciones recientes y rápidas.

Aunque con la secuenciación tipo *Hyb-Seq* es posible recuperar prácticamente el genoma completo de los cloroplastos, como se hizo en el **CAPÍTULO 2**, éste no se volvió a analizar en estudios posteriores debido a: (1) Un bajo poder resolutivo. Se obtuvo una señal filogenética mucho más débil, por ejemplo, un 5,5% de posiciones informativas con los datos de regiones codificantes del cloroplasto comparado con un 38% con los datos de los COS loci; y (2) Una mayor propensión o sensibilidad en mostrar los efectos del ILS (*incomplete lineage sorting*) o eventos de evolución reticulada, mostrando agrupaciones sesgadas por posibles eventos pasados de captura cloroplástica, como la agrupación de *Saussurea* con *Arctium-Cousinia* en vez de *Jurinea* (ver **CAPÍTULO 2**); y (3) topologías incongruentes entre los géneros del complejo *Saussurea-Jurinea* (ver **CAPÍTULO 3**) si se compara el análisis de todo el genoma cloroplástico *vs.* solo las regiones codificantes de proteínas (Xu et al., 2019; Zhang et al., 2019).

Por último, en cuanto a la técnica, es importante destacar su alta efectividad, no solo con muestras frescas o secadas con gel de sílice, sino también con aquellas procedentes de herbario, algunas de ellas muy antiguas. Por ejemplo, se han llegado a secuenciar más de 10 muestras del siglo XIX, la más antigua del 1836: *Jurinea chaetocarpa* del Altai, Rusia, conservada en el herbario de Viena (W0022735). Otros autores también han mostrado que usando muestras de individuos de entre 40–120 años de antigüedad no se producen errores de secuenciación debido al estado de conservación y la antigüedad de las muestras (Hart et al., 2016; Brewer et al., 2019; Forrest et al., 2019). Este hecho está siendo sin duda una de las mayores ventajas de la técnica *Hyb-Seq* para estudios de sistemática molecular.

Diseño de un índice de estabilidad climática

¿Qué se ha hecho?

Dada la importancia que han tenido, y seguirán teniendo en el futuro, los cambios climáticos como moldeadores de la biodiversidad, se ha diseñado un índice de estabilidad climática a escala global que permite localizar las zonas más y menos estables en términos de variaciones de temperatura y precipitación (**CAPÍTULO 4**).

¿Cuáles son las nuevas aportaciones?

Con la emergente publicación de reconstrucciones del clima pasado cada vez más antiguas y de mayor resolución (p.ej. Brown et al., 2018; Gamisch, 2019), cartografiar y conocer cuáles son las zonas climáticamente más estables del planeta era una rama que faltaba por explorar dentro del ámbito de las ciencias de la tierra. Por ello, en los últimos años varios estudios han propuesto estimaciones de la variabilidad climática a escala global (Fordham et al., 2019; Brown et al., 2020a, 2020b).

Hemos descrito un índice de estabilidad climática (ver **CAPÍTULO 4**) que presenta mejoras respecto a estos trabajos previos en los siguientes aspectos: (1) El rango temporal: los anteriores abarcan solo desde el presente hasta 21.000 años atrás, mientras que el índice que aquí presentamos permite llegar hasta 3,3 Ma atrás en el Plioceno. Además, permite estimaciones de la estabilidad climática en el rango pasado-presente y también en el presente-futuro, teniendo en cuenta nueve modelos diferentes de predicciones futuras y cuatro escenarios socioeconómicos de cambio global (SSP); (2) La resolución espacial: pasamos de 2,5 º (278 Km) a 2,5 arc-min (5 Km); (3) El uso de variables: pasamos de utilizar solo una variable (temperaturas medias mensuales) o dos (temperaturas y precipitaciones medias mensuales) a utilizar 10–12 variables, dependiendo del conjunto de datos; estas variables recogen no solo métricas referentes a la media, sino también valores máximos o mínimos de temperatura y precipitación o su estacionalidad; y (4) La adaptabilidad: se trata de un índice adaptable, es decir, según los criterios y objetivos de los trabajos que quieran utilizarlo se pueden seleccionar unas variables u otras que también se muestran disponibles ya de manera independiente: 14 variables bioclim para el conjunto pasado-presente y 19 para el conjunto presente-futuro. Tener este tipo de estimaciones de la variación paleoclimática es útil y aplicable para muchas ramas de la ciencia, como el que se ha aplicado aquí en el caso de la historia evolutiva y biogeográfica de los géneros de plantas (**CAPÍTULOS 5** y **6**).

Nivel taxonómico

Nuevas clasificaciones taxonómicas en distintos niveles

¿Qué se ha hecho?

Se han construido árboles filogenéticos con altos valores de apoyo de sus ramas para grupos extremadamente complejos y diversos: (1) las subtribus dentro de las *Cardueae* (**CAPÍTULO 2**); (2) el complejo *Saussurea-Jurinea* (**CAPÍTULO 3**); y (3) las especies de *Jurinea* y *Saussurea* (**CAPÍTULOS 5** y **6**). La clasificación de estos grupos había supuesto un desafío constante para los botánicos especialistas, dada la baja resolución de las filogenias Sanger y la falta de muestreos completos. Estos dos factores limitantes se han mejorado aquí, empleando la secuenciación masiva y recopilando muestras, por un lado, mediante la colaboración con botánicos especialistas de las diferentes regiones geográficas, y por otro, mediante visitas a los principales herbarios con grandes colecciones del grupo de estudio (p.ej. Edimburgo, E; Moscú, MW; o Viena, W).

¿Cuáles son las nuevas aportaciones?

En términos generales se han secuenciado un total de 856 táxones dentro del grupo de estudio. Al provenir los datos de COS loci extraídos mediante la técnica *Hyb-Seq*, podrán ser reutilizados para reconstrucciones filogenéticas futuras dentro del grupo de estudio y la familia de las Compuestas.

Los trabajos que se muestran en los **CAPÍTULOS 2** y **3** fueron los primeros para la familia de las Compuestas en los que se formalizan propuestas taxonómicas basadas en filogenias inferidas con los COS loci. En ellos se remarca la gran utilidad de esta aproximación metodológica para fines taxonómicos, teniendo presentes estudios morfológicos previos.

En cuanto a las relaciones entre complejos y subtribus dentro de las *Cardueae*, en el **CAPÍTULO 1** se pudo confirmar por primera vez que el grupo *Arctium* y el grupo *Saussurea* constituyen dos complejos separados, y a su vez son grupos hermanos: por un lado, *Arctium-Cousinia*, y por otro lado *Saussurea-Jurinea*. En el **CAPÍTULO 2** se muestra cómo el salto a la secuenciación masiva (de cinco marcadores con secuenciación Sanger a 1142, incluyendo los COS loci y los marcadores cloroplásticos, en nuestro trabajo actual) ha hecho posible una nueva clasificación subtribal de las *Cardueae*, que pasan de cinco a 12 subtribus. El tratamiento taxonómico previo más aceptado (Susanna & Garcia-Jacas, 2007, 2009) arrastraba el problema de cómo delimitar la subtribu parafilética *Carduinae*, que albergaba más del 70% de la diversidad de la tribu (1700 especies). Los autores reconocían agrupaciones morfológicas como grupos informales, pero las filogenias Sanger no mostraban suficiente apoyo estadístico para ayudar a confirmarlas y reconocerlas formalmente. Ahora, las antiguas *Carduinae* sí han podido ser fragmentadas en ocho subtribus monofiléticas con un amplio apoyo: *Carduinae* s.str., *Dipterocominae*, *Xerantheminae*, *Berardiinae*, *Staehelininae*, *Onopordinae*, *Arctiinae* (antiguo complejo *Arctium-Cousinia*) y *Saussureinae* (antiguo complejo *Saussurea-Jurinea*). En este trabajo también se ha presentado una nueva datación molecular de la tribu. Por esta razón, además de las aportaciones taxonómicas, nuestros resultados también pueden ser útiles para futuros estudios de sistemática y evolución en la tribu *Cardueae* que

precisen de un marco de evolución temporal o puntos de calibración secundaria para la datación de filogenias.

En lo tocante a la delimitación genérica y las relaciones entre géneros cercanos, en el **CAPÍTULO 3** se realizó una evaluación de la monofilia y la validez taxonómica de los géneros que incluye la recientemente descrita subtribu *Saussureinae*. Este trabajo constituye el primer estudio filogenético en el que se secuencian miembros de todos los géneros satélite descritos en el complejo *Saussurea-Jurinea*, un total de 15 géneros segregados. De todos ellos, solo se recuperaron tres clados principales, que se proponen reconocer como géneros *Dolomiaea*, *Saussurea* y *Jurinea*, congruentes con hipótesis morfológicas anteriores. Los representantes de los demás géneros quedarían incluidos en alguno de estos tres clados, y por lo tanto en su sinonimia, según corresponda. Basándonos en estos resultados, se presenta una descripción general para cada uno de estos tres géneros según la nueva circunscripción propuesta y una clave dicotómica de los caracteres que los diferencian, y se formalizan nuevas combinaciones para las especies de los géneros satélite que son transferidas a alguno de estos tres géneros. Cabe destacar que durante el transcurso y publicación de este estudio surgieron dos trabajos con el mismo objetivo de esclarecer los límites del complejo (Szukala et al., 2019; Kasana et al., 2020). Sin embargo, como en casos anteriores, el muestreo de los géneros satélite era insuficiente y las filogenias se basaban solamente en cuatro marcadores moleculares (ITS, ETS, *trnK/matK, trnL-F*; Szukala et al., 2019) o solo en uno (ITS en Kasana et al., 2020).

Por último, en los **CAPÍTULOS 5** y **6** se presentan las primeras filogenias de alta cobertura taxonómica de los géneros *Jurinea* (incluyendo 187 especies) y *Saussurea* (324 especies), con cerca del 77% y 70%, respectivamente, de su diversidad muestreada. Aunque ambos estudios no tengan un enfoque taxonómico *per se*, los resultados filogenéticos que se muestran pueden servir como base para futuros trabajos centrados en su complicada y muy discutida clasificación infragenérica y seccional. Durante el transcurso de la tesis doctoral, surgieron varios estudios que presentaban filogenias de ambos géneros, pero con muestreos más limitados y un menor número de marcadores moleculares: en el caso de *Saussurea*, filogenias realizadas con el genoma cloroplástico completo, como los trabajos de Xu et al. (2019) con 136 especies y Zhang et al. (en revisión) con 199 especies, y en *Jurinea* con 81 especies y cuatro marcadores Sanger (Szukala et al., 2019).

Nivel evolutivo

Historia evolutiva de la diversificación de *Jurinea* y *Saussurea*

¿Qué se ha hecho?

Se han investigado dos de las grandes radiaciones de especies dentro de la tribu *Cardueae* y de la familia *Compositae*, los géneros *Jurinea* (200 especies; **CAPÍTULO 5**) y *Saussurea* (400–500 especies; **CAPÍTULO 6**). El estudio se ha realizado mediante la extracción de datos filogenómicos (técnica *Hyb-Seq* con los COS loci), morfológicos (en el caso de *Jurinea*), y ecológicos y paleoclimáticos. Los puntos fuertes de estos dos trabajos son: (1) Muestreo extensivo de especies; (2) Apoyo robusto de los principales

clados, que da como resultado filogenias estadísticamente respaldadas; (3) Árboles filogenéticos calibrados en una escala temporal; (4) Clasificación fenotípica basada en formas vitales con diferentes adaptaciones al medio (para *Jurinea*); y (5) Recopilación de registros geolocalizados de las especies para reconstrucciones de nicho climático presente y pasado.

¿Cuáles son las nuevas aportaciones?

Los estudios de radiaciones presentados aquí (**CAPÍTULOS 5** y **6**) pueden servir como referencia metodológica para resolver otras radiaciones de géneros de plantas o animales hiperdiversos cuya exploración previa con métodos moleculares Sanger no haya dado buena resolución. El flujo de trabajo seguido en la presente tesis doctoral ha dado como resultado una visión general de cuál podría haber sido la historia evolutiva de ambos géneros (Tabla 4), para *Jurinea* en el **CAPÍTULO 5**, y para *Saussurea* en el **CAPÍTULO 6**. Ambos estudios representan los primeros muy completos realizados en estos grupos de especies, los cuales abordan desde aspectos biogeográficos hasta evidencias de cuándo surgieron la mayor parte las especies y qué factores pudieron desencadenar esa explosión.

Tabla 4. Resumen de los principales aspectos evolutivos que se hipotetizan para cada género objeto de estudio.

Cuestiones evolutivas	*Jurinea*	*Saussurea*
Origen temporal y área ancestral	Mioceno tardío hace 10,7 Ma (9,5–12,0 Ma) y el área ancestral se situaría alrededor de la región de la Mesopotamia y la meseta iraní.	Mioceno medio hace 15,3 Ma (11,4–17,5 Ma) y área ancestral incierta, las reconstrucciones muestran un origen probable alrededor de la región iranoturania y las montañas Hengduan.
Grupos monofiléticos principales	Principalmente se segregaron tres clados durante la transición Mioceno-Plioceno; por regiones biogeográficas: (1) este de Asia; (2) región circumboreal; (3) Asia central.	Principalmente se segregaron tres clados durante el Tortoniano (Mioceno tardío; 11,6–7,3 Ma), por regiones biogeográficas: (1) iranoturanio; (2) circumboreal y sino-japonesa; (3) Tibet-Himalaya-Hengduan.
Cambio en la tasa de diversificación y factores propuestos como desencadenantes	Cambio en la tasa de diversificación hacia los 3 Ma (transición Plioceno-Pleistoceno), cuando prácticamente la tasa se dobló (λ1= 0,93 spp/Myr, λ2=1,77 spp/Myr). Como desencadenante, factores extrínsecos, el cambio climático hacia condiciones más frías y áridas. Una pre-adaptación a este tipo de condiciones habría facilitado su supervivencia y diversificación.	Cambio en la tasa de diversificación hacia los 6 Ma (Mioceno tardío), cuando la tasa prácticamente se dobló (λ1= 0,63 spp/Myr, λ2=1,14 spp/Myr). Tres principales factores extrínsecos: (1) aumento de la actividad tectónica, coincidiendo con uno de los levantamientos de las montañas Hengduan; (2) intensificación del monzón del Este Asiático; (3) notable descenso de las temperaturas.
Modo de especiación predominante	Alopátrica sería el principal, aunque también divergencia ecológica local a lo largo de gradientes altitudinales.	Alopátrica sería el principal, aunque también divergencia ecológica local a lo largo de gradientes altitudinales.

Cambio de nicho ecológico a escala regional	Divergencia de nicho hacia zonas más húmedas y/o de baja altitud con mayor cobertura arbórea, tanto en la región Circumboreal como en las zonas más al límite de su distribución hacia el este de Ásia.	Divergencia de nicho hacia zonas más húmedas y de menor altitud con mayor cobertura arbórea, tanto en la región Circumboreal como en la Sino-Japonesa.
Efecto de las oscilaciones climáticas	La notable estabilidad climática detectada en la meseta iraní explicaría su alta riqueza de especies. Las fluctuaciones moderadas en las regiones circumboreal y de Asia Central fomentarían la especiación alopátrica.	Una fluctuación climática de baja a moderada podría haber contribuido a un aumento de la diversificación en las regiones montañosas escarpadas y de amplios gradientes altitudinales.

Para ambos géneros se han detectado dos patrones evolutivos comunes relacionados con el clima que podrían haber contribuido a su éxito evolutivo, en cuanto a la alta riqueza de especies que observamos en el presente. Los paralelismos encontrados entre *Jurinea* y *Saussurea* serían:

(1) Los cambios climáticos pasados, relacionados con **descensos importantes de la temperatura y precipitación** (aumento de la aridez y condiciones frías), que habrían contribuido a la formación de numerosas especies y linajes en el Hemisferio Norte. Aquí hemos encontrado evidencias de que esto habría sucedido en ambos géneros, un hecho que concuerda con casos previamente documentados, como el orden *Saxifragales* (Folk et al., 2019). Tanto en *Jurinea* como en *Saussurea*, la conservación del nicho ancestral, presumiblemente adaptado ya a este tipo de ambientes según las reconstrucciones de nicho (hipótesis de pre-adaptación; Wiens et al., 2010), habría proporcionado una ventaja evolutiva respecto a otros linajes adaptados a condiciones climáticas anteriores más húmedas y cálidas, que no habrían podido prosperar en el nuevo entorno climático de enfriamiento global. Para cada género se detectó un notable incremento en su diversificación, de casi el doble en ambos casos, pero en tiempos geológicos distintos. La explosión en la diversificación habría sucedido primero en *Saussurea* (**CAPÍTULO 6**), en el Mesiniano hacia los 6 Ma (Mioceno tardío), y posteriormente en *Jurinea* (**CAPÍTULO 5**), en la transición Plioceno-Pleistoceno hacia los 3 Ma.

(2) Las **oscilaciones climáticas moderadas** han podido contribuir al incremento de la diversidad en regiones alpinas actuando como motor de especiación, tal y como habían sugerido previamente Mosbrugger et al. (2018). Una de las hipótesis evolutivas tradicionales postula que la estabilidad climática continuada en el tiempo promueve la conservación de las especies y linajes (Pianka, 1966; Marin & Hedges, 2016). Por ello, las regiones a escala global que han sido climáticamente muy estables en los últimos millones de años son hoy en día muy diversas, como las regiones tropicales (Sosa & Loera, 2017). Gracias a la creación del índice de estabilidad climática presentado en el **CAPÍTULO 4**, se ha podido observar que las regiones más diversas de *Jurinea* (**CAPÍTULO 5**) y *Saussurea* (**CAPÍTULO 6**) no son las climáticamente más estables, sino que la riqueza de especies se maximiza en las zonas de moderada inestabilidad. De acuerdo con la idea propuesta por Mosbrugger et al. (2018), hemos podido comprobar que, en zonas montañosas como las regiones de Asia Central y las montañas del Himalaya y Hengduan, un cierto nivel de variación en el clima podría haber determinado que las especies que se dispersaron a lo largo del gradiente altitudinal pudieran establecerse en su óptimo ecológico, diferenciándose de otras poblaciones mediante alopatría a

pequeña escala y especialización a los recursos y las condiciones locales (aislamiento geográfico y divergencia ecológica).

Por último, queremos recalcar la importancia de la obtención de material, el cual sería uno de los puntos clave para el buen desarrollo de estudios enfocados en la diversificación de las especies. En la presente tesis doctoral, gran parte de las especies que se han incluido ha sido obtenidas gracias a los testigos identificados por botánicos especialistas depositados en herbarios. La inclusión y colaboración con investigadores locales de la zona iranoturania y China especializados en ambos géneros también ha sido un elemento imprescindible para la obtención de material y la interpretación de los resultados.

Nivel biogeográfico

Patrones biogeográficos y evolutivos en las regiones iranoturania y Tibet-Himalaya-Hengduan

¿Qué se ha hecho?

Se ha explorado dónde y cuándo las especies de *Jurinea* y *Saussurea* pudieron divergir en un contexto geográfico y temporal, mediante métodos de inferencia biogeográfica y contraste de hipótesis de diversificación. Con los resultados obtenidos se contribuye al conocimiento de la historia evolutiva de la biota, particularmente en dos áreas hiperdiversas, pero aún poco exploradas, como son la región Iranoturania (en el caso *Jurinea*; **CAPÍTULO 5**) y el Tibet-Himalaya-Hengduan (en el caso de *Saussurea*; **CAPÍTULO 6**).

¿Cuáles son las nuevas aportaciones?

Para cada una de las regiones se corroboran algunas de las hipótesis que han sido previamente formuladas sobre la evolución de su flora, y también se aportan algunas nuevas sugeridas a lo largo de los **CAPÍTULOS 5** y **6**, tomando como modelos de estudio *Jurinea* (Tabla 5) y *Saussurea* (Tabla 6). Cabe destacar que, como paralelismo entre ambas regiones y para ambos géneros, las especies habrían establecido un núcleo de diversificación en la región Circumboreal. Ésta habría sido colonizada de manera exitosa, ya que aproximadamente el 30% de la diversidad de *Jurinea* y el 20% de la diversidad de *Saussurea* se encuentra actualmente en la región Circumboreal.

Tabla 5. Compilación de las hipótesis evaluadas y formulación de otras nuevas en la región iranoturaniana en el estudio de diversificación de *Jurinea*.

Región biogeográfica Iranoturania	
Hipótesis	**Detalles**
El origen de los elementos xéricos típicos iranoturanios se situaría en el Mioceno medio o tardío (Lauterbach et al. 2019; Peterson et al. 2019).	**Confirmación** de una hipótesis previamente propuesta. Para *Jurinea* se infirió un origen alrededor de 10,7 Ma (9,5–12,0 Ma).
La mayor diversificación de elementos xéricos iranoturanios se remontaría al Plioceno-Pleistoceno (Manafzadeh et al. 2017; Moharrek et al., 2019; Mahmoudi Shamsabad et al., 2020)	**Confirmación** de una hipótesis previamente propuesta. Para *Jurinea* se estima que la tasa de diversificación se habría prácticamente doblado alrededor de los 3 Ma.
La región iranoturania constituyó una fuente importante de linajes xerofíticos, no solo hacia regiones adyacentes como la Mediterránea como ya había sido previamente documentado (Manafzadeh et al. 2017), sino también hacia otras regiones de Eurasia como la Circumboreal.	**Nueva** hipótesis propuesta aquí.
La elevada riqueza de especies y la persistencia de linajes antiguos en la meseta iraní se explicaría, en parte, por la elevada estabilidad climática desde el Plioceno hasta la actualidad.	**Nueva** hipótesis propuesta aquí.
Las oscilaciones climáticas moderadas habrían estimulado la diversificación alopátrica y ecológica en las montañas de Asia Central.	**Nueva** hipótesis propuesta aquí.

Tabla 6. Compilación de las hipótesis evaluadas y formulación de nuevos factores a considerar en la región Tibet-Himalaya-Hengduan en el estudio de diversificación de *Saussurea*.

Región biogeográfica Tibet-Himalaya-Hengduan	
Hipótesis	**Detalles**
La región habría sido el punto de origen de dispersiones de lineajes pre-adaptados al frío hacia el resto del Hemisferio Norte, conocida como "out-of-Tibet or out-of-QTP" (Wen et al., 2014; Wang et al., 2015)	**Confirmación** de una hipótesis previamente propuesta.
La tasa de diversificación de elementos alpinos templados en la región habría aumentado durante el Mioceno tardío (Ding et al., 2020).	**Confirmación** de una hipótesis previamente propuesta.
Los procesos que habrían contribuido a la explosión de especies en la región serían el levantamiento de las montañas (Xing & Ree, 2017), el aumento de las lluvias monzónicas (Ding et al., 2020; Stokstad, 2020) y el enfriamiento del clima (Meng et al., 2017).	**Confirmación** de una hipótesis previamente propuesta.
La alta diversidad florística encontrada hoy en día en la región se explicaría, además de por los factores históricos (levantamientos de montañas y cambios paleoclimáticos), por cuatro factores adicionales (*mountain geo-diversity*; Mosbrugger et al., 2018; Muellner-Riehl, 2019): zonificación pronunciada, oscilaciones climáticas, terreno accidentado-abrupto y dinámica de alta conectividad.	**Confirmación** de una hipótesis previamente propuesta y **nuevo factor** a añadir en la formulación de la hipótesis: una precipitación moderada (400–800 mm) durante el trimestre o período más cálido del año.

Referencias bibliográficas

Allen, J.M., Huang, D.I., Cronk, Q.C., Johnson, K.P. 2015. ATRAM - automated target restricted assembly method: a fast method for assembling loci across divergent taxa from next-generation sequencing data. *BMC Bioinformatics* 16: 98.

Brewer, G.E. et al. 2019. Factors affecting targeted sequencing of 353 nuclear genes from herbarium specimens spanning the diversity of angiosperms. *Frontiers in Plant Science* 10: 1102.

Brown, J.L., Hill, D.J., Dolan, A.M., Carnaval, A.C., Haywood, A.M. 2018. PaleoClim, high spatial resolution paleoclimate surfaces for global land areas. *Scientific Data* 5: 1–9.

Brown, S.C., Wigley, T.M., Otto-Bliesner, B.L., Rahbek, C., Fordham, D.A. 2020b. Persistent Quaternary climate refugia are hospices for biodiversity in the Anthropocene. *Nature Climate Change* 10: 244–248.

Brown, S.C., Wigley, T.M.L., Otto-Bliesner, B.L., Fordham, D.A. 2020a. StableClim, continuous projections of climate stability from 21000 BP to 2100 CE at multiple spatial scales. *Scientific Data* 7: 335.

Ding, W.N., Ree, R.H., Spicer, R.A., Xing, Y.W. 2020. Ancient orogenic and monsoon-driven assembly of the world's richest temperate alpine flora. *Science* 369: 578–581.

Faircloth, B.C. 2015. PHYLUCE is a software package for the analysis of conserved genomic loci. *Bioinformatics* 32: 786–788.

Fér, T., Schmickl, R.E., 2018. HybPhyloMaker: Target enrichment data analysis from raw reads to species trees. *Evolutionary Bioinformatics* 14: 1–9.

Folk, R.A. et al. 2019. Rates of niche and phenotype evolution lag behind diversification in a temperate radiation. *Proceedings of the National Academy of Sciences USA* 116: 10874–0882.

Fordham, D.A., Brown, S.C., Wigley, T.M., Rahbek, C. 2019. Cradles of diversity are unlikely relics of regional climate stability. *Current Biology* 29: R356–R357.

Forrest, L.L., Hart, M.L., Hughes, M., Wilson, H.P., Chung, K.F., Tseng, Y.H., Kidner, C.A. 2019. The limits of Hyb-Seq for herbarium specimens: impact of preservation techniques. *Frontiers in Ecology and Evolution* 7: 439.

Fragoso-Martínez, I. et al. 2017. A pilot study applying the plant Anchored Hybrid Enrichment method to New World sages (*Salvia* subgenus *Calosphace*; Lamiaceae). *Molecular Phylogenetics and Evolution* 117: 124–134.

Gamisch, A. 2019. Oscillayers: A dataset for the study of climatic oscillations over Plio-Pleistocene time-scales at high spatial-temporal resolution. *Global Ecology and Biogeography* 28: 1552–1560.

Hart, M.L. et al. 2016. Retrieval of hundreds of nuclear loci from herbarium specimens. *Taxon* 65: 1081–1092.

Johnson, M.G. et al. 2016. HybPiper: Extracting coding sequence and introns for phylogenetics from high-throughput sequencing reads using target enrichment. *Applications in Plant Science* 4: 1600016.

Jones, K. E. et al. 2019. An empirical assessment of a single family-wide hybrid capture locus set at multiple evolutionary timescales in Asteraceae. *Applications in Plant Sciences* 7: e11295.

Kasana, S., Dwivedi, M.D., Uniyal, P.L., Pandey, A.K. 2020. An updated circumscription of *Saussurea* (Cardueae, Asteraceae) and allied genera based on morphological and molecular data. *Phytotaxa* 450: 173–187.

Kubatko, L.S., Degnan, J.H. 2007. Inconsistency of phylogenetic estimates from concatenated data under coalescence. *Systematic Biology* 56: 17–24.

Lauterbach, M., Veranso-Libalah, M.C., Sukhorukov, A.P., Kadereit, G. 2019. Biogeography of the xerophytic genus *Anabasis* L. (Chenopodiaceae). *Ecology and Evolution* 9: 3539–3552.

Liu, L., Xi, Z., Wu, S., Davis, C.C., Edwards, S.V. 2015. Estimating phylogenetic treesfrom genome-scale data. *Annals of the New York Academy of Sciences* 1360: 36–53.

Mahmoudi Shamsabad, M., Moharrek, F., Assadi, M., Nieto Feliner, G. 2020. Biogeographic history and diversification patterns in the Irano-Turanian genus *Acanthophyllum* s.l. (Caryophyllaceae). *Plant Biosystems* 155: 425–435. https://doi.org/10.1080/11263504.2020.1756974

Manafzadeh, S., Staedler, Y.M., Conti, E. 2017. Visions of the past and dreams of the future in the Orient: the Irano-Turanian region from classical botany to evolutionary studies. *Biological Reviews* 92: 1365–1388.

Mandel, J.R. et al. 2014. A target enrichment method for gathering phylogenetic information from hundreds of loci: An example from the Compositae. *Applications in Plant Science* 2: 1300085.

Marin, J., Hedges, S.B. 2016. Time best explains global variation in species richness of amphibians, birds and mammals. *Journal of Biogeography* 43: 1069–1079.

Meng, H.H., Su, T., Gao, X.Y., Li, J., Jiang, X.L., Sun, H., Zhou, Z.K. 2017. Warm-cold colonization: response of oaks to uplift of the Himalaya-Hengduan Mountains. *Molecular Ecology* 26: 3276–3294.

Moharrek, F., Sanmartín, I., Kazempour-Osaloo, S., Nieto Feliner, G. 2019. Morphological innovations and vast extensions of mountain habitats triggered rapid diversification within the species-rich Irano-Turanian genus *Acantholimon* (Plumbaginaceae). *Frontiers in Genetics* 9: 698.

Mosbrugger, V., Favre, A., Muellner-Riehl, A. N., Päckert, M., Mulch, A. 2018. Cenozoic evolution of geo-biodiversity in the Tibeto-Himalayan region. In: Hoorn, C., Perrigo, A., Antonelli, A. (eds.). Mountains, Climate and Biodiversity. Hoboken, NJ: Wiley-Blackwell, 429–448.

Muellner-Riehl, A.N. 2019. Mountains as evolutionary arenas: patterns, emerging approaches, paradigm shifts, and their implications for plant phylogeographic research in the Tibeto-Himalayan region. *Frontiers in Plant Science* 10: 195.

Peterson, A., Harpke, D., Peterson, J., Harpke, A., Peruzzi, L. 2019. A pre-Miocene Irano-Turanian cradle: Origin and diversification of the species-rich monocot genus *Gagea* (Liliaceae). *Ecology and Evolution* 9: 5870–5890.

Pianka, E.R. 1966 Latitudinal gradients in species diversity: a review of concepts. *American Naturalist* 100: 33–46.

Siniscalchi, C. et al. 2021. Lineage-specific vs. universal: a comparison of the Compositae1061 and Angiosperms353 enrichment panels in the sunflower family. *Applications in Plant Science* (in press).

Sosa, V., Loera, I. 2017. Influence of current climate, historical climate stability and topography on species richness and endemism in Mesoamerican geophyte plants. *PeerJ* 5: e3932.

Stokstad, E., 2020. Mountains and monsoons created Tibetan biodiversity. *Science* 369: 493.

Susanna, A., Garcia-Jacas, N. 2007. Tribe *Cardueae* Cass. (1819). Pp. 123–146. En: Kadereit, J.W., Jeffrey, C. (eds.), *The families and genera of vascular plants*, vol. 8. Springer, Berlin & Heidelberg.

Susanna, A., Garcia-Jacas, N. 2009. *Cardueae* (Carduoideae). Pp. 293–313. En: Funk, V.A., Susanna, A., Stuessy, T.F. Bayer, R.J. (eds.), *Systematics, evolution, and biogeography of Compositae*. International Association for Plant Taxonomy, Vienna.

Szukala, A. et al. 2019. Phylogeny of the Eurasian genus *Jurinea* (Asteraceae: Cardueae): Support for a monophyletic genus concept and a first hypothesis on overall species relationships. *Taxon* 68: 112–131.

Wang, X., et al. 2015. Cenozoic vertebrate evolution and paleoenvironment in Tibetan Plateau: Progress and prospects. *Gondwana Research* 27: 1335–1354.

Watson, L.E., Siniscalchi, C.M., Mandel, J.R. 2020. Phylogenomics of the hyperdiverse daisy tribes: Anthemideae, Astereae, Calenduleae, Gnaphalieae, and Senecioneae. *Journal of Systematics and Evolution* 58: 841–852.

Wen, J., et al. 2014. Evolutionary diversifications of plants on the Qinghai-Tibetan Plateau. *Frontiers in Genetics* 5: 4.

Wiens J.J. et al. 2010. Niche conservatism as an emerging principle in ecology and conservation biology. *Ecology Letters* 13:1310–1324.

Xing, Y., Ree, R.H. 2017. Uplift-driven diversification in the Hengduan Mountains, a temperate biodiversity hotspot. *Proceedings of the National Academy of Sciences of the United States of America* 114: E3444–E3451.

Xu, L.S., Herrando-Moraira, S., Susanna, A., Galbany-Casals, M., Chen, Y.S. 2019. Phylogeny, origin and dispersal of *Saussurea* (Asteraceae) based on chloroplast genome data. *Molecular Phylogenetics and Evolution* 141: 106613.

Zhang, X. et al. 2019. Plastome phylogenomics of *Saussurea* (Asteraceae: Cardueae). *BMC Plant Biology* 19: 1–10.

Conclusiones

1. La secuenciación y posterior análisis filogenético de un total de 856 táxones vegetales confirma que el panel de genes 1061-Compositae (o COS loci) proporciona información filogenética para un amplio rango taxonómico dentro de la familia de las Compuestas (nivel tribal, subtribal y genérico). En general, se recuperan relaciones dicotómicas altamente apoyadas para la mayoría de clados, especialmente bajo el enfoque de concatenación, cuando con reconstrucciones Sanger formaban grandes politomias con relaciones no resueltas. Sin embargo, en el nivel de especie, para grupos de reciente diversificación la resolución no es óptima.

2. El método HybPiper destaca como método óptimo de extracción de secuencias en comparación con PHYLUCE. Se considera un paso clave la detección y eliminación de loci potencialmente afectados por paralogía. En cuanto al filtrado posterior de secuencias extraídas, la eliminación de posiciones hipervariables no afectaría a los resultados obtenidos, especialmente en aquellos grupos de reciente diversificación.

3. La técnica de sequenciacion *Hyb-Seq* es muy efectiva, no solo con muestras frescas o secadas con gel de sílice, sino también con aquellas procedentes de herbario que pueden ser muy antiguas o mal conservadas, y en consecuencia estar altamente degradadas.

4. Se proponen nuevas clasificaciones taxonómicas: una nueva clasificación subtribal de la tribu *Cardueae*, pasando de cinco a 12 subtribus, y una nueva delimitación genérica dentro de la subtribu *Saussureinae*, pasando de hasta 17 géneros descritos, a un total de tres (*Dolomiaea*, *Saussurea* y *Jurinea*).

5. Los factores extrínsecos son los mayores responsables de la diversificación de *Saussurea* y *Jurinea*. Como impulsores de la especiación se sugieren principalmente los cambios climáticos y los levantamientos de montañas.

6. Los cambios climáticos pasados hacia condiciones más frías y áridas promovieron un aumento considerable de la diversificación de especies en zonas montañosas y templadas del Hemisferio Norte. Estas aceleraciones en la formación de especies pudieron tener lugar en distintos períodos geológicos, como durante el Mesiniano en el caso de *Saussurea* (cambio en la tasa de diversificación detectado hacia los 6 Ma), o durante la transición Plioceno-Pleistoceno en *Jurinea* (hacia los 3 Ma). El aumento de la diversificación en estos períodos de enfriamiento y desecación global pudo darse gracias a la conservación del nicho ecológico ancestral, es decir, en linajes ya previamente adaptados o tolerantes a estas condiciones.

7. También se han detectado cambios de nicho ecológico en ambos géneros. Se observa que una divergencia de nicho puede dar lugar a un linaje de alto éxito evolutivo en el número de especies, como sería la diversificación en zonas más húmedas y/o de baja altitud en la región Circumboreal, o, por el contrario, puede desembocar en un linaje empobrecido en cuanto al número de especies como *Jurinea* en el este asiático.

8. Las oscilaciones climáticas moderadas han podido contribuir a la especiación de *Saussurea* y *Jurinea* en regiones montañosas templadas. En un contexto de fluctuaciones climáticas en zonas montañosas de amplios gradientes altitudinales, los linajes que se dispersaron pudieron encontrar su óptimo ecológico a lo largo del gradiente. Estos movimentos habrían favorecido la especiación por alopatría y la divergencia ecológica en condiciones locales o microclimas.

9. Las cordilleras montañosas continuas como el corredor formado por las montañas de Asia Central, Himalaya y las montañas Hengduan habrían promovido dispersiones de especies e intercambios de linajes, aumentando así la diversidad que encontramos hoy en día en estas regiones.

10. Los resultados del estudio de *Jurinea* confirman el papel crucial de la región Irano-Turania como centro de origen y diversificación de especies y como fuente de diversidad para las regiones adyacentes.

11. En el caso de *Saussurea*, los resultados obtenidos confirman el papel de las montañas Hengduan como centro de diversificación y fuente de diversidad para las regiones colindantes.

Anexos

Contribuciones de la doctoranda

En la Tabla 7 se detallan las contribuciones específicas por cada fase o tarea de la investigación desarrollada en la presente tesis doctoral.

Tabla 7. Resumen de las tareas y contribuciones de la doctoranda y los miembros del proyecto. En casos específicos se concreta entre paréntesis sobre qué capítulo de la tesis se ha contribuido y para las tareas de laboratorio se especifican los porcentajes aproximados de muestras procesadas por cada contribuyente/s. Los puntos marcados con un asterisco fueron aprendidos durante la estancia de 5 semanas en el laboratorio de la doctora Jennifer R. Mandel (University of Memphis, USA). Abreviaturas: C (CAPÍTULO de la tesis doctoral). Abreviaturas de la doctoranda y directores: SH (Sonia Herrando), NGJ (Núria Garcia Jacas, MGC (Mercè Galbany Casals), AS (Alfonso Susanna). Abreviaturas de investigadores colaboradores: ANS (Alexander N. Sennikov), CR (Cristina Roquet), EP (Elizaveta Pyak), FC (Fernando Castro), HTI (Hyoung-Tak Im), IM (Iraj Mehregan), JLP (Jordi López Pujol), JQL (Jian-Quan Liu), KF (Kazumi Fujikawa), LB (Laia Barres), LSX (Lian-Sheng Xu), MLG (Maria Luisa Gutiérrez), NI (Neus Ibáñez), RV (Roser Vilatersana), SCK (Seung-Chul Kim), SM (Sergi Massó), YSC (You-Sheng Chen).

Tareas		Contribución	
Generales	**Específicas**	**Doctoranda**	**Otros participantes**
Recopilación de muestras	Campañas de campo	SH, AS suspensión de la campaña en las montañas Hengduan por motivos de salud (C6)	ANS, IM (C5) EP, HTI, JLP, JQL, KF, LSX, SCK, SM, YSC (C6)
	Visitas a herbarios	SH	AS, JLP, NGJ, NI, RV
Recopilación de datos no genéticos	Búsqueda de localidades	SH (C5)	JQL (C6)
	Clasificación de morfotipos	SH (C5)	MGC (C5), LSX, YSC (C6)
	Revisión nomenclatural y taxonómica	SH	AS, NGJ, MGC, ANS
Laboratorio	Extracción de ADN	SH (20%)	FC, MLG (80 %)
	Protocolo Hyb-Seq*	SH (70 %)	LB, LSX (30%)
Análisis bioinformáticos	Extracción de secuencias e inferencia filogenética*	SH	–
	Biogeográficos, datación	SH	–
	Diversificación	SH	CR
	Paleoclima y nicho ecológico	SH	JLP
Manuscritos	Interpretación de resultados y redacción	SH	MGC, NGJ, AS, CR
	Revisión	Todos los participantes del proyecto	

Cronograma del desarrollo de la tesis

A continuación, se detallan cronológicamente las principales tareas desarrolladas para cada uno de los seis capítulos de la tesis doctoral (Tabla 8), marcadas en cuatro períodos trimestrales: (1) enero-marzo; (2) abril-junio; (3) julio-septiembre; y (4) octubre-diciembre. Además, también se incluyen otras actividades con relación directa con la tesis o colaboraciones con otros proyectos e investigadores (Tabla 9).

Tabla 8. Cronograma de desarrollo de las tareas principales por capítulos de la tesis doctoral. Durante los trimestres 1,2 y 3 del 2019 la doctoranda estuvo de baja por maternidad.

Año	16	2017				2018				2019				2020				2021	
Trimestre	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2
Capítulo 1										Baja mat.									
Muestreo																			
Laboratorio																			
Análisis de datos																			
Redacción																			
Envío																			
Publicación																			
Capítulo 2																			
Muestreo																			
Laboratorio																			
Análisis de datos																			
Redacción																			
Envío																			
Publicación																			
Capítulo 3																			
Muestreo																			
Laboratorio																			
Análisis de datos																			
Redacción																			
Envío																			
Publicación																			
Capítulo 4																			
Muestreo																			
Laboratorio																			
Análisis de datos																			
Redacción																			
Envío																			
Publicación																			
Capítulo 5																			
Muestreo																			
Laboratorio*																			
Análisis de datos																			
Redacción																			
Envío																			
Publicación																			
Capítulo 6																			
Análisis de datos																			
Redacción																			
Envío																			
Publicación																			

*el procesamiento de las muestras en el laboratorio fue cojunto para el estudio de *Jurinea* (CAPÍTULO 5) y el estudio de *Saussurea* (CAPÍTULO 6).

Tabla 9. Trimestre 1-3 (2019) baja por maternidad. *Estancias breves en la Universidad de Memphis (5 semanas de noviembre a diciembre de 2016) y Missouri Botanical Garden (1 semana en julio de 2018). ** La numeración especificada en colaboración en docencia representa el número de asignaturas en las que se ha participado. Abreviaturas: c. (co-autora), colab. (colaboración), 1ª (primera autora).

Año	**16**	**2017**				**2018**				**2019**				**2020**				**2021**	
Trimestre	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2
Otras publicaciones																			
Publicaciones 1ª (8)	1			1				1			2	1		1		1			
Publicaciones c. (11)		1			2		1		2	1		2		1		1			
Otras actividades																			
Congresos	1			1															
Seminariosimpartidos			3					1									1		
Estancias breves*	1							1											
Cursos	1	1		2				1											
Colab. Docencia**					2		3												