Org Divers Evol (2017) 17:323–349
DOI 10.1007/s13127-016-0316-0

ORIGINAL ARTICLE

ORGANISMS DIVERSITY & EVOLUTION

Against all odds: reconstructing the evolutionary history of *Scrophularia* (Scrophulariaceae) despite high levels of incongruence and reticulate evolution

Agnes Scheunert[1] · **Günther Heubl**[1]

Received: 4 August 2016 / Accepted: 30 November 2016 / Published online: 5 January 2017

Abstract The figwort genus *Scrophularia* (Scrophulariaceae), widespread across the temperate zone of the Northern Hemisphere, comprises about 250 species and is a taxonomically challenging lineage displaying large morphological and chromosomal diversity. *Scrophularia* has never been examined in a large-scale phylogenetic and biogeographic context and represents a useful model for studying evolutionary history in the context of reticulation. A comprehensively sampled phylogeny of *Scrophularia* was constructed, based on nuclear ribosomal (ITS) and plastid DNA sequences (*trnQ-rps16* intergenic spacer, *trnL-trnF* region) of 147 species, using Bayesian inference and maximum likelihood approaches. Selected individuals were cloned. A combination of coding plastid indels and ITS intra-individual site polymorphisms, and applying Neighbor-Net and consensus network methods for adequate examination of within-dataset uncertainty as well as among-dataset incongruence, was used to disentangle phylogenetic relationships. Furthermore, divergence time estimation and ancestral area reconstruction were performed to infer the biogeographic history of the genus. The analyses reveal significant plastid-nuclear marker incongruence and considerable amounts of intra-individual nucleotide polymorphism in the ITS dataset. This is due to a combination of processes including reticulation and incomplete lineage sorting, possibly complicated by inter-array heterogeneity and pseudogenization in ITS in the presence of incomplete concerted evolution. Divergence time estimates indicate that *Scrophularia* originated during the Miocene in Southwestern Asia, its primary center of diversity. From there, the genus spread to Eastern Asia, the New World, Europe, Northern Africa, and other regions. Hybridization and polyploidy played a key role in the diversification history of *Scrophularia*, which was shaped by allopatric speciation in mountainous habitats during different climatic periods.

Keywords *Scrophularia* · Incongruence · Reticulate evolution · Intra-individual polymorphism · 2ISP · Allopatric speciation

Introduction

In recent years, an increasing number of phylogenetic studies in plants, based on molecular sequence information from numerous independent loci, have revealed discordance among chloroplast and nuclear gene trees as well as gene trees in general. Although methodological issues in data collection or analysis might be responsible for some of these observations, incongruence may also be due to conflicting genealogical histories (e.g., Rokas et al. 2003; van der Niet and Linder 2008). These can be caused by gene duplications or losses, or by incomplete lineage sorting (ILS; Maddison 1997; Degnan and Rosenberg 2009). Furthermore, processes involving reticulation, i.e., gene flow among species, have been identified, e.g., horizontal gene transfer, introgression, and homo- or polyploid hybridization. These phenomena are common in plants (Rieseberg and Wendel 1993; Rieseberg et al. 1996; Wendel and Doyle 1998; Mason-Gamer 2004; Richardson and Palmer 2007); reticulation is now even regarded as a

Electronic supplementary material The online version of this article (doi:10.1007/s13127-016-0316-0) contains supplementary material, which is available to authorized users.

✉ Agnes Scheunert
agnes.scheunert@gmx.net

[1] Systematic Botany and Mycology, Department Biology I, and GeoBio-Center LMU, Ludwig-Maximilians-University Munich, Menzinger Strasse 67, 80638 Munich, Germany

major driving force in the evolution and diversification of plant lineages (Ellstrand and Schierenbeck 2000; Seehausen 2004; Soltis and Soltis 2009), as is polyploidization (Leitch and Bennett 1997; Abbott et al. 2013).

Researchers have aimed to trace reticulations within the evolutionary history of plants by examining incongruences among gene trees, often a combination of uniparentally (usually plastid) and biparentally (nuclear) inherited molecular markers (e.g., Rieseberg 1991; Baumel et al. 2002; Okuyama et al. 2005; Marhold and Lihová 2006; Fehrer et al. 2007; Scheunert and Heubl 2011, 2014). However, it has to be noted that the information content of phylogenetic reconstructions can be severely impeded in the presence of hybridization. Additionally, ILS/deep coalescences lead to incongruent patterns identical to those from hybridization and introgression. The use of nuclear ribosomal internal transcribed spacer (ITS) sequence information, although most popular due to its considerable advantages (Baldwin et al. 1995), can possibly mislead phylogenetic inference (reviewed by Álvarez and Wendel 2003; Nieto Feliner and Rosselló 2007). The arrangement of the 35S rDNA genes in tandem arrays at (sometimes several) separate ribosomal loci results in thousands of repeats per genome. Consequently, any ITS sequence will contain the summarized signal from many, often non-orthologous, not necessarily identical units. This leads to intra-individual site polymorphism, which can have a variety of origins (see "Discussion" section). Although these variabilities in rDNA units are often homogenized by the process of concerted evolution (Brown et al. 1972; Arnheim et al. 1980; Nei and Rooney 2005), several cases are known where it is slowed down, incomplete or non-operational (Grimm et al. 2007; Rosselló et al. 2007; Grimm and Denk 2008; Xiao et al. 2010; Hodač et al. 2014). All of the phenomena described above can have substantial impact on phylogenetic reconstruction (Álvarez and Wendel 2003; Linder and Rieseberg 2004; Beiko et al. 2008; Degnan and Rosenberg 2009), challenging traditional approaches which aim to depict evolutionary history and its underlying events using dichotomously branching phylogenetic trees. In such cases, the incorporation of network construction methods (e.g., Huson and Bryant 2006) into the analyses is a useful alternative.

The genus *Scrophularia* L., 1753 (Lamiales: Scrophulariaceae) represents a useful model for studying the influence of reticulation on evolutionary history and how it affects the inference of the latter. The genus consists of about 250 species, with estimates ranging from about 200 species in Mabberley (1997) and Fischer (2004), to more than 300 in Willis (1973). Plants are mostly herbaceous perennials, less often suffrutescent perennials or subshrubs and, more rarely, biennial or annual herbs. They are characterized by quadrangular, sometimes winged stems and considerably variable, undivided to 3-pinnatisect, generally opposite leaves. The inflorescence typically is a thyrse or a raceme with chasmogamous flowers. The five lobes of the calyx frequently possess a scarious margin, the often brownish, purplish or greenish corolla is sympetalous, bilabiate, and generally tubular or ventricose in shape. Apart from the four fertile stamens, the fifth (adaxial) stamen, if not completely absent, is generally sterile and forms a scale-like staminode of various shapes. The fruit is a septicidal, globose to subconical capsule containing numerous small seeds.

The most recent taxonomic treatment by Stiefelhagen (1910) divided the genus into two sections, *Scrophularia* sect. *Anastomosantes* Stiefelh. (= *S.* sect. *Scrophularia*) and *Scrophularia* sect. *Tomiophyllum* Benth. (= *S.* sect. *Canina* G.Don). *Scrophularia* extends throughout the temperate zone of the Northern Hemisphere (Asia, Europe and Northern America), with very few species expanding into tropical regions (e.g., the Greater Antilles). Based on floristic studies, species are concentrated in the Irano-Turanian floristic region sensu Takhtajan (1986), with particularly high species diversity found in Iran and Turkey (42 and 59 species according to Lall and Mill 1978; Grau 1981; Davis et al. 1988), but a large number also accounted for in the Flora of the USSR (Gorschkova 1997). Secondary centers of species richness are located in the Himalayan region and China (more than 36 species; Hong et al. 1998), as well as the Iberian Peninsula and adjacent areas including Macaronesia (28 species; Dalgaard 1979; Ortega Olivencia 2009). Representatives of the genus mainly inhabit highland plateaus and mountainous regions (all centers of species diversity comprise mountainous regions) but also coastal and lowland areas. Many species prefer shady and/or moist habitats, while others are xerophytic (especially within *S.* sect. *Tomiophyllum*) and can tolerate drier conditions; true desert plants are however rare. Most importantly, the genus is characterized by frequent natural hybridization, expressed in a variety of polyploid species and also linked to great morphological plasticity (Stiefelhagen 1910; Grau 1981). Natural hybrids in *Scrophularia* have been reported or even described as species by Menezes (1903, 1908), Stiefelhagen (1910), Pennell (1943) or Grau (1981). Artificial crossings were successful according to Goddijn and Goethart (1913), Shaw (1962), Carlbom (1964), Grau (1976), and Dalgaard (1979), and several cases of homoploid and allopolyploid speciation have been reported by Scheunert and Heubl (2011, 2014), who also found evidence for substantial tree incongruence.

Until now, phylogenetic relationships were mainly addressed on a restricted geographical scale, regarding e.g., species of the New World (Scheunert and Heubl 2011), Iran (Attar et al. 2011), and the Mediterranean and Macaronesia (Scheunert and Heubl 2014). Based on a time-calibrated phylogeny, Navarro Pérez et al. (2013) recently found support for monophyly of the genus and its divergence in the Miocene. However, a robust phylogenetic framework comprising all important lineages has been missing to date.

Here, we use sequences from the nuclear ribosomal ITS region and two plastid DNA regions (the *trnQ-rps16* intergenic spacer and the *trnL-trnF* region) to infer phylogenetic relationships. We aim to test the extent to which it is possible to reconstruct evolutionary history in the face of reticulation and incomplete lineage sorting without resorting to cloning or whole genome analyses. The main objectives of our study are (1) to establish a comprehensive evolutionary framework for *Scrophularia* based on a broad taxon sampling; (2) to assess the amount of intra-individual polymorphisms in ITS sequence data and to explore their possible causes; (3) to identify inconsistencies between nuclear and plastid DNA phylogenies and to examine their relation to reticulate evolution; (4) to reconstruct the biogeographic history of *Scrophularia* and to reveal which processes account for its current distribution patterns and species diversity.

Materials and methods

A broad range of methods has been applied in the present work; these are generally outlined below. Additionally, as the intent of this study is also to provide a workflow for researchers dealing with similarly complex groups, the information provided here is complemented by detailed descriptions including settings and procedures, available from Online Resource 1.

Taxon sampling

The taxon sampling is the most comprehensive presented so far in a molecular study on the genus and comprises 147 of the approximately 250 extant *Scrophularia* species. Sampled taxa include representatives from throughout the distribution area and cover all proposed sections and subsections. Known hybrid taxa were only exceptionally included to avoid unnecessary introduction of further conflicts into the dataset. Five widespread or morphologically diverse species (*S. vernalis* L., 1753; *S. scopolii* Hoppe ex Pers., 1806; *S. heterophylla* Willd., 1800; *S. canina* L., 1753; *S. variegata* M.Bieb., 1798) were sampled with additional subspecies and/or varieties. To investigate intraspecific variability, five species were included with up to four representatives (*S. auriculata* L., 1753; *S. lyrata* Willd., 1805; *S. arguta* Sol., 1789; *S. nodosa* L., 1753; *S. olympica* Boiss., 1844). Altogether, the *Scrophularia* ingroup consisted of 162 accessions. Based on previously established relationships, 18 taxa, from the Scrophulariaceae (represented by five species) and other families within Lamiales (Calceolariaceae, Gesneriaceae, Plantaginaceae, Stilbaceae, Bignoniaceae, Verbenaceae), were selected as outgroups (Kornhall et al. 2001; Albach et al. 2005; Oxelman et al. 2005; Nie et al. 2006; Datson et al. 2008; Schäferhoff et al. 2010). In addition, the sampling included one species from *Oreosolen* Hook.f., 1884 (Scrophulariaceae), a genus which comprises one to four species endemic to the Himalayas and the Tibetan Plateau and was found to be most closely related to *Scrophularia* (Albach et al. 2005; Oxelman et al. 2005). Complete information on voucher specimens is provided in Online Resource 2 alongside accession numbers for all analyzed sequences.

DNA extraction, PCR, sequencing, and cloning

Leaf material for DNA extraction was obtained from herbarium specimens (169 accessions from collections in A, B, E, GH, HAL, HU/HZU, HSNU, KUN, KSC, LE, M, MA, MSB, W, WU, and WUK) and in nine cases from plants cultivated by the authors in the greenhouse of the Botanical Garden Munich (vouchers deposited in MSB). DNA extraction, PCR, purification and sequencing reactions were performed according to methods described in Scheunert and Heubl (2011, 2014). Two well-established loci from these studies were used, the non-coding chloroplast ("cp") *trnQ-rps16* intergenic spacer and the nuclear ("nr") ribosomal ITS region (internal transcribed spacer 1, 5.8S rRNA gene, internal transcribed spacer 2). Additionally, the plastid *trnL-trnF* region (consisting of the *trnL* intron, the *trnL* 3′ exon, and the *trnL-trnF* intergenic spacer; Taberlet et al. 1991) was used (see also Navarro Pérez et al. 2013). All primer sequences alongside references are provided in Online Resource 3. DNA sequences generated by Scheunert and Heubl (2011, 2014), Navarro Pérez et al. (2013) and others, as well as sequences from selected outgroup taxa were downloaded from NCBI's GenBank (http://www.ncbi.nlm.nih.gov, accessed 10 January 2014; see Online Resource 2). For further investigation of the considerable amount of intra-individual nr DNA variability and to support identification of putative hybrid species, six selected individuals (from *S. auriculata*; *S. incisa* Weinm., 1810; *S. lyrata*; *S. musashiensis* Bonati, 1911; *S. ruprechtii* Boiss., 1879; and *S. villosa* Pennell, 1923) were additionally cloned (for detailed information and PCR protocols see Online Resource 1). All clones were included into a separate phylogenetic analysis together with uncloned sequences (see *Phylogenetic inference*).

Data matrix composition and coding of chloroplast indels and ITS intra-individual site polymorphisms

Raw DNA sequence reads were edited and, where necessary, assembled into contigs with the CLC Main Workbench v. 6 (CLC Bio, Aarhus, Denmark). Ambiguously specified basepairs (due to poor sequence read quality or site polymorphism) were recorded using IUPAC ambiguity codes. Contigs were aligned using MAFFT v.7.110 (Katoh and Standley 2013; online version available at http://mafft.cbrc.jp/alignment/server/, accessed 13 October 2013); used

settings are reported in Online Resource 1. Manual refinements were done in BioEdit v. 7.1.11 (Hall 1999). Mononucleotide repeats and ambiguously aligned regions were excluded from further analysis. ITS sequences were checked for potential pseudogeny according to J-S Liu and Schardl (1994), Jobes and Thien (1997), and Bailey et al. (2003). Chloroplast indels, which have been shown to contain phylogenetic information in *Scrophularia* (Scheunert and Heubl 2011, 2014), were coded as binary states for the ingroup only, using the simple indel coding algorithm (Simmons and Ochoterena 2000) as implemented in SeqState v. 1.4.1 (Müller 2005).

In order to make optimal use of the information contained in ITS sequence data, all sequences were examined for the presence of polymorphic sites (PS). Then, two methods were applied (with minor adaptations), which are intended to incorporate information from PS into phylogenetic analyses. Using the ad hoc 2ISP-informative maximum likelihood (ML) approach from Potts et al. (2014), all IUPAC codes including polymorphic sites (there termed 2ISPs, intra-individual site polymorphisms) are treated as unique characters, by recoding the complete alignment as a standard matrix, which is then analyzed using the multistate analysis option for categorical data in RAxML (see below). This method is similar to some approaches described in Grimm et al. (2007); we complemented the Potts et al. (2014) method by the application of Bayesian inference (BI) to our coded dataset as well, based on Grimm et al. (2007). Detailed information on the coding procedure and the original methods can be found in Online Resource 1.

A different approach, pursued by Fuertes Aguilar and Nieto Feliner (2003), concentrates on "Additive Polymorphic Sites" while ignoring the remainder of intra-individual polymorphisms. According to their definition, a site is referred to as an "APS" when both bases involved in the polymorphism can each be found separately at the same site in at least one other accession. To investigate the usefulness of APS in phylogenetic reconstruction, these were also extracted from the dataset. A subset of 17 alignment positions with high numbers of APS, here termed "highly polymorphic alignment positions" ("HPPs"), was then recoded according to the procedure described above and added to the original DNA alignment. This data matrix was likewise analyzed using ML and BI. More detailed explanations are available in Online Resource 1; used codes for all HPPs are listed in Online Resource 4.

Phylogenetic inference

Five datasets were used, one containing the combined *trnQ-rps16* and *trnL-trnF* region data and coded indels from both markers ("cp dataset"), one based on uncoded nr DNA sequence data ("uncoded dataset"), one with nr sequence data alongside coded HPPs ("APS-coded dataset"), one with the complete nr sequence alignment recoded following the 2ISP-informative approach ("2ISP-coded dataset"), and one corresponding to the 2ISP-coded dataset, but also comprising cloned sequences ("nr+clones dataset"). Datasets were analyzed separately using ML and BI. For comparison purposes, additional analyses of the uncoded dataset were conducted, once excluding the 17 highly polymorphic positions themselves and once excluding 17 accessions with high amounts of APS in their sequences (see "Results" section). Appropriate nucleotide substitution models were selected using MrModelTest v. 2.3 (Nylander 2004), which suggested GTR+Γ with four rate categories as best fit to the data according to the Akaike information criterion, adding a proportion of invariant characters for the ITS dataset only (GTR+I+Γ). Bayesian analyses were performed with MrBayes v. 3.2.2 (Ronquist et al. 2012) on a local PC. Maximum likelihood analyses were carried out with RAxML v. 7.4.2 (Stamatakis 2006) on a local PC using raxmlGUI v. 1.3 (Silvestro and Michalak 2012), and RAxML v. 8.0.24 (Stamatakis 2014) on the CIPRES Science Gateway (MA Miller et al. 2010; available at https://www.phylo.org/portal2/login!input.action, accessed 02 December 2015). Settings for all runs and the different datasets are described in Online Resource 1.

Identifying phylogenetic conflict and uncertainty within and among datasets

The amount of ambivalent signal contained within the ITS raw data was illustrated using SplitsTree v. 4.12.3. (Huson and Bryant 2006). To this end, a Neighbor-Net splits graph (NN; Bryant and Moulton 2004) was created, based on a pairwise distance matrix obtained from the ITS sequence alignment (non-excluded characters, see Table 1). In order to incorporate information from polymorphic sites, polymorphism *p*-distances (see Potts et al. 2014) were calculated for the ingroup. Online Resource 1 provides information on required software and optimal settings. Additional networks were computed excluding columns containing gaps and/or accessions with high numbers of "?" and "N" in their sequences, to assess the impact of missing data on the results.

For the visualization of incongruence among datasets, consensus networks (CN; Holland and Moulton 2003; Holland et al. 2004) were constructed in SplitsTree, from each 1001 trees of both BI runs of the cp dataset and the 2ISP-coded dataset. Importantly, edge lengths in these CNs are proportional to the split frequency within the sampled topologies ("COUNT" option). A trees threshold of 0.33 was used for displaying splits, and splits were transformed

Table 1 Alignment characteristics and Maximum Likelihood-based tree statistics for the plastid *trnQ-rps16* and *trnL-trnF* markers and nuclear ITS. Average G+C contents and parsimony-(un)informative characters calculated excluding outgroups, other characteristics include outgroup values. Percentage of parsimony-informative characters referable to non-excluded characters; the latter inclusive of "highly polymorphic alignment positions." Alpha, tree score, and length in ITS given for the uncoded and 2ISP-coded dataset. Values marked with *asterisks* are based on the nr+clones dataset; values with *circles* refer to results from the concatenated cp dataset. *Alpha* the alpha value of the gamma shape parameter, *avg* average, *bp* basepairs, *no* number

	trnQ-rps16	trnL-trnF region	ITS	ITS (clones)
No. of taxa (including outgroups)	180	94	181	62
Sequence length (avg.)	427–1584 bp(961 bp)	100–818 bp(725 bp)	247–667 bp(580 bp)	436–602 bp (571 bp)
Aligned length	2063 bp	946 bp	763 bp	763 bp
Non-excluded characters	2046 bp	930 bp	738 bp	738 bp
Parsimony-uninformative characters	139 bp	51 bp	69 bp	98 bp*
Parsimony-informative characters	134 bp (6.55%)	32 bp (3.44%)	125 bp (16.94%)	150 bp (20.33%)*
Average G+C content	27.00%	33.3%	60.12%	60.87%
ML tree score	−15,802.897°	–	−9001.089/−15,043.732	–
ML tree length	1.835°	–	6.273/26.544	–
Alpha	0.979°	–	0.718/1.089	–

as outlined in Online Resource 1. Congruence among the sequence datasets was also tested with the Incongruence Length Difference (ILD) test (Farris et al. 1995) implemented as the Partition Homogeneity Test in PAUP v. 4.0b10 (Swofford 2003). Accessions or clades exhibiting hard incongruence (HI) were identified by visual inspection of the cp and (2ISP-coded) nr phylogenetic trees for well-supported conflicting placements (Mason-Gamer and Kellogg 1996), using a threshold of ≥0.90 Bayesian posterior probability (PP) in both topologies. Further information can be found in Online Resource 1.

Molecular clock analyses

Divergence times were estimated for the cp dataset using BEAST v. 1.8 (Drummond and Rambaut 2007). For information on input matrices and how information from coded binary indels was incorporated see Online Resource 1.

Divergence dating using fossils was impossible for this dataset of *Scrophularia* (possible reasons for the failure of fossil-based dating and additional information on the used fossils are given in Online Resource 1). Thus, rather than completely relying on secondary calibration constraints for relaxed clock analyses, an approach which is generally error-prone (see e.g., Hipsley and Müller 2014), we decided to resort to a first-step strict clock analysis in order to obtain information about a reasonable age for the ingroup which could be used as a starting point for further analyses. For the strict clock, a fixed substitution rate of $8.1E^{-4}$ per site per million years was used (Lavin et al. 2005). Then, divergence times of the ingroup and the closest outgroup genus only (reduced dataset) were inferred under a relaxed clock with log-normally distributed rates, using the estimated ingroup age from the strict clock run as secondary calibration point. Analyses were performed with BEAST v. 1.8 on the CIPRES Science Gateway; exact procedures, prior values, and settings are provided in Online Resource 1. The relaxed clock analysis was repeated without data (prior-only) on a local PC, to review effective prior distributions and assess the decisiveness of the data.

Biogeographic analyses

Ancestral area optimization relied on the Bayesian Binary Markov Chain Monte Carlo (BBM) algorithm as implemented in RASP v. 2.0b (Yu et al. 2010, 2014). Biogeographic areas are mapped on the world map in Fig. 1b; detailed definitions can be found in Online Resource 5. Distributions of species (including those of known synonyms) were then assigned to the respective areas. Further information on the classification of areas, the determination of species distributions, and the RASP analyses is available from Online Resource 1.

"Maxareas," the maximum number of ancestral areas inferred at each node, was constrained following Ronquist (1997). We assumed equal dispersal ability for ancestors and their present-day descendants (Sanmartín 2003) and therefore set maxareas to two (83% of the *Scrophularia* ingroup species occur in one or two areas only). The number of maximum areas was kept at five during one additional run for comparison purposes. Inferred ancestral distributions were mapped on the majority-rule consensus tree from Bayesian analysis in Fig. 1a, using a threshold of 25% marginal probability (frequency of occurrence of the respective range over the Bayesian sample of trees).

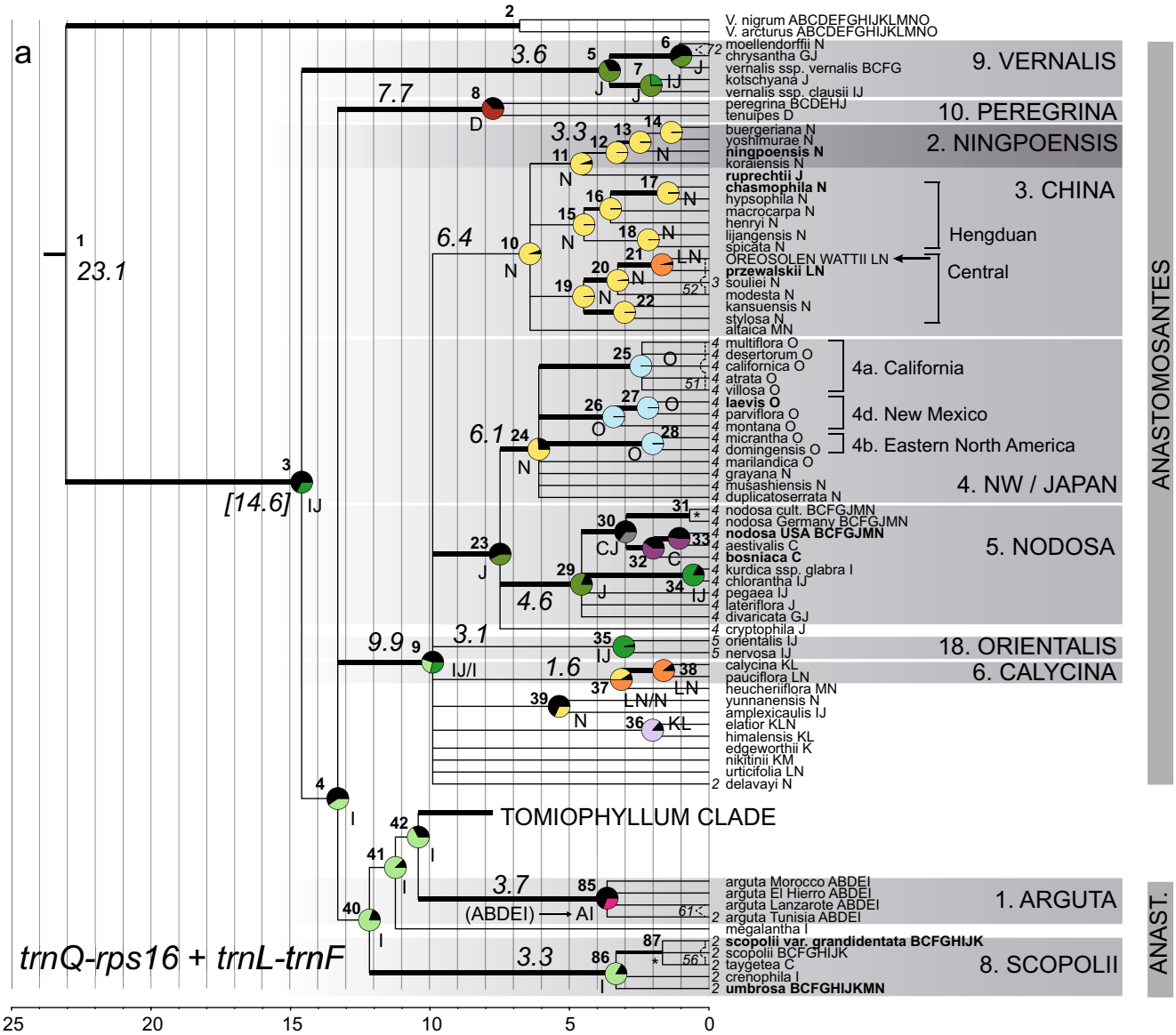


Fig. 1 **a** Dated phylogeny and ancestral area reconstruction for 147 *Scrophularia* species, on a majority-rule consensus tree obtained from Bayesian analysis of combined plastid *trnQ-rps16* intergenic spacer and *trnL-trnF* region alongside coded indels. Branches indicate levels of support, based on posterior probabilities (*PP*) and plotted bootstrap support values (*BS*) from Maximum Likelihood optimization; *bold* PP ≥ 95 or BS ≥ 85, *semi-bold* PP ≥ 90 or BS ≥ 75, *thin* PP < 90/BS < 75. Seven additional nodes only supported by ML (BS ≥ 50) were added manually but not incorporated into further analyses. *Gray bars on the right* denote Clades 1–18 and main species groups as discussed in the text. An *arrow* indicates the position of the Himalayan-Tibetan endemic genus *Oreosolen*. Single accessions displaying hard incongruence among (2ISP-coded) nuclear and plastid trees are marked in *bold*; Clades 7 and 5 (excluding *S. chlorantha*; plus *S. cryptophila*) as a whole are also hardly incongruent. The occurrence of large indels as defined in Table 2 is indicated next to each accession with the respective length type number; *no number* means length type = 1. *Node heights* represent mean ages and were inferred under a Bayesian relaxed clock with log-normally distributed rates, using one calibration point at node 3. Important clade crown ages are given in million years. Ancestral area optimization is based on 9002 trees from the Bayesian analysis, distribution ranges of single taxa are provided after the respective names, area codes and colors are as defined in (**b**) and Online Resource 5. *Pie charts* at nodes indicate inferred distributions of MRCAs from run four of four RASP runs (maxareas = 2); *asterisks* mark nodes where no ancestral distribution reached a marginal probability of 25%. At nodes 9, 37, 61, and 63, separate runs differed with respect to the most probable ancestral area. Alternative distributions shown in brackets derive from an additional analysis with maxareas = 5. Exact values for each node including highest posterior density intervals for inferred ages are listed in Online Resource 7. *NW* New World, *V Verbascum*, *cult* cultivated. **b** World map showing areas defined for ancestral area reconstruction analyses

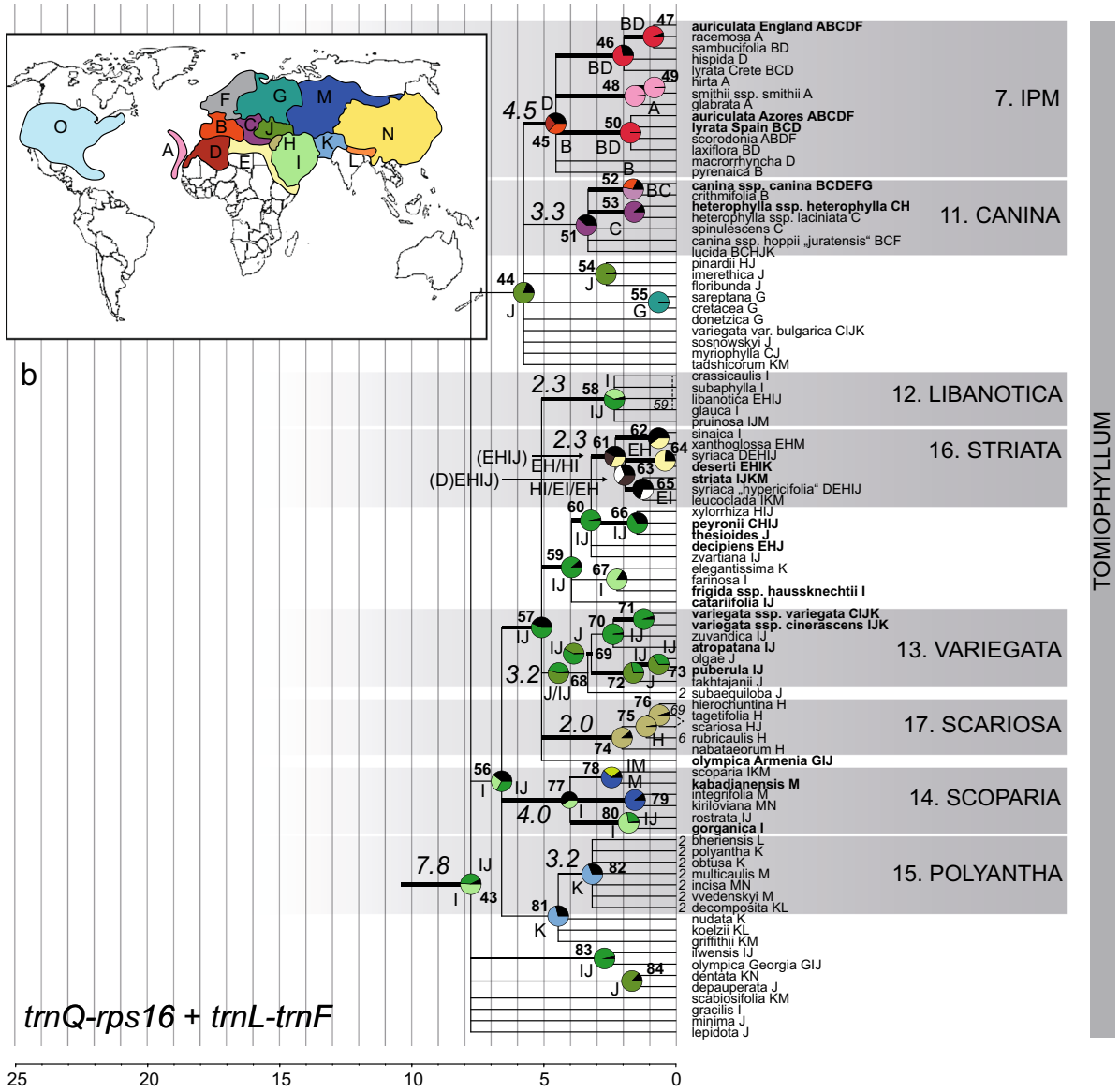


Fig. 1 (continued)

Results

Alignment characteristics

Altogether, 281 sequences were generated for this study, 125 for *trnQ-rps16*, 123 for ITS, and 33 for the *trnL-trnF* region. A total of 62 cloned ITS sequences was obtained from six accessions (available from GenBank under accession numbers KY067709–KY067770) and included into phylogenetic reconstructions, raising the number of ITS sequences to 243 in the nr+clones dataset. Data coverage for *trnQ-rps16* and ITS is complete for all accessions except *Aragoa*, where sequencing of *trnQ-rps16* failed and no sequence was available in GenBank; the genus was coded as missing for the respective marker but contributed a *trnL-trnF* sequence to the plastid alignment. For the *trnL-trnF* region, which was used as supplementary marker, only selected accessions were sequenced; the remainder was likewise coded as missing for *trnL-trnF*. Detailed information on sequence and alignment statistics including average G+C contents and proportions of parsimony-informative characters is given in Table 1. Twenty-nine ingroup indels were coded for the *trnL-trnF* region and 84 for *trnQ-rps16*. The latter is characterized by the occurrence of larger indels ranging from 312 to 839 characters in length; Table 2 shows the positions of all observed indels >300 characters (the respective accessions are marked in Fig. 1a).

PCR products of ITS showed clear single bands in most cases. However, polymorphic sites were present in almost three quarters of the ingroup accessions (121 of 163). Additive polymorphic sites (APS) as defined above were recorded in 95 accessions (58%), from which 17 had five or

more APS in their sequence ("APS-rich accessions," see Fig. 2a). Fourteen of these APS-rich accessions are members of the "Tomiophyllum" clade (see below). Generally, an unequivocal differentiation between artificial double peaks, PS and APS, was not always possible, which means that the reported numbers rather represent best possible estimates.

Phylogenetic relationships

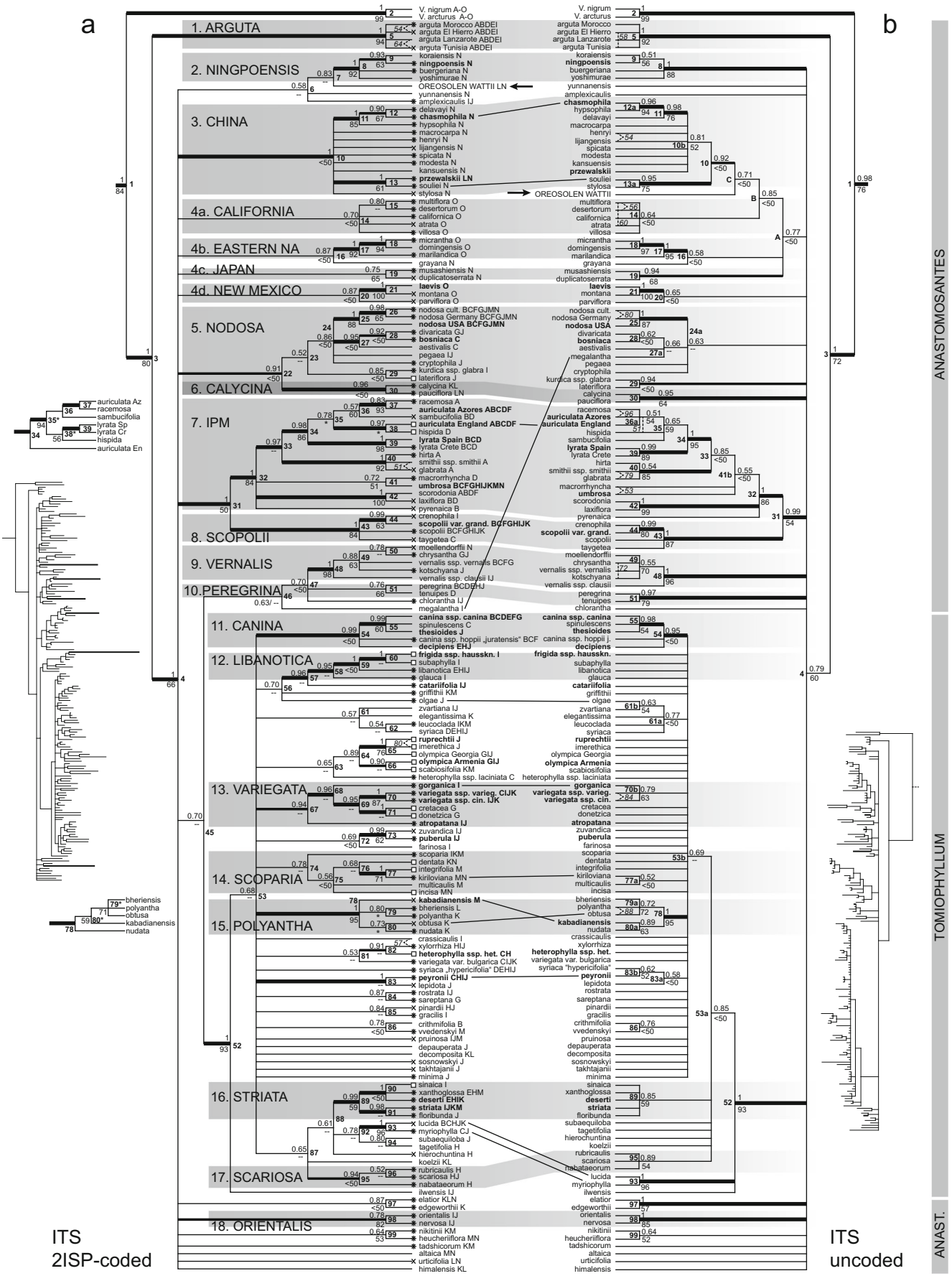
Bayesian analyses of the cp and nr datasets had reached convergence at the end of the runs (standard deviation of split frequencies below 0.01). The majority-rule consensus topologies, with all outgroups except *Verbascum* L., 1753 pruned, are shown in Figs. 1a and 2. Statistics on ML analyses are given in Table 1. The best ML trees only very rarely contradicted the Bayesian consensus trees with respect to nodes with bootstrap support ("BS") ≥ 50; supports were generally lower when using ML. Relative to the Bayesian tree from the uncoded ITS dataset (Fig. 2b, 64 nodes with PP ≥ 0.5), removal of the 17 HPPs yielded a tree with 18 nodes collapsed and weakened support values in 22 cases; removal of the APS-rich accessions from the uncoded dataset had less impact but led to generally lower supports in basal nodes (results not shown). Conversely, including the coded HPP matrix into calculations (APS-coded dataset) resulted in a tree with 19 new nodes (PP ≥ 0.5; not shown); using the 2ISP-informative approach and coding all double peaks present in the sequences (2ISP-coded dataset, Fig. 2a) yielded a tree with 35 new nodes relative to the uncoded phylogeny. By contrast, resolution did not change much in ML trees and was even slightly reduced.

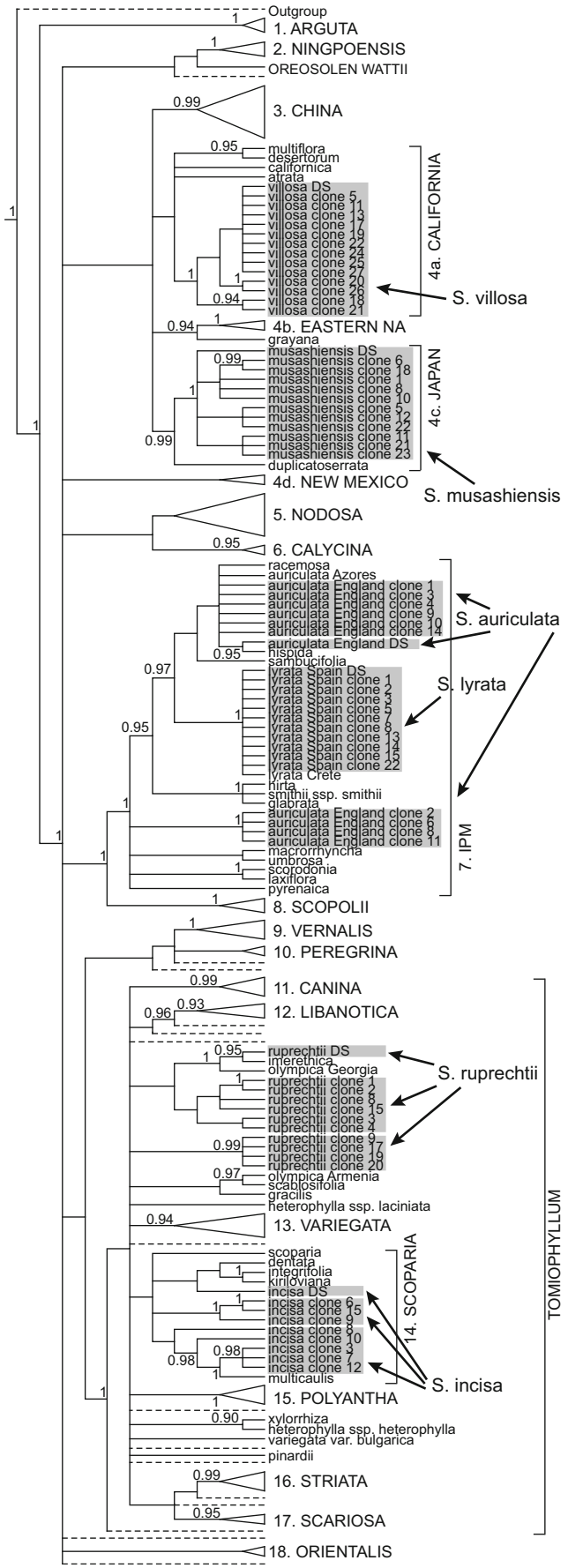
Both 2ISP-coded nuclear and plastid tree (Figs. 2a and 1a) support a monophyletic clade of all accessions from *Scrophularia* with moderate to maximum support (cp PP 1.00, BS 100/nr PP 1.00, BS 80). However, the accession from *Oreosolen wattii* Hook.f. is deeply nested within the ingroup at similar positions depending on the dataset, rendering *Scrophularia* paraphyletic with respect to this genus. For convenience and to avoid confusion with other clade names, Stiefelhagen's (1910) section names are adopted here to name the two main phylogenetic entities: the highly supported Tomiophyllum clade (cp PP 1.00, BS 99/nr PP 1.00, BS 93) largely corresponds to, but is not exactly identical with, *Scrophularia* sect. *Tomiophyllum*. The remainder of the

Fig. 2 Majority-rule consensus trees (cladograms) for 147 *Scrophularia* species, obtained from Bayesian analysis of the nuclear internal transcribed spacer region (ITS), with **a** intra-individual site polymorphisms (IUPAC codes) treated as informative using 2ISP coding or **b** using uncoded, unmodified sequences. Downsized phylograms are shown beside each tree, with six taxa having exceptionally long branches in the 2ISP-coded dataset marked in *bold*. Outgroup taxa are reduced to the closest genus. Posterior probabilities (*PP*) are given above branches, plotted bootstrap support values (*BS*) from Maximum Likelihood (*ML*) optimization below. Double dashes state that the node was not present in the fully resolved best-scoring ML tree. *Asterisks* denote cases with deviating ML topologies illustrated beside the tree; five/twelve additional nodes only supported by ML (BS ≥ 50) were added manually. Branches indicate levels of support as defined in Fig. 1. *Gray bars* denote Clades 1–18 and main species groups. *Arrows* indicate the position of the Himalayan-Tibetan endemic genus *Oreosolen*. Single accessions displaying hard incongruence among (2ISP-coded) nuclear and plastid trees are marked in bold; Clades 7 and 5 (excluding *S. calycina* and *S. pauciflora* Benth., 1835) as a whole are also hard incongruent. Accessions obtaining different positions in both phylogenies are connected across trees. Amounts of intra-individual polymorphism are indicated to the left of each accession using the following symbols: *no symbol* no polymorphic sites (*PS*), *cross* PS but no APS (additive polymorphic sites, see "Materials and methods" section), *star* APS present in the sequence, *square* ≥ 5 APS present, "APS-rich accession." Distribution ranges of single taxa are provided after the respective names, area codes are as defined in Fig. 1b and Online Resource 5. *NA* North America, *V Verbascum*, *cult* cultivated, *grand grandidentata*, *hausskn haussknechtii*, *varieg variegata*, *cin cinerascens*, *het heterophylla*

Table 2 Nine characteristic indels, corresponding to eight sequence length types, in the *trnQ-rps16* intergenic spacer alignment created from sequences of 162 *Scrophularia* accessions. Length type "1" no larger indels present, full alignment length 3325 basepairs. "Indel position" is referable to aligned length. Clade numbers as defined in the main text. Assessment of phylogenetic value as diagnostic character is given for each indel type. *No acc* number of accessions possessing the respective indel, *bp* basepairs

Length type	Indel position	Indel length	No acc	Species/clade	Diagnostic?
1	–	–	118	–	No
2a	485–796	312 bp	7	Clade 15	Yes
2b	518–845	328 bp	1	Delavayi	No
2c	527–845	319 bp	1	Subaequiloba	No
2d	528–845	318 bp	5	Clade 8	Yes?
	527–844	318 bp	1	Arguta El Hierro	No
3	688–1118	431 bp	1	Souliei	No
4	551–1147	597 bp	25	Clades 4+5	Yes
5	251–1014	764 bp	2	Clade 18	Yes
6	207–1045	839 bp	1	Rubricaulis	No


a
b
1. ARGUTA
2. NINGPOENSIS
3. CHINA
4a. CALIFORNIA
4b. EASTERN NA
4c. JAPAN
4d. NEW MEXICO
5. NODOSA
6. CALYCINA
7. IPM
8. SCOPOLII
9. VERNALIS
10.PEREGRINA
11. CANINA
12. LIBANOTICA
13. VARIEGATA
14. SCOPARIA
15. POLYANTHA
16. STRIATA
17. SCARIOSA
18. ORIENTALIS
ANASTOMOSANTES
TOMIOPHYLLUM
ANAST.
ITS
2ISP-coded
ITS
uncoded



◀ **Fig. 3** Majority-rule consensus tree (cladogram) for 224 accessions of 147 *Scrophularia* species, including 62 cloned sequences from six accessions, and obtained from Bayesian analysis of 2ISP-coded nuclear ITS. Detailed relationships are shown for clades which include clones and where deviating from Fig. 2a; *dashed lines* represent one or more accessions removed for clarity. Posterior probabilities ≥ 0.90 are given above branches. Clones and corresponding direct sequences (*DS*) are highlighted in *gray*; *brackets on the right* denote the clade they belong to. Distinct positions of clones and DS for each species are indicated by *arrows*. *NA* North America

species mainly belong to *Scrophularia* sect. *Anastomosantes*; these species do not form a monophyletic clade, but are paraphyletic with respect to the Tomiophyllum clade. Nevertheless, they are united by a number of shared characteristics and are therefore referred to as the "Anastomosantes group."

Several clades can be recognized in both trees (see boxes in Fig. 1a and 2a): for example, "Arguta" (Clade 1) comprises all four accessions of this annual species and receives high to maximum support in the analyses. A clade of mainly Chinese endemics ("China," Clade 3) receives maximum support by BI in ITS only. In the cp tree, its composition is slightly different and includes the "Ningpoensis" clade (Clade 2). The most widespread representative of the genus, the type species *S. nodosa*, forms a monophyletic clade with seven to nine other taxa ("Nodosa," Clade 5) and is moderately to highly supported by BI. In the cp tree, it includes the Southwestern Asian/Turkish - Caucasian *S. chlorantha* Kotschy & Boiss., 1879, in ITS the Turkish endemic *S. cryptophila* Boiss. & Heldr., 1853, and the mainly Southern Asian Calycina clade ("Calycina," Clade 6). Clade 7, the "IPM" clade ("Iberian Peninsula–Macaronesia" as introduced by Scheunert and Heubl 2014), is supported in all analyses and differs between cp and nr only with regard to *S. umbrosa* Dumort., 1827. Within the Tomiophyllum clade, fewer clades are present in both trees, and consistency regarding their members is much less pronounced. Taxa from the New World group differently in the analyses: the plastid tree supports monophyly of all New World species and three "Japanese taxa" (*S. grayana* Maxim. ex Kom., 1907; *S. duplicatoserrata* Makino, 1906; *S. musashiensis*) in a "New World ("NW")/Japan" clade (Fig. 1a, Clade 4). Within the clade, three subclades from the New World are supported, one with mainly lowland taxa centered in California ("California," 4a), one with subalpine taxa distributed in New Mexico and Arizona ("New Mexico," 4d), and one comprising species from the Greater Antilles, and in the nr tree the mainly lowland Eastern North American *S. marilandica* L., 1753 ("Eastern North America," 4b). In the nr trees, the NW/Japan clade collapses into its subclades, with *S. grayana* being associated with Clade 4b and two of the Japanese taxa forming Clade 4c ("Japan"). Clades 4a–d remain unconnected using the 2ISP-coded dataset, but based on

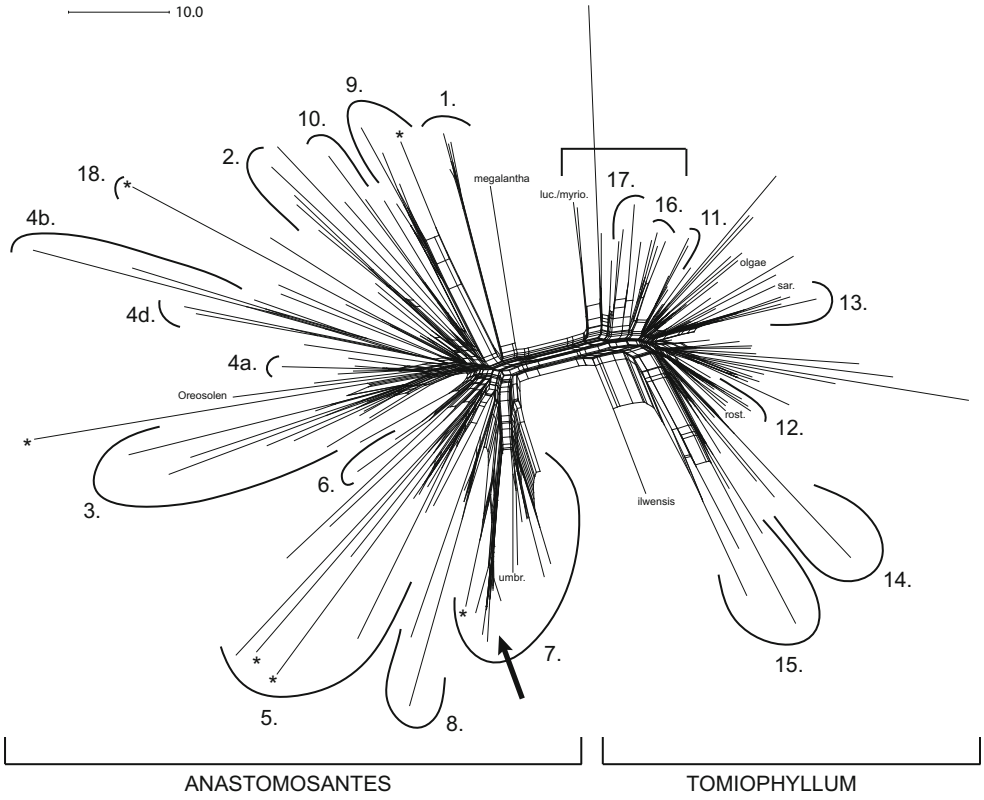


Fig. 4 Neighbor-Net splits graph for 163 accessions of *Scrophularia*, based on polymorphism *p*-distances (i.e., treating intra-individual site polymorphisms as informative) inferred from the ITS sequence data set. *Scale bar* corresponds to split weights based on ordinary least squares estimates. Clades 1–18 are marked, *square brackets* below denote main species groups. *Asterisks* highlight six taxa having exceptionally long branches in the 2ISP-coded dataset, e.g., *S. auriculata* from England, the accession with the highest number of polymorphic sites in this study, within Clade 7. The Auriculata subclade of Clade 7 is indicated by an *arrow*, an assemblage of taxa (node 87 in Fig. 2a) which also includes the Striata (16) and Scariosa (17) clades is marked by the square bracket above. *sar sareptana*, *rost rostrata*, *luc/myrio lucida/ myriophylla*, *umbr umbrosa*

the uncoded dataset, three of them form a weakly to very weakly supported grade leading towards the China clade (Fig. 2). By contrast, in the cp tree the NW/Japan clade is sister to the Nodosa clade and *S. cryptophila* with moderate to maximum support. Altogether, the members of 11 of the 18 main clades vary due to low resolution and/or incongruence. Relationships among clades and the backbones of the trees are only poorly resolved.

Bayesian analysis of the nr+clones dataset resulted in the majority-rule consensus presented in Fig. 3. The topology is largely similar to the 2ISP-coded ITS tree, with generally equal or slightly lowered support values in clades without clones. The clade containing *S. ruprechtii* (Fig. 2a, node 63) collapsed into its subclades when clones were included.

ITS raw data network and incongruence among chloroplast and nuclear markers

The Neighbor-Net splits graph computed from polymorphism *p*-distances (Fig. 4) is extremely non-treelike; high complexity and substantial ambiguity are present throughout the whole ingroup. Its dumbbell-shaped structure traces the two main phylogenetic groups also seen in tree reconstructions. Accessions recovered in well-supported clades in Fig. 2a mostly group together in the NN. Although distance-based methods are often sensitive to missing data, their exclusion did not improve the resulting network distinctly; rather, negative effects were observed as the amount of discarded data increased. Thus, the NN in Fig. 4 is based on the full ingroup dataset.

Although several clades are found in both plastid and nuclear trees, their composition is often considerably incongruent. Considering only cases of well-supported incongruence (among the plastid and 2ISP-coded nr dataset) however, only 28 single accessions display hard incongruence as defined above (HI accessions; highlighted in bold in Figs. 1a and 2). These conflicts are based on switches within the same clade in almost 60%; swaps among Clades 1–18 are observed in *S. umbrosa, S. kabadianensis* L.Fedtsch., 1913, and *S. gorganica* Rech.f., 1955. In *S. ruprechtii*, incongruence involves a switch from a clade within Tomiophyllum in ITS towards the Ningpoensis clade within a large clade of Anastomosantes taxa in cp. In addition to the single HI accessions, the IPM clade (Clade 7) as a whole is incongruent as well: it is embedded in the Tomiophyllum clade (cp) while being sister to the "Scopolii" clade (Clade 8) of Anastomosantes in ITS (see also its position in Fig. 5a). Similarly, *S. cryptophila* and the Nodosa clade as shown in the cp tree (excluding *S. chlorantha*) are also hardly incongruent. The detected HI accessions thus sum up to a total of 47.

The ILD test did not detect significant incongruence between the two chloroplast markers at the $p = 0.05$ confidence level ($p = 0.180$), so combined analysis of the plastid data was justified. However, as expected from visual inspection of the trees, the test revealed severe incongruence among nuclear and plastid datasets ($p = 0.002$). Deleting the 47 HI accessions

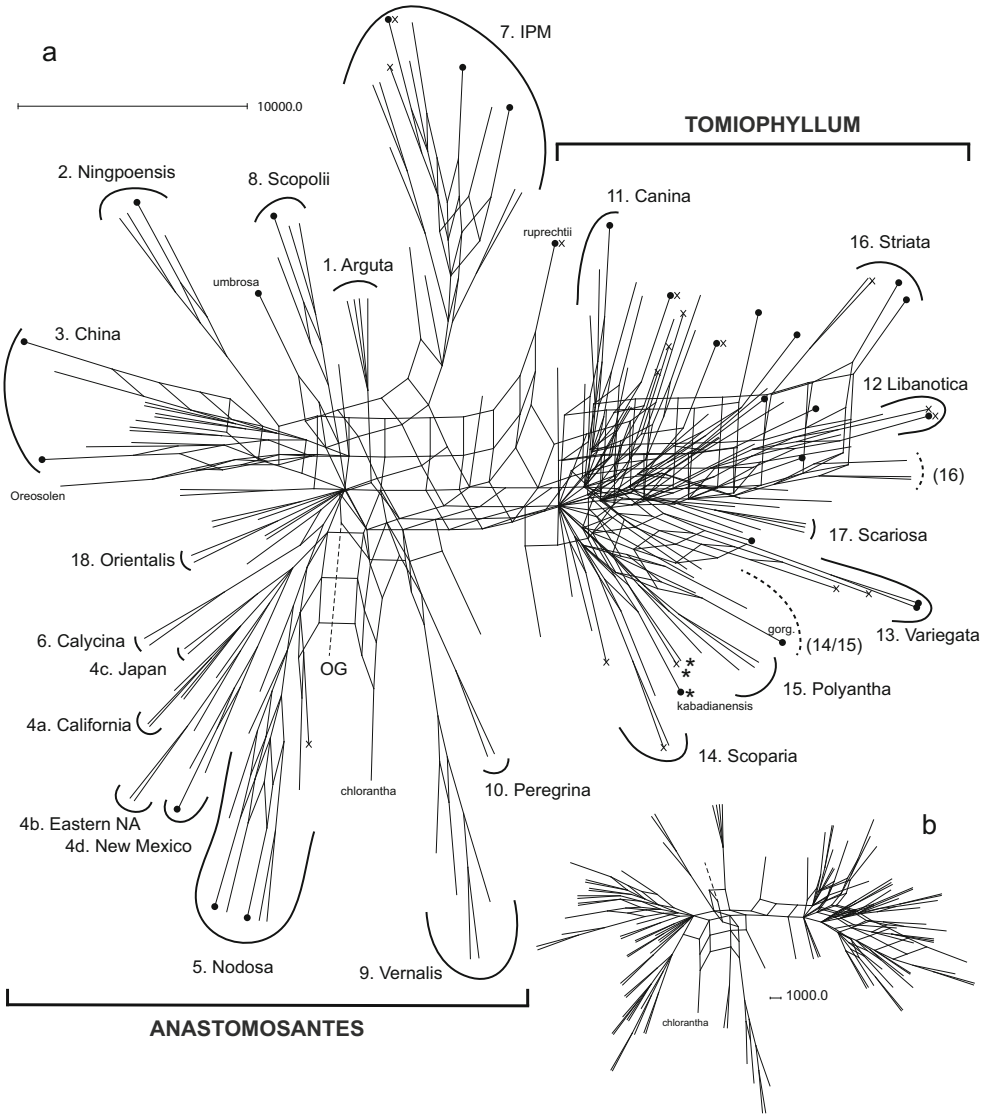


Fig. 5 Consensus networks (CNs) based on each 1001 trees from both runs of both single marker Bayesian analyses of the 2ISP-coded nuclear ITS and plastid datasets (which yielded the consensus trees shown in Figs. 2a and 1a). *Scale bar units* reflect the frequency of splits within the entered trees (COUNT option), trees proportion threshold for displaying splits = 0.33. **a** CN containing all accessions. *Square brackets* denote main species groups, Clades 1–18 are marked. *Dashed lines* highlight exemplary taxa supported within the indicated clade by one (the plastid) dataset only. *Asterisks* mark three species with intermediate positions between Clades 14 and 15. Hard incongruent accessions are highlighted by *dots*, APS-rich accessions by *crosses* (terms as defined in "Materials and methods" and "Results" sections). **b** CN with 28 hard incongruent accessions, the IPM clade, the Nodosa clade without *S. calycina*/*S. pauciflora*/*S. chlorantha*, and 17 APS-rich accessions removed using the "exclude selected taxa" option. *gorg gorganica*

from the matrix did not change the result ($p = 0.002$); incongruence between datasets apparently is not limited to well-supported conflicts. This is evident also from the consensus network constructed with Splitstree (Fig. 5a). Although the clades defined above can, at least partly, be traced in the CN, the relationships among them are inextricably entangled, often with similar numbers of trees supporting each of the conflicting splits. Bundles of parallel edges indicate conflicting signals or uncertainty also within several clades, and some are even split in the CN (examples see dashed lines in Fig. 5a). When HI accessions (labeled by black dots) were removed from the network, the number of parallel edges decreased considerably; the effect was much less pronounced when APS-rich accessions (indicated by black crosses) were excluded (not shown). But even upon exclusion of all marked accessions, a large number of reticulations remained in the consensus network (Fig. 5b, 99 non-trivial splits left compared to 167 in Fig. 5a).

Divergence time estimation

The chronogram from the strict clock analysis is available in Online Resource 6. Compared with the results from the BEAST run with the prior only, posterior distributions differed from prior distributions in most cases, indicating that the priors did not illegitimately influence the result. The effective prior distributions of the prior-only run matched the expected ones without any conflicts among priors. Figure 1a shows the respective chronogram, inferred by relaxed clock analysis of the reduced dataset, on the majority-rule consensus topology derived from the MrBayes runs. Node ages for all nodes alongside 95% highest posterior density (HPD) intervals are provided in Online Resource 7. The mean substitution rate was calculated at $7.8592E^{-4}$. The standard deviation of the uncorrelated log-normal clock and the coefficient of variation were 0.47 and 0.49 (95% HPD 0.26–0.71), indicating medium rate heterogeneity

within the dataset. The origins of the genus and its basal lineages can be traced back to the Miocene (23.1–11.2 million years ago, "mya", according to Fig. 1a); however, 11 of the 18 clades did not start to diversify before the beginning of the Pliocene (4.6–3.1 mya), and four clades even date to the Pleistocene (2.3–1.6 mya). Three of those belong to the Tomiophyllum clade, while old clades with crown ages in the late Miocene (7.7–6.1 mya) include the "Peregrina" (Clade 10), China and NW/Japan clades of the Anastomosantes group.

Ancestral area optimization

Results of the four independent RASP runs were largely similar; all runs, including that with maxareas set to five, reached convergence (standard deviations of split frequencies 0.0013–0.0018). In Fig. 1a, pie charts at nodes represent the inferred distributions; exact marginal probabilities for each range are available from Online Resource 7. Southwestern Asia, Turkey, and the Caucasus were inferred as the most likely ancestral areas of *Scrophularia* (I and J, Fig. 1a, node 3) with low marginal probability (32.70). The same ancestral distribution was inferred for the Tomiophyllum clade (node 43) with high probability. Some clades within the Tomiophyllum clade have other most probable ranges of origin, including the Levant which is here defined as present-day Lebanon, Syria, Israel, and Jordania ("Scariosa" clade 17, node 74), or Afghanistan to the Western Himalayas ("Polyantha" clade 15, node 82) with high frequencies of occurrence (89.22/68.72, respectively). The ancestor(s) of the New World taxa were reconstructed as distributed in Eastern Asia (node 24) with high marginal probability. The uncertainty in ancestral areas at nodes 61 and 63 argues for a more widespread occurrence of the most recent common ancestor, as also inferred by the five maxareas analysis (results in brackets in Fig. 1a).

Discussion

Implementation of ITS intra-individual polymorphism and chloroplast indel information

The genus *Scrophularia* provides a striking example of incongruence and ambiguity among, but also within, gene trees and the corresponding DNA sequences. Although tree reconstruction per se seems appropriate to adequately illustrate the tree-like parts of the genus' phylogenetic history, it is hampered by the widespread abundance of intra-individual site polymorphisms in the ITS sequences, which introduces considerable conflict into the dataset. The problem of reduced tree resolution when analyzing polymorphic sequences has been known for some time and has been noticed by several authors (e.g., Eidesen et al. 2007; Grimm et al. 2007), especially in cases where hybrids (or more generally, taxa bearing APS) were included into tree reconstructions (McDade 1995; Campbell et al. 1997; Whittall et al. 2000; see also the weakly resolved ITS phylogeny from dataset B in Fuertes Aguilar and Nieto Feliner 2003). Several suggestions have been made on how to deal with intra-individual polymorphisms. Among the most-used is pruning the respective taxa (e.g., Whittall et al. 2000; Fuertes Aguilar and Nieto Feliner 2003); another possibility is exclusion of polymorphic alignment positions (e.g., Scherson et al. 2008). However, in *Scrophularia*, attempts to infer a stable backbone using a representative subset of completely monomorphic sequences did not improve the result (not shown). Neither did the removal of APS-rich accessions and exemplary deletion of the 17 HPPs considerably reduced the resolution of the resulting phylogeny (see "Results" section). Further strategies include the replacement of polymorphisms by missing data or the most common nucleotide, their resolution in favor of the stronger signal (Fehrer et al. 2009), statistical haplotype phasing methods (e.g., Stephens et al. 2001, employed in Lorenz-Lemke et al. 2005), or cloning (see Nieto Feliner and Rosselló 2007; however, cloning all accessions may not be feasible in species-rich genera).

Rather than discarding or substituting sequence site variabilities, including them as phylogenetically informative characters seems to be a better solution. This can be achieved by the approach employed here, using 2ISP coding and the ad hoc implementation in ML from Potts et al. (2014) with additional adaptation of the method for BI. The procedure is straightforward to apply, does not require the use of step matrices, and allows the choice of appropriate DNA substitution models. As in Grimm et al. (2007) and Potts et al. (2014), recoding resulted in considerably increased resolution and enhanced support values in the phylogenetic tree: for instance, the "Libanotica" clade (Clade 12), corroborated by morphological similarities of its taxa (Boissier 1879; Grau 1981) is only recovered in the Bayesian topology using 2ISP coding (see Fig. 2a, node 58). Furthermore, the results obtained only very rarely contradict the topology generated without additional coding, but mostly strengthen existing clades. This corroborates the applicability of the method to the *Scrophularia* dataset.

The latter seems to be characterized by polymorphisms derived from several processes, including hybridization resulting in APS according to Fuertes Aguilar and Nieto Feliner (2003) but also others like e.g., inherited ancestral polymorphism (which might lead to ILS). Furthermore, homoeologous rDNA arrays might be subjected to differential silencing after interspecific hybridization. This can produce pseudogenes (Bailey et al. 2003; Volkov et al. 2007) recognizable by certain characteristics (Mayol and Rosselló 2001; Grimm and Denk 2008). Although G+C contents were slightly lowered in nine uncloned accessions, only one of them (*S. nervosa* Benth., 1846) showed a long edge in the NN

indicating increased distance (Fig. 4, Clade 18, see asterisk), and no ingroup accession had substitutions or length changes in the conserved parts of ITS1 or the 5.8 rDNA. However, it is still possible that incipient pseudogeny accounts for some of the observed sequence ambiguities. Finally, the rDNA marker used here can be affected by recombination (see Álvarez and Wendel 2003). No clear evidence of recombination was found in the sequences; however, respective patterns would be difficult to detect in a complex dataset like this (see below), which is additionally characterized by low sequence divergence. Consequently, an influence of recombination processes cannot be ruled out.

Due to the large number of polymorphisms, the processes involved in their formation cannot be distinguished in the *Scrophularia* dataset: informative 2ISPs are scattered across the alignment, with very few coherent mutation patterns identifiable among them. This also excludes the explicit detection of hybrids by comparing APS patterns as done by Fuertes Aguilar and Nieto Feliner (2003). Instead, APS were tentatively used for character coding, to assess their influence on the result. However, as reliable detection of all APS could not be achieved, only selected putative APS within 17 HPPs were coded. Analysis of the APS-coded dataset (results not shown) yielded an intermediate topology with respect to those from the 2ISP-coded and uncoded dataset. Supports of individual nodes matched those of one or the other dataset (e.g., nodes 10, 20, 68, 95, see Online Resource 7), were intermediate between them (nodes 19, 28, 40) or were worse or better (nodes 4, 33, 61b, 80a). This suggests that the approach is biased, and the results are dependent on the chosen subset of alignment columns and their APS. We thus conclude that restricting the coding procedure to APS only is not possible in *Scrophularia*. By contrast, the 2ISP-informative approach is particularly suitable for such datasets, as it does not discard polymorphisms based on their origin. The improvement in phylogenetic results seen here suggests that the information contained in intra-individual polymorphisms can be used even in cases where their sources are not exactly known, as also emphasized by Potts et al. (2014).

An inherent disadvantage of the approach is that if artificial ambiguities due to bad read quality are present, these are also coded and could possibly blur relationships. In the present study, discrimination of artificial double peaks from "real" polymorphism cannot be assumed to be completely reliable (although low-quality sequences are not too frequent), which was why all ambiguous bases were subjected to coding. In such cases, results regarding accessions with potential data quality issues, which have long branches in the coded tree compared to the uncoded tree, should be regarded with caution. From the six long-branch accessions marked in Fig. 2 (thumbnail trees; the respective accessions are marked by asterisks in the NN in Fig. 4), four can be assumed to be influenced by artificial ambiguities, while long branches seen in *S. lateriflora* Trautv., 1866, and *S. auriculata* from England rather also reflect real polymorphism. However, although cautionary interpretation regarding taxa with many artificial characters seems justified, the fact that both trees are largely congruent and that nodes strongly supported in the uncoded tree are only rarely weakened using 2ISP coding, indicates that the method is robust to a certain amount of "noise." Thereby, the tips of the phylogeny seem to benefit most from coding, while more basal relationships remain unchanged or may even collapse (Fig. 2b, nodes A and B, node 53b).

It is important to note that although 2ISP coding improves the resolution of the phylogenetic tree, it will not solve the problem of inherent conflict present within the data, among others incompatible signals after hybridization (Potts et al. 2014). Here, examination of the NN is useful as it represents all information and all conflict contained in the sequences (Bryant and Moulton 2004; Morrison 2010). Its highly interwoven structure as presented in Fig. 4 suggests that irrespective of the method applied, any bifurcating tree will suffer from an insufficiently resolved backbone as well as an amount of weakly supported nodes regarding certain relationships. Clades showing high amounts of ambiguity in the NN (Fig. 4) are likely to be sensitive to the type of analysis conducted (BI vs. ML), and will often remain insufficiently resolved (see Fig. 2a with inserts: Clade 15, Clade 7 with the "Auriculata" subclade, Clade 3, clade node 87, highlighted by a square bracket in Fig. 4, *S. ilwensis* K. Koch, 1844). This also means that no conclusions should be made based on weak Bayesian support values in these cases (compare e.g., the weakly supported sister clade of *S. rostrata* Boiss. & Buhse, 1860, and *S. sareptana* Kleop. ex Ivanina, 1972, to their positions in the NN, or the weak association of *S. olgae* Grossh., 1932, with the Libanotica clade 12).

It is remarkable that generally, results from ML are much less resolved (see Fig. 2); many nodes supported by BI are not found in the best-scoring ML tree or are supported by BS < 50. Furthermore, 2ISP coding does not improve the situation, it seems to have only very little effect. One possible reason for this is that ML reconstruction in RAxML per se is not completely naive concerning polymorphic sites as outlined in Potts et al. (2014), and that coding thus does not make much difference. The different way site ambiguities are handled are also visible from the different positions of the highly polymorphic *S. auriculata* from England. Another explanation might be that in a dataset with not too many, but clearly structured informative polymorphisms, bootstrapping more often produces deviating topologies while BI quickly achieves one "optimal solution."

Apart from the coding of ITS polymorphisms, additional information can also be drawn from plastid *trnQ-rps16* indels. This marker is particularly suitable for *Scrophularia*, regarding both its high information content in terms of parsimony-

informative characters (Table 1) as well as the occurrence of diagnostic indels as already described in Scheunert and Heubl (2011). Both NW/Japan (Clade 4) and Nodosa (Clade 5) clades are exclusively characterized by one particular indel of 597 bp length (Table 2). The same accounts for Clade 18, the "Orientalis" clade (764 bp), while a shorter indel (312 bp) occurs in Clade 15 (Polyantha). These length differences constitute a reliable diagnostic tool for the respective clades; quick discrimination can be achieved by a simple PCR reaction (see Fig. 2 in Scheunert and Heubl 2011). Other indels might be of limited phylogenetic value: the Scopolii clade is characterized by a 318-bp indel, however, in the respective region of the alignment (length types 2b–d, ranging roughly from position 518 to 845), indels also seem to arise spontaneously in single unrelated taxa or even accessions (Table 2, Fig. 1a).

Utility of gene tree discordance, cloned sequences, and Neighbor-Net splits graphs for tracing reticulate events in *Scrophularia*

In addition to intra-individual variability, many taxa obtained conflicting positions among nr and cp phylogenies. Moreover, the number of incongruent accessions is probably underestimated. As hard incongruence is by definition dependent on high supports of the associated nodes, it must be assumed that lowered resolution, caused by large amounts of PS in certain ITS sequences, prevented taxa from being recognized as hard incongruent. This is corroborated by the fact that the removal of HI accessions did not render the ILD test insignificant. Accordingly, the consensus network of both markers showed a considerable amount of reticulation even after excluding all HI accessions and APS-rich accessions (Fig. 5b). In consequence, weakly supported cases of incongruence should not be ignored completely (see for example the position of *S. chlorantha* between Clades 5 and 9/10).

Conflicts among datasets can be dealt with in various ways: ignoring the incongruence altogether (concatenation approach; Gadagkar et al. 2005; L-Y Chen et al. 2014; but see Rokas et al. 2003; Kubatko and Degnan 2007; Weisrock et al. 2012), pruning conflicting taxa prior to combined analysis (Huelsenbeck et al. 1996), or duplicating them (Pirie et al. 2008, 2009). While incongruence in the present dataset is far too widespread for pruning or using the taxon duplication approach, it is also quite obvious from the CN from ITS and plastid trees that combining both datasets in a concatenation approach would yield a rather uninformative tree. Comparison of individual plastid and nuclear gene trees including phylogenetic relationships of cloned sequences, and additional examination of the corresponding networks seems more suitable in this case.

Similar to the ambiguity present within the ITS dataset, the complex relationships found among the phylogenetic trees suggest that a combination of different processes has influenced the evolutionary history of *Scrophularia*. With the exception of *S. arguta*, conspecific accessions are never monophyletic, and sequences from three of the six cloned accessions group with other taxa. Unequivocal identification of evolutionary events like ILS and reticulation is not possible based on the current dataset, but some inferences can nevertheless be made. For example, while earlier studies have determined the placements of *S. auriculata* and *S. lyrata* within the Auriculata subclade (as in Fig. 2a, node 34), two of the accessions unexpectedly group with *S. scorodonia* L., 1753 and *S. laxiflora* Lange, 1878, in the cp tree. Although lineage sorting effects cannot be ruled out in this relatively young group, the observed pattern is more likely to reflect geographic proximity: on the Azores, where the accession from *S. auriculata* was sampled, *S. scorodonia* occurs as an introduced species, and the accessions from *S. lyrata* from Spain and *S. laxiflora* were collected app. 150 km apart, while *S. lyrata* (Crete) was collected as far as 2500 km away. Together with the fact that hybridization of *S. scorodonia* and *S. auriculata* is possible (Grau 1976; Dalgaard 1979), this suggests that the observed pattern might be due to introgressive hybridization. Interestingly, no APS were found in the ITS sequences of both accessions, which might be explained by repeated backcrossing towards *S. auriculata*/*S. lyrata* (chloroplast capture). Geographic patterns in plastid phylogenies, as opposed to those in ITS which are often morphology-corroborated, have been found in several other plant genera, including *Phlomis* (Lamiaceae; Albaladejo et al. 2005), *Mitella* (Saxifragaceae; Okuyama et al. 2005), or *Antirrhinum* (Plantaginaceae; Vargas et al. 2009). However, analysis of plastid data may also correctly infer relationships which are blurred in ITS. This is the case in e.g., the Scariosa clade, which is expected to include *S. hierochuntina* Boiss., 1853 based on morphological evidence (Boissier 1879, *Flora Orientalis*; Eig 1944; Grau 1980). Correct interpretation of phylogenetic relationships in *Scrophularia* therefore requires careful comparisons of all results.

Recent or ancient hybridization can result in deviating ITS copies (see above) as well as incongruence among markers (Fuertes Aguilar and Nieto Feliner 2003; Vriesendorp and Bakker 2005; Peng and Wang, 2008; Vargas et al. 2009), and obviously had an important impact on the speciation process in *Scrophularia* (as demonstrated in Scheunert and Heubl 2014). In such cases, cloned sequences are particularly useful as they can provide information on putative parent lineages (Fig. 3). In the species studied here, clones in general contained all variation corresponding to the extracted APS from the "direct sequences" (i.e., obtained from direct sequencing), and often even more. This means that cloning should be favored over direct sequencing as far as possible, especially when high polyploids are present and polymorphic sites are more easily missed (Joly et al. 2006). *Scrophularia auriculata* (with 2n = 84 chromosomes) was already proposed

to be an allopolyploid resulting from hybridization between *S. lyrata* or *S. hispida* Desf., 1798 (both 2n = 58) and *S. umbrosa* (2n = 26) or their ancestors (Grau 1979; Scheunert and Heubl 2014). The English voucher specimen of *S. auriculata*, with its chromosome number determined to match the typical number for the species, has as much as 13 PS in the direct sequence. Cloned sequences are separated into two distinct clades: six clones obtain a position similar to that of the accession in the cp tree; they are part of a clade also containing both direct sequences of *S. auriculata*, *S. racemosa* Lowe, 1831, and *S. hispida*. The fact that the direct sequence of *S. auriculata* England is sister to *S. hispida* using the 2ISP-coded dataset, and that none of the clones is found within the monophyletic clade of clones from *S. lyrata*, argues for *S. hispida* as potential parent. The remaining ITS clones are situated in a monophyletic clade at the basal polytomy of the IPM clade, but without depicting a sister relationship to *S. umbrosa*. This might suggest that the hybridization event is more ancient, with more time for accumulation of autapomorphisms in *S. auriculata*. In such cases, parental lineages would be more difficult to trace due to a greater accumulation of autapomorphies (Wolfe and Elisens 1994; Baumel et al. 2002; Vargas et al. 2009). The distinct status of *S. auriculata* England is also illustrated in the NN, where it is isolated in the Auriculata subclade and positioned closer towards the Scopolii clade (Clade 8), another potential parental lineage related to *S. umbrosa* (Fig. 4, species marked by an asterisk within the subclade). It remains unclear why traces of the hybrid ancestry of *S. auriculata* can be found in only one of the sampled accessions. However, the variable, intermingled phylogenetic relationships and weakly defined species boundaries among the closely related taxa of the IPM clade have not been fully recovered yet.

The Caucasian endemic *S. ruprechtii* is part of a predominantly Turkish-Caucasian clade in ITS. It was cloned to elucidate a possible hybrid origin concerning the Ningpoensis clade, with which it is closely associated in the cp tree, resulting in a clearly intermediate placement in the CN (Fig. 5a). However, although two diverging ITS ribotypes were found, neither of them bear signs of relationships with lineages from outside Tomiophyllum. Close relationships with *S. olympica* and *S. imerethica* Kem.-Nath., 1955, are corroborated by six of the clones, a clade of four clones remains unresolved.

The six species of the "Scoparia" clade (Clade 14) in ITS occur in partly overlapping distribution areas ranging from Central Asia across Afghanistan and the Western Himalayas to China and Siberia. They represent a rather homogeneous assemblage of subshrubs featuring few-flowered cymes, exserted stamens, linear staminodes, and leaves which are divided to various extents. *Scrophularia multicaulis* Turcz., 1840, a perennial species with included stamens, does not fit into this pattern, and also is the only species without any PS in the ITS sequences, while two (in *S. kiriloviana* Schischk., 1955) to eight (in *S. incisa*) PS are found in all other members. The clones from *S. incisa* reveal two main ribotypes, one of which is associated with *S. multicaulis*. Three of the clones remain unresolved within the clade. It seems reasonable to assume that *S. incisa* was the result of a hybridization event that involved *S. multicaulis* and a second member from the Scoparia clade, which would explain the morphological similarities. Shared PS in the ITS sequences connect *S. incisa* to *S. scoparia* Pennell, 1943, but also to *S. dentata* Royle ex Benth., 1835. Further evidence can be drawn from the occurrence of a sequence length polymorphism within ITS2, which is characteristic for the Scoparia clade and was excluded from calculations. The morphologically very variable *S. scoparia* (as all other sampled *Scrophularia* accessions) has a clear sequence containing a GTG motif, while *S. multicaulis* shows a clear sequence with a GTGTG motif at the respective position. *Scrophularia incisa* (as well as all other members of the clade) features a length polymorphism, with some ITS copies having GTG (here present in clones 6 and 15) and others having GTGTG. This might support the hypothesis that *S. scoparia*, which is likewise unresolved in Fig. 3, acted as second parent for *S. incisa*. Apart from the complex relationships within the clade, the Scoparia clade is also entangled with the Polyantha clade, whose species according to the ITS tree are distributed from Central Asia and Afghanistan southeastwards as far as the Eastern Himalayas. Three taxa (among those *S. incisa* as well as *S. multicaulis*) switch positions between the two clades in plastid and nuclear reconstructions, resulting in their intermediate position in the CN (Fig. 5a, marked by asterisks).

The origin of *Scrophularia*

Although divergence dating should rely on results from different cell compartments whenever possible, we refrained from performing molecular clock analyses on the ITS dataset. In *Scrophularia*, the respective sequences are highly polymorphic and considerably influenced by reticulation. This means that they cannot provide an entirely tree-like signal, which makes them unsuitable for molecular clock or ancestral area inferences. Results are therefore based on one, the plastid, marker only. However, it still should be kept in mind that estimation of divergence times might be impaired by the presence of a reticulate history. For example, for a hybrid lineage, one distinct time of divergence may not represent its true history which might have involved several periods of gene flow (Payseur and Rieseberg 2016) or several independent hybridization events. Furthermore, gene flow between two species might lead to underestimation of their divergence time, as demonstrated by Leaché et al. (2014) for species trees.

According to the molecular dating analysis, *Scrophularia* diverged from *Verbascum* around the Oligocene/Miocene

boundary (approximately 23 mya), with diversification of major lineages starting in the Miocene, within approximately the last 15 my. These results are comparable to those of a recently published time-calibrated phylogeny based on *ndhF* sequence data of Lamiales, where the authors used a combination of several fossil and secondary calibrations (Navarro Pérez et al. 2013). Ages as inferred here are in clear contrast with considerably older divergence times obtained in a small-scale study on New World species based on secondary calibration of the root only (Scheunert and Heubl 2011). This is probably due to too small taxon sample size and the choice of method and information source for obtaining secondary calibration points. Ancestral area reconstructions revealed that *Scrophularia* originated in a region comprising Southwestern Asia and Turkey (Fig. 1a, node 3), which corresponds to its present-day primary center of diversity. This contradicts Stiefelhagen (1910), who promoted the Himalaya as ancestral region for the genus.

In the Miocene, Southwestern Asia was under strong influence of processes initiated by the collision of the Arabian with the Eurasian plate, which had started in the late Eocene and led to the formation of the main mountain ranges within the region (Djamali et al. 2012a). During that period, Asian biota also were affected by substantial climatic and environmental changes, including a shift towards increased seasonality (An et al. 2001) and the aridification of interior Asia (Guo et al. 2002, 2004; Fortelius et al. 2006; Miao et al. 2012). A combination of these processes is likely to have triggered the divergence and diversification of *Scrophularia*. At the time inferred for the origin of the genus, the final closure of the Tethyan seaway and the retreat of the Paratethys ocean (Ramstein et al. 1997; Mouthereau et al. 2012) created additional suitable land area for plant colonization in Southwestern Asia, e.g., in the Zagros region. While the Greater Caucasus was still largely submerged at that time, the Alborz and Anatolian highlands already existed for a longer time (Popov et al. 2004). Also, according to time estimates presented here, diversification of *Scrophularia* into its main lineages began during a period where new suitable habitats for species adapted to montane environments came into existence in its ancestral region; by the emergence of the Greater Caucasus and its transformation into a mountain chain (approximately 14–13 mya and 9 mya; Meulenkamp and Sissingh 2003; Popov et al. 2004; Olteanu and Jipa 2006), the uplift of the Iranian Plateau (from 15–12 mya; Mouthereau et al. 2012; Djamali et al. 2012a and references therein), and the formation of the Western Alborz (approximately 12–10 mya) and Zagros (approximately 15–10 mya) mountains (Dercourt et al. 1986; Popov et al. 2004; Guest et al. 2007; Mouthereau et al. 2012). Additionally, the establishment of continental climate conditions (Berberian and King 1981) with increasing aridity (Ballato et al. 2010) during this period could have triggered the colonization of more moderate, humid habitats of higher elevations (Roe 2005).

Establishment of the genus in mountainous regions might have been accelerated by its potential for frequent and successful interspecific hybridization and (allo-)polyploidization. Both evolutionary processes enable adaptation to new or hitherto unsuitable environments by the rapid formation of new phenotypic combinations, promoting adaptive radiation and colonization, and thus have been associated with speciation events and diversification (Stebbins 1959; Seehausen 2004; Mallet 2007; Estep et al. 2014). While *Verbascum* resembles its sister *Scrophularia* in many respects including habitat requirements and a Southwestern Asian center of diversity, the former genus apparently lacks the aforementioned peculiarities. Although species easily hybridize wherever they grow together, the resulting offspring is always sterile according to Murbeck (1933) and Huber-Morath (1978). Reported chromosome numbers within *Verbascum* mostly range between 2n = 30–36, with the highest count being 2n = c. 64 (Goldblatt and Johnson 1979–). In contrast, *Scrophularia* features numbers from 2n = 18 (in *S. altaica* Murray, 1781) up to 2n = 96 (in the NW/Japan clade), with several clades being characterized by (high) polyploidy (Shaw 1962; Goldblatt and Johnson 1979–; Scheunert and Heubl 2014). It seems worth mentioning that the highest polyploids within *Scrophularia* are concentrated in regions where only few or no *Verbascum* species occur (i.e., China and the New World). Carlbom (1969) concluded that polyploidy in *Scrophularia* has triggered migration into the rough environments of higher altitudes as well as regions of higher latitudes. The predisposition for polyploidy and hybridization found in *Scrophularia* is shared by other lineages within Scrophulariaceae (e.g., *Diascia* or *Nemesia*; Steiner 1996; Datson et al. 2006) and related families (e.g., Clay et al. 2012 and references therein; Rojas Andrés et al. 2015).

Apart from that, mountainous regions themselves promote plant species diversification: diversification rates have been shown to be elevated in regions subjected to active orogeny (for a short overview see Hoorn et al. 2013 and references therein), and mountain ranges provide complex, heterogeneous habitats on small geographical scales, which facilitate adaptive speciation and allopatric divergence (Lobo 2001; Djamali et al. 2012a and b and references therein; Wen et al. 2014), and can also stabilize new hybrids by isolating them from their parents. These conditions might be assumed to also have produced the high levels of endemicity in *Scrophularia* as mentioned by Vaarama and Hiirsalmi (1967). Even in the absence of detailed distribution mapping, it seems likely that a combination of geographic isolation/habitat fragmentation and successful hybridization has been the key factor in the diversification of *Scrophularia*. This has also been suggested for other genera occurring in the Tibetan Plateau or mountainous regions of the Mediterranean (e.g., Senecioneae,

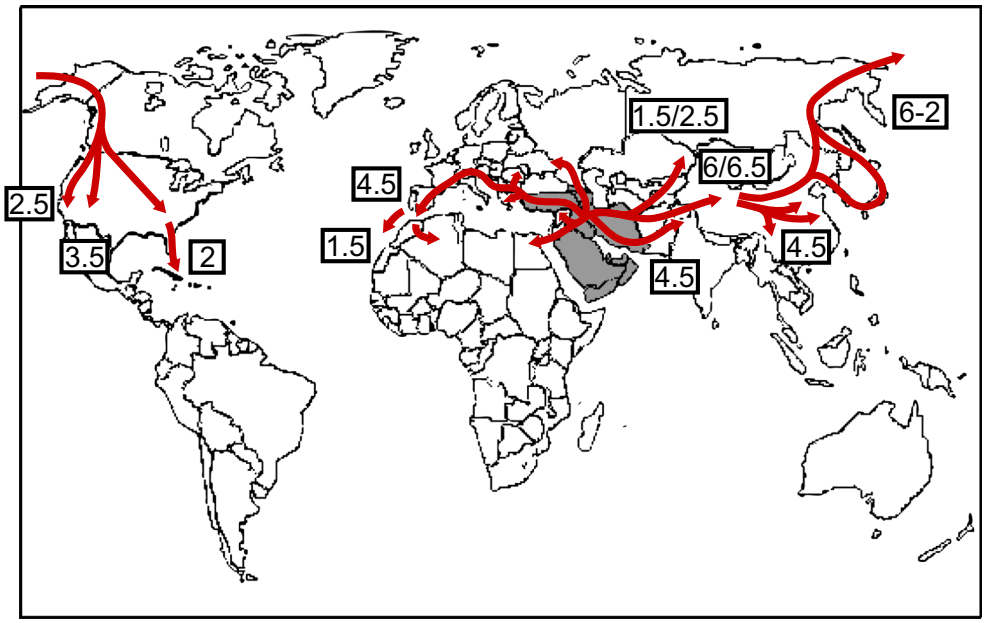


Fig. 6 Biogeographic history of *Scrophularia*. The ancestral region of the genus is marked in *gray*; *red arrows* indicate main dispersal routes as inferred from plastid DNA data. Note that arrows in some cases may illustrate more than one dispersal. *Boxes next to arrows* denote the approximate time of the migration event in million years ago, drawn from crown ages of the respective clades

Asteraceae, J-Q Liu et al. 2006; *Linaria*, Plantaginaceae, Blanco-Pastor et al. 2012; *Meconopsis*, Papaveraceae, Yang et al. 2012; *Rhodiola*, Crassulaceae, Zhang et al. 2014).

A remarkable result in the ITS tree is the basal position of the annual *S. arguta* as sister to the remainder of the generally persistent *Scrophularia* (Fig. 2a, node 3). However, upon a closer look, this position might likely be artificial. Potentially higher evolutionary rates in annuals can result in support of a sister relationship of the former to their perennial relatives (e.g., Laroche and Bousquet 1999; Andreasen and Baldwin 2001; Tank and Olmstead 2008; Müller and Albach 2010; J-X Yue et al. 2010). But, although rarely, other annual species are found in *Scrophularia* which obtain inconspicuous placements, e.g., in Clade 6 or 10 (*S. calycina* Benth., 1835, *S. peregrina* L., 1753). More importantly however, and unlike most of the other *Scrophularia* species, self-pollination is common in *S. arguta*, and the species is unique within the genus in possessing a mixed mating system of chasmogamous flowers occasionally complemented by small cleistogamous flowers near the ground. High substitution rates have been correlated with a selfing breeding system (Glémin et al. 2006), although this is subject of debate (Wright et al. 2002; Müller and Albach 2010). Moreover, mating shifts seem to create strong interspecies isolation barriers which effectively prevent hybridization (Wright et al. 2013). The Arguta lineage might thus owe its isolated position to the accumulation of mutations due to breeding system effects. It is likely that the particular reproductive traits of *S. arguta* have enabled colonization of habitats otherwise unsuitable for *Scrophularia*. In the hot and dry environments of e.g., the Sudan, Eritrea, Somalia and Oman, *S. arguta* consequently is the only representative of the genus. In the plastid tree, the species is shown as an earlier diverging lineage and is sister to the Tomiophyllum clade. Its ancestral area could not be satisfactorily determined; analyses with a maximum of five areas yielded an ancestral range corresponding to its present-day distribution (areas ABDEI; Fig. 1a, node 85). A recent study on *S. arguta*, based on representatives from several populations although not covering the whole distribution range, found evidence for a westward expansion from the east of its distribution range (Valtueña et al. 2016), which reaches its limit on the Arabian Peninsula.

Reconstruction of major evolutionary events

Based on individual clade ages and ancestral areas (Fig. 1a), several expansions of the genus in various directions can be hypothesized (Fig. 6). Eastward migration events to China and the Tibetan Plateau region are inferred at approximately 6 mya (nodes 10 and 24) and later. Given the preference for mountainous habitats in *Scrophularia*, dispersal along higher mountain chains as corridors for colonization seems reasonable. Evidence for growth of the Kunlun Shan in the northern part of the Tibetan Plateau has been reported since Eocene times (see Yuan et al. 2013), while uplifts of the Tian Shan located to the northwest presumably occurred from the late Oligocene through the Miocene until approximately 7 mya or later (Abdrakhmatov et al. 1996; Sobel et al. 2006; citations in Miao et al. 2012). The Hindu Kush, which spreads farther west into Afghanistan, underwent uplift around the Oligocene-Miocene boundary (Hildebrand et al. 2000), with tectonic processes in the region occurring much earlier (Dercourt et al. 1986; Hildebrand et al. 2001). The Kopet Dag finally, roughly situated between the Hindu Kush in the east and the Alborz of Iran in the west, only re-emerged as a mountain

chain at approximately 10 mya or later, after submergence following higher-altitude phases from the late Oligocene until the early Miocene (Dercourt et al. 1986; Popov et al. 2004). Altogether, this means that by the time inferred for the first eastward migrations of figworts, a more or less continuous mountain belt should have existed, which connected the Southwestern Asian Alborz to the Himalayas and the Tibetan Plateau, and presumably provided a suitable pathway for dispersal of *Scrophularia*. Transitions of this kind from Western/Central Asia to Eastern Asia (or vice versa) have been reported in several other genera, including *Incarvillea* (Bignoniaceae; S-T Chen et al. 2005), *Rhodiola* (Zhang et al. 2014), or *Solmslaubachia* (Brassicaceae; J-P Yue et al. 2009); the latter genus inhabits alpine scree-slope habitats similar to several *Scrophularia* species.

China has been colonized at various times and by different lineages; the most important secondary diversity center of *Scrophularia* harbors, among others, species from the "Vernalis" (Clade 9) and Calycina clades, *S. umbrosa*, and species from the Scoparia/Polyantha clade of Tomiophyllum. These mainly alpine taxa have extended their distribution areas from Central Asia, Siberia, or Southern Asia into China. In contrast, the China and Ningpoensis clades almost exclusively consist of Chinese species. When disregarding altitude, Chinese *Scrophularia* have rather similar habitat requirements (especially within the China clade s.str. as shown in Fig. 2); they occur in humid conditions in mountainous forests or grasslands, often in crevices and among rocks.

The Ningpoensis clade (Clade 2), with its crown age estimated at approximately 3 my, constitutes an eastern group of mainly non-alpine taxa including the pharmaceutically important *S. ningpoensis* Hemsl., 1899. Its distribution extends from the eastern parts of China to Korea, Japan, and Taiwan; repeated contacts between those landmasses from the late Miocene onwards offered opportunities for dispersal (references in Qiu et al. 2011). In contrast, the taxa of the China clade s.str. (Clade 3 in the ITS tree) are mostly alpine species with mostly narrow distributions. Most species from the "Hengduan" subclade resolved in the plastid tree are restricted to the Hengduan mountains (located in parts of Sichuan and Yunnan provinces with adjacent Tibet), which are considered one of the world's biodiversity hotspots (Mittermeier et al. 2004). The "Central" subclade comprises high-alpine and subalpine species distributed in central parts of China.

Crown ages of both subclades were estimated at about 4.5 my during the Pliocene. Many authors have attributed diversification events in the region to the uplift of the Tibetan Plateau (see review by Qiu et al. 2011 and references in Y-S Sun et al. 2012), often without checking for exact spatial and temporal concordance among inferred divergence times and geological events. This however seems to be indispensible given the complex geological history of the Tibetan Plateau and the ongoing debates on appropriate models for its formation (Yuan et al. 2013; C-S Wang et al. 2014; J-J Li et al. 2015). Diversification of the Hengduan subclade from the early Pliocene on coincides with a period of uplift hypothesized for the Hengduan Shan by B-N Sun et al. (2011) and Ming (2007). However, it may not be necessary to invoke uplift as a cause for diversification when its result seems to be more important: the extreme topography with alternating high peaks and deep ridges and the variety of vegetation types and climatic conditions which characterize this biodiversity hotspot (see Boufford et al. in Mittermeier et al. 2004) effectively stimulates speciation and leads to high species diversity and endemism. Generally, high species numbers have been linked with the "extreme physiographical heterogeneity of temperate eastern Asia" by Qian and Ricklefs (2000). Finally, the present-day distribution patterns of the Hengduan and Central subclade species might also be the result of recolonization after the Last Glacial Maximum (approximately 24,000–18,000 years ago) from suitable refugial areas in the region; such have been recognized in the Hengduan Shan (e.g., for *Metagentiana*, Gentianaceae, S-Y Chen et al. 2008; *Angelica*, Apiaceae, Feng et al. 2009; *Lepisorus*, Polypodiaceae, L Wang et al. 2011) and, for the Central subclade, in the Qinling mountains or, more generally, the "Northeast Qinghai-Tibetan Plateau edge" (Qiu et al. 2011).

According to our reconstructions, North America was colonized up to three times independently from Eastern Asia (Figs. 1a, 2). The Japanese and most of the New World taxa are connected to the Nodosa clade (plastid tree) as well as taxa from the China clade, with whom they share a clade in the uncoded ITS tree (Fig. 2b, node A) and the coded ITS+clones tree (Fig. 3). Divergence from the Asian ancestors has happened approximately 6 mya at the earliest (Fig. 1a, node 24). This corresponds to a general Eastern Asian-North American disjunct pattern found in many plant genera, e.g., *Picea* (Pinaceae; Lockwood et al. 2013), *Gleditsia* (Fabaceae; Schnabel et al. 2003), *Triosteum* (Caprifoliaceae; Gould and Donoghue 2000), and also Scrophulariaceae (Hong 1983), among many others (see H-L Li 1972; Boufford and Spongberg 1983; Hong 1993; Xiang et al. 1998; Wen et al. 2010, 2014). Similar divergence times of North American from Eastern Asian lineages have been reported in *Kellogia* (Rubiaceae, 5.42 ± 2.32 mya; Nie et al. 2005) and *Rhodiola* (5.3 mya, 95% HPD 2.3–9.1 mya; Zhang et al. 2014). Both studies suggest long-distance dispersal for colonization of North America. Zhang et al. (2014) additionally hypothesize migration across the Bering Land Bridge (BLB; Tiffney and Manchester 2001); this was also proposed for *Angelica genuflexa* max. 4.3 mya, by Liao et al. (2012). During which periods and how long the BLB was available for plant migrations remains a matter of debate. Estimates for the opening of the Bering Strait range from 3.3 to 9 mya (Brigham-Grette

2001; Denk et al. 2011); however, even after the final flooding, intermittent short-time closures of the Strait have been assumed, among others at 4.9, 4.0, 3.3, and 2.5 mya (KG Miller et al. 2005). Cold-adapted *Scrophularia* taxa could have spread to the New World during times when landmasses were connected; long-distance dispersal however is likely to have played a role as well, especially during later periods (see Fig. 1a, nodes 25, 26, 28). *Scrophularia* seeds do not possess special adaptations favoring any mode of dispersal; however, they are easily dispersed by wind due to their small size and weight, and thus may not have been dependent on suitable land bridges.

In accordance with results obtained by Scheunert and Heubl (2011), the New World taxa of *Scrophularia* are divided into three geography-based clades, whose distribution ranges do not overlap much and which receive stronger supports in the plastid tree. Most of the North American species are closely related; they have been successfully hybridized (Shaw 1962) or even intergrade naturally in contact zones (e.g., *S. parviflora* Wooton & Standl., 1913 and *S. californica* Cham. & Schltdl., 1827; Kearney and Peebles 1951). Shaw (1962) emphasized that rather than reproductive isolation, geographic barriers seem to play an important role in maintaining the species (compare Carlbom 1969), an assumption that exactly fits the general diversification mechanisms discussed above and is corroborated by the characteristics of the three clades found here. A similar situation was reported for *Jamesbrittenia*, another Scrophulariaceae genus prone to successful interspecific hybridization, where geography helps to maintain species identity (Verboom et al. 2016). Species diversity is greater in the west (10 species) than in the east of the North American mainland (two species), possibly due to greater geographic heterogeneity in the former (Qi and Yang 1999) but also the California floristic province biodiversity hotspot located there (Mittermeier et al. 2004).

Apart from eastward migrations by *Scrophularia*, resulting in the China and New World clades, westward movements from the ancestral region led to the colonization of the Mediterranean, Northern Africa, and Europe. Scheunert and Heubl (2014) recently found that the IPM clade (Fig. 2a, node 32/Fig. 1a, node 45), which comprises the majority of Iberian and Macaronesian species, is of hybrid origin, involving progenitors both of the Scopolii clade or *S. umbrosa*, and the "Canina" clade (Clade 11) or allies (Fig. 2a, node 31 or 32/Fig. 1a, node 44). According to reconstructions using the plastid dataset, these ancestors were distributed in Southwestern Asia and in the Turkey-Caucasus region, respectively (Fig. 1a, node 40 or 86/44). The approximate time of the hybridization event, limited by divergence from the parental lineage and diversification of the hybrid lineage, is assumed at around 5 mya; the ancestor of the IPM clade was most likely distributed in the Western Mediterranean and diversified from about 4.5 mya (Fig. 1a, node 45; Scheunert and Heubl 2014). The role of the Irano-Turanian floristic region as a key source for colonization of the Mediterranean has been emphasized (Comes 2004; Djamali et al. 2012b), especially for temperate elements (Quézel 1985; Thompson 2005; Mansion et al. 2008).

The Tomiophyllum clade

The relative ages of the two main lineages within *Scrophularia* have been discussed by various authors. *Scrophularia* sect. *Tomiophyllum* might be regarded as "primitive" due to putatively ancestral traits like the often xerophytic, subshrubby habit of its members and the general lack of polyploid chromosome numbers (Carlbom 1969). The section is centered in the Caucasus, Iran, Iraq, and Turkey, with approximately half of the sampled taxa distributed there, and is completely absent from Macaronesia and the New World. On the other hand, the mainly herbaceous, richly foliated, often meso- or hygrophytic members of *S.* sect. *Anastomosantes* (Stiefelhagen 1910) are characterized by a large number of polyploid species, a wide ecological amplitude and a geographic distribution which exceeds that of *S.* sect. *Tomiophyllum* by far. Regarding molecular results, the Tomiophyllum lineage is highly supported as a distinct clade in both analyses and is nested within clades of Anastomosantes taxa (Figs. 1a and 2a). This reveals *S.* sect. *Tomiophyllum* to be derived from within *S.* sect. *Anastomosantes*.

The factors leading to the evolutionary success of the Tomiophyllum lineage remain uncertain. One might speculate that changes in aridity in its ancestral region during the second half of the Miocene and later (Ballato et al. 2010) had an influence on its divergence and diversification (which started approximately 8 mya). While ecological preferences of members of *Scrophularia* sects. *Anastomosantes* and *Tomiophyllum* overlap, their habitats reveal a certain shift from moist sites on riverbanks and in forests (in the former) towards rock crevices and gravelly substrates with low humidity in the latter, illustrating a greater tolerance of dry conditions. This is also reflected in differences in their respective distributions, with Tomiophyllum species predominantly inhabiting the dry parts of e.g., Iran and Turkey, while not necessarily being absent from other areas. Several of the xerophytic species also take advantage of the lower temperatures at higher altitudes; the few desert representatives tend to flower during early spring, with their rhizomes persisting in crevices during hot periods.

It is noteworthy that the phylogenetic (and also morphological) boundaries between the Anastomosantes and Tomiophyllum groups are also not altogether strict. For example, the position of *S. megalantha* Rech.f., 1955 in the NN, with no unequivocal connection to any of the Anastomosantes clades and closer to the center of the graph (Fig. 4), indicates a certain affinity to the Tomiophyllum clade. This is also supported by the trees in Figs. 1a and 2a, however, with

consistently weak supports. On the other hand, similarities to *S.* sect. *Anastomosantes* can be found in some members of the Tomiophyllum clade, e.g., in the mainly Turkish and Caucasian *S. ilwensis*, which was considered part of *S.* sect. *Anastomosantes* by Stiefelhagen (1910) and occurs in more humid habitats like forests or near water. While plastid reconstructions of the Tomiophyllum clade support a main split into two, once more geography-based, main clades, this species is sister to the remainder of Tomiophyllum in the ITS trees (Fig. 2). Its distinct status is reflected also in its intermediate position in the NN (Fig. 4). Apart from *S. ilwensis*, 14 further taxa (including the "Striata" (Clade 16) and Scariosa clades) obtain basal positions within the Tomiophyllum clade in the uncoded ITS tree. In the 2ISP-coded tree, these species form a clade; both placements are however weakly supported (Fig. 2a, node 87; Fig. 2b, nodes 52 and 53a) due to inherent ambiguity in the data as discussed above. The NN (Fig. 4, square bracket) clearly depicts the complex relationships of this assemblage and also a certain shift towards Anastomosantes. *Scrophularia nabataeorum* Eig, 1944 from the Scariosa clade indeed features morphological traits typical for *S.* sect. *Anastomosantes* as well: Eig (1944) described its ambiguous characteristics in between the two main sections and stated himself being "undecided as to the affinity of this species." Interestingly, the Striata clade comprises some of the few species that have colonized truly arid environments, occurring in steppe and desert habitats.

Both Striata and Scariosa clade give evidence for migration into areas west and southwest of present-day Iran, which was facilitated since the dry-up of the Mesopotamian Basin in the late Miocene (Popov et al. 2004). This has led to the colonization of the Levant, the Arabian Peninsula south of Iraq, and also Eastern North Africa by members of these clades. Notably, while several IPM clade species also extend into (or are endemic for) western regions of North Africa, the three species confined to its eastern part are found within the Striata and Libanotica clades. Other species occur throughout Northern Africa (*S. canina*, *S. peregrina*, *S. syriaca* Benth., 1846, *S. arguta*).

Conclusions

The present paper represents the first comprehensive phylogenetic study of the genus *Scrophularia*, based on a broad taxon sampling including representatives of all sections. This study has confirmed the monophyly of the genus but has also provided evidence for significant phylogenetic incongruence and ambiguity among and within sequence datasets. Exemplary cloning of taxa showed that intra-individual site polymorphism in ITS is widespread. Our study suggests that conflicting signals in *Scrophularia* derive from a variety of sources, most importantly reticulation, due to frequent hybridization and introgression. The methodical workflow as presented here is suitable for any plant group where similar problems are encountered and laborious search for (potentially likewise problematic; Nieto Feliner and Rosselló 2007) low-copy nuclear markers or cloning of all taxa cannot be considered.

The molecular phylogenies revealed two large groups of species (of which only one is monophyletic) corresponding to previously described taxonomic entities. The emergence of *Scrophularia* in the Miocene and its diversification are closely linked to geological and climatic events in the Irano-Turanian region and Central/Eastern Asia. Most diversification events as well as further successful dispersals to other regions were dated to the colder Pliocene-Pleistocene period.

The inferred spatio-temporal framework will provide a solid basis for future studies focusing on specific clades or morphological questions. It can be assumed that the considerable morphological variability is linked to the complex evolutionary history of the genus; this is relevant also regarding previous taxonomic concepts, which need to be re-evaluated. A survey of the relevant morphological traits, together with karyological analyses, will complement the present study from a more taxonomic perspective, also with respect to the small Himalayan genus *Oreosolen*, which has to be transferred to *Scrophularia* (Scheunert and Heubl in preparation).

Acknowledgements The authors wish to thank the herbaria and curators of A, B, E, GH, HAL, KUN, M, MA, MSB, W, and WU, for providing loans, for access to the specimens for study and for help in obtaining leaf material. Dirk Albach, Cheng-Xin Fu, Mark Mayfield, Søren Rosendal Jensen, Andrej Sytin, and Lang-Ran Xu are acknowledged for sending plant material from specimens deposited in GOET, HU/HZU, KSC, HSNU, LE, and WUK. Christian Bräuchler, Matthias Erben, and the Botanical Garden Munich kindly contributed seeds or plants required for study. We are grateful to Dirk Albach for sending photos and specimens and for continued support and helpful suggestions; to Guido Grimm for the generous time given for help with and discussion of the ITS data; to Susanne Renner and Lars Nauheimer for advice in divergence dating and BEAST; and to Ingo Michalak for help with RAxML and substitution model details. Alastair Potts and three anonymous reviewers are acknowledged for helpful comments which improved the manuscript. We thank Tanja Ernst for invaluable help in the lab and Wei Jie for providing translations from Chinese language.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflicts of interest.

This article does not contain any studies with human participants or animals performed by any of the authors.

References

Abbott, R., Albach, D., Ansell, S., Arntzen, J. W., Baird, S. J. E., Bierne, N., et al. (2013). Hybridization and speciation. *Journal of Evolutionary Biology, 26*, 229–246.

Abdrakhmatov, K. Y., Aldazhanov, S. A., Hager, B. H., Hamburger, M. W., Herring, T. A., Kalabaev, K. B., et al. (1996). Relatively

recent construction of the Tien Shan inferred from GPS measurements of present-day crustal deformation rates. *Nature, 384*, 450–453.

Albach, D. C., Meudt, H. M., & Oxelman, B. (2005). Piecing together the "new" Plantaginaceae. *American Journal of Botany, 92*(2), 297–315.

Albaladejo, R. G., Fuertes Aguilar, J., Aparicio, A., & Nieto Feliner, G. (2005). Contrasting nuclear-plastidial phylogenetic patterns in the recently diverged Iberian *Phlomis crinita* and *P. lychnitis* lineages (Lamiaceae). *Taxon, 54*, 987–998.

Álvarez, I., & Wendel, J. F. (2003). Ribosomal ITS sequences and plant phylogenetic inference. *Molecular Phylogenetics and Evolution, 29*, 417–434.

An, Z.-S., 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.

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. *Molecular Biology and Evolution, 18*(6), 936–944.

Arnheim, N., Krystal, M., Schmickel, R., Wilson, G., Ryder, O., & Zimmer, E. (1980). Molecular evidence for genetic exchanges among ribosomal genes on nonhomologous chromosomes in man and apes. *Proceedings of the National Academy of Sciences of the United States of America, 77*(12), 7323–7327.

Attar, F., Riahi, M., Daemi, F., & Aghabeigi, F. (2011). Preliminary molecular phylogeny of Eurasian *Scrophularia* (Scrophulariaceae) based on DNA sequence data from *trnS-trnG* and ITS regions. *Plant Biosystems, 145*(4), 857–865.

Bailey, C. D., Carr, T. G., Harris, S. A., & Hughes, C. E. (2003). Characterization of angiosperm nrDNA polymorphism, paralogy, and pseudogenes. *Molecular Phylogenetics and Evolution, 29*, 435–455.

Baldwin, B. G., Sanderson, M. J., Porter, J. M., Wojciechowski, M. F., Campbell, C. S., & Donoghue, M. J. (1995). The its region of nuclear ribosomal DNA: a valuable source of evidence on angiosperm phylogeny. *Annals of the Missouri Botanical Garden, 82*(2), 247–277.

Ballato, P., Mulch, A., Landgraf, A., Strecker, M. R., Dalconi, M. C., Friedrich, A., et al. (2010). Middle to late Miocene Middle Eastern climate from stable oxygen and carbon isotope data, southern Alborz mountains, N Iran. *Earth and Planetary Science Letters, 300*, 125–138.

Baumel, A., Ainouche, M. L., Bayer, R. J., Ainouche, A. K., & Misset, M. T. (2002). Molecular phylogeny of hybridizing species from the genus *Spartina* Schreb. (Poaceae). *Molecular Phylogenetics and Evolution, 22*(2), 303–314.

Beiko, R. G., Doolittle, W. F., & Charlebois, R. L. (2008). The impact of reticulate evolution on genome phylogeny. *Systematic Biology, 57*(6), 844–856.

Berberian, M., & King, G. C. P. (1981). Towards a paleogeography and tectonic evolution of Iran. *Canadian Journal of Earth Sciences, 18*, 210–265.

Blanco-Pastor, J. L., Vargas, P., & Pfeil, B. E. (2012). Coalescent simulations reveal hybridization and incomplete lineage sorting in Mediterranean *Linaria*. *PloS One, 7*(6), e39089.

Boissier, P. E. (1879). *Flora orientalis, sive enumeratio plantarum in Oriente a Graecia et Aegypto ad Indiae fines hucusque observatarum, 4, 2*. Geneva: H. Georg.

Boufford, D. E., & Spongberg, S. A. (1983). Eastern Asian – Eastern North American phytogeographical relationships - a history from the time of Linnaeus to the twentieth century. *Annals of the Missouri Botanical Garden, 70*, 423–439.

Brigham-Grette, J. (2001). New perspectives on Beringian Quaternary paleogeography, stratigraphy, and glacial history. *Quaternary Science Reviews, 20*, 15–24.

Brown, D. D., Wensink, P. C., & Jordan, E. (1972). A comparison of the ribosomal DNA's of *Xenopus laevis* and *Xenopus mulleri*: the evolution of tandem genes. *Journal of Molecular Biology, 63*, 57–73.

Bryant, D., & Moulton, V. (2004). Neighbor-Net: an agglomerative method for the construction of phylogenetic networks. *Molecular Biology and Evolution, 21*(2), 255–265.

Campbell, C. S., Wojciechowski, M. F., Baldwin, B. G., Alice, L. A., & Donoghue, M. J. (1997). Persistent nuclear ribosomal DNA sequence polymorphism in the *Amelanchier* agamic complex (Rosaceae). *Molecular Biology and Evolution, 14*, 81–90.

Carlbom, C. G. (1964). *An experimental taxonomic study of a Northwest American polyploid species, Scrophularia lanceolata Pursh*. Corvallis: Oregon State University.

Carlbom, C. G. (1969). Evolutionary relationships in the genus *Scrophularia* L. *Hereditas, 61*, 287–301.

Chen, L.-Y., Zhao, S.-Y., Mao, K.-S., Les, D. H., Wang, Q.-F., & Moody, M. L. (2014). Historical biogeography of Haloragaceae: an out-of-Australia hypothesis with multiple intercontinental dispersals. *Molecular Phylogenetics and Evolution, 78*, 87–95.

Chen, S.-T., Guan, K.-Y., Zhou, Z.-K., Olmstead, R., & Cronk, Q. (2005). Molecular phylogeny of *Incarvillea* (Bignoniaceae) based on ITS and *trnL-F* sequences. *American Journal of Botany, 92*, 625–633.

Chen, S.-Y., Wu, G.-L., Zhang, D.-J., Gao, Q.-B., Duan, Y.-Z., Zhang, F.-Q., et al. (2008). Potential refugium on the Qinghai–Tibet Plateau revealed by the chloroplast DNA phylogeography of the alpine species *Metagentiana striata* (Gentianaceae). *Botanical Journal of the Linnean Society, 157*, 125–140.

Clay, D. L., Novak, S. J., Serpe, M. D., & Tank, D. C. (2012). Homoploid hybrid speciation in a rare endemic *Castilleja* from Idaho (*Castilleja christii*, Orobanchaceae). *American Journal of Botany, 99*(12), 1976–1990.

Comes, H. P. (2004). The Mediterranean region - a hotspot for plant biogeographic research. *New Phytologist, 164*, 11–14.

Dalgaard, V. (1979). Biosystematics of the Macaronesian species of *Scrophularia*. *Opera Botanica, 51*, 1–64.

Datson, P. M., Murray, B. G., & Hammett, K. R. W. (2006). Pollination systems, hybridization barriers and meiotic chromosome behaviour in *Nemesia* hybrids. *Euphytica, 151*, 173–185.

Datson, P. M., Murray, B. G., & Steiner, K. E. (2008). Climate and the evolution of annual/perennial life-histories in *Nemesia* (Scrophulariaceae). *Plant Systematics and Evolution, 270*, 39–57.

Davis, P. H., Mill, R. R., & Tan, K. (1988). *Flora of Turkey and the East Aegean Islands, 10* (Suppl. 1). Edinburgh: Edinburgh University Press.

Degnan, J. H., & Rosenberg, N. A. (2009). Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends in Ecology and Evolution, 24*(6), 332–340.

Denk, T., Grimsson, F., Zetter, R., & Simonarson, L. A. (2011). *Late Cainozoic floras of Iceland: 15 million years of vegetation and climate history in the northern North Atlantic*. Dordrecht: Springer.

Dercourt, J., Zonenshain, L. P., Ricou, L.-E., Kazmin, V. G., Le Pichon, X., Knipper, A. L., et al. (1986). Geological evolution of the Tethys belt from the Atlantic to the Pamirs since the Lias. *Tectonophysics, 123*, 241–315.

Djamali, M., Baumel, A., Brewer, S., Jackson, S. T., Kadereit, J. W., López-Vinyallonga, S., 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, 207*, 237–249.

Drummond, A. J., & Rambaut, A. (2007). BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology, 7*, 214.

Eidesen, P. B., Alsos, I. G., Popp, M., Stensrud, O., Suda, J., & Brochmann, C. (2007). Nuclear vs. plastid data: complex Pleistocene history of a circumpolar key species. *Molecular Ecology, 16*, 3902–3925.

Eig, A. (1944). Revision of the oriental *Scrophularia* species of the herbarium of the Hebrew University. *Palestine. Journal of Botany, 3*, 79–93.

Ellstrand, N. C., & Schierenbeck, K. A. (2000). Hybridization as a stimulus for the evolution of invasiveness in plants? *Proceedings of the National Academy of Sciences of the United States of America, 97*(13), 7043–7050.

Estep, M. C., McKain, M. R., Vela Diaz, D., Zhong, J.-S., Hodge, J. G., Hodkinson, T. R., et al. (2014). Allopolyploidy, diversification, and the Miocene grassland expansion. *Proceedings of the National Academy of Sciences of the United States of America, 111*(42), 15149–15154.

Farris, J. S., Källersjö, M., Kluge, A. G., & Bult, C. (1995). Testing significance of incongruence. *Cladistics, 10*, 315–319.

Fehrer, J., Gemeinholzer, B., Chrtek Jr., J., & Bräutigam, S. (2007). Incongruent plastid and nuclear DNA phylogenies reveal ancient intergeneric hybridization in *Pilosella* hawkweeds (*Hieracium*, Cichorieae, Asteraceae). *Molecular Phylogenetics and Evolution, 42*, 347–361.

Fehrer, J., Krak, K., & Chrtek Jr., J. (2009). Intra-individual polymorphism in diploid and apomictic polyploid hawkweeds (*Hieracium*, Lactuceae, Asteraceae): disentangling phylogenetic signal, reticulation, and noise. *BMC Evolutionary Biology, 9*, 239.

Feng, T., Downie, S. R., Yu, Y., Zhang, X.-M., Chen, W.-W., He, X.-J., et al. (2009). Molecular systematics of *Angelica* and allied genera (Apiaceae) from the Hengduan Mountains of China based on nrDNA ITS sequences: phylogenetic affinities and biogeographic implications. *Journal of Plant Research, 122*(4), 403–414.

Fischer, E. (2004). Scrophulariaceae. In J. W. Kadereit (Ed.), *The families and genera of vascular plants*, vol. 7 (pp. 333–432). Heidelberg: Springer.

Fortelius, M., Eronen, J., Liu, L.-P., Pushkina, D., Tesakov, A., Vislobokova, I., et al. (2006). Late Miocene and Pliocene large land mammals and climatic changes in Eurasia. *Palaeogeography, Palaeoclimatology, Palaeoecology, 238*, 219–227.

Fuertes Aguilar, J., & Nieto Feliner, G. (2003). Additive polymorphisms and reticulation in an ITS phylogeny of thrifts (*Armeria*, Plumbaginaceae). *Molecular Phylogenetics and Evolution, 28*, 430–447.

Gadagkar, S. R., Rosenberg, M. S., & Kumar, S. (2005). Inferring species phylogenies from multiple genes: concatenated sequence tree versus consensus gene tree. *Journal of Experimental Zoology, 304B*, 64–74.

Glémin, S., Bazin, E., & Charlesworth, D. (2006). Impact of mating systems on patterns of sequence polymorphism in flowering plants. *Proceedings of the Royal Society of London B, 273*, 3011–3019.

Goddijn, W. A., & Goethart, J. W. C. (1913). Ein künstlich erzeugter Bastard. *Scrophularia neesii Wirtg. x S. vernalis L. Mededeelingen van's Rijks Herbarium Leiden, 15*, 1–10.

Goldblatt, P., & Johnson, D. E., Eds. (1979). Index to plant chromosome numbers. Missouri Botanical Garden. http://www.tropicos.org/Project/IPCN. Accessed 27 November 2015.

Gorschkova, S. G. (1997). *Scrophularia*. In B. K. Schischkin, & E. G. Bobrov (Eds.), *Flora of the USSR, translated from Russian, vol. 22* (pp. 205–274). New Delhi: Amerind.

Gould, K. R., & Donoghue, M. J. (2000). Phylogeny and biogeography of *Triosteum* (Caprifoliaceae). *Harvard Papers in Botany, 5*, 157–166.

Grau, J. (1976). Die Cytologie südwestmediterraner *Scrophularia*-Arten. *Mitteilungen der Botanischen Staatssammlung München, 12*, 609–654.

Grau, J. (1979). The probable allopolyploid origin of *Scrophularia auriculata* and *S. pseudoauriculata*. *Webbia, 34*, 497–499.

Grau, J. (1980). *Scrophularia*. In P. Mouterde (Ed.), *Nouvelle flore du Liban et de la Syrie, vol. 3* (pp. 235–241). Beyrouth: Dar el-Machreq Éditeurs.

Grau, J. (1981). Scrophularia. In K. H. Rechinger (Ed.), *Flora Iranica, cont. 147* (pp. 213–284). Graz: Akademische Druck- und Verlagsanstalt.

Grimm, G. W., & Denk, T. (2008). ITS evolution in *Platanus* (Platanaceae): homoeologues, pseudogenes and ancient hybridization. *Annals of Botany, 101*, 403–419.

Grimm, G. W., Denk, T., & Hemleben, V. (2007). Coding of intraspecific nucleotide polymorphisms: a tool to resolve reticulate evolutionary relationships in the ITS of beech trees (*Fagus* L., Fagaceae). *Systematics and Biodiversity, 5*(3), 291–309.

Guest, B., Guest, A., & Axen, G. (2007). Late tertiary tectonic evolution of northern Iran: a case for simple crustal folding. *Global and Planetary Change, 58*, 435–453.

Guo, Z.-T., Peng, S.-Z., Hao, Q.-Z., Biscaye, P. E., An, Z.-S., & Liu, T.-S. (2004). Late Miocene–Pliocene development of Asian aridification as recorded in the Red-Earth Formation in northern China. *Global and Planetary Change, 41*, 135–145.

Guo, Z.-T., Ruddiman, W. F., Hao, Q.-Z., Wu, H.-B., Qiao, Y.-S., Zhu, R.-X., et al. (2002). Onset of Asian desertification by 22 Myr ago inferred from loess deposits in China. *Nature, 416*, 159–163.

Hall, T. A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for windows 95/98/NT. *Nucleic Acids Symposium Series, 41*, 95–98.

Hildebrand, P. R., Noble, S. R., Searle, M. P., Waters, D. J., & Parrish, R. R. (2001). Old origin for an active mountain range: geology and geochronology of the eastern Hindu Kush, Pakistan. *Geological Society of America Bulletin, 113*(5), 625–639.

Hildebrand, P. R., Searle, M. P., Shakirullah, K. Z. A., & van Heijst, H. J. (2000). Geological evolution of the Hindu Kush, NW frontier Pakistan: active margin to continent-continent collision zone. In M. A. Khan, P. J. Treloar, M. P. Searle, & M. Q. Jan (Eds.), *Tectonics of the Nanga Parbat Syntaxis and the western Himalaya. The Geological Society of London special publications, vol. 170* (pp. 277–293). London: The Geological Society of London.

Hipsley, C. A., & Müller, J. (2014). Beyond fossil calibrations: realities of molecular clock practices in evolutionary biology. *Frontiers in Genetics, 5*, 138.

Hodač, L., Scheben, A. P., Hojsgaard, D., Paun, O., & Hörandl, E. (2014). ITS polymorphisms shed light on hybrid evolution in apomictic plants: a case study on the *Ranunculus auricomus* complex. *PloS One, 9*(7), e103003.

Holland, B. R., Huber, K. T., Moulton, V., & Lockhart, P. J. (2004). Using consensus networks to visualize contradictory evidence for species phylogeny. *Molecular Biology and Evolution, 21*(7), 1459–1461.

Holland, B. R., & Moulton, V. (2003). Consensus networks: a method for visualising incompatibilities in collections of trees. In G. Benson & R. Page (Eds.), *Algorithms in bioinformatics, WABI 2003* (pp. 165–176). Berlin: Springer.

Hong, D.-Y. (1983). The distribution of Scrophulariaceae in the Holarctic with special reference to the floristic relationships between eastern Asia and eastern North America. *Annals of the Missouri Botanical Garden, 70*(4), 701–712.

Hong, D.-Y. (1993). Eastern Asian-North American disjunctions and their biological significance. *Cathaya, 5*, 1–39.

Hong, D.-Y., Yang, H.-B., Jin, C.-L., & Holmgren, N. H. (1998). Scrophulariaceae. In Z.-Y. Wu (Ed.), *Flora of China, vol. 18* (pp. 1–212). St. Louis: Missouri Botanical Garden.

Hoorn, C., Mosbrugger, V., Mulch, A., & Antonelli, A. (2013). Biodiversity from mountain building. *Nature Geoscience, 6*, 154.

Huber-Morath, A. (1978). *Verbascum*. In P. H. Davis (Ed.), Flora of Turkey and the East Aegean Islands, vol. 6 (pp. 461–603). Edinburgh: Edinburgh University Press.

Huelsenbeck, J. P., Bull, J. J., & Cunningham, C. W. (1996). Combining data in phylogenetic analysis. *Trends in Ecology and Evolution, 11*, 152–158.

Huson, D. H., & Bryant, D. (2006). Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution, 23*, 254–267.

Jobes, D. V., & Thien, L. B. (1997). A conserved motif in the 5.8S ribosomal RNA (rRNA) gene is a useful diagnostic marker for plant internal transcribed spacer (ITS) sequences. *Plant Molecular Biology Reporter, 15*, 326–334.

Joly, S., Starr, J. R., Lewis, W. H., & Bruneau, A. (2006). Polyploid and hybrid evolution in roses east of the Rocky Mountains. *American Journal of Botany, 93*, 412–425.

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.

Kearney, T. H., & Peebles, R. H. (1951). *Arizona Flora*. Berkeley, Los Angeles, London: University of California Press.

Kornhall, P., Heidari, N., & Bremer, B. (2001). Selagineae and Manuleeae, two tribes or one? Phylogenetic studies in the Scrophulariaceae. *Plant Systematics and Evolution, 228*, 199–218.

Kubatko, L. S., & Degnan, J. H. (2007). Inconsistency of phylogenetic estimates from concatenated data under coalescence. *Systematic Biology, 56*(1), 17–24.

Lall, S. S., & Mill, R. R. (1978). *Scrophularia*. In P. H. Davis (Ed.), Flora of Turkey and the East Aegean Islands, vol. 6 (pp. 603–647). Edinburgh: Edinburgh University Press.

Laroche, J., & Bousquet, J. (1999). Evolution of the mitochondrial *rps3* intron in perennial and annual angiosperms and homology to *nad5* intron 1. *Molecular Biology and Evolution, 16*(4), 441–452.

Lavin, M., Herendeen, P. S., & Wojciechowski, M. F. (2005). Evolutionary rates analysis of Leguminosae implicates a rapid diversification of lineages during the Tertiary. *Systematic Biology, 54*(4), 575–594.

Leaché, A. D., Harris, R. B., Rannala, B., & Yang, Z.-H. (2014). The influence of gene flow on species tree estimation: a simulation study. *Systematic Biology, 63*(1), 17–30.

Leitch, I. J., & Bennett, M. D. (1997). Polyploidy in angiosperms. *Trends in Plant Science, 2*(12), 470–476.

Li, H.-L. (1972). Eastern Asia-Eastern North America species-pairs in wide ranging genera. In A. Graham (Ed.), *Floristics and paleofloristics of Asia and Eastern North America* (pp. 65–78). Amsterdam: Elsevier.

Li, J.-J., Zhou, S.-Z., Zhao, Z.-J., & Zhang, J. (2015). The Qingzang Movement: the major uplift of the Qinghai-Tibetan Plateau. *Science China Earth Sciences, 58*(11), 2113–2122.

Liao, C.-Y., Downie, S. R., Yu, Y., & He, X.-J. (2012). Historical biogeography of the *Angelica* group (Apiaceae tribe Selineae) inferred from analyses of nrDNA and cpDNA sequences. *Journal of Systematics and Evolution, 50*(3), 206–217.

Linder, C. R., & Rieseberg, L. H. (2004). Reconstructing patterns of reticulate evolution in plants. *American Journal of Botany, 91*(10), 1700–1708.

Liu, J.-Q., Wang, Y.-J., Wang, A.-L., Hideaki, O., & Abbott, R. J. (2006). Radiation and diversification within the *Ligularia–Cremanthodium–Parasenecio* complex (Asteraceae) triggered by uplift of the Qinghai-Tibetan Plateau. *Molecular Phylogenetics and Evolution, 38*, 31–49.

Liu, J.-S., & Schardl, C. L. (1994). A conserved sequence in internal transcribed spacer 1 of plant nuclear rRNA genes. *Plant Molecular Biology, 26*, 775–778.

Lobo, J. M. (2001). Spatial and environmental determinants of vascular plant species richness distribution in the Iberian Peninsula and Balearic Islands. *Biological Journal of the Linnean Society, 73*, 233–253.

Lockwood, J. D., Aleksić, J. M., Zou, J.-B., Wang, J., Liu, J.-Q., & Renner, S. S. (2013). A new phylogeny for the genus *Picea* from plastid, mitochondrial, and nuclear sequences. *Molecular Phylogenetics and Evolution, 69*, 717–727.

Lorenz-Lemke, A. P., Muschner, V. C., Bonatto, S. L., Cervi, A. C., Salzano, F. M., & Freitas, L. B. (2005). Phylogeographic inferences concerning evolution of Brazilian *Passiflora actinia* and *P. elegans* (Passifloraceae) based on ITS (nrDNA) variation. *Annals of Botany, 95*, 799–806.

Mabberley, D. J. (1997). *The plant-book, 2nd edn*. Cambridge: Cambridge University Press.

Maddison, W. P. (1997). Gene trees in species trees. *Systematic Biology, 46*, 523–536.

Mallet, J. (2007). Hybrid speciation. *Nature, 446*, 279–283.

Mansion, G., Rosenbaum, G., Schoenenberger, N., Bacchetta, G., Rosselló, J. A., & Conti, E. (2008). Phylogenetic analysis informed by geological history supports multiple, sequential invasions of the Mediterranean basin by the angiosperm family Araceae. *Systematic Biology, 57*, 269–285.

Marhold, K., & Lihová, J. (2006). Polyploidy, hybridization and reticulate evolution: lessons from the Brassicaceae. *Plant Systematics and Evolution, 259*, 143–174.

Mason-Gamer, R. J. (2004). Reticulate evolution, introgression, and intertribal gene capture in an allohexaploid grass. *Systematic Biology, 53*(1), 25–37.

Mason-Gamer, R. J., & Kellogg, E. A. (1996). Testing for phylogenetic conflict among molecular data sets in the tribe Triticeae (Gramineae). *Systematic Biology, 45*(4), 524–545.

Mayol, M., & Rosselló, J. A. (2001). Why nuclear ribosomal DNA spacers (ITS) tell different stories in *Quercus*. *Molecular Phylogenetics and Evolution, 19*(2), 167–176.

McDade, L. A. (1995). Hybridization and phylogenetics. In P. C. Hoch & A. G. Stephenson (Eds.), *Experimental and molecular approaches to plant biosystematics* (pp. 305–331). St. Louis: Missouri Botanical Garden.

Menezes, C. A. (1903). Diagnoses d'algumas plantas novas ou pouco conhecidas da ilha da Madeira. In A. Nobre (Ed.), *Annaes de Sciencias Naturaes (Porto), vol. 8* (pp. 95–99). Porto: Tipographia Occidental.

Menezes, C. A. (1908). *Notice sur les espèces madériennes du genre* Scrophularia. Funchal: "Diario Popular".

Meulenkamp, J. E., & Sissingh, W. (2003). Tertiary palaeogeography and tectonostratigraphic evolution of the northern and southern Peri-Tethys platforms and the intermediate domains of the African-Eurasian convergent plate boundary zone. *Palaeogeography, Palaeoclimatology, Palaeoecology, 196*, 209–228.

Miao, Y.-F., Herrmann, M., Wu, F.-L., Yan, X.-L., & Yang, S.-L. (2012). What controlled mid-late Miocene long-term aridification in Central Asia?—global cooling or Tibetan Plateau uplift: a review. *Earth-Science Reviews, 112*, 155–172.

Miller, K. G., Kominz, M. A., Browning, J. V., Wright, J. D., Mountain, G. S., Katz, M. E., et al. (2005). The Phanerozoic record of global sea-level change. *Science, 310*, 1293–1298.

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 2010) New Orleans, (pp. 1–8).

Ming, Q.-Z. (2007). A study on the neotectonic division and environment evolution of Qing-zang Plateau and three parallel rivers area. *Yunnan. Geology, 26*(4), 387–396.

Mittermeier, R. A., Gil, R. R., Hoffman, M., Pilgrim, J., Brooks, T., Mittermeier, C. G., Lamoreux, J., & da Fonseca, G. A. B. (Eds.)

(2004). *Hotspots revisited: Earth's biologically richest and most threatened terrestrial ecoregions*. CEMEX: Mexico D.F.

Morrison, D. A. (2010). Using data-display networks for exploratory data analysis in phylogenetic studies. *Molecular Biology and Evolution, 27*(5), 1044–1057.

Mouthereau, F., Lacombe, O., & Vergés, J. (2012). Building the Zagros collisional orogen: timing, strain distribution and the dynamics of Arabia / Eurasia plate convergence. *Tectonophysics, 532-535*, 27–60.

Müller, K. (2005). SeqState - primer design and sequence statistics for phylogenetic DNA data sets. *Applied Bioinformatics, 4*(1), 65–69.

Müller, K., & Albach, D. C. (2010). Evolutionary rates in *Veronica* L. (Plantaginaceae): disentangling the influence of life history and breeding system. *Journal of Molecular Evolution, 70*, 44–56.

Murbeck, S. (1933). *Monographie der Gattung* Verbascum. Lund: H. Ohlsson.

Navarro Pérez, M. L., López, J., Fernández Mazuecos, M., Rodríguez Riaño, T., Vargas, P., & Ortega Olivencia, A. (2013). The role of birds and insects in pollination shifts of *Scrophularia* (Scrophulariaceae). *Molecular Phylogenetics and Evolution, 69*, 239–254.

Nei, M., & Rooney, A. P. (2005). Concerted and birth-and-death evolution of multigene families. *Annual Review of Genetics, 39*, 121–152.

Nie, Z.-L., Sun, H., Beardsley, P. M., Olmstead, R. G., & Wen, J. (2006). Evolution of biogeographic disjunction between Eastern Asia and Eastern North America in *Phryma* (Phrymaceae). *American Journal of Botany, 93*(9), 1343–1356.

Nie, Z.-L., Wen, J., Sun, H., & Bartholomew, B. (2005). Monophyly of *Kellogia* Torrey ex Benth. (Rubiaceae) and evolution of its intercontinental disjunction between Western North America and Eastern Asia. *American Journal of Botany, 92*(4), 642–652.

Nieto Feliner, G., & Rosselló, J. A. (2007). Better the devil you know? Guidelines for insightful utilization of nrDNA ITS in species-level evolutionary studies in plants. *Molecular Phylogenetics and Evolution, 44*, 911–919.

Nylander, J. A. A. (2004). MrModeltest version 2. https://www.abc.se/~nylander. Accessed 23 April 2013.

Okuyama, Y., Fujii, N., Wakabayashi, M., Kawakita, A., Ito, M., Watanabe, M., et al. (2005). Nonuniform concerted evolution and chloroplast capture: heterogeneity of observed introgression patterns in three molecular data partition phylogenies of Asian *Mitella* (Saxifragaceae). *Molecular Biology and Evolution, 22*, 285–296.

Olteanu, R., & Jipa, D. C. (2006). Dacian basin environmental evolution during upper Neogene within the Paratethys domain. *Geo-Eco-Marina, 12*, 91–105.

Ortega Olivencia, A. (2009). *Scrophularia*. In C. Benedí, E. Rico, J. Güemes, & A. Herrero (Eds.), *Flora Iberica, vol. 13* (pp. 97–134). Madrid: Real Jardín Botánico, CSIC.

Oxelman, B., Kornhall, P., Olmstead, R. G., & Bremer, B. (2005). Further disintegration of Scrophulariaceae. *Taxon, 54*, 411–425.

Payseur, B. A., & Rieseberg, L. H. (2016). A genomic perspective on hybridization and speciation. *Molecular Ecology, 25*(11), 2337–2360.

Peng, D., & Wang, X.-Q. (2008). Reticulate evolution in *Thuja* inferred from multiple gene sequences: implications for the study of biogeographical disjunction between eastern Asia and North America. *Molecular Phylogenetics and Evolution, 47*, 1190–1202.

Pennell, F. W. (1943). *The Scrophulariaceae of the Western Himalayas: The Academy of Natural Sciences of Philadelphia Monographs number 5*. Lancaster, Pennsylvania: Wickersham Printing Company.

Pirie, M. D., Humphreys, A. M., Barker, N. P., & Linder, H. P. (2009). Reticulation, data combination, and inferring evolutionary history: an example from Danthonioideae (Poaceae). *Systematic Biology, 58*, 612–628.

Pirie, M. D., Humphreys, A. M., Galley, C., Barker, N. P., Verboom, G. A., Orlovich, D., et al. (2008). A novel supermatrix approach improves resolution of phylogenetic relationships in a comprehensive sample of danthonioid grasses. *Molecular Phylogenetics and Evolution, 48*, 1106–1119.

Popov, S. V., Rögl, F., Rozanov, A. Y., Steininger, F. F., Shcherba, I. G., Kovac, M. (Eds.) (2004). *Lithological-Paleogeographic maps of Paratethys, 10 maps Late Eocene to Pliocene, Courier Forschungsinstitut Senckenberg 250*. Stuttgart: E. Schweizerbart.

Potts, A. J., Hedderson, T. A., & Grimm, G. W. (2014). Constructing phylogenies in the presence of intra-individual site polymorphisms (2ISPs) with a focus on the nuclear ribosomal cistron. *Systematic Biology, 63*(1), 1–16.

Qi, Y., & Yang, Y.-H. (1999). Topographic effect on spatial variation of plant diversity in California. *Geographic Information Sciences, 5*(1), 39–46.

Qian, H., & Ricklefs, R. E. (2000). Large-scale processes and the Asian bias in species diversity of temperate plants. *Nature, 407*, 180–182.

Qiu, Y.-X., Fu, C.-X., & Comes, H. P. (2011). Plant molecular phylogeography in China and adjacent regions: tracing the genetic imprints of Quaternary climate and environmental change in the world's most diverse temperate flora. *Molecular Phylogenetics and Evolution, 59*, 225–244.

Quézel, P. (1985). Definition of the Mediterranean region and origin of its flora. In C. Gomez-Campo (Ed.), *Plant conservation in the Mediterranean area* (pp. 9–24). Dordrecht: Kluwer.

Ramstein, G., Fluteau, F., Besse, J., & Joussaume, S. (1997). Effect of orogeny, plate motion and land-sea distribution on Eurasian climate change over the past 30 million years. *Nature, 386*, 788–795.

Richardson, A. O., & Palmer, J. D. (2007). Horizontal gene transfer in plants. *Journal of Experimental Botany, 58*(1), 1–9.

Rieseberg, L. H. (1991). Homoploid reticulate evolution in *Helianthus* (Asteraceae): evidence from ribosomal genes. *American Journal of Botany, 78*(9), 1218–1237.

Rieseberg, L. H., & Wendel, J. F. (1993). Introgression and its consequences in plants. In R. Harrison (Ed.), *Hybrid zones and the evolutionary process* (pp. 70–109). New York: Oxford University Press.

Rieseberg, L. H., Whitton, J., & Linder, C. R. (1996). Molecular marker incongruence in plant hybrid zones and phylogenetic trees. *Acta Botanica Neerlandica, 45*(3), 243–262.

Roe, G. H. (2005). Orographic precipitation. *Annual Review of Earth and Planetary Sciences, 33*, 645–671.

Rojas Andrés, B. M., Albach, D. C., & Martínez Ortega, M. M. (2015). Exploring the intricate evolutionary history of the diploid–polyploid complex *Veronica* subsection *Pentasepalae* (Plantaginaceae). *Botanical Journal of the Linnean Society, 179*, 670–692.

Rokas, A., Williams, B. L., King, N., & Carroll, S. B. (2003). Genome-scale approaches to resolving incongruence in molecular phylogenies. *Nature, 425*, 798–804.

Ronquist, F. (1997). Dispersal-vicariance analysis: a new approach to the quantification of historical biogeography. *Systematic Biology, 46*, 195–203.

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D. L., Darling, A., Höhna, S., et al. (2012). MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology, 61*, 539–542.

Rosselló, J. A., Lázaro, A., Cosín, R., & Molins, A. (2007). A phylogeographic split in *Buxus balearica* (Buxaceae) as evidenced by nuclear ribosomal markers: when ITS paralogues are welcome. *Journal of Molecular Evolution, 64*, 143–157.

Sanmartín, I. (2003). Dispersal vs. vicariance in the Mediterranean: historical biogeography of the Palearctic Pachydeminae (Coleoptera, Scarabaeoidea). *Journal of Biogeography, 30*, 1883–1897.

Schäferhoff, B., Fleischmann, A., Fischer, E., Albach, D. C., Borsch, T., Heubl, G., et al. (2010). Towards resolving Lamiales relationships: insights from rapidly evolving chloroplast sequences. *BMC Evolutionary Biology, 10*, 352.

Scherson, R. A., Vidal, R., & Sanderson, M. J. (2008). Phylogeny, biogeography, and rates of diversification of New World *Astragalus* (Leguminosae) with an emphasis on South American radiations. *American Journal of Botany, 95*, 1030–1039.

Scheunert, A., & Heubl, G. (2011). Phylogenetic relationships among New World *Scrophularia* L. (Scrophulariaceae): new insights inferred from DNA sequence data. *Plant Systematics and Evolution, 291*, 69–89.

Scheunert, A., & Heubl, G. (2014). Diversification of *Scrophularia* (Scrophulariaceae) in the Western Mediterranean and Macaronesia - phylogenetic relationships, reticulate evolution and biogeographic patterns. *Molecular Phylogenetics and Evolution, 70*, 296–313.

Schnabel, A., McDonel, P. E., & Wendel, J. F. (2003). Phylogenetic relationships in *Gleditsia* (Leguminosae) based on ITS sequences. *American Journal of Botany, 90*(2), 310–320.

Seehausen, O. (2004). Hybridization and adaptive radiation. *Trends in Ecology and Evolution, 19*, 198–207.

Shaw, R. J. (1962). The biosystematics of *Scrophularia* in Western North America. *Aliso, 5*, 147–178.

Silvestro, D., & Michalak, I. (2012). RaxmlGUI: a graphical front-end for RAxML. *Organisms Diversity & Evolution, 12*, 335–337.

Simmons, M. P., & Ochoterena, H. (2000). Gaps as characters in sequence-based phylogenetic analyses. *Systematic Biology, 49*(2), 369–381.

Sobel, E. R., Chen, J., & Heermance, R. V. (2006). Late Oligocene–Early Miocene initiation of shortening in the Southwestern Chinese Tian Shan: implications for Neogene shortening rate variations. *Earth and Planetary Science Letters, 247*, 70–81.

Soltis, P. S., & Soltis, D. E. (2009). The role of hybridization in plant speciation. *Annual Review of Plant Biology, 60*, 561–588.

Stamatakis, A. (2006). RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics, 22*(21), 2688–2690.

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics, 30*(9), 1312–1313.

Stebbins, G. L. (1959). The role of hybridization in evolution. *Proceedings of the American Philosophical Society, 103*(2), 231–251.

Steiner, K. E. (1996). Chromosome numbers and relationships in tribe Hemimerideae (Scrophulariaceae). *Systematic Botany, 21*(1), 63–76.

Stephens, M., Smith, N. J., & Donnelly, P. (2001). A new statistical method for haplotype reconstruction from population data. *The American Journal of Human Genetics, 68*(4), 978–989.

Stiefelhagen, H. (1910). Systematische und pflanzengeographische Studien zur Kenntnis der Gattung *Scrophularia*. *Engler's. Botanische Jahrbücher für Systematik, Pflanzengeschichte und Pflanzengeographie, 44*, 406–496.

Sun, B.-N., Wu, J.-Y., Liu, Y.-S. C., Ding, S.-T., Li, X.-C., Xie, S.-P., et al. (2011). Reconstructing Neogene vegetation and climates to infer tectonic uplift in western Yunnan, China. *Palaeogeography, Palaeoclimatology, Palaeoecology, 304*, 328–336.

Sun, Y.-S., Wang, A.-L., Wan, D.-S., Wang, Q., & Liu, J.-Q. (2012). Rapid radiation of *Rheum* (Polygonaceae) and parallel evolution of morphological traits. *Molecular Phylogenetics and Evolution, 63*, 150–158.

Swofford, D. L. (2003). PAUP*: Phylogenetic Analysis Using Parsimony (* and other methods) version 4. http://www.paup.csit.fsu.edu/index.html. Accessed 23 April 2013.

Taberlet, P., Gielly, L., Pautou, G., & Bouvet, J. (1991). Universal primers for amplification of three non-coding regions of chloroplast DNA. *Plant Molecular Biology, 17*, 1105–1109.

Takhtajan, A. (1986). *Floristic regions of the World*. Berkeley: University of California Press.

Tank, D. C., & Olmstead, R. G. (2008). From annuals to perennials: phylogeny of subtribe Castillejinae (Orobanchaceae). *American Journal of Botany, 95*(5), 608–625.

Thompson, J. D. (2005). *Plant evolution in the Mediterranean*. Oxford: Oxford University Press.

Tiffney, B. H., & Manchester, S. R. (2001). The use of geological and paleontological evidence in evaluating plant phylogeographic hypotheses in the Northern Hemisphere Tertiary. *International Journal of Plant Sciences, 162*(S6), S3–S17.

Vaarama, A., & Hiirsalmi, H. (1967). Chromosome studies on some Old World species of the genus *Scrophularia*. *Hereditas, 58*, 333–358.

Valtueña, F. J., López, J., Álvarez, J., Rodríguez Riaño, T., & Ortega Olivencia, A. (2016). *Scrophularia arguta*, a widespread annual plant in the Canary Islands: a single recent colonization event or a more complex phylogeographic pattern? *Ecology and Evolution, 6*(13), 4258–4273.

van der Niet, T., & Linder, H. P. (2008). Dealing with incongruence in the quest for the species tree: a case study from the orchid genus *Satyrium*. *Molecular Phylogenetics and Evolution, 47*, 154–174.

Vargas, P., Carrió, E., Guzmán, B., Amat, E., & Güemes, J. (2009). A geographical pattern of *Antirrhinum* (Scrophulariaceae) speciation since the Pliocene based on plastid and nuclear DNA polymorphisms. *Journal of Biogeography, 36*, 1297–1312.

Verboom, G. A., Herron, M. L., Moncrieff, G. R., & Slingsby, J. A. (2016). Maintenance of species integrity in the context of a recent radiation: the case of *Jamesbrittenia* (Scrophulariaceae: Limoselleae) in southern Africa. *Botanical Journal of the Linnean Society, 182*, 115–139.

Volkov, R. A., Komarova, N. Y., & Hemleben, V. (2007). Ribosomal DNA in plant hybrids: inheritance, rearrangement, expression. *Systematics and Biodiversity, 5*(3), 261–276.

Vriesendorp, B., & Bakker, F. T. (2005). Reconstructing patterns of reticulate evolution in angiosperms: what can we do? *Taxon, 54*, 593–604.

Wang, C.-S., Dai, J.-G., Zhao, X.-X., Li, Y.-L., Graham, S. A., He, D.-F., et al. (2014). Outward-growth of the Tibetan Plateau during the Cenozoic: a review. *Tectonophysics, 621*, 1–43.

Wang, L., Wu, Z.-Q., Bystriakova, N., Ansell, S. W., Xiang, Q.-P., Heinrichs, J., et al. (2011). Phylogeography of the Sino-Himalayan fern *Lepisorus clathratus* on "The Roof of the World". *PloS One, 6*(9), e25896.

Weisrock, D. W., Smith, S. D., Chan, L. M., Biebouw, K., Kappeler, P. M., & Yoder, A. D. (2012). Concatenation and concordance in the reconstruction of mouse lemur phylogeny: an empirical demonstration of the effect of allele sampling in phylogenetics. *Molecular Biology and Evolution, 29*(6), 1615–1630.

Wen, J., Ickert-Bond, S. M., Nie, Z.-L., & Li, R. (2010). Timing and modes of evolution of Eastern Asian – North American biogeographic disjunctions in seed plants. In M.-Y. Long, H.-Y. Gu, & Z.-H. Zhou (Eds.), *Darwin's heritage today: proceedings of the Darwin 200 Beijing International Conference* (pp. 252–269). Beijing: Higher Education.

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.

Wendel, J. F., & Doyle, J. J. (1998). Phylogenetic incongruence: window into genome history and molecular evolution. In D. E. Soltis, P. S. Soltis, & J. J. Doyle (Eds.), *Molecular systematics of plants II: DNA sequencing* (pp. 265–296). Dordrecht: Kluwer.

Whittall, J., Liston, A., Gisler, S., & Meinke, R. J. (2000). Detecting nucleotide additivity from direct sequences is a SNAP: an example from *Sidalcea* (Malvaceae). *Plant Biology, 2*, 211–217.

Willis, J. C. (1973). *A dictionary of the flowering plants and ferns, 8th edn*. Cambridge: Cambridge University Press.

Wolfe, A. D., & Elisens, W. J. (1994). Nuclear ribosomal DNA restriction-site variation in *Penstemon* section *Peltanthera*

(Scrophulariaceae): an evaluation of diploid hybrid speciation and evidence for introgression. *American Journal of Botany, 81*(12), 1627–1635.

Wright, S. I., Kalisz, S., & Slotte, T. (2013). Evolutionary consequences of self-fertilization in plants. *Proceedings of the Royal Society of London B, 280*(*1760*), 20130133.

Wright, S. I., Lauga, B., & Charlesworth, D. (2002). Rates and patterns of molecular evolution in inbred and outbred *Arabidopsis*. *Molecular Biology and Evolution, 19*(9), 1407–1420.

Xiang, Q.-Y., Soltis, D. E., & Soltis, P. S. (1998). The Eastern Asian and Eastern and Western North American floristic disjunction: congruent phylogenetic patterns in seven diverse genera. *Molecular Phylogenetics and Evolution, 10*(2), 178–190.

Xiao, L.-Q., Möller, M., & Zhu, H. (2010). High nrDNA ITS polymorphism in the ancient extant seed plant *Cycas*: incomplete concerted evolution and the origin of pseudogenes. *Molecular Phylogenetics and Evolution, 55*, 168–177.

Yang, F.-S., Qin, A.-L., Li, Y.-F., & Wang, X.-Q. (2012). Great genetic differentiation among populations of *Meconopsis integrifolia* and its implication for plant speciation in the Qinghai-Tibetan Plateau. *PloS One, 7*(5), e37196.

Yu, Y., Harris, A. J., & He, X.-J. (2010). S-DIVA (statistical dispersal–vicariance analysis): a tool for inferring biogeographic histories. *Molecular Phylogenetics and Evolution, 56*, 848–850.

Yu, Y., Harris, A. J., & He, X.-J. (2014). RASP (Reconstruct Ancestral State in Phylogenies) version 2. http://mnh.scu.edu.cn/soft/blog/RASP. Accessed 23 April 2013.

Yuan, D.-Y., Ge, W.-P., Chen, Z.-W., Li, C.-Y., Wang, Z.-C., Zhang, H.-P., et al. (2013). The growth of northeastern Tibet and its relevance to large-scale continental geodynamics: a review of recent studies. *Tectonics, 32*, 1358–1370.

Yue, J.-P., Sun, H., Baum, D. A., Li, J.-H., Al-Shehbaz, I. A., & Ree, R. (2009). Molecular phylogeny of *Solms-laubachia* (Brassicaceae) s.l., based on multiple nuclear and plastid DNA sequences, and its biogeographic implications. *Journal of Systematics and Evolution, 47*(5), 402–415.

Yue, J.-X., Li, J.-P., Wang, D., Araki, H., Tian, D.-C., & Yang, S.-H. (2010). Genome-wide investigation reveals high evolutionary rates in annual model plants. *BMC Plant Biology, 10*, 242.

Zhang, J.-Q., Meng, S.-Y., Allen, G. A., Wen, J., & Rao, G.-Y. (2014). Rapid radiation and dispersal out of the Qinghai-Tibetan Plateau of an alpine plant lineage *Rhodiola* (Crassulaceae). *Molecular Phylogenetics and Evolution, 77*, 147–158.