Acta Universitatis Agriculturae Sueciae

Doctoral Thesis No. 2022:68
Faculty of Landscape Architecture, Horticulture and Crop Production Science

Phylogenetic studies of *Elymus s.s.* L. (Poaceae)

Plant systematics as a prerequisite for efficient conservation and utilization

Jonatan Leo

Doctoral Thesis No. 2022:68 • Phylogenetic studies of *Elymus* s.s. L. (Poaceae) • Jonatan Leo

Phylogenetic studies of *Elymus s.s.* L. (Poaceae)

Plant systematics as a prerequisite for efficient conservation and utilization

Jonatan Leo

Faculty of Landscape Architecture, Horticulture and Crop Production Science

Department of Plant Breeding

Alnarp

SLU

SWEDISH UNIVERSITY OF AGRICULTURAL SCIENCES

DOCTORAL THESIS

Alnarp 2022

Acta Universitatis Agriculturae Sueciae
2022:68

Cover: *Elymus glaucescens*
(photo: Jonatan Leo)

ISSN 1652-6880
ISBN (print version) 978-91-8046-012-5
ISBN (electronic version) 978-91-8046-013-2

Swedish University of Agricultural Sciences, Department of Plant Breeding, Alnarp, Sweden

Print: Media-Tryck, Lund University, Lund, 2022

Phylogenetic studies of *Elymus s.s.* L. (Poaceae)

Abstract

Elymus (wild-rye), in the tribe Triticeae, is a diverse genus and contains about 150 species growing in a wide range of habitats in temperate areas all over the world. The species are wild relatives to some of the most important cereal crops like bread wheat, durum wheat, rye, and barley. The genus is considered a genetic resource and contains many valuable traits that could be used in future crop breeding for sustainable agriculture. Several species are used today as forage grasses and others could potentially be introduced as new crops. To fully explore, exploit and conserve the genetic diversity of *Elymus*, basic knowledge of the systematics of the genus is needed. This thesis focus on *Elymus s.s.*, including only species with an StH genome combination, and investigates the phylogeny and phylogeography on different taxonomical levels: within a species (Paper I), within a region (Paper II) and within a genus (Paper III). Genotyping-By-Sequencing techniques were used to explore the relationships, diversification, generic structure, migration, and origin of the genus. The results indicate separate origins of American and Eurasian species and show a complex relationship between *Elymus* in South and North America with multiple migration events. The thesis gives insights into the evolutionary history of *Elymus s.s.* and provides phylogenetic data for further taxonomical work within the genus.

Keywords: DArTseq, *Elymus*, genetic resources, Genotyping-By-Sequencing, phylogenetics, Poaceae, Triticeae

Phylogenetic studies of *Elymus s.s.* L. (Poaceae)

Abstract

Elymus (elmar), i tribus Triticeae, är ett mycket variationsrikt släkte och innehåller cirka 150 arter som växer i en rad olika habitat i tempererade områden över hela världen. Arterna är vilda släktingar till några av de viktigaste spannmålsgrödorna som brödvete, durumvete, råg och korn. Släktet utgör en genetisk resurs och innehåller många värdefulla egenskaper som skulle kunna användas i framtida växtförädling för ett hållbart jordbruk. Flera arter används idag som fodergräs och andra skulle potentiellt kunna introduceras som nya grödor. För att till fullo utforska, utnyttja och bevara den genetiska mångfalden hos *Elymus* behövs grundläggande kunskaper om släktets systematik. Den här avhandlingen fokuserar på *Elymus s.s.*, vilket inkluderar endast arter med en StH-genomkombination, och undersöker fylogenin och fylogeografin på olika taxonomiska nivåer: inom en art (Paper I), inom en region (Paper II) och inom ett släkte (Paper III). Genotypning-genom-sekvensering (GBS) användes för att utforska släktskap, diversifiering, generisk struktur, migration och ursprung. Resultaten indikerar separata ursprung för amerikanska och eurasiska arter och visar ett komplext förhållande mellan *Elymus* i Syd- och Nordamerika med flera migrationer. Avhandlingen ger en fördjupad förståelse av *Elymus s.s.* evolutionära historia och tillhandahåller fylogenetiska data för vidare taxonomiskt arbete inom släktet.

Nyckelord: DArTseq, *Elymus*, fylogeni, genetiska resurser, Genotypning-genom-sekvensering, Poaceae, Triticeae

In memory of Dr. Björn Salomon

Contents

List of publications

This thesis is based on the work presented in the following papers, referred to by Roman numerals in the text:

I. Jonatan Leo, Therése Bengtsson, Anders S. Carlsson, Jonathan Brassac, Roland von Bothmer (2022). Population structure and phylogeography of *Elymus mutabilis* and its genetic relationships with *E. transbaicalensis* (Poaceae). Nordic Journal of Botany. Issue 4. e03520. https://doi.org/10.1111/njb.03520

II. Jonatan Leo, Therése Bengtsson, Arturo Morales, Anders S. Carlsson, Roland von Bothmer. Phylogenetic relationships and phylogeography of the genus *Elymus* L. (Poaceae) in North and South America

III. Jonatan Leo, Therése Bengtsson, Arturo Morales, Anders S. Carlsson, Roland von Bothmer. Phylogenetic relationships among *Elymus s.s.* (Poaceae)

Paper I is reproduced with the permission of the publisher.

Abbreviations

ABS	Access and Benefit-sharing
BI	Bayesian Inference
CBD	Convention of Biological Diversity
CBM	Coalescence-Based Methods
CWR	Crop Wild Relatives
DArTseq	Diversity Arrays Technology by sequencing
GBS	Genotyping-By-Sequencing
GWAS	Genome-Wide Association Study
ILS	Incomplete Lineage Sorting
IGP	Introgression Gene Pools
ISSR	Inter Simple Sequence Repeat
ML	Maximum Likelihood
MP	Maximum Parsimony
NGS	Next Generation Sequencing
NJ	Neighbour Joining
PCA	Principal Component Analysis
RADseq	Restriction-site Associated DNA sequencing
RAPD	Random Amplified Polymorphic DNA
RFLP	Restriction Fragment Length Polymorphism

RE	Restriction Enzyme
RGP	Recombination Gene Pools
s.l.	*sensu lato*; meaning “in the broad sense”, indicating a circumscription before a split of a taxon
s.s.	sensu stricto; meaning “in a strict sense”, indicating a circumscription after a split of a taxon
SNP	Single Nucleotide Polymorphism
SSR	Simple Sequence Repeat

1 Introduction

Plant systematics, including taxonomy and phylogeny, constitute the framework for all knowledge about plants. The recognition and organization of individuals into larger discrete entities based on evolutionary relationships not only acknowledge plant diversity but also makes it possible to draw scientific generalizations and link information to specific groups. Continuous research in plant systematics, and the development of associated skills, are important for both plant conservation and plant use. Poaceae, the grass family, is one of the largest plant families in the world including more than 11 600 species dispersed on all continents in a wide range of habitats (Stevens 2001). Grasses are important for many ecosystems and 41% of the earth's land surface is covered by grass biomes (Lehmann et al. 2019). Even though grasses constitute three percent of the total number of angiosperm species, the family is responsible for 28% of primary gross productivity and 25% of terrestrial photosynthesis (Beer et al. 2010; Lehmann et al. 2019). We are completely dependent on the grasses for our survival and well-being, and knowledge of this group of plants is essential for a sustainable future. Grasses were among the first crops to be domesticated and they are still the main staple food of the world with rice, maize, wheat, and other cereals making up the bulk of humanity's calorie intake (Lu & Ellstrand 2014). Grasses also contribute to the world food supply by feed for milk and meat production and sugar cane as the leading global source of sugar. In addition, grasses in horticultural landscapes, such as ornamental grasses, turf grasses, and lawns, are a large industry with high economic and cultural value.

The world is changing. Crop improvement and adaptation are needed to meet harsher growing conditions in the course of climate change, to encourage crop diversification to establish resilient food systems, and for yield efficiency to maximize output from more restricted access to arable land and a growing population. It is necessary to intensify the breeding efforts of established crops, as well as to initiate the domestication processes of new species. Crop wild relatives (CWR) contain genes for a multitude of useful traits that, if transferred to cultivated crops, can improve cultivar performance by boosting yields, widening the adaptation range, and increasing the overall efficiency of agricultural production.

The tribus Triticeae Dumort. comprises some of the oldest and most economically important cereal crops such as bread wheat, durum wheat, rye and, barley, as well as several perennial forage grasses (e.g. *Psathyrostachys juncea* and *Elymus trachycaulus*), weeds (e.g. *Elytrigia repens*), sand binders (*Agropyron fragile* and *Leymus* spp.) and ornamentals (e.g. *Hordeum jubatum* and *Elymus magellanicus*) (Bothmer et al. 1992, Barkworth and von Bothmer 2009). The tribus contains many wild species considered a genetic resource for crop improvement and domestication. *Elymus* L. is the largest genus in the tribe harboring a variety of agronomical desirable traits for potential use in cereal and forage grass breeding (Salomon et al. 1997). Fundamental knowledge of taxonomy, phylogeny, genetic variation, population structures, and gene flow is needed to assess the usefulness of species and populations as genetic resources, to create efficient and sustainable breeding programs, as well as to create informed conservation strategies. *Elymus* is also a biologically interesting group for general research in polyploidization, hybridization, evolutionary patterns, adaptation, and domestication (von Bothmer & Salomon 1994).

There is a long tradition of Triticeae research at SLU and the Department of Plant Breeding with Prof. Roland von Bothmer and the late Dr. Björn Salomon (Figure 1) as world-leading experts in their fields. The research has been focused on classical taxonomy, biosystematics and gene transfer, cytogenetics, molecular genetics, and plant breeding and pre-breeding. Several thesis's from the department have been concerning Triticeae and two have focused on *Elymus*: genetic diversity of *Elymus* in Northern Europe by Oscar Diaz (1999) and biosystematics studies of Asian *Elymus* species by Bao-Rong Lu (1993). The present thesis is partly linked to these previous dissertations in investigating *E. mutabilis* diversity in Eurasia (Paper I), but

it also explores new areas by focusing on the relationships between species in North and South America (Paper II) and the world (Paper III). The research projects within this thesis have been conducted in collaboration with IPK, Gatersleben, Germany (paper I) and INIA, Chile (papers II and III) in form of project development and discussions, utilization of research infrastructure, joint collection expeditions, knowledge exchange, and article writing.

The objective of this summary of the thesis is to put the research into a wider scientific and social context, put the results into an international context, and discuss the contribution to the state of knowledge. The purpose is also to identify strengths and weaknesses with the contribution and discuss future research projects. The Background (chapter 2) gives an overview of Triticeae in general and *Elymus* in particular followed by a discussion of CWR and their conservation and use. I also want to highlight the importance of *Elymus* as an organism group and why plant systematic research is important, not only for its own sake but for all botanical and agricultural research. The method (chapter 4) is discussed with an elaborate description of the plant material and the choices of technology. A short discussion of ethical concerns is also included. In the results and discussion part (chapter 5), the synthesis of the three Papers is presented with a deeper account of the taxonomical insinuation of the phylogenetic results. The summary is finalized with a conclusion (chapter 6) and a discussion of future research projects (chapter 7).

Figure 1. Björn Salomon on a collection mission in Russia in 1990.

2 Background

2.1 Crop wild relatives (CWR)

Our cultivated crops originate from wild populations that have undergone generations of adaptation to human cultivation regimes and local environments to become domesticated. Domestication is a dynamic process and CWR can be continuously used as a natural genetic resource, if traits are transferable, for further crop adaptation and improvement. To do so, we need to know as much as possible about the genetic variation and structure of CWR. The classification of cultivated plants in the concept of gene pools (Figure 2), coined by Harlan and de Wet (1971), has been particularly applicable for plant breeders in their efforts of using CWR in breeding programs and as a tool for prioritization in collecting wild genetic resources (von Bothmer & Seberg 1995). The system is designed based on three groupings with different abilities to incorporate genes into the conventionally grown crop. The primary gene pool includes taxa without sterility barriers like landraces, breeding lines, varieties, wild forms, and weedy forms. Taxa in the secondary gene pool will cross with the crop while gene transfer is difficult. The tertiary gene pool includes taxa for which gene transfer is difficult due to strong sterility barriers but possible with techniques other than traditional crossings. The whole gene pool should be considered a unit of study and must therefore be the target of research, collection, and conservation. The knowledge and structure of crop gene pools are completely dependent on a sound taxonomical framework including

cytogenetic and biosystematic data, genetic relationships, and species delimitations. In order to exploit the diversity of plant genetic resources, we need to: 1) know how many taxa there are within a group, 2) be able to distinguish different taxa, 3) know the characteristics of different taxa, and 4) understand the distribution of diversity and which taxa are the closest relatives to the crops (Maxted et al. 2020).

Figure 2. Gene Pool concept based on Harlan and de Wet (1971) where GP1A includes cultivars and cultivated forms or lineages of the crop, and GP1B includes wild or weedy forms of the crop. GP2 includes more or less close wild relatives and GP3 more distant wild relatives.

Wild species within Triticeae constitute a gene pool for the important cereals wheat, barley, and rye, as well as for several forage crops (for example Psathyrostachys juncea, Elymus trachycaulus, Leymus angustus, Elytrigia intermedia, and Agropyron cristatum). The tribe harbors useful genes for potential crop adaptation and improvement, with traits such as pest and disease resistance, wider adaptation and stress resistance (for example salt tolerance and cold hardiness), morphological modification (for example

short stature and coloration), yield, improved quality and nutrient content, modes of reproduction (including apomixes), and perennial life form (Harlan 1976; Salomon & Lu 1992; von Bothmer et al. 1992; Diaz 1999). The rather weak hybridization barriers between genera make it possible to induce new natural and domesticated variation by combination breeding. The breeding efforts involving wild material from Triticeae have mainly focused on improving wheat, rye, and forage grasses, but there is also a great potential for producing new food and forage crops (Lu 1993). Elymus is considered to belong to the tertiary gene pool of barley with the potential of transferring genes by wide crosses and advanced techniques for progeny rescue (Salomon et al. 1997).

Maxted et al. (2006) introduced the Taxon Group Concept to define CWR based on five taxonomical hierarchies: 1a – the crop, 1b – the same species as the crop, 2 – the same series or section as the crop, 3 – the same subgenus as the crop, 4 – the same genus as the crop, and 5 – the same tribe but different genus to the crop. They argue that the taxonomical closeness can be used for setting conservation priorities in groups with limited information about crossability and genetic diversity. Based on this classification, Elymus belongs to the Taxon Group 5 of wheat and barley.

2.2 Triticeae

The tribe Triticeae constitutes both annual and perennial species distributed in almost all temperate regions of the world with the highest number of species diversity in East Asia (Barkworth & von Bothmer 2005). The basic chromosome number is seven (x=7) and both diploid and polyploid species exist. Triticeae is morphologically characterized by a simple spike with one or two, rarely more, sessile spikelets per node, open leaf sheaths, membranous ligules, and ovaries with a hairy top (Barkworth & von Bothmer 2009). Phylogenetic relationships within the tribe are complicated. It is difficult to establish a comprehensive, accessible, and phylogenetically accurate classification system due to the high number of species, high degree of polyploidization and hybridization, large distribution areas, and a lack of universally accepted principles for species delimitation (Kellogg 1989; Frederiksen & Seberg 1992). The high degree of intergeneric hybridization within Triticeae even led to the suggestion by Stebbins and Snyder (1956) to lump all species in Triticeae into a single genus. The estimated number of

species in the genus ranges from 350 to 500, depending on the author (Dewey 1984; Löve 1984; West et al. 1988; Tzvelev 1989). This is a short overview of the historical treatment of Triticeae and a more comprehensive review can be found in *Taxonomy of the Triticeae: a historical perspective* by Barkworth (1992).

The taxonomic treatment of Triticeae has developed together with access to new plant materials and novel data from technological advancements. There is still no consensus on taxonomic treatment on neither the generic nor the species level due to differences in method evaluation and priorities of taxonomic characters (Wang et al. 1994; Yen et al. 2005; Baum et al. 2011). Pre-Darwinian taxonomists based their classification of the tribe primarily on morphological traits, often involving only a few inflorescence characters. Morphological data is, however, often considered inadequate for phylogenetic reconstruction in Triticeae due to instability and a high degree of homoplasy (Kellogg 1989). The taxonomical treatment of the tribe at the beginning of the 21st century differed between geographical regions and there was a need for a unifying systematic principle for genera delimitation (Barkworth & von Bothmer 2005). During the 1960s and 1970s, a large amount of cytogenetic data was accumulated which resulted in two independent works by Löve (1984) and Dewey (1984). They proposed that genomic constitutions should be the basis for an international generic circumscription of Triticeae, which is the now prevailing basis for the majority of contemporary treatments. According to this classification system, a genus is defined as monogenomic if all species have the same haplome or haplome combinations, and where a haplome is defined as the basic or monoploid set of chromosomes.

Most genera have one type of haplome, for example, *Agropyron* (**P**), *Pseudoroegneria* (**St**) and *Secale* (**R**). Some have several haplomes such as *Aegilops* (**B, C, D, M, N, U** or **X**) and *Hordeum* (**H, Xa, Xu, I**). Allopolyploid genera are often defined as a specific combination of two or more parental haplomes, for example, the allohexaploid *Anthosachne* (**StHW**) is derived from the combination of the *Pseudoroegneria* (**St**), *Hordeum* (**H**) and *Australopyrum* (**W**) genomes. In addition, some genera are defined based on a common haplome in combination with one or more different ones, thus intrinsically polyphyletic. *Thinopyrum* has one unifying haplome (**E**) without or together with one or several additional haplomes (**P**, **St** or **L**), and *Elymus s.l.* has one unifying haplome (**St**) together with several

additional haplomes (**H**, **Y** or **W**). Variation within haplomes is sometimes noted, as in *Psathyrostachys* ($\mathbf{N^{j,\ f,\ or\ h}}$). Most publications of today follow the standardization of genome designation in Triticeae suggested by Wang *et al.* (1994).

The genome-based classification is theoretical, methodological, and biological problematic (Baum et al. 1987; Gupta & Baum 1989; Seberg 1989; Seberg & Frederiksen 2001; Al-Saghir 2016). Critics oppose to the "single character" classification of genera and advocate a "maximum information" based approach where genome analyses are used as a part of a more inclusive taxonomy. The character of the genome constitution is not universal and cannot be used in all groups on all taxonomical levels. The possibility of polyphyletic origins and potential homoplasies is contradictive to the prevailing assumption of monophyletic clades as the only accepted entities in a phylogenetic classification (Petersen et al. 2011). Genome combinations as the main character of generic delimitation should be considered a convention, and the system has proven useful for understanding the current segregation of taxa as an important evolutionary mechanism.

Most Triticeae treatments are incomplete in only taking regional scopes or including a limited number of taxa (Barkworth & von Bothmer 2009). The work by Löve (1984) is an exception and includes a checklist with genera as well as species. New NGS (Next Generation Sequencing) techniques have gained a lot of new information with further insights into the evolution of Triticeae, but multiple polyploidizations and reticulate evolution still make it difficult to define monophyletic groups (Barkworth 1992; Dvorak & Zhang 1992; Brassac & Blattner 2015; Bernhardt et al. 2017; Edet et al. 2018). Baum *et al.* (2015) argue that molecular markers should primarily be used for species circumscriptions within genome combinations.

2.3 Elymus

Elymus is the largest genus in Triticeae with approximately 150 species (Löve 1984). The number of recognized species in the genus varies between different authors depending on the author's assumed generic delimitation and species recognition. Frequent hybridization resulting in interspecific gene flow affects species and population structures, contributing to the difficulties of solving the evolutionary history of *Elymus*. The large number of taxa in Central Asia and North America have been intensively covered and the

majority of publications and gene bank material originates from these areas. *Elymus* in South America is comparatively less studied with a large diversity and many uncollected areas.

2.3.1 Morphology

The genus is morphologically diverse, both within and between species, and shows a strong phenotypic plasticity (Sun & Salomon 2009). Spikelets are arranged in spikes with one-two, occasionally up to five, spikelets per rachis node and with rachis internodes less than half as long as the spikelets (Barkworth & von Bothmer 2009) (Figure 3). All species are perennials with densely to loosely caespitose, occasionally rhizomatous, growth habit and mainly self-pollinating with short anthers (Löve 1984; Diaz 1999; Sun & Salomon 2009). Based on own general observations, characters like stature, spike color, spike length, spike density and awn length often varies between populations or among individuals within a population.

Figure 4. Spikes of *Elymus angulatus* (left) and *E. glaucescens* (right) with multiple and single spikes per rachis node, respectively.

2.3.2 Ploidy and genomes

All species in *Elymus* are allopolyploids with the majority (75%) being tetraploids with two non-homologous sets of chromosomes (2n = 4x = 28), some (20%) hexaploids (2n = 6x =42) and a few (5%) octaploids (2n = 8x =56) (Dewey 1984; Löve 1984; Svitashev et al. 1996; Diaz 1999). The genome combinations of species are generally well documented through decades of studies of meiotic chromosome pairing in intergeneric and interspecific hybrids (Seberg & von Bothmer 1991; Jensen 1993), karyotype comparisons by Giemsa C- and N-banding (Linde-Laursen et al. 1994; Linde-Laursen & Seberg 1999), variations in repeated nucleotide sequences (Dubcovsky et al. 1997), molecular cytogenetics and *in situ* hybridization (FISH-GISH) (Tomas et al. 2012), genome-specific molecular markers such as RFLP (Restriction Fragment Length Polymorphism), RAPD (Random Amplified Polymorphic DNA), and 5S NTS sequences (Svitashev et al. 1998; Baum et al. 2015). Five different basic haplomes have so far been identified in *Elymus*, following genome designations by Wang *et al.* (1994): **St** (*Pseudoroegneria*), **H** (*Hordeum*), **P** (*Agropyron*), **W** (*Australopyron*), **Y** (unknown origin) (Dewey 1984; Svitashev et al. 1996; Sun et al. 1997, 2008; Liu et al. 2006; Petersen et al. 2011) (Figure 4.). The **St**-genome is the only haplome represented in all species and tetraploids are either having the **StH, StY**, or **StP** combination. The other haplomes and duplication of haplomes are found in hexaploids or octaploids, for example, **StHH**, **StStH**, S**tHW**, **StHY**, and **StStHH**. It is difficult to distinguish between the **StH** and **StY** combinations based on morphology. A study by Salomon and Lu (1992) shows that only two morphological characters correlate with genome combination, palea apex shape and size of cilia of palea, while characters often used for species delimitation, such as shape and length of the lemma, spike orientation, and number of spikelets per rachis node, does not. The evolutionary implication of different genome combinations is shown from crossing experiments with high crossability between species within the same and no inter-fertility between species from different genome combinations (Salomon & Lu 1992). The method of genome studies is, however, insufficient for further exploration of species origins and relationships.

2.3.3 Ecology and biogeography

The genus grows mainly in temperate areas, in both the northern (about 90% of the species) and southern hemispheres, on grasslands, semi-deserts, riverbanks, mountain slopes and valleys, among bushes, in forests and in forest margins (Lu 1993; Salomon et al. 1997). The main species diversity center is in the mountains of Central Asia with about 50% of the species in the genera (Salomon et al. 1997). The **StH** group contains approximately 50 species (Sun & Salomon 2009) distributed all over the world, from sea level up to 5000 m in altitude and in subtropical to arctic habitats (Dewey 1984; Löve 1984; Sun & Ma 2009). The diverse habitat preferences in the **StH** group makes them interesting from a genetic resource perspective (Sun & Salomon 2009). The main species diversity of the group is the northern hemisphere, but a high diversity can also be found in the southern hemisphere, in Chile and Argentina. The distribution area of different species varies, for example, *E. sibiricus* is restricted to eastern Asia, *E. caninus* to western Eurasia, and *E. mutabilis* is present throughout the area (Salomon et al. 1997). Others are endemic to restricted areas and even locations. The **StY** genome species are restricted to Central and East Asia (Dewey 1984; Löve 1984). Knowledge of species distribution and population density is often deficient with a bias in data availability favoring certain regions and easily accessed areas.

2.3.4 Taxonomy

It is difficult to establish a sound taxonomy of *Elymus* and there are still disagreements on both generic and species delimitations (Lu 1993). The classification of the genus has traditionally been based on morphological traits and Linnaeus (1753) was the first to describe the genus. The large number of species, the low number of comparable morphological characters, the large morphological variation within species and the high degree of phenotypic plasticity make it difficult to delimit species based on morphological characters alone (McMillan & Sun 2004; Petersen et al. 2011). Inflorescence with single versus multiple spikelets per rachis node has traditionally been used as a decisive character to differentiate *Elymus* from closely related genera. Some authors have a narrow circumscription of *Elymus* and only include species with multiple spikelets per rachis node while species with single spikelets were placed in *Agropyron* or *Roegneria*

(Bentham 1881; Hitchcock 1951; Baum 1983; Keng 1989). Others have taken a wider circumscription and consider the number of spikelets as an insignificant character (Tzvelev 1976; Meledris 1980; Löve 1984). It has later been shown that the separation between species with single and multiple spikelets per rachis node has no support in phylogenetic studies (McMillan & Sun 2004).

Different taxonomic classifications have been followed in different geographical areas: Hitchcock's in North America, Meledris' and Tzvelev's in Europe and Russia, and Keng's in China (Barkworth 1992). The proposal by Löve (1984) and Dewey (1984) to determine generic delimitation in Triticeae based on cytogenetics, contributed to a stabilization of taxonomy and standardization of nomenclature in *Elymus*. The combination of **StH**, **StY**, and **StP** with their respective hexa- and octaploid derivatives is considered by some authors as a broad circumscription of the genus, *Elymus sensu lato* (*s.l.*, meaning "in the broad sense"), while only including **StH** is considered to be the narrow sense, *Elymus sensu stricto* (*s.s.*, meaning "in the narrow sense") (Barkworth & von Bothmer 2009). *Elymus s.l.* is sometimes divided into strict generic classes based on genomic constitution (Figure 4): *Roegneria* K. Koch (**StY**), *Kengyilia* Yen & Yang (**StPY**), *Campeiostachys* Drobov (**StHY**), *Anthosachne* Steudel (**StYW**), *Stenostachys* Turcz. (**StHW**), *Douglasdeweya* C. Yen, J. L. Yang & B. R. Baum (**StP**) and *Elymus s.s.* (**StH**) (Yen et al. 2005; Yen & Yang 2009). Several of the names are considered as sections by other authors (Salomon & Lu 1992). Even though molecular data has been introduced, the generic delimitations in most contemporary studies are still based on genome combinations (Baum et al. 2015). Petersen *et al.* (2011) argue that a combination of morphology, cytology and molecular data should be used to establish a classification of *Elymus*. In a diversity study of 24 **StH** *Elymus* species, Baum *et al*. (2016) used canonical discriminant analysis based on nr5SDNA and concluded that *Elymus* classification could be defined based on ploidy and geographical regions.

Taxonomical classifications change depending on the accumulation of new data and the adopted species concept. In *Elymus*, several disputable taxonomic names have been proposed often referring to morphological variants in a continuous range of variation or to biologically insignificant forms (von Bothmer & Seberg 1995). The settlement of taxonomic units and

genetic relationships is necessary for clear establishment of habitat preferences, distribution and for optimal use and conservation.

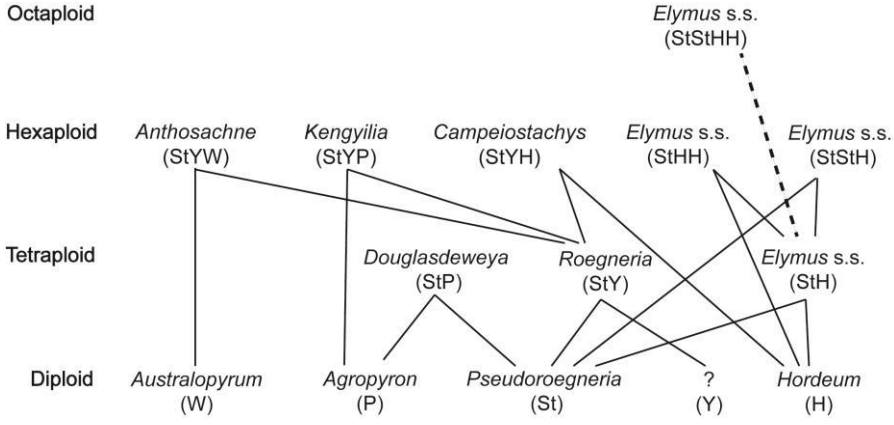

Figure 4. Generic delimitation and relationships based on genome combinations proposed by Fan et al. (Fan et al. 2013), Yen et al. (Yen et al. 2005), and Yen and Yang (Yen & Yang 2009). Lines indicate allopolyploid origins from diploid and tetraploid progenitors. Dashed line indicates a potential autopolyploid.

2.3.5 Phylogeny

Repeated reticulate evolution between independent evolutionary lineages is difficult to handle within a conventional phylogenetic framework based on the presumption of exclusively divergent evolution. Frequent introgressive hybridization and polyploidization make it difficult to resolve the phylogeny also within smaller species groups of the genus (Sun 2014; Baum et al. 2015; Wu et al. 2016). Several studies investigate the phylogeny of *Elymus*, often together with diploid progenitors (Mason-Gamer 2001; Mason-Gamer et al. 2002; Helfgott & Mason-Gamer 2004; Sun et al. 2008; Baum et al. 2015, 2016). These studies mainly include a few samples and one or a few gene sequences, leading to incomplete phylogenies with a high degree of unresolved polytomies and/or poorly supported branches. Several studies show that the **H**-genome donor differ between species from America and species from Eurasia, suggesting separate origins of the **StH** genome combination in North America and Eurasia (Jaaska 1992; Linde-Laursen et

al. 1994; Sun et al. 2008; Yan & Sun 2012). However, polymorphic **H** genomes as a consequence of reciprocal hybrids, introgression and subsequent backcrossing, cannot be excluded (Zuo et al. 2015). In contrast, other studies advocate a single origin (Mason-Gamer et al. 2010; Wang et al. 2011; Dong et al. 2015). It is common that polyploids with the same parental background arise multiple times (Soltis & Soltis 1993). *Pseudoroegneria* (**St**) is likely the female genome donor for both the **StH** and **StY** combinations (McMillan & Sun 2004; Mason-Gamer et al. 2010; Petersen et al. 2011), but it remains unknown if the **St** genome derives from one or several *Pseudoroegneria* species (Liu et al. 2006, Sun et al. 2008, Gao et al. 2015, Zuo et al. 2015, Lei et al. 2020). Multiple origins of morphologically similar species, sometimes called cryptospecies, have been documented (Agafonov 2004; Liu et al. 2006; Yan & Sun 2012; Gao et al. 2015; Zuo et al. 2015; Leo et al. 2022), and karyotype rearrangements by inter-genomic translocations have been reported in the Triticeae (Devos & Gale 2000; Wang et al. 2010), including *Elymus scabrifolius* (Tomas et al. 2012). This further complicates both the choice of investigated taxa and the interpretation of phylogenetic results.

2.4 Conservation

Conservation of CWR is necessary to guarantee a continuous access to genetic resources for future crop adaptation and improvement. However, it is not possible to conserve all genetic variability (allelic variants) in a natural population or species due to high demand of resources, human labor and financial limitations. However, it is preferable to preserve as much variability as possible both inter- and intrapopulational. There is a need to minimize the number of accessions and samples needed for each taxon and at the same time maximize the conserved genetic diversity. To make informed selections, we should advantageously know the genetic structure and phenotypic variability within species. In conservation of genetic resources, a special attention should be paid to isolated populations and those on the edges of the ecogeographical distribution (von Bothmer & Seberg 1995). There is a higher chance that these populations harbor unique traits, even though they tend to be genetically more uniform than central populations.

Frankel *et al.* (1995) advocate *in situ* conservation prior to *ex situ*. The main advantage, they claim, is that populations *in situ* are still subject to the evolutionary process through local selection and adaptation. In addition, populations on-site may include a higher number and a wider diversity of genotypes and species, and continuous coevolution of inter- and intraspecific interactions, including species associates like pests, diseases, and mutualistic species. *Ex situ* collections should be considered as a valuable complement to *in situ* conservation and needs to be readily available to biologists and plant breeders for efficient research and use. The static conservation of *ex situ* populations is exposed to long-term risks due to repeated regeneration. Frankel *et al*. (1995) point out the importance to avoid or reduce natural selection in order to conserve the original genetic diversity. The risk of genetic drift increases with smaller number of individuals in the regeneration and contamination from other accessions or adjacent wild populations will alter the genetic composition. In addition, there is always a risk of accession loss due to pests, climatic conditions, or human error. *In situ* populations are also vulnerable to pests, climatic conditions and human exploitation, further motivating *ex situ* collections. It is expensive and time consuming to keep a collection with repeated regenerates, but it is also resource-intensive to monitor wild stands of populations.

In the mission to collect wild genetic resources, Frankel *et al.* (1995) argue that endangered species and populations, species easily used in breeding programs, and populations with the likelihood of hosting significant genetic resources for the improvement of cultivated species should be prioritized. In the case of *Elymus*, some species are endemic or have restricted distribution areas and some are threatened by extinction. Others have wider distribution ranges, but could have valuable marginal populations worth preserving. Knowledge of species distributions and ecogeography is needed to assess the risk of extinction or endangerment as well as to document populations or ecotypes with valuable traits.

The classification of organisms is usually based on phylogenetic data with the virtue of being more informative compared to overall similarity. Genetic relationships helps us understand the evolution of biodiversity and allow us to infer traits among closely related taxa as genetic closeness often reflects phenotypic similarity. In the context of conservation, taxonomy can help us decide which taxa to choose in order to maximize the overall range of genetic diversity (Maxted et al. 2020). To preserve taxa that are distantly related

ensures the broadest range of genetic diversity and vouch for comprehensive gene pools representing the phylogenetic diversity of the genus as well as the ecological and genetic diversity of species (Frankel et al. 1995). Taxonomy can also assist conservationists to select those taxa most likely to be used by breeders. Relatedness is often connected to successful crossability, which is still needed to get novel adaptive traits into domesticated crops (Maxted et al. 2020).

Salomon *et al.* (1997) proposed an *Elymus* core collection with the objective of building an available standard for research and breeding. They suggested that species selection for such a collection should be based on primarily taxonomy, ecogeography and cytogenetic data. However, the project was unfortunately not set up.

2.5 Utilization of crop wild relatives

Elymus species have long been of interest to crop and forage breeders. A range of favorable traits, like perenniality, stress tolerance, apomixis, nutritional content and disease resistance could potentially be transferred into domesticated crops to create a more sustainable food production. Multiple *Elymus* species have for research and pre-breeding purposes, been hybridized with mainly wheat using embryo rescue (Salomon et al. 1988; Cox et al. 2002). However, gene transfer by crossing over is very low, because of little or no homology between the different genome sets, and other advanced methods of gene transfer are needed. Both *Agrobacterium*-mediated transformation and particle bombardment are for example routinely used to transfer genes into barley and wheat (Kumar et al. 2020).

Several species are used today as forage grasses in North America (for example E. trachycaulus, E. glaucus, E. lanceolatus, E. canadensis, and E. virginicus) and eastern Asia (for example E. dahuricus, E. excelsus, E. tangutorum, E. nutans, and E. sibiricus) and many cultivars have been developed (Hu et al. 1992; Aubry et al. 2005). An analysis conducted in northeast Mississippi, USA, showed similarities in forage quality between eight Elymus species and other domesticated and commercially grown grass cultivars (Rushing & Baldwin 2013). Elymus still harness unexplored qualities when it comes to the introduction of new forage grasses. Harlan (1983) points out the difficulties in forage breeding including the complex interaction between multiple species in a pasture mix, the range of grazing

and harvesting regimes and adaptation to the local environment. He believed that the most important characters in developing a new forage variety are domesticability, ease of establishment, ease of maintenance, and production of livestock. Species with a weedy or colonizer behavior are easier to establish and do not tend to decrease during use. One strategy in the introduction of new forage species is to use genotypes from productive ecotypes. This should be considered as a selection procedure rather than domestication and breeding. This stresses the importance of finding, collecting and conserving wild genetic resources, as well as plant communities, that could be advantageous in future forage crop development. Many Elymus species grow and successfully compete in natural pastures and indigenous species in the genus could be interesting for new and locally adapted forage crops.

Elymus has a rich ethnobotanical record with documented use as food, forage, medicine, and constructing material in both North America and Eurasia (Frawley et al. 2020). Characters like a compact and determinate inflorescence structure, a predominantly self-pollinating reproduction behavior and the ability to hybridize with close relatives make *Elymus* promising also for *de novo* domestication as a food crop (Frawley et al. 2020). However, seed shattering is a problem that needs to be solved (Zhao et al. 2019).

3 Research aims

3.1 General aims

The general aim of this thesis is to investigate the dynamic nature of *Elymus* (*s.s.*) evolution including genetic relationships, genetic diversity, within species variation and phylogeography. The Papers are planned and structured based on three taxonomical levels of diversity: variation within a species (Paper I), variation within a region (Paper II), and variation within a genus (Paper III). An understanding of the species distribution in combination with intra- and interspecific genetic diversity is important for determining evolutionary adaptive potential, predicting limiting factors for the further spread, and assessing threatened species groups. The thesis stresses the importance of having both a holistic perspective and detailed knowledge in plant systematics.

3.2 Specific aims

Paper I included a study of a species complex in Eurasia with a focus on differentiation and speciation. The study investigated the variation within *E. mutabilis* over the whole distribution area and included the morphologically similar *E. transbaicalensis* for comparison. The objectives were to study the phylogeographic population structure of *E. mutabilis* and find molecular evidence of diversity and eventual differentiation between Nordic and Asian *E. mutabilis* populations. An additional purpose was to emphasize the

importance of population studies with wide taxonomical and geographical scopes. The objectives of paper II were to investigate the genetic relationships between American *Elymus* species, and the evolutionary history, migration routes and the phylogeographic pattern of the species in South America. This will provide a basis for further, ongoing studies of *Elymus* diversity in South America for conservation and utilization purposes. The objectives of paper III were to study the genetic relationships of *Elymus s.s.* (**StH**) of the world and to gather phylogenetic evidence for further taxonomical work.

4 Material and methods

4.1 Plant material

The plant material used in Papers I-III comes from wild populations, from novel seed and leaf collections, from gene bank's original collections, or collections that have been regenerated and/or multiplied. In genetic diversity studies, a natural population is often defined as a local group of individuals of a specific taxon that is isolated from other similar groups and has a free exchange of genes (von Bothmer & Seberg 1995). The lack of knowledge about gene flow often results in a more pragmatic definition where populations are considered as all individuals of a specific taxon found in a definite, ecologically homogeneous, place at a particular time. The leaf material represents the actual genotypes present in the field while seeds represent the genetic "potential". Accessions with a wild origin that have been regenerated or propagated in a gene bank should correspond to the original genetic setting if selection has been avoided. Seberg (1991) raises some issues that often occur with Triticeae gene bank material. There is a risk of contamination through seeds and pollen when accessions are recycled and multiplied repeatedly, especially if the curators are not specialists in Triticeae. Many accessions also lack the important passport data and are often incorrectly determined due to insufficient or outdated taxonomic treatment.

All gene bank material in this thesis comes from SLU Triticeae germplasm collection (STGC). The seeds of *Elymus* are orthodox and can be

stored for a long time in a cold and dry environment, making them suitable for preservation in gene banks. The material in STGC originates from SLU collections of wild populations or from collaboration partners that are specialized in Triticeae. The material has been collected with the attention of taxonomic and phylogenetic research as well as for genetic diversity studies and conservation of genetic resources. The accessions represent the whole distribution area of *Elymus* and the purpose is to have species and populations readily available for internal projects and to maintain reference material for published work. Parts of the collection have been transferred to NordGen (The Nordic Genetic Resource Center) with the purpose of making seed material available for internal and external research and breeding projects. Based on own experience, seed samples that are kept in −18°C freezers are most often viable and germinate with ease in 20°C. Seed stored in 4°C have low or no germination.

A 'sufficient' or 'appropriate' taxon sampling is dependent on the scientific questions being addressed, but has to be carefully considered as an important determinant for accurate phylogenetic estimation (Young & Gillung 2020). A denser taxon sampling improves phylogenetic accuracy and reduces the risk of systematic bias (Heath et al. 2008). All subordinate taxa should ideally be sampled in order to unravel the phylogeny of an entire taxonomic unit. However, the sampling is dependent on the access and quality of the plant material as well as limited financial resources for genotyping. When information on the genetic structure of a species is scarce, populations that are geographically well separated should be sampled to obtain representative investigative taxonomical units.

Accessions included in the thesis were selected based on taxonomical and geographical distance and coverage with a higher priority of collections from expeditions arranged by SLU with comprehensive passport data. Some accessions were stored as bulk collections but with intact spikes. In these cases, one seed per spike was randomly selected. In the case of bulk collections without separate spikes, seeds were randomly selected. Other accessions were stored as 'single spike collections' with individual spikes in separate bags in a series derived from a collection transect. To avoid closely related individuals, spikes were selected evenly separated from each other in the series and three seeds per spike were taken for germination and later reduced to one individual per spike.

Seeds were germinated and grown in a greenhouse at SLU Alnarp at a temperature around 20°C with ventilation hatches opening at 23°C. A standard peat-based potting mix (SW Horto, Hammenhög, Sweden) was used without additional fertilizers. Leaf samples were taken when first leaf reached five to ten cm. Plants were repotted after sampling for further cultivation and collection of vouchers. Vouchers were taken continuously during flowering. Individuals that did not flower were kept outside during winter to facilitate flowering following year. Plants that still did not flower were identified from the original spike collection. The key from Seberg and Petersen (1998) and Flora of North America (Barkworth et al. 1993) was used for confirming South American and North American species, respectively. The Flora of China (eFloras 2008) and Flora of Siberia (Peschkova 1990) were used to confirm Asian species including the determination and separation of *E. mutabilis* and *E. transbaicalensis* in Paper I. Von Bothmer and Seberg (1995) pointed out that the taxonomic literature only rely on plant material available to the author. Moreover, distribution maps often reflect collecting routes of botanist rather than the actual distribution of the species.

4.2 Genotyping

Phylogenetic studies have historically been relying on phenotypic characters for evolutionary inference, including morphology, anatomy, biosystematics and cytogenetics. The introduction of Sanger sequencing and Polymerase Chain Reaction (PCR) in the late 1970s made genotypic data available and the field of molecular phylogenetics has since made rapid progress in the development of new affordable techniques with an increasing accumulation of informative data from a greater number of genotypes. A genetic marker is a detectable alteration in the DNA sequence that can be used to identify individuals, populations, species or genes. Earlier phylogenetic studies of *Elymus* have been using nuclear and chloroplast genome sequences (Mason-Gamer 2001, 2013; Mason-Gamer et al. 2002, 2005, 2010; Liu et al. 2006; Sun 2007; Sun et al. 2008; Sun & Ma 2009; Zhang et al. 2009; Wang et al. 2010, 2011; Dizkirici et al. 2010; Hodge et al. 2010; Sun & Sun 2012; Yan & Sun 2012; Gao et al. 2014, 2015; Baum et al. 2015, 2016; Dong et al. 2015). These studies include a relatively small set of loci with a limited number of phylogenetically informative gene markers, thus sensitive to

stochastic and/or sampling errors (Young & Gillung 2020). Genetic markers, like morphological data, may be subject to convergence and parallelism, and single gene sequences are often not informative enough to be used for resolving phylogenetic relationships (Patwardhan et al. 2014; Choi et al. 2019). Different gene trees may be contradictive and create inconsistent tree topologies arising from incomplete lineage sorting and/or ongoing hybridization, which will give an incorrect representation of the species tree (Choi et al. 2019). Increasing sequence length, larger amount of sequencing data, or a higher number of markers often have a beneficial effect on the accuracy of phylogenetic analyses (Heath et al. 2008).

The introduction of Next Generation Sequencing (NGS) in the last decade has revolutionized molecular phylogenetics by reducing the sequencing cost per nucleotide and increasing the generation of informative data. Instead of analyzing a few genes, phylogeneticists now have access to potentially thousands of loci from across the genome, reducing the impact of stochastic errors and data availability (Young & Gillung 2020). NGS can be used for reduced-representation genome sequencing (also called target enrichment) and shotgun sequencing, the two main sequencing methods used in phylogenomics (Young & Gillung 2020). However, the new techniques are still prone to systematic errors and highly dependent on data quality, inference methods or biased taxon and gene sampling (Delsuc et al. 2005). The analyses require complex bioinformatic procedures with a large variety of methods for acquisition, manipulation, analysis and interpretation of the extensive datasets (Young & Gillung 2020).

A single nucleotide polymorphism (SNP) is a nucleotide difference between two DNA sequences at a specific position in the genome. These differences accumulate through spontaneous mutations over time and can, for example, be used for comparisons of individuals and for association studies. SNPs have become the most popular molecular markers to use in plant genetic analyses (Edwards et al. 2007). They are the most frequent form of genetic variation, widely distributed over the genome, inheritable and with a relatively low mutation rate, which makes them useful on more or less all taxonomic levels in both phylogenetic and diversity studies. By analyzing thousands of SNPs distributed across the whole genome, it is possible to decrease the risk of parallelism and convergence.

Genotyping-by-sequencing (GBS) is a high-throughput method to discover thousands to millions of SNPs that can be used for various genetic

studies. The method was first described by Elshire *et al.* (2011). GBS uses restriction enzymes (RE) to digest DNA. The sticky ends of the restriction sites are ligated with barcode adapters and amplified by PCR for the GBS library construction. The sequencing results in about 100 bp long single-end reads. This approach reduces genome complexity compared to whole-genome sequencing and generates genome-wide SNP markers for further diversity analyses. The technique is relatively cheap with the possibility to simultaneously sequencing several DNA samples. In subsequent bioinformatic analyses, raw sequence data are filtered and aligned, and SNPs are identified and scored for various coverage, depth and genotypic statistics (see further comments below). The technique is highly reproducible with the possibility to call known SNPs in the sequenced samples. GBS is suitable for studies of species with high diversity and large genomes. The method does not require a reference genome for sequence tag mapping, which makes it suitable for *de novo* sequencing of less studied species without prior knowledge of its genome structure. GBS has been used in many genomic diversity studies including phylogenetic and population genetic studies. The method can also be used for pre-breeding and breeding purposes with the possibility of implementing genome-wide association studies (GWAS), genetic linkage analysis, and genomic selection (GS) (He et al. 2014). An often addressed disadvantage of GBS is the extensive production of missing data due to low coverage of sequencing, lack of the restriction sites or polymorphism in the restriction site in a particular sample (Favre et al. 2021). The bioinformatic analyses are also highly complex although streamlined open access analyses pipelines have been developed to cope with these difficulties, making GBS as a method more accessible for both researchers and breeders without requiring in-depth knowledge of bioinformatics. GBS has been used for phylogenetic studies of *Triticum* (Hyun et al. 2020) and is also suitable for population diversity studies (Favre et al. 2021). Earlier studies of *Elymus* population diversity have been using isozyme, allozyme, RAPD, microsatellite (SSR), ISSR, and seed storage protein markers (Sun et al. 1998b, 1999, 2001, 2002, 2006a, 1997, 1998a; Agafonov et al. 1998; Diaz et al. 1998, 1999a, 2000; Zhang et al. 2000, 2002; Sun & Salomon 2003; Gaudett et al. 2005; Stevens et al. 2007; Ma et al. 2012; Wu et al. 2016; Xiong et al. 2019). The advantage of GBS is the higher resolution of fine scale diversity due to the high number of genetic markers and the method has earlier been used in diversity studies in Triticeae: *Elymus lanceolatus* ssp.

lanceolatus (Li et al. 2018) and *Agropyron cristatum* (Baral et al. 2018). GBS is used in Paper I.

DArTseq technology is another method using genotyping-by-sequencing, with the same basic idea of reducing the genome complexity, but with a different way of preparing the library (www.diversityarrays.com/). It uses a tailored combination of RE to provide a selection of genome fractions from low-copy sequences. These correspond predominantly to active genes while repetitive and less informative fractions of the genome are avoided. The method allows cost- and labor-efficient whole genome SNP analysis and is a useful DNA polymorphism identification method highly suitable for non-model species. DArTseq has earlier been used in phylogenetic analysis of *Triticum* and *Aegilops* (Edet et al. 2018) and *Secale* (Al-Beyroutiová et al. 2016). DArTseqLD™ is a variation of DArTseq with lower marker density, which further lower the cost of the assays. The method is capable of providing up to at least 10,000 markers per analyzed line. DArTseqLD™ is used in Paper II and III.

Other comparable methods are Whole-genome Re-sequencing, Whole-exome-capture Sequencing, RNA-sequencing (RNA-seq), Amplicon Sequencing, and Restriction-site-associated DNA Sequencing (RADseq) which all differ in cost-effectiveness, ease of use, required sample quality and bioinformatic workflow (Onda & Mochida 2016; Young & Gillung 2020). GBS and DArTseq were chosen in this thesis because of the low cost, the availability of a well-developed and easily accessible downstream workflow, and the requirements of sample quality. The acquired amount of informative data is comparable with RNA-seq and RADseq, but much greater compared to older methods like SSR and Sanger sequencing, resulting in more accurate and less ambiguous representative phylogenetic tree.

4.3 Data analysis

4.3.1 Data filtering

Accurate and bias-free phylogenetic inference is highly dependent on informative and reliable genetic markers. Filtering of NGS raw sequence data is necessary to remove ‘noise’ such as highly variable loci at fast

evolving sites and sequencing errors containing non-phylogenetic signals (Jeffroy et al. 2006; Townsend et al. 2012). Because of the large number of informative genetic markers in NGS approaches, a certain amount of missing data can be tolerated without losing accuracy in the phylogenetic analysis (Philippe et al. 2004). The goal is to obtain a dataset with as many accurate and informative markers or loci as possible.

Several assembly workflows are available for determining various factors like coverage, depth and genotypic statistics prior to SNP discovery. *ipyrad* is used in the study of *E. mutabilis* (Paper I), which is an assembly pipeline originally developed for RADseq but applicable also for GBS (Eaton 2014; Eaton & Overcast 2020). The pipeline is easily used with documentation, tutorials, and an associated analysis toolkit for various diversity study purposes. To avoid sequencing errors, reads are filtered on quality, minimum depth and maximum alleles per individual. Maximum depth is used to remove highly repetitive fragments and missing data is reduced by controlling the minimum number of samples that must have data at a given locus. Another issue is the identification of orthologous loci and reduction of poor alignment. The data is filtered using a clustering threshold that determines the level of similarity at which sequences are considered homologous. A low clustering indicates paralogous sites. The set of maximum number of SNPs and indels (insertions and deletions) allowed in a final locus reduces potential effects of poor alignment. Paralogs are identified by high fractions of heterozygous bases, indicating fixed difference, as well as a large number of shared polymorphic sites in a locus.

The two phylogenetic studies (Paper II and III) used the R package *dartR*, recommended by Diversity Arrays Technology Pty Ltd, as the downstream pipeline for filtering the data sets. The quality of the raw data was controlled prior to analysis by filtering using *dartR*. The data sets were filtered on reproducibility, which is a measure of the consistency of scoring technical replicates, to remove unreliable loci. All but one of the multiple SNPs in the same fragment were randomly removed to avoid the potential influence of linkage. The data sets were also filtered on locus and individual call rate to remove samples with excessive missing data. Finally, monomorphic loci and loci with a minor allele frequency (MAF) ≥2% were removed. The *dartR* package also provides tools for population statistical analyses and visualization using Principal Component Analysis (PCA) and Neighbour-joining (NJ) trees.

4.3.2 Model of evolution

The accuracy of the phylogenetic inference is not only dependent on informative characters but also on the quality of the applied model of evolution. Several methods of phylogenetic inference have been put forward like Neighbor joining (NJ), Maximum parsimony (MP), Maximum likelihood (ML) and Bayesian Inference (BI). MP recovers one or more optimal trees based on the simplest relationship of discrete characters that require the smallest number of evolutionary change for a certain group of taxa. NJ uses a clustering algorithm to establish relationships between sequences according to their genetic distance. The number of mutations that have accumulated since a split from a common ancestor is used as an estimate to generate genetic distances and infer phylogenetic relationships. The method generates a single tree without the possibility of comparing it to other potential trees. The method is fast and more robust than MP, but considered statistically weaker compared to ML and BI. The NJ based setting Neighbour-Net in SplitsTree is used in Paper I-II, which gives a graphic depiction of the distribution of genetic divergence in the samples.

Whelan *et al*. (2001) argue that the probabilistic methods, such as ML and BI, should be preferred. Methods based on probability explicitly incorporate the processes of sequence evolution in the models that they use and are considered as more complex and accurate in their statistical assumptions compared to more simplistic methods. Neither NJ nor MP require a model of evolutionary change. Ml and BI are, however, also subject to inconsistency under some conditions if assumptions are not met, for example in the case of independent evolutionary changes at different sites and the homogeneity of the nucleotide-substitution process (Delsuc et al. 2005). This could potentially generate erroneous signals that will compete with the true phylogenetic signal. MP and BI investigates a set of possible phylogenetic trees, trying to identify those that are best, i.e. those that are associated with the highest likelihood (ML) or prior probability (BI).

The presence of incomplete lineage sorting (ILS) can be challenging in phylogenetic and population studies making gene trees different from the species tree estimation. Methods like ML and BI use concatenated approaches in which loci are aggregated into a super matrix. These methods are sensitive to ILS resulting in the return of incorrect trees with high statistic support (Chou et al. 2015). Coalescence-based methods (CBM) take another approach by first estimating gene trees and then combine gene trees into a

species tree. CBM are therefore less affected by ILS by showcasing gene tree conflicts (Young & Gillung 2020). A lot of effort is put into the development of better reconstruction methods for using genomic data for phylogenetic purposes (Delsuc et al. 2005). In this thesis, ML, BI as well as a CBM were used for comparison and scientific transparence.

4.4 Ethical and legal considerations

Genetic resources, including all plant material, shall be considered national property, which raises both ethical and legal questions about plant collection and plant research. The *Convention on Biological Diversity* (CBD) is a multilateral treaty with the objective to develop national strategies for the conservation and sustainable use of biological diversity, and a fair and equitable sharing of benefits arising from the utilization of genetic resources. The *Nagoya Protocol on Access and Benefit-sharing* (ABS) is a supplementary agreement to CBD with the objective to provide a transparent legal framework for the implementation of a fair and equitable sharing of benefits agreed upon in the CBD. The novel collections in this thesis comes from Chile, which is not a party to the Nagoya Protocol. There is also no national regulation of plant collecting other than in protected areas. However, a request of access to genetic material *in situ* was sent to and approved by the Genetic Resource Coordinator at the Agricultural Research Institute (INIA). Permissions for collecting in national parks were issued by The National Forestry and Parks Agency (CONAF). The collaboration, cooperation, and contribution to scientific research and publications were considered eligible measures for benefit sharing among all parties. Seed material and herbarium vouchers are kept in Chile for conservation purposes and future investigations. Only dry leaf material was transferred to Sweden for genotyping. The digital sequence information has been deposited in public databases for scientific reproducibility and for others to use in further research. The saving of raw data is extra important as the development of analysis methods progresses and older versions of programs become inaccessible (Young & Gillung 2020). The gene bank material used in this thesis comes from previous collections in countries where some now are parties of the Nagoya Protocol. The protocol entered into force on 12th of October 2014 and material collected before this date is not considered included in the agreement. *The International Treaty on Genetic Resources*

for Food and Agriculture (ITPGRFA) is an international agreement to have an open exchange of food crops and their genetic materials. *Elymus*, as a wild relative to wheat, is included in the Annex I of ITPGRFA as an exception from ABS. However, *in situ* material is not included in the treaty, which means that new collections of wild material are subject to national ABS regulation.

5 Results and discussion

This thesis provides a large data set, including many taxa as well as a high number of informative phylogenetic characters, giving a deepened understanding of the intricate nature of *Elymus* evolution. The taxonomical progression benefits from more advanced and explicit genetic methods shown by the phylogenetic results in Paper I-III. The NGS techniques confirm the genome based revision of the genus by Dewey (1984) and Löve (1984), and support the importance of recognizing genome combinations in delimitation of genera. This chapter aims for lifting the main and combined results from the three Papers as well as elaborate further on the discussion and implications of the research.

5.1 Phylogeny and origin of *Elymus s.s.*

Paper III investigates the evolutionary relationships of *Elymus s.s.* and comprise 36 taxa from the whole distribution area as well as non-target taxa. The wide geographical and taxonomical inclusion of samples vouches for an adequate representation of the group. The separation of **StH** and **StY** genome combinations is evident in Paper III and has been previously documented by molecular markers (Svitashev et al. 1996; Sun et al. 1997). The investigated **StH** taxa cluster into two main clades (Fig. 5), with one clade comprising predominantly American species (American clade) and the other Eurasian species (Eurasian clade). The Eurasian clade can be divided into three subclades: 1. the *sibiricus-confusus* clade, 2. the *mutabilis-fibrosus-caninus* clade, and 3. the *alaskanus-macrourus-transbaicalensis* clade. The American clade can further be divided into two subclades with *E. lanceolatus*

as a more or less unresolved sister species to the other clades: 1. the *trachycaulus* clade, and 2. the "transcontinental" clade.

Papers II and III confirm the affiliation of the included species to the previously assigned **StH** genome combination. The nesting of both *Hordeum* (**H**) and *Pseudoroegneria* (**St**) within *Elymus s.s.* once again affirm the allopolyploid origin of the group. The placement of different species of *Pseudoroegneria* in the two major clades suggest a polyphyletic origin also of the **StH** genome combination. The positioning of the *Hordeum* species in the SplitsTree analysis further suggest the involvement of several different *Hordeum* species. These results support a subdivision of the **StH** genome into $\mathbf{St_1H_1}$ and $\mathbf{St_2H_2}$. However, the results cannot exclude the possibility of several independent origins also within each subgroup.

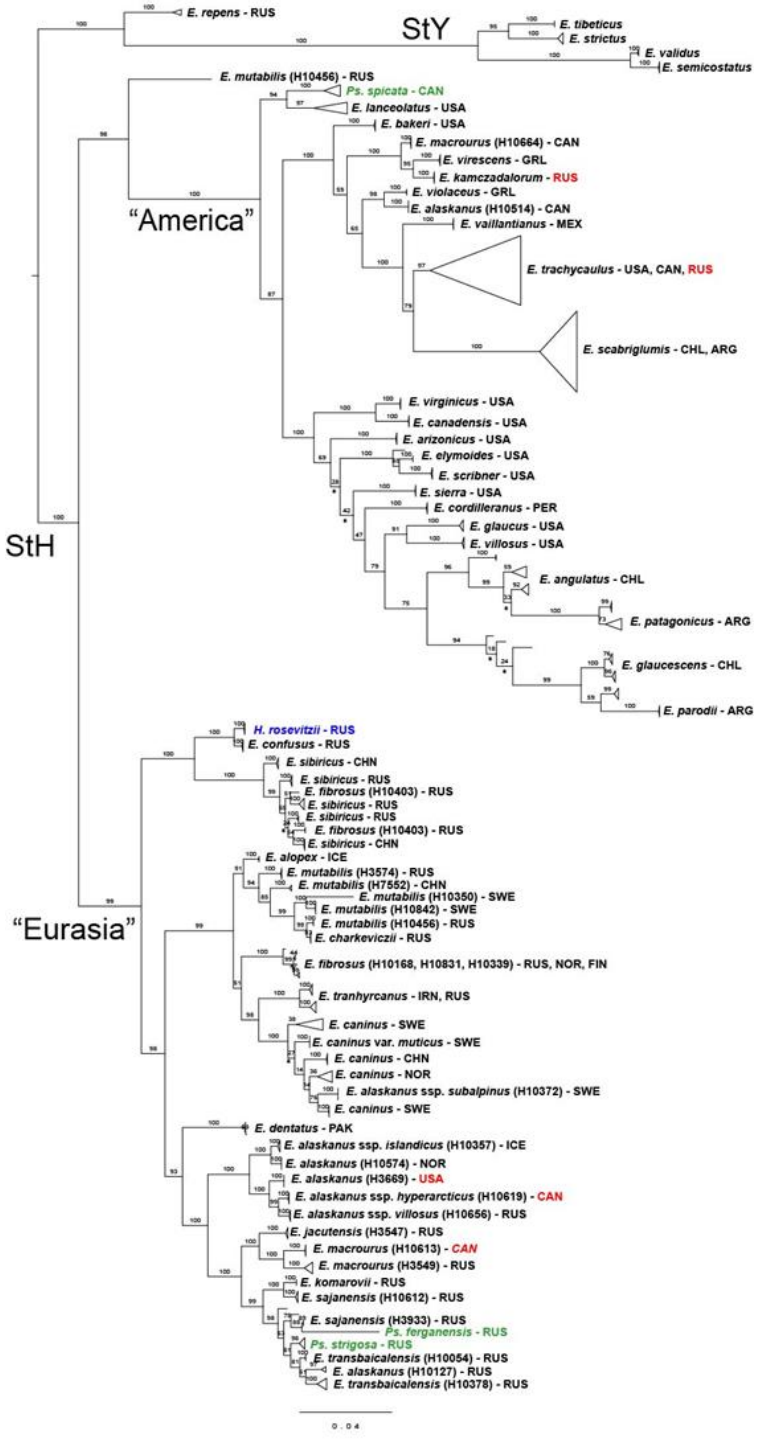
E. repens - RUS
StY
E. tibeticus
E. strictus
E. validus
E. semicostatus
E. mutabilis (H10456) - RUS
Ps. spicata - CAN
E. lanceolatus - USA
E. bakeri - USA
E. macrourus (H10664) - CAN
E. virescens - GRL
E. kamczadalorum - RUS
E. violaceus - GRL
E. alaskanus (H10514) - CAN
E. vaillantianus - MEX
"America"
E. trachycaulus - USA, CAN, RUS
E. scabriglumis - CHL, ARG
E. virginicus - USA
E. canadensis - USA
E. arizonicus - USA
E. elymoides - USA
E. scribner - USA
E. sierra - USA
E. cordilleranus - PER
E. glaucus - USA
E. villosus - USA
StH
E. angulatus - CHL
E. patagonicus - ARG
E. glaucescens - CHL
E. parodii - ARG
H. rosevitzii - RUS
E. confusus - RUS
E. sibiricus - CHN
E. sibiricus - RUS
E. fibrosus (H10403) - RUS
E. sibiricus - RUS
E. sibiricus - RUS
E. fibrosus (H10403) - RUS
E. sibiricus - CHN
E. alopex - ICE
E. mutabilis (H3574) - RUS
E. mutabilis (H7552) - CHN
E. mutabilis (H10350) - SWE
E. mutabilis (H10842) - SWE
E. mutabilis (H10456) - RUS
E. charkeviczii - RUS
E. fibrosus (H10168, H10831, H10339) - RUS, NOR, FIN
E. tranhyrcanus - IRN, RUS
E. caninus - SWE
E. caninus var. muticus - SWE
E. caninus - CHN
E. caninus - NOR
E. alaskanus ssp. subalpinus (H10372) - SWE
E. caninus - SWE
"Eurasia"
E. dentatus - PAK
E. alaskanus ssp. islandicus (H10357) - ICE
E. alaskanus (H10574) - NOR
E. alaskanus (H3669) - USA
E. alaskanus ssp. hyperarcticus (H10619) - CAN
E. alaskanus ssp. villosus (H10656) - RUS
E. jacutensis (H3547) - RUS
E. macrourus (H10613) - CAN
E. macrourus (H3549) - RUS
E. komarovii - RUS
E. sajanensis (H10612) - RUS
E. sajanensis (H3933) - RUS
Ps. ferganensis - RUS
Ps. strigosa - RUS
E. transbaicalensis (H10054) - RUS
E. alaskanus (H10127) - RUS
E. transbaicalensis (H10378) - RUS
0.04

Figure 5. Phylogenetic tree from Paper III based on Maximum Likelihood (ML), using 10 000 replicates and the GTR + G substitution model, and Bayesian inference (BI). Both analyses included 239 genotypes and 4220 binary SNPs. Bootstrap values are indicated above branches. Branches with a posterior probability <0.95 are marked with *, and accession numbers in (). Red colour indicates geographical exceptions from the general origin of the clade. *Pseudoroegneria* species are marked in green and *Hordeum* species in blue.

5.1.1 American species – the *trachycaulus* subclade

The *trachycaulus* subclade is highly supported in the ML and BI analyses (BS = 100%; PP = 1), but with weak internal branches. Four internal groups can be identified, but the relationships among those are uncertain.

Elymus trachycaulus has often been described in terms of a species complex greatly affected by ecotypic differentiation, hybridization, introgression and genetic continuity between species. The species complex has been suggested to include North American as well as circumpolar taxa such as *E. alaskanus*, *E. macrourus*, *E. scribneri*, *E. sierrae*, *E. latiglumis* and *E. subsecundum* (Barkworth 1994). The results in Paper III separates most of these taxa on a phylogenetic basis. The group is also morphologically similar to *E. caninus* (Sun & Li 2005; Sun et al. 2006a), and the two have been treated as conspecific subspecies (Jozwik 1966; Hitchcock 1969), even though most authors consider them as distinct (Hitchcock 1951; Dewey 1975). The separation of *E. trachycaulus*, *E. caninus* and *E. alaskanus* is also supported by Sun et al (2006a) who used RAPD markers and by Larson *et al.* (2003) who used AFLP markers. ***Elymus bakeri*** is sometimes considered a subspecies of *E. trachycaulus* (Löve 1984), but is recognized as a distinct species in Flora of North America (Barkworth et al. 1993) and is placed as a sister species to the rest of the subclade.

All *E. trachycaulus* populations, except H10664 from Kamchatka, form a highly supported monophyletic group together with *E. scabriglumis* and *E. vaillantianus*. This suggests that ***E. vaillantianus*** from Mexico should not be considered conspecific to *E. repens* as listed in the Catalogue of New World Grasses (Poaceae): IV. Subfamily Pooideae (Campbell & Soreng 2003), but rather as a sister species or a part of *E. trachycaulus* or the *E. trachycaulus* complex. The hexaploid ***E. scabriglumis*** turns out as separate from the other South American species and most likely derive from *E. trachycaulus*. The taxon only occurs in a limited area in Northern Chile and Argentina (Seberg & Petersen 1998) and the complete origin is still unknown. As discussed in

Paper II, the PCA analysis indicates a certain similarity between *E. scabriglumis* and the other South American species, which could be interpreted as potential introgression or even a genome contribution from a South American *Hordeum* species.

The accession H10664, identified as *E. trachycaulus,* group with *E. virescens* and *E. kamczadalorum* in a highly supported clade (BS = 100%, PP = 1). The treatment of ***E. virescens*** is summarized by Sun *et al.* (2006b). They conclude, based on AFLP and morphological similarity, that the taxon should be considered a subspecies to *E. trachycaulus,* and Flora of North America treats the taxon as *E. trachycaulus* subsp. *virescens* (Barkworth et al. 1993). Biosystematic analyses show that the Kamchatkan ***E. kamczadalorum*** is differentiated from other sympatric **StH** species (Agafonov & Salomon 2002), and the phylogenetic results from Paper III suggest a close relationship to the North American *E. virescens*. According to Flora of Siberia, only *E. trachycaulus* subsp. *novae-angliae* is found in Russia (Peschkova 1990), and the Russian samples included in the study group together with the North American accessions of *E. trachycaulus*. More comparative studies of *kamczadalorum, E. virescens,* and *E. trachycaulus* H10664 are desirable to distinguish between taxa and clarify relationships to other *E. trachycaulus* accessions. If following the classification of Flora of North America, the provided results suggests that E. *kamczadalorum* should be treated as *E. trachycaulus* ssp. *kamczadalorum* (Nevski) Tzvelev in accordance with Löve (1984).

There is great confusion about the taxonomic treatment of ***E. violaceus***, and the name has been used for different materials from Greenland, North America and Northern Europe (Sun et al. 2006b). In line with RAPD data from Sun *et al.* (2006a), the results from Paper III suggest that accessions of *E. violaceus* from Greenland and Canada are closely related to *E. trachycaulus*. Harrison and Hebda (2011) recommend treating the taxon as a subspecies to *E. alaskanus* and to be included in *E. alaskanus* ssp. *latiglumis*. However, this treatment would make *E. alaskanus* polyphyletic according to the phylogenetic results (see discussion about *E. alaskanus* below). Sun and Ma (2009) concluded, based on nucleotide variation in chloroplast Asp(GUC)–Thr(GGU) intergenic regions, that *E. violaceus* is genetically distinct from both *E. alaskanus* and *E. trachycaulus*. However, the recognized distinction of *E. violaceus* makes *E. trachycaulus* a paraphyletic

group. If *E. violaceus* should remain as a distinct species, the group *kamczadalorum-virescens-trachycaulus* should be revised.

5.1.2 American species – the transcontinental subclade

The transcontinental subclade consists of both North American and South American species with many unresolved branches. A polytomic pattern in the SplitsTree analyses in Paper II and III indicates incomplete lineage sorting and potential introgression due to recent and radial speciation events in the subclade. Several introductions from North to South America may have occurred and/or recolonization from South to North America, which is further discussed in Paper II. Diversification can be found among species in South America, also through polyploidization.

The close relationship between ***E. virginicus*** and ***E. canadensis*** in the phylogenetic analyses in Paper II and III is supported by frequent findings of hybrids occurring in areas were these species grow in sympatry and introgression is assumed (Pohl 1959; Nelson 1978). The two species ***E. villosus*** and ***E. glaucus***, from eastern respectively western North America, group together with ***E. angulatus*** and ***E. patagonicus*** from southern South America in a well-supported monophyletic group in Paper II.

The morphologically diverse ***Elymus glaucescens*** is often divided into several species (Parodi 1940; Nicora 1978; Moore 1983; Löve 1984; Seberg 1989). A multivariate analysis of the morphology shows a clinal variation pattern between two "extremes" (Seberg 1989), but this should preferably be confirmed by genetic data. Populations of *E. glaucescens* grow in a wide range of habitats, from dry steppe to wet meadows and saline pebble beaches. Observations from the field suggest rather distinct ecotypes explaining the desire to split the taxon. The more northern occurring ***E. parodii*** is closely related to *E. glaucescens*. The H-genome has been documented in *E. parodii*, but the **St**-genome has only been deduced based on morphology and not confirmed by other methods (Linde-Laursen & Seberg 2001; Tomas et al. 2012). The results in this thesis suggest a similar genome combination for *E. parodii* as for the other *Elymus s.s.*

Elymus scribneri was treated as *E. trachycaulus* ssp. *scribneri* by Löve (1984), and has been suggested to be a hybrid between *E. trachycaulus* or *E. alaskanus* and *E. elymoides* (Bowden 1965; Hitchcock 1969). This thesis shows that *E. scribneri* is nested within ***E. elymoides***, supporting the

involvement of *E. elymoides* as well as the close relationship to *E. trachycaulus*. This would suggest a hybridization between the two subclades of the American clade.

5.1.3 American species – *E. lanceolatus*

The phylogenetic analyses in Paper II place *E. lanceolatus* into the American clade, but the subclade affiliation is unresolved with different placement depending on method of choice. The species groups as a sister species to the other American subclades in Paper III, but with an unsatisfying branch support (BS = 87%). The origin of this strongly rhizomatous and highly outcrossing species and the relationship to the rest of the American *Elymus* species needs further investigation including more accessions.

5.1.4 Eurasian species – *sibiricus-confusus* subclade

The two species, ***E. sibiricus*** and ***E. confusus,*** with multiple and single spikelets per rachis node, respectively, are closely related and demonstrate the inappropriateness of using spikelet numbers as a diagnostic character for genus delimitation and evolutionary relationships.

5.1.5 Eurasian species – *mutabilis-fibrosus-caninus* subclade

Populations of ***E. mutabilis*** from Sweden, Far East Russia as well as Central Russia and China studied in Paper I are also included in the phylogenetic analysis in Paper III. The results from the two papers are in concordance and show that populations from Northern Europe and Far East Russia are more closely related than they are to populations in Central Russia and China. The two studies also separate the morphologically similar *E. mutabilis* and ***E. transbaicalensis*** as two distinct taxa. The phylogenetic tree in Paper III shows that the taxa belong to two different clades (Fig 5) with *E. mutabilis* more closely related to *E. fibrosus*, *E. transhyrcanus* and *E. caninus*, while *E. transbaicalensis* is more closely related to *E. alaskanus*, *E. macrourus* and *E. sajanensis*. This explains the low degree of introgression between the two species and highlight the problem of cryptic speciation in *Elymus*. One individual of *E. mutabilis* from accession H10456, appears at the base of the "American" clade in Paper III and is likely of hybrid origin. It originates

from Kamchatka and could possibly be a hybrid between *E. mutabilis* and E. *kamczadalorum*. If so, it suggests a hybridization, and potential introgression, between the American and Eurasian clades.

The hexaploid ***E. transhyrcanus*** has a **StStH** genome combination, but chromosome pairing has indicated a different variant of the combination compared to *E. repens* (Assadi & Runemark 1995). Crossing experiments by Dewey (1972) show that artificial amphiploid hybrids between *E. caninus* (**StH**) and *Ps. libanotica* (St) were biologically equivalent to the naturally occurring hexaploid *E. transhyrcanus*. The close relationship between *E. caninus* and *E. transhyrcanus* in Paper III supports this hypothesis.

In the phylogenetic analyses in Paper III, ***E. caninus*** var. *muticus* is nested among the other accessions of *E. caninus*, in accordance with allozyme studies by Diaz et al (1999b). The variety is morphologically similar to the autonym, but lacks the lemma awns and its habitat differs since it thrives in more open environments (Hård av Segerstad 1943). This could be an example of ecotypic differentiation often found within species of *Elymus*.

5.1.6 Eurasian species – *alaskanus-macrourus-transbaicalensis* subclade

The circumpolar ***E. alaskanus*** is morphologically homogenous within regions and populations, but variable over its whole distribution area (Wu et al. 2016). In this thesis, samples of *E. alaskanus* from Iceland, Norway, Russia, Canada, and USA (Alaska) turn out as a monophyletic group with high branching support (BS = 100%; PP = 1) within the Eurasian clade. The taxon *E. alaskanus* subsp. *latiglumis* (Scribn. & J.G.Sm.) Á. Löve was included in both Paper II and III, but there treated as *E. violaceus* according to Flora of North America (Barkworth et al. 1993) and groups with *E. violaceus* from Greenland in the American clade. These results are in accordance with RAPD and SSR marker studies by Zhang et al. (2002) and Sun and Salomon (2003), respectively, where *E. alaskanus* ssp. *latiglumis* is more closely related to *E. violaceus*, and differs from other *E. alaskanus* accessions, which are more closely related to *E. komarovii* and *E. sajanensis*. Results from a RAPD marker study by Sun and Li (2005) show an *E. alaskanus* accession from USA related to *E. trachycaulus* complex while a second accession appear more closely related to the Eurasian *E. mutabilis*,

but the lower ranks were not indicated. This confusion demonstrates the importance of careful taxonomical scrutiny and invokes further genetic studies with a wide selection of wild collected and morphologically well-described populations.

The circumpolar ***Elymus macrourus*** is morphologically differentiated from *E. alaskanus* in the shape of their glumes and their narrower glume margins (Barkworth et al. 1993). Agafonov and Salomon (2002) concluded that *E. macrourus* and *E. jacutensis* form a single recombination gene pool (RGP) and should be considered a complex, which is further supported by the close relationship observed in Paper III. Agafonov (2008) suggests, based on morphological, electrophoretic and crossing data, that *E. jacutensis* should be treated as the variety *E. macrourus* var. *jacutensis*.

Paper III shows a close relationship between ***E. komarovii***, ***E. transbaicalensis***, and ***E. sajanensis***, which is in agreement with biosystematic, morphological and isozyme data (Agafonov et al. 1998; Agafonov & Salomon 2002; Agafonov 2004). The microevolutionary level of differentiation between *E. komarovii, E. transbaicalensis, E. kronokensis,* and *E. sajanensis* has been studied by Agafonov *et al.* (2019). The species are morphologically variable with many different biotypes showing complex genetic relationships. The phylogenetic analyses in Paper III indicate that *E. komarovii* is differentiated from *E. transbaicalensis* and *E. sajanensis,* but a weak branch support makes the relationship between *E. transbaicalensis* and *E. sajanensis* unresolved. The three species could possibly be included in a larger circumscription or considered parts of a species complex.

The species ***E. dentatus***, endemic to Pakistan, appears as a sister species to the rest of the *alaskanus-macrourus-transbaicalensis* subclade. This could indicate a common origin of the subclade in Central Asia.

5.1.7 Elymus repens

According to the SplitsTree analysis in Paper III, the hexaploid ***E. repens*** groups as an intermediate between the **StY** and **StH** *Elymus* species indicating a close relationship with these two groups. *Elymus repens* has earlier been assigned a **StStH** genome combination (Assadi & Runemark 1995; Ørgaard & Anamthawat-Jónsson 2001; Mahelka & Kopecký 2010). However, the genetic origin is complex and additional contribution through introgression from several different genera, also outside the tribe, has been

suggested (Mason-Gamer 2004; Dixelius et al. 2008). In a phylogenetic study by Mason-Gamer (2008), two low-copy nuclear genes encoding phosphoenolpyruvate carboxylase (pepC) and β-amylase suggest that *E. repens* originates from the combination of *Hordeum* (**H**), *Pseudoroegneria* (**St**) and an unknown donor. Based on present results, together with the results from Mason-Gamer (2008), it can be hypothesized that *E. repens* shares common ancestors from both **StH** and **StY** *Elymus*, and could possibly be a hybrid between a **StY** species and a *Hordeum* species also involved in the **StH** formation. Further investigations are needed to fully understand the origin of *E. repens*.

5.2 Phylogeography of *Elymus*

The three Papers (I-III) discuss the phylogeography of *Elymus s.s.* on three different taxonomical levels and demonstrate the dynamic nature of *Elymus* evolution and migration. Some species are restricted to a limited area, some are widely distributed and some have managed to spread also between continents and become circumpolar.

Paper I shows that the phylogeographic pattern of *E. mutabilis* follows the geographical distribution of the species. Populations from Northern Europe, Southern Siberia and Far East Russia form a separate clade distinct from central Asia, indicating a common ancestry of the peripheral populations. The phylogenetic analyses show a radiation pattern among populations in northern Europe indicating a founding followed by rapid dispersal. The phylogeographic pattern is confirmed by the phylogenetic analyses in Paper III where samples from northern Europe and eastern Russia group together.

Paper II shows a dynamic phylogeography of *Elymus s.s* in America. South American species are assumed to originate from North America due to the lack of a *Pseudoroegneria* progenitor species on the continent. The polyphyletic structure of the South American species suggests multiple transcontinental dispersal between South and North America, further discussed in Paper II. No species occurs naturally on both continents.

Paper III show that the main phylogenetic division is correlated with geographical origin with one group of predominantly Eurasia species and the

other with American species. Two species, *E. macrourus* and *E. alaskanus* are truly circumpolar and derives from Eurasia, but the exact migration route cannot be established. The species could have migrated over the Bering Strait or the Greenland Sea in the northern Atlantic Ocean, or both. The main distribution area of *E. trachycaulu*s is in North America but it does occur in Eurasia. The close relationship between *E. kamczadalorum* from eastern Russia and the *E. trachycaulus* complex from North America also suggests a transcontinental migration event over the Bering Strait. In summary, the results suggest that at least two migration events have occurred from Eurasia to North America (*E. alaskanus* and *E. macrourus* from the Eurasian clade), and at least two in the opposite direction (*E. trachycaulus* and *E. kamczadalorum* in the American clade).

5.3 Taxonomy of *Elymus s.s.*

Phylogenies are used to understand evolutionary relationships, but the information also has taxonomical implications. Based on the findings of this thesis, it would be appropriate to form a classification of subgenera and/or sections on the phylogenetic relationships and monophyletic groups. The division of *Elymus s.l.* into several genera based on genome constitution is supported by the results from Paper III with *Elymus* s.s. restricted to species with the **StH** genome. However, as discussed in 5.1, the pronounced distinction between the American and the Eurasian *Elymus s.s.* (**StH**) species suggests a further taxonomical division, if not into different genera, then possibly into different subgenera. However, the difficulties in easily distinguishing between the two groups morphologically and the potential polyphyletic origin complicates this division. The identified monophyletic subclades could further be divided into sections even though the problem of distinction and origin remains. Morphological data need to be assessed together with genetic and preferably biosystematic data, including crossbreeding experiments. The classification of sections is not accompanied by morphological descriptions in Löve (1984), but is well described in Yen and Yang (2022). Neither of the classifications are, however, in accordance with the phylogenetic results. For example, *E. caninus* and *E. trachycaulus* are placed in section Goulardia by Löve (1984) and Yen and Yang (2022),

but the two taxa falls out in different clades in the phylogenetic analyses. A proposition of taxonomical classification is given in Table 1 and 2. Phylogenetic data should be used in order to revise morphologically significant characters.

A subdivision of the **StH** genome into $\mathbf{St_1H_1}$ and $\mathbf{St_2H_2}$ does complicate the delimitation of genera based on strict genome combinations, like *Elymus s.s.*, which is emphasized by Yen and Yang (2022). It could be the case that the Eurasian St progenitor is more closely related to the St progenitor of the **StY** species (*Roegneria*), described in Figure 6. If so, the recognition of *Elymus s.l.* including both **StH** and **StY** species is more logical from a monophyletic perspective than separating the genome combinations into two genera. The occurrence of hexaploid species with the inclusion of additional genome sets could further complicate the taxonomy, which favors a wider circumscription of the genus.

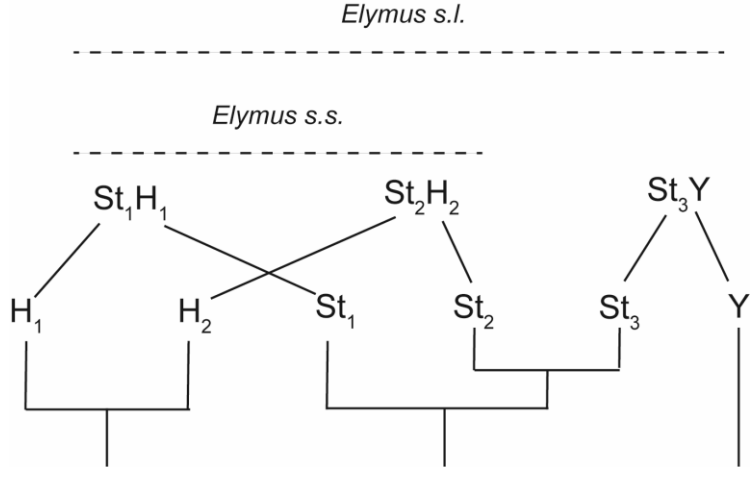

Figure 6. Hypothetical relationships between American $\mathbf{St_1H_1}$, Eurasian $\mathbf{St_2H_2}$, and $\mathbf{St_3Y}$ species of *Elymus s.l.*. *Elymus s.s.* includes species with a **StH** combination. *Hordeum* genomes are indicated with **H** and *Pseudoroegneria* genomes with **St**. **Y** = unknown progenitor.

Table 1. Elymus s.s. species included in this thesis sorted on species with assigned sections according to Löve (1984) and Yen and Yang (2022). Sections are: Macrolepis (M), Elymus (El), Elytrigia (Et), Goulardia (G), Sitanion (S), Dasystachyae (D), and Hystrix (H). Sections proposed by Leo are based on the clades (A = American clade, B = Eurasian clade) and subclades (roman numbers) identified in Paper III.

Species	**Author(s)**	**Sections**			**Ploidy**
		Löve	Yen & Yang	Leo	
E. alaskanus	(Scribn. & Merr.) Á. Löve	G	Et	BIII	4x
E. alopex	B. Salomon			BII	4x
E. angulatus	J. Presl	D	El	AIII	4x
E. arizonicus	(Scribn. & J. G. Sm.) Gould	D	G	AIII	4x
E. bakeri	(E. E. Nelson) Á. Löve	G	G	AII	4x
E. canadensis	L.	M	El	AIII	4x
E. caninus	(L.) L.	G	G	BII	4x
E. charkeviczii	Prob.		G	BII	4x
E. confusus	(Roshev.) Tzvelev	El	G	BI	4x
E. cordilleranus	Davidse & R. W. Pohl		Et	AIII	4x
E. dentatus	(Hook.f.) Tzvelev	G	G	BIII	4x
E. elymoides	(Raf.) Swezey	S	S	AIII	4x
E. fibrosus	(Schrenk) Tzvelev	G	G	BII	4x
E. glaucescens	Seberg	D	G	AIII	4x
E. glaucus	Buckley	El	El	AIII	4x
E. jacutensis	(Drobow) Tzvelev	G	G	BIII	4x
E. komarovii	(Nevski) Tzvelev	H	G	BIII	4x
E. lanceolatus	(Scribn. & J. G. Sm.) Gould	D	Et	AI	4x
E. macrourus	(Turcz. ex Steud.) Tzvelev	G	Et	BIII	4x
E. mutabilis	(Drobow) Tzvelev	G	G	BII	4x
E. parodii	Seberg & G. Petersen	D	El	AIII	4x
E. patagonicus	Speg.	D	El	AIII	6x
E. sajanensis	(Nevski) Tzvelev	G	G	BIII	4x
E. scabriglumis	(Hack.) Á. Löve	D	G	AII	6x
E. scribneri	(Vasey) M. E. Jones	G	G	AIII	4x
E. sibiricus	L.	El	El	BI	4x
E. sierra	Gould	G	G	AIII	4x
E. trachycaulus	(Link) Gould	G	G	AII	4x
E. transbaicalense	(Nevski) Tzvelev	G	G	BIII	4x
E. transhyrcanus	(Nevski) Tzvelev	G	G	BII	6x
E. vaillantianus	(Wulfen ex Schreb.) K. B. Jensen		Et	AII	4x
E. villosus	Muhl. ex Willd.		El	AIII	4x
E. violaceus	(Hornem.) J. Feilberg	G	G	AII	4x
E. virescens	Piper	G	El	AII	4x
E. virginicus	L.	M	El	AIII	4x
E. kamczadalorum	(Nevski) Tzvelev	G	G	AII	4x

Table 2. Elymus s.s. species included in this thesis sorted on species with assigned sections according to Löve (1984) and Yen and Yang (2022). Sections are: Macrolepis (M), Elymus (El), Elytrigia (Et), Goulardia (G), Sitanion (S). Dasystachyae (D) and Hystrix (H). Sections proposed by Leo are based on the clades (A = American clade, B = Eurasian clade) and subclades (roman numbers) identified in Paper III.

Species	**Author(s)**	**Sections**			**Ploidy**
		Löve	Yen & Yang	Leo	
E. lanceolatus	(Scribn. & J. G. Sm.) Gould	D	Et	AI	4x
E. bakeri	(E. E. Nelson) Á. Löve	G	G	AII	4x
E. scabriglumis	(Hack.) Á.Löve	D	G	AII	6x
E. trachycaulus	(Link) Gould	G	G	AII	4x
E. vaillantianus	(Wulfen ex Schreb.) K. B. Jensen		Et	AII	4x
E. violaceus	(Hornem.) J.Feilberg	G	G	AII	4x
E. virescens	Piper	G	El	AII	4x
E. kamczadalorum	(Nevski) Tzvelev	G	G	AII	4x
E. angulatus	J. Presl	D	El	AIII	4x
E. arizonicus	(Scribn. & J. G. Sm.) Gould	D	G	AIII	4x
E. canadensis	L.	M	El	AIII	4x
E. cordilleranus	Davidse & R. W. Pohl		Et	AIII	4x
E. elymoides	(Raf.) Swezey	S	S	AIII	4x
E. glaucescens	Seberg	D	G	AIII	4x
E. glaucus	Buckley	El	El	AIII	4x
E. parodii	Seberg & G. Petersen	D	El	AIII	4x
E. patagonicus	Speg.	D	El	AIII	6x
E. scribneri	(Vasey) M. E. Jones	G	G	AIII	4x
E. sierra	Gould	G	G	AIII	4x
E. villosus	Muhl. ex Willd.		El	AIII	4x
E. virginicus	L.	M	El	AIII	4x
E. confusus	(Roshev.) Tzvelev	El	G	BI	4x
E. sibiricus	L.	El	El	BI	4x
E. alopex	B. Salomon			BII	4x
E. caninus	(L.) L.	G	G	BII	4x
E. charkeviczii	Prob.		G	BII	4x
E. fibrosus	(Schrenk) Tzvelev	G	G	BII	4x
E. mutabilis	(Drobow) Tzvelev	G	G	BII	4x
E. transhyrcanus	(Nevski) Tzvelev	G	G	BII	6x
E. alaskanus	(Scribn. & Merr.) Á. Löve	G	Et	BIII	4x
E. dentatus	(Hook.f.) Tzvelev	G	G	BIII	4x
E. jacutensis	(Drobow) Tzvelev	G	G	BIII	4x
E. komarovii	(Nevski) Tzvelev	H	G	BIII	4x
E. macrourus	(Turcz. ex Steud.) Tzvelev	G	Et	BIII	4x
E. sajanensis	(Nevski) Tzvelev	G	G	BIII	4x
E. transbaicalense	(Nevski) Tzvelev	G	G	BIII	4x

6 Conclusions

6.1 General conclusions

This thesis investigates genetic relationships on three taxonomic levels by presenting inferred phylogenetic relationships of the **StH** genome, the phylogeography of the species as well as the genetic diversity within widely distributed species. The findings give further understanding of the evolutionary complexity of *Elymus* by giving evidence for multiple origins of the **StH** genome with several involved progenitor species, and the intercontinental migration between *Elymus* in North and South America, and America and Eurasia. This thesis also contributes to an increased understanding of within species variation of widely distributed species as well as identifying groups with need of taxonomical revision. The thesis also stresses the importance of studying closely related and morphologically similar species, as well as species with a high level of morphological intraspecific variation, in order to unravel cryptic speciation and clinal variation. The phylogenies should be considered a framework for further investigations at a lower taxonomical level, within subclades, species complexes and species. Next generation sequencing, including both GBS and DArTseq, is a well-suited approach for the purpose of both phylogenetic and population studies and is recommended for future research of *Elymus* systematics. Highly informative data such as SNPs are necessary to go further into the evolutionary history and phylogeography of the genus *Elymus*.

6.2 Specific conclusions

Paper I shows a clear phylogeographic structure with a pattern of variation corresponding to the geographical distribution of *E. mutabilis* and where the main genetic diversity in central Asia is evident. The genetic variation of populations in northern Europe are distinct from Asian populations, but shows a high genetic similarity to populations from Far East Russia and Southern Siberia. The *E. mutabilis* populations in the peripheral regions most likely originated in Altai and southern Siberia from where they spread to other areas. A phylogenetic radiation pattern among populations in northern Europe indicates a founding followed by rapid dispersal. Paper I and III provides molecular evidence for considering *E. mutabilis* and *E. transbaicalensis* as two distinct species.

Paper II shows that the six included species from South America do not form a monophyletic group, indicating multiple dispersal events from North to South America and that regional affiliation does not fully correlate with evolutionary history. The analyses place *E. scabriglumis* in a clade separate from the other five South American species. Unresolved internal branches in the later clade make the origin and migration of the five South American *Elymus* species uncertain. The two North American species *E. villosus* and *E. glaucus* group together with *E. angulatus* and *E. patagonicus* from southern South America in a well-supported monophyletic subclade. Thus, also the five South American species *E. angulatus, E. patagonicus, E. cordilleranus, E. glaucescens,* and *E. parodii* constitute a non-monophyletic group indicating multiple introductions of the genus to South America.

Paper III presents a phylogenetic reconstruction of *Elymus s.s.* from the whole genus distribution range. The separation of **StH** and **StY** genome combinations is demonstrated as well as a differentiation between two major clades within the **StH** group. The two clades highly correlates with geographical distribution and can be identified as the "American" clade, and the "Eurasian" clade. The placement of different *Pseudoroegneria* and *Hordeum* species in the different clades suggests the involvement of several progenitor species, making *Elymus s.s.* (**StH**) a polyphyletic group. The "Eurasian" clade can be divided into three subclades and the "American" clade can be divided into two subclades. The transcontinental and circumpolar species shows that the Bering Sea is a possible route of dispersal but it is still unknown if species have migrated over the Greenland Sea in

northern Atlantic Ocean. A new classification of *Elymus* sections, where phylogenetic relationships are taken into account, is needed.

7 Future research

Future research of *Elymus* systematics, conservation and utilization should focus on: 1. further investigations of species diversity, species relationships and related taxonomical work, 2. utilization of *Elymus* for agricultural purposes, and 3. assessment of further need of conservation, both *in situ* and *ex situ*. The studies in this thesis provide phylogenetic data for future taxonomical work. The discovery of multiple **StH** genome origins must be accompanied by morphological data to create a more representative classification of the genus. For example, it is desirable to define sections based on both phylogeny and morphology. More population diversity studies are needed to gain knowledge on species delimitations and suitable taxa for inclusion in phylogenetic analyses. This is also essential for efficient and well-informed conservation strategies and effective breeding programs. Studies of population structures and gene flow between populations help us understand the evolution of diversification of polyploid species. Continued work to explore the large variation within species over wide distribution areas should also be prioritized, including addressing clinal variation with a gradual change in morphology so often found in *Elymus*.

Future studies need a broader inclusion of parental species in phylogenetic analyses. The determination of *Elymus* origin would contribute to our understanding of polyploid speciation and give further information on *Elymus* classification. With continued development of phylogenomics, more genetic information will be available with the possibility to establish more robust evolutionary relationships, as well as to investigate gene family evolution, lateral gene transfer and prediction of gene function in *Elymus*. Whole genome sequencing of an *Elymus* species has not yet been done. The

availability of the whole genome sequence enables studies of genome structure such as gene order, intron positions, insertions and deletions (indels), retrotransposon integrations, and gene fusion and fission (Delsuc et al. 2005). All these techniques could potentially give further knowledge of *Elymus* evolution.

The link between an accession and its phenotypic characteristics as well as trials for evaluation of agricultural properties is necessary to utilize *Elymus* as a genetic resource for future plant breeding and/or selection of suitable genotypes. The collections from this thesis could be used in a first utility screening for future supplementary collections. Further collections of genetic resources could benefit of a clear goal based on a need from a crop improvement perspective and could focus on collection areas with ecologically interesting populations. A taxonomical utility and risk assessment of clades and subclades should be conducted to identify favorable accessions for future plant breeding and possible threats to taxonomical diversity, for example, how different groups will be affected by climate change.

Southern South America is a hotspot for Triticeae and *Elymus* in particular. An extensive collection of *Elymus* plant material from Chile been carried out in connection to the acquisition of material for this thesis. Future studies will focus on *Elymus* diversity in Chile with a focus on the Magallanes region, including investigations of population structures, gene flow and species delimitations. These studies will contribute to future conservation strategies and utilization efforts.

References

Agafonov, A. V (2004). Intraspecific structure and reproductive relationships between *Elymus mutabilis* and *E. transbaicalensis* (Poaceae) in southern Siberia from the viewpoint of taxonomical genetics. *Russian Journal of Genetics*, 40 (11), 1229–1238. https://doi.org/10.1023/B:RUGE.0000048665.40740.2c

Agafonov, A. V, (Gerus), D.E.N. & (Kobozeva), E.V.S. (2019). The variability and specificity of the histone H1 among the Siberian species *Elymus transbaicalensis, E. komarovii, E. sajanensis, E. kronokensis* (Poaceae) and some morphologically deviating forms. *Turczaninowia*, 22 (1), 5–18. https://doi.org/10.14258/turczaninowia.22.1.1

Agafonov, A. V & Salomon, B. (2002). Genepools among SH genome *Elymus* species in boreal Eurasia. In: Hernándes, P. et. al. (ed.) *Triticeae IV. Consejeria de Agricultura y Pesca*. Sevillia, Spain. 37–41.

Agafonov, A. V, Salomon, B., Kostina, E. V & Diaz, O. (1998). *Biosystematic relationships between Elymus komarovii (Nevski) Tzvel. and related species*. Jaradat, A.A. (ed.) (Jaradat, A. A., ed.) *Triticeae III*. Novosibirsk: Science Publishers Inc.

Agafonov, A. V (2008). Биосистематическое исследование комплекса *Elymus macrourus-E. jacutensis* и критического таксона *E. ircutensis* (Triticeae: Poaceae). *Растительный мир Азиатской России*, (2), 20–32

Al-Beyroutiová, M., Sabo, M., Sleziak, P., Dušinský, R., Birčák, E., Hauptvogel, P., Kilian, A. & Švec, M. (2016). Evolutionary relationships in the genus *Secale* revealed by DArTseq DNA polymorphism. *Plant Systematics and Evolution*, 302 (8), 1083–1091. http://www.jstor.org/stable/44853272

Al-Saghir, M. (2016). Taxonomy and Phylogeny in Triticeae: A Historical Review and Current Status. *Advances in Plants & Agriculture Research*, 3 (5), 139–143. https://doi.org/10.15406/apar.2016.03.00108

Assadi, M. & Runemark, H. (1995). Hybridisation, genomic constitution and generic delimitation in *Elymus s. l.* (Poaceae: Triticeae). *Plant Systematics and Evolution*, 194 (3), 189–205. https://doi.org/10.1007/BF00982855

Aubry, C., Shoal, R. & Erickson, V. (2005). *Grass cultivars: their origins, development, and use on national forests and grasslands in the Pacific Northwest*. https://www.fs.fed.us/wildflowers/Native_Plant_Materials/documents/cultivars_maindoc_040405_appendices.pdf [2022-05-31]

Baral, K., Coulman, B., Biligetu, B. & Fu, Y.-B. (2018). Genotyping-by-Sequencing Enhances Genetic Diversity Analysis of Crested Wheatgrass [*Agropyron cristatum* (L.) Gaertn.]. *International journal of molecular sciences*, 19 (9), 2587. https://doi.org/10.3390/ijms19092587

Barkworth, M.E. (1992). Taxonomy of the Triticeae - A Historical Perspective. *Hereditas*, 116 (1–2), 1–14. https://doi.org/10.1111/j.1601-5223.1992.tb00197.x

Barkworth, M.E. (1994). The *Elymus trachycaulus* Complex in North America: More Questions than Answers. In: Wang, R.R.C., Jensen, L.B., & Jaussi, C. (eds) *Preceedings of the 2nd International Triticeae Symposium*. Utha: Logan Press. 189–198.

Barkworth, M.E. & von Bothmer, R. (2005). Twenty-one years later: the impact of Löve and Deweys genomic classification proposal. *Czech Journal of Genetics and Plant Breeding*, 41, 3–9

Barkworth, M.E. & von Bothmer, R. (2009). *Scientific names in the Triticeae*. Feuillet, C. & Muehlbauer, G.J. (eds) (Feuillet, C. & Muehlbauer, G. J., eds) *Genetics and Genomics of the Triticeae* New York, NY: Springer. https://doi.org/10.1007/978-0-387-77489-3_1

Barkworth, M.E., Campbell, J.J.N. & Salomon, B. (1993). *Elymus* L. In: Flora of North America Editorial Committee (ed.) *Flora of North America North and Mexico, Volume 24*. New York and Oxford. 288–289.

Baum, B.R. (1983). A phylogenetic analysis of the tribe Triticeae (Poaceae) based on morphological characters of the genera. *Canadian Journal of Botany-Revue Canadienne De Botanique*, 61 (2), 518–535. https://doi.org/10.1139/b83-059

Baum, B.R., Edwards, T. & Johnson, D.A. (2015). Diversity within the genus *Elymus* (Poaceae: Triticeae) as investigated by the analysis of the nr5S rDNA variation in species with St and H haplomes. *Molecular Genetics and Genomics*, 290 (1), 329–342. https://doi.org/10.1007/s00438-014-0907-4

Baum, B.R., Edwards, T. & Johnson, D.A. (2016). Diversity within the genus *Elymus* (Poaceae: Triticeae) II: analyses of variation within 5S nrDNA restrict membership in the genus to species with StH genomes. *Mol Genet Genomics*, 291 (1), 217–225. https://doi.org/10.1007/s00438-015-1096-5

Baum, B.R., Estes, J.R. & Gupta, P.K. (1987). Assessmant of the genomic system of classification in the Triticeae. *American Journal of Botany*, 74 (9), 1388–1395. https://doi.org/10.2307/2444316

Baum, B.R., Yang, J.-L., Yen, C. & Agafonov, A. V. (2011). A taxonomic synopsis of the genus *Campeiostachys* Drobov. *Journal of Systematics and Evolution*, 49 (2), 146–159. https://doi.org/https://doi.org/10.1111/j.1759-6831.2010.00106.x

Beer, C., Reichstein, M., Tomelleri, E., Ciais, P., Jung, M., Carvalhais, N., Rödenbeck, C., Arain, M.A., Baldocchi, D., Bonan, G.B., Bondeau, A., Cescatti, A., Lasslop, G., Lindroth, A., Lomas, M., Luyssaert, S., Margolis, H., Oleson, K.W., Roupsard, O., Veenendaal, E., Viovy, N., Williams, C., Woodward, I.F. & Papale, D. (2010). Terrestrial Gross Carbon Dioxide Uptake: Global Distribution and Covariation with Climate. *Science*, 329 (5993), 834–838. https://doi.org/10.1126/science.1184984

Bentham, G. (1881). Notes on Gramineæ. *Botanical Journal of the Linnean Society*, 19 (115–1160), 14–134

Bernhardt, N., Brassac, J., Kilian, B. & Blattner, F.R. (2017). Dated tribe-wide whole chloroplast genome phylogeny indicates recurrent hybridizations within Triticeae. *BMC Evolutionary Biology*, 17, 141. https://doi.org/10.1186/s12862-017-0989-9

von Bothmer, R. & Salomon, B. (1994). Triticeae: a tribe for food, feed and fun. In: Wang, R.R.C., Jensen, K.B., & Jaussi, C. (eds) *Proceedings of the 2nd International Triticeae Symposium*. Logan: Utha state University Press. 1–12.

von Bothmer, R., Seber, O., Jacobsen, N. & von Bothmer O.; Jacobsen, N., R.. S. (1992). Genetic resources in the Triticeae. *Hereditas*, 116, 141–150

von Bothmer, R. & Seberg, O. (1995). Strategies for collecting of wild species. In: Guarino, L., Rao, V.R., & Reid, R. (eds) *Collecting plant genetic diversity*. Wallingford: Cab International. 93–111.

Bothmer, R. von, Seberg, O. & Jacobsen, N. (1992). Genetic resources in the Triticeae. *Hereditas*, 116 (s1), 141–150. https://doi.org/https://doi.org/10.1111/j.1601-5223.1992.tb00814.x

Bowden, W.M. (1965). Cytotaxonomy of the species and interspecific hybrids of the genus *Agropyron* in Canada and neighboring areas. *Canadian Journal of Botany*, 43 (11), 1421–1448. https://doi.org/10.1139/b65-151

Brassac, J. & Blattner, F.R. (2015). Species-level phylogeny and polyploid relationships in *Hordeum* (Poaceae) inferred by Next-Generation Sequencing and In Silico Cloning of multiple nuclear loci. *Systematic Biology*, 64 (5), 792–808. https://doi.org/10.1093/sysbio/syv035

Campbell, J.J.N. & Soreng, R.J. (2003). *Elymus*. In Catalogue of New World Grasses (Poaceae): IV. Subfamily Pooideae. *Contributions from the United States National Herbarium*, 48, 279–307

Choi, B., Crisp, M.D., Cook, L.G., Meusemann, K., Edwards, R.D., Toon, A. & Külheim, C. (2019). Identifying genetic markers for a range of phylogenetic utility - From species to family level. *PloS one*, 14 (8), e0218995–e0218995. https://doi.org/10.1371/journal.pone.0218995

Chou, J., Gupta, A., Yaduvanshi, S., Davidson, R., Nute, M., Mirarab, S. & Warnow, T. (2015). A comparative study of SVDquartets and other

coalescent-based species tree estimation methods. *BMC Genomics*, 16 (10), S2. https://doi.org/10.1186/1471-2164-16-S10-S2

Cox, T.S., Bender, M., Picone, C., Tassel, D.L. Van, Holland, J.B., Brummer, E.C., Zoeller, B.E., Paterson, A.H. & Jackson, W. (2002). Breeding Perennial Grain Crops. *Critical Reviews in Plant Sciences*, 21 (2), 59–91. https://doi.org/10.1080/0735-260291044188

Delsuc, F., Brinkmann, H. & Philippe, H. (2005). Phylogenomics and the reconstruction of the tree of life. *Nature Reviews Genetics*, 6 (5), 361–375. https://doi.org/10.1038/nrg1603

Devos, K.M. & Gale, M.D. (2000). Genome relationships: The grass model in current research. *Plant Cell*, 12 (5), 637–646. https://doi.org/10.1105/tpc.12.5.637

Dewey, D.R. (1972). The Origin of *Agropyron leptourum*. *American Journal of Botany*, 59 (8), 836–842. https://doi.org/10.2307/2441088

Dewey, D.R. (1975). Introgression between *Agropyron dasystachyum* and *A. trachycaulum*. *Botanical Gazette*, 136 (1), 122–128. http://www.jstor.org/stable/2473802

Dewey, D.R. (1984). The genomic system of classification as a guide to intergeneric hybridization with the perennial Triticeae. *Gene manipulation in plant improvement*, 209–279

Diaz, O. (1999). *Genetic diversity in* Elymus *species (Triticeae) with emphasis on the Nordic region (Doctoral thesis)*. *Department of Plant Breeding Research Svalöv*. Swedish Univerity of Agricultural Sciences.

Diaz, O., Salomon, B. & von Bothmer, R. (1998). *Description of isozyme polymorphisms in Elymus species using starch gel electrophoresis*. Jaradat, A.A. (ed.) (Jaradat, A. A., ed.) *Triticeae III*. Swedish Univ Agr Sci, Dept Plant Breeding Res, S-26831 Svalov, Sweden. Diaz, O (reprint author), Swedish Univ Agr Sci, Dept Plant Breeding Res, S-26831 Svalov, Sweden.: Science Publishers Inc.

Diaz, O., Salomon, B. & von Bothmer, R. (1999a). Genetic variation and differentiation in Nordic populations of Elymus alaskanus (Scrib. ex Merr.) Löve (Poaceae). *Theoretical and Applied Genetics*, 99 (1–2), 210–217. https://doi.org/10.1007/s001220051227

Diaz, O., Salomon, B. & Von Bothmer, R. (1999b). Genetic diversity and structure in populations of *Elymus caninus* (L.) L. (Poaceae). *Hereditas*, 131 (1), 63–74. https://doi.org/10.1111/j.1601-5223.1999.00063.x

Diaz, O., Sun, G.L., Salomon, B. & von Bothmer, R. (2000). Levels and distribution of allozyme and RAPD variation in populations of *Elymus fibrosus* (Schrenk) Tzvel. (Poaceae). *Genetic Resources and Crop Evolution*, 47 (1), 11–24. https://doi.org/10.1023/a:1008719119756

Dixelius, C., Fahleson, J., Okori, P. & Åkerblom-Espeby, L. (2008). Genetic variability and genomic divergence of *Elymus repens* and related species.

Plant Systematics and Evolution, 271 (3/4), 143–156. http://www.jstor.org/stable/23657315

Dizkirici, A., Kaya, Z., Cabi, E. & Dogan, M. (2010). Phylogenetic relationships of *Elymus* L. and related genera (Poaceae) based on the nuclear ribosomal internal transcribed spacer sequences. *Turkish Journal of Botany*, 34 (6), 467–478. https://doi.org/10.3906/bot.0912.249

Dong, Z.Z., Fan, X., Sha, L.N., Wang, Y., Zeng, J., Kang, H.Y., Zhang, H.Q., Wang, X.L., Zhang, L., Ding, C.B., Yang, R.W. & Zhou, Y.H. (2015). Phylogeny and differentiation of the St genome in *Elymus* L. sensu lato (Triticeae; Poaceae) based on one nuclear DNA and two chloroplast genes. *BMC Plant Biol*, 15, 179. https://doi.org/10.1186/s12870-015-0517-2

Dubcovsky, J., Schlatter, A.R. & Echaide, M. (1997). Genome analysis of South American *Elymus* (Triticeae) and *Leymus* (Triticeae) species based on variation in repeated nucleotide sequences. *Genome*, 40 (4), 505–520. https://doi.org/10.1139/g97-067

Dvorak, J. & Zhang, H.B. (1992). Application of melecular tools for study of phylogeny of diploid and polyploid taxa in Triticeae. *Hereditas*, 116 (1–2), 37–42. https://doi.org/10.1111/j.1601-5223.1992.tb00202.x

Eaton, D.A.R. (2014). PyRAD: assembly of de novo RADseq loci for phylogenetic analyses. *Bioinformatics*, 30 (13), 1844–1849. https://doi.org/10.1093/bioinformatics/btu121

Eaton, D.A.R. & Overcast, I. (2020). ipyrad: Interactive assembly and analysis of RADseq datasets. *Bioinformatics*, 36 (8), 2592–2594

Edet, O.U., Gorafi, Y.S.A., Nasuda, S. & Tsujimoto, H. (2018). DArTseq-based analysis of genomic relationships among species of tribe Triticeae. *Scientific Reports*, 8, 11. https://doi.org/10.1038/s41598-018-34811-y

Edwards, D., Forster, J.W., Chagné, D. & Batley, J. (2007). What Are SNPs? BT - Association Mapping in Plants. In: Oraguzie, N.C., Rikkerink, E.H.A., Gardiner, S.E., & De Silva, H.N. (eds). New York, NY: Springer New York. 41–52. https://doi.org/10.1007/978-0-387-36011-9_3

eFloras (2008). *Flora of China. Missouri Botanical Garden, St. Louis, MO & Harvard University Herbaria, Cambridge, MA.* http://www.efloras.org/florataxon.aspx?flora_id=2&taxon_id=242320323 [2021-04-14]

Elshire, R.J., Glaubitz, J.C., Sun, Q., Poland, J.A., Kawamoto, K., Buckler, E.S. & Mitchell, S.E. (2011). A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. *Plos One*, 6 (5), 10. https://doi.org/10.1371/journal.pone.0019379

Fan, X., Sha, L.N., Yu, S.B., Wu, D.D., Chen, X.H., Zhuo, X.F., Zhang, H.Q., Kang, H.Y., Wang, Y., Zheng, Y.L. & Zhou, Y.H. (2013). Phylogenetic reconstruction and diversification of the Triticeae (Poaceae) based on

single-copy nuclear Acc1 and Pgk1 gene data. *Biochemical Systematics and Ecology*, 50, 346–360. https://doi.org/10.1016/j.bse.2013.05.010
Favre, F., Jourda, C., Besse, P. & Charron, C. (2021). Genotyping-by-Sequencing Technology in Plant Taxonomy and Phylogeny. *Methods in molecular biology (Clifton, N.J.)*, 2222, 167–178. https://doi.org/10.1007/978-1-0716-0997-2_10
Frankel, O.H., Brown, A.H.D. & Burdon, J.J. (1995). *The conservation of plant biodiversity*. Cambridge: Camebridge University Press.
Frawley, E.S., Ciotir, C., Micke, B., Rubin, M.J. & Miller, A.J. (2020). An Ethnobotanical Study of the Genus *Elymus*. *Economic Botany*, 74 (2), 159–177. https://doi.org/10.1007/s12231-020-09494-0
Frederiksen, S. & Seberg, O. (1992). Phylogenetic analysis of the Triticeae (Poaceae). *Hereditas*, 116 (s1), 15–19. https://doi.org/https://doi.org/10.1111/j.1601-5223.1992.tb00793.x
Gao, G., Deng, J., Gou, X., Wang, Q., Ding, C., Zhang, L., Zhou, Y. & Yang, R. (2015). Phylogenetic relationships among *Elymus* and related diploid genera (Triticeae: Poaceae) based on nuclear rDNA ITS sequences. *Biologia*, 70 (2), 183–189. https://doi.org/10.1515/biolog-2015-0019
Gao, G., Gou, X., Wang, Q., Zhang, Y., Deng, J., Ding, C., Zhang, L., Zhou, Y. & Yang, R. (2014). Phylogenetic relationships and Y genome origin in Chinese *Elymus* (Triticeae: Poaceae) based on single copy gene DMC1. *Biochemical Systematics and Ecology*, 57, 420–426. https://doi.org/10.1016/j.bse.2014.09.019
Gaudett, M., Salomon, B. & Sun, G. (2005). Molecular variation and population structure in *Elymus trachycaulus* and comparison with its morphologically similar *E. alaskanus*. *Plant Systematics and Evolution*, 250 (1–2), 81–91. https://doi.org/10.1007/s00606-004-0199-y
Gupta, P.K. & Baum, B.R. (1989). Stable classification and nomenclature in the Triticeae - desirability, limitations and prospects. *Euphytica*, 41 (3), 191–197
Hård av Segerstad, F. (1943). En anmärkningsvärd *Agropyron* från Värmland. *Acta Horti Gotob.*, 15, 163–174
Harlan, J.R. (1976). Genetic resources in wild relatives of crops. *Crop Science*, 16, 329–333
Harlan, J.R. (1983). The scope for collection and improvement of forage plants. In: McIvor, J.G. & Bray, R.A. (eds) *Genetic rersources of forage plants*. Melbourne: CSIRO. 3–14.
Harlan, J.R. & de Wet, J.H.J. (1971). Toward a rational classification of cultivated plants. *TAXON*, 20 (4), 509–517
Harrison, K. & Hebda, R.J. (2011). A morphometric analysis of variation between *Elymus alaskanus* and *Elymus violaceus* (Poaceae): Implications for

recognition of taxa. *Madroño*, 58 (1), 32–49. http://www.jstor.org/stable/41425906
He, J., Zhao, X., Laroche, A., Lu, Z.-X., Liu, H. & Li, Z. (2014). Genotyping-by-sequencing (GBS), an ultimate marker-assisted selection (MAS) tool to accelerate plant breeding. *Frontiers in Plant Science*. https://www.frontiersin.org/article/10.3389/fpls.2014.00484
Heath, T.A., Hedtke, S.M. & Hillis, D.M. (2008). Taxon sampling and the accuracy of phylogenetic analyses. *Journal of Systematics and Evolution*, 46 (3), 239–257. https://doi.org/10.3724/sp.j.1002.2008.08016
Helfgott, D.M. & Mason-Gamer, R.J. (2004). The Evolution of North American *Elymus* (Triticeae, Poaceae) Allotetraploids: Evidence from Phosphoenolpyruvate Carboxylase Gene Sequences. *SystBotany*, 29 (4), 850–861
Hitchcock, A.S. (1951). Tribe 3. Hordeae. In: *Manual of the Grasses of the United States.* Washington, D.C.: U.S. Government Printing Office. 230–274.
Hitchcock, C.L. (1969). Gramineaee. In: Hitchcock, C.L., Cronquist, A., Ownbey, M., & Thompson, J.W. (eds) *Vascular plants of the Pacific Northwest, part I.* Seattle: University of Washington Press. 384–726.
Hodge, C.D., Wang, H. & Sun, G. (2010). Phylogenetic analysis of the maternal genome of tetraploid StStYY *Elymus* (Triticeae: Poaceae) species and the monogenomic Triticeae based on rps16 sequence data. *Plant Science*, 178 (5), 463–468. https://doi.org/https://doi.org/10.1016/j.plantsci.2010.03.002
Hu, S.T., Hannaway, D.B. & Youngberg, H.W. (1992). *Forage resources of China*. Wageningen: Pudoc.
Hyun, D.Y., Sebastin, R., Lee, K.J., Lee, G.-A., Shin, M.-J., Kim, S.H., Lee, J.-R. & Cho, G.-T. (2020). Genotyping-by-Sequencing Derived Single Nucleotide Polymorphisms Provide the First Well-Resolved Phylogeny for the Genus *Triticum* (Poaceae). *Frontiers in Plant Science*. https://www.frontiersin.org/article/10.3389/fpls.2020.00688
Jaaska, V. (1992). Isoenzyme variation in the grass genus *Elymus* (Poaceae). *Hereditas*, 117 (1), 11–22. https://doi.org/10.1111/j.1601-5223.1992.tb00002.x
Jeffroy, O., Brinkmann, H., Delsuc, F. & Philippe, H. (2006). Phylogenomics: the beginning of incongruence? *Trends in Genetics*, 22 (4), 225–231. https://doi.org/https://doi.org/10.1016/j.tig.2006.02.003
Jensen, K.B. (1993). Cytogenetics of *Elymus magellanicus* and its intrageneric and intergeneric hybrids with *Pseudoroegneria spicata, Hordeum violaceum, Elymus trachycaulus, E. lanceolatus* and *E. glaucus* (Poaceae, Triticeae). *Genome*, 36 (1), 72–76. https://doi.org/10.1139/g93-010
Jozwik, F.X. (1966). *A biosystematic analysis of slender wheatgrass complex. Doctoral dissertation.* University of Wyoming, Laramie.

Kellogg, E.A. (1989). Comments on genomic genera in the Triticeae (Poaceae). *American Journal of Botany*, 76 (6), 796–805. https://doi.org/10.2307/2444536

Keng, Y.L. (1989). Tribe Hordeae. In: Keng, Y.L. (ed.) *Flora Illustralis Plantarum Primarum Sinicarum (Garmineae)*. Beijing: Science Press. 340–451.

Kumar, K., Gambhir, G., Dass, A., Tripathi, A.K., Singh, A., Jha, A.K., Yadava, P., Choudhary, M. & Rakshit, S. (2020). Genetically modified crops: current status and future prospects. *Planta*, 251 (4), 91. https://doi.org/10.1007/s00425-020-03372-8

Larson, S.R., Jones, T.A., McCracken, C.L. & Jensen, K.B. (2003). Amplified fragment length polymorphism in *Elymus elymoides, Elymus multisetus*, and other *Elymus* taxa. *Canadian Journal of Botany*, 81 (8), 789–804. https://doi.org/10.1139/b03-077

Lehmann, C.E.R., Griffith, D.M., Simpson, K.J., Anderson, T.M., Archibald, S., Beerling, D.J., Bond, W.J., Denton, E., Edwards, E.J., Forrestel, E.J., Fox, D.L., Georges, D., Hoffmann, W.A., Kluyver, T., Mucina, L., Pau, S., Ratnam, J., Salamin, N., Santini, B., Smith, M.D., Spriggs, E.L., Westley, R., Still, C.J., Strömberg, C.A.E. & Osborne, C.P. (2019). Functional diversification enabled grassy biomes to fill global climate space. *bioRxiv*, 583625. https://doi.org/10.1101/583625

Lei, Y.-X., Fan, X., Sha, L.-N., Wang, Y., Kang, H.-Y., Zhou, Y.-H. & Zhang, H.-Q. (2020). Phylogenetic relationships and the maternal donor of *Roegneria* (Triticeae: Poaceae) based on three nuclear DNA sequences (ITS, Acc1, and Pgk1) and one chloroplast region (trnL-F). *Journal of Systematics and Evolution*, 60, 305–318. https://doi.org/https://doi.org/10.1111/jse.12664

Leo, J., Bengtsson, T., Carlsson, A.S., Brassac, J. & von Bothmer, R. (2022). Population structure and phylogeography of Elymus mutabilis and its genetic relationships with E. transbaicalensis (Poaceae). *Nordic Journal of Botany*, 2022 (4), e03520. https://doi.org/https://doi.org/10.1111/njb.03520

Li, P.C., Bhattarai, S., Peterson, G.W., Coulman, B., Schellenberg, M.P., Biligetu, B. & Fu, Y.B. (2018). Genetic Diversity of Northern Wheatgrass (*Elymus lanceolatus* ssp *lanceolatus*) as Revealed by Genotyping-by-Sequencing. *Diversity-Basel*, 10 (2), 14. https://doi.org/10.3390/d10020023

Linde-Laursen, I.B. & Seberg, O. (1999). The karyotype of *Elymus mendocinus* (Poaceae; Triticeae) from Argentina studied by banding techniques. *Hereditas*, 131 (3), 247–252. https://doi.org/10.1111/j.1601-5223.1999.00247.x

Linde-Laursen, I.B. & Seberg, O. (2001). Karyotypes of *Elymus scabrifolius* (Poaceae : Triticeae) from South America studied by banding techniques and in situ hybridization. *Hereditas*, 135 (1), 41–50. https://doi.org/10.1111/j.1601-5223.2001.00041.x

Linde-Laursen, I.B., Seberg, O. & Salomon, B. (1994). Comparison of the Giemsa C-banded and N-banded karyotypes of two *Elymus* species, *E. dentatus* and *E. glaucescens* (Poaceae, Triticeae). *Plant Systematics and Evolution*, 192 (3–4), 165–176. https://doi.org/10.1007/bf00986250

Linnaeus, C. (1753). *Species Plantarum*. Holmiae: Imprensis Laurentii Salvii.

Liu, Q., Ge, S., Tang, H., Zhang, X., Zhu, G. & Lu, B.R. (2006). Phylogenetic relationships in *Elymus* (Poaceae: Triticeae) based on the nuclear ribosomal internal transcribed spacer and chloroplast trnL-F sequences. *New Phytol*, 170 (2), 411–420. https://doi.org/10.1111/j.1469-8137.2006.01665.x

Löve, A. (1984). Conspectus of the Triticeae. *Feddes Repertorium*, 95 (7/8), 425–521

Lu, B.R. (1993). *Biosystematic Investigations of Asiatic Wheatgrasses* - Elymus *L. (Triticeae: Poaceae) (Doctoral thesis)*. *Department Breeding Research Svalöv*. Swedish University of Agricultural Sciences.

Lu, B.R. & Ellstrand, N. (2014). World food security and the tribe Triticeae (Poaceae): Genetic resources of cultivated, wild, and weedy taxa for crop improvement. *Journal of Systematics and Evolution*, 52 (6), 661–666. https://doi.org/10.1111/jse.12131

Ma, X., Chen, S., Zhang, X., Bai, S. & Zhang, C. (2012). Assessment of worldwide genetic diversity of Siberian wild rye (*Elymus sibiricus* L.) germplasm based on gliadin analysis. *Molecules*, 17 (4), 4424–4434. https://doi.org/10.3390/molecules17044424

Mahelka, V. & Kopecký, D. (2010). Gene Capture from across the Grass Family in the Allohexaploid *Elymus repens* (L.) Gould (Poaceae, Triticeae) as Evidenced by ITS, GBSSI, and Molecular Cytogenetics. *Molecular Biology and Evolution*, 27 (6), 1370–1390. https://doi.org/10.1093/molbev/msq021

Mason-Gamer, R.J. (2001). Origin of North American *Elymus* (Poaceae: Triticeae) Allotetraploids Based on Granule-Bound Starch Synthase Gene Sequences. *Systematic Botany*, 26 (4), 757–768. http://www.jstor.org/stable/3093857

Mason-Gamer, R.J. (2004). Reticulate Evolution, Introgression, and Intertribal Gene Capture in an Allohexaploid Grass. *Systematic Biology*, 53 (1), 25–37. https://doi.org/10.1080/10635150490424402

Mason-Gamer, R.J. (2008). Allohexaploidy, introgression, and the complex phylogenetic history of *Elymus repens* (Poaceae). *Molecular Phylogenetics and Evolution*, 47 (2), 598–611. https://doi.org/https://doi.org/10.1016/j.ympev.2008.02.008

Mason-Gamer, R.J. (2013). Phylogeny of a genomically diverse group of *Elymus* (Poaceae) allopolyploids reveals multiple levels of reticulation. *PLoS One*, 8 (11), e78449. https://doi.org/10.1371/journal.pone.0078449

Mason-Gamer, R.J., Burns, M.M. & Naum, M. (2005). Polyploidy, Introgression, and Complex Phylogenetic Patterns within *Elymus*. *Czech Journal of Genetics and Plant Breeding*, 41 (Special Issue), 21–26

Mason-Gamer, R.J., Burns, M.M. & Naum, M. (2010). Phylogenetic relationships and reticulation among Asian *Elymus* (Poaceae) allotetraploids: Analyses of three nuclear gene trees. *Molecular Phylogenetics and Evolution*, 54 (1), 10–22. https://doi.org/10.1016/j.ympev.2009.10.002

Mason-Gamer, R.J., Orme, N.L. & Anderson, C.M. (2002). Phylogenetic analysis of North American *Elymus* and the monogenomic Triticeae (Poaceae) using three chloroplast DNA data sets. *Genome*, 45 (6), 991–1002. https://doi.org/10.1139/g02-065

Maxted, N., Ford-Lloyd, B. V., Jury, S., Kell, S. & Scholten, M. (2006). Toward a definition of a crop wild relative. *Biodiversity and Conservation*, 15, 2673–2685

Maxted, N., Hunter, D. & Ortiz Ríos, R. (2020). *Plant genetic resources*. Cambridge: Cambridge University Press.

McMillan, E. & Sun, G. (2004). Genetic relationships of tetraploid *Elymus* species and their genomic donor species inferred from polymerase chain reaction-restriction length polymorphism analysis of chloroplast gene regions. *Theor Appl Genet*, 108 (3), 535–542. https://doi.org/10.1007/s00122-003-1453-3

Meledris, A. (1980). *Elymus* L. In: Tutin V. H.; Burges, N. A.; Moore, D. M.; Valentine, D. H., Walters, S. M.; Webb D. A., T.G.. H. (ed.) *Flora Europea vol. 5*. Cambridge, New York, Port Chester, Melbourne, Sydney: University Press. 192–198.

Moore, D.M. (1983). *Flora of Tierra del Fuego*. Oswestry & St. Louis.

Nelson, E.N. (1978). Hybridization and Introgression between *Elymus canadensis* and *Elymus virginicus* (Poaceae). *Proceedings of the Oklahoma Academy of Science*, 58, 32–34

Nicora, E.G. (1978). *Flora Patagonica. Parte III. Gramineae*. Correa, M.. N. (ed.) (Correa, M. . N., ed.). Buenos Aires: Instituto Nacional de Tecnologfa Agropecuaria.

Onda, Y. & Mochida, K. (2016). Exploring Genetic Diversity in Plants Using High-Throughput Sequencing Techniques. *Current genomics*, 17 (4), 358–367. https://doi.org/10.2174/1389202917666160331202742

Ørgaard, M. & Anamthawat-Jónsson, K. (2001). Genome discrimination by in situ hybridization in Icelandic species of *Elymus* and *Elytrigia* (Poaceae: Triticeae). *Genome*, 44 (2), 275–283. https://doi.org/10.1139/g00-109

Parodi, L.R. (1940). Estudio crítico de las gramíneas Austral-Americanas del género *Agropyron*. *Revista del Museo de La Plata. Seccion Botanica.*, 3, 1–63

Patwardhan, A., Ray, S. & Roy, A. (2014). Molecular Markers in Phylogenetic Studies - A Review. *Journal of Phylogenetics & Evolutionary Biology*, 2 (2), 1–9

Peschkova, G.A. (1990). *Elymus*, L. In: Malyschev, L.I. & Peschkova, G.A. (eds) *Flora of Siberia, vol. 2, Poaceae (Gramineae)*. Enfield (NH), USA, Plymouth, UK: Scientific Publishers, Inc. 361.

Petersen, G., Seberg, O. & Salomon, B. (2011). The origin of the H, St, W, and Y genomes in allotetraploid species of *Elymus* L. and *Stenostachys* Turcz. (Poaceae: Triticeae). *Plant Systematics and Evolution*, 291 (3–4), 197–210. https://doi.org/10.1007/s00606-010-0381-3

Philippe, H., Snell, E.A., Bapteste, E., Lopez, P., Holland, P.W.H. & Casane, D. (2004). Phylogenomics of eukaryotes: impact of missing data on large alignments. *Molecular biology and evolution*, 21 (9), 1740–1752. https://doi.org/10.1093/molbev/msh182

Pohl, R.W. (1959). Morphology and Cytology of Some Hybrids Between *Elymus canadensis* and *E. virginicus*. *Proceedings of the Iowa Academy of Science*, 66 (1), 155–159

Rushing, J.B. & Baldwin, B.S. (2013). Evaluation of wildrye (*Elymus* spp.) as a potential forage and conservation planting for northeast Mississippi. *Native Plants Journal*, 14 (3), 192–204. https://doi.org/10.3368/npj.14.3.192

Salomon, B., von Bothmer, R. & Seberg, O. (1997). A proposal for an *Elymus* core collection. *Plant Genetic Resources Newsletter (IPGRI/FAO)*, 111, 77–81

Salomon, B., Bothmer, R. von, Yang, J.L. & Lu, B.R. (1988). Notes on perennial Triticeae – species in northern Pakistan. *Botanische Jahrbücher für Systematik, Pflanzengeschichte und Pflanzengeographie*, 110, 7–15

Salomon, B. & Lu, B.R. (1992). Genomic groups, morphology and sectional delimitaion in Eurasian *Elymus* (Poaceae, Triticeae). *Plant Systematics and Evolution*, 180 (1–2), 1–13. https://doi.org/10.1007/bf00940394

Seberg, O. (1989). A biometrical analysis of the South American *Elymus glaucescens* complex (Poaceae, Triticeae). *Plant Systematics and Evolution*, 166 (1–2), 91–104. https://doi.org/Doi 10.1007/Bf00937878

Seberg, O. & von Bothmer, R. (1991). Genome analysis of *Elymus angulatus* and *E. patagonicus* (Poaceae) and their hybrids with North American and South American Hordeum spp. *Plant Systematics and Evolution*, 174 (1–2), 75–82. https://doi.org/10.1007/bf00937695

Seberg, O. & Frederiksen, S. (2001). A phylogenetic analysis of the monogenomic Triticeae (Poaceae) based on morphology. *Botanical Journal of the Linnean Society*, 136 (1), 75–97. https://doi.org/10.1006/bojl.2000.0415

Seberg, O. & Petersen, G. (1998). A taxonomic revision of the genus *Elymus* L. s. l. (Poaceae, triticeae) in South America. *Botanische Jahrbücher fur Systematik, Pflanzengeschichte und Pflanzengeographie*, 120 (4), 503–543

Soltis, D.E. & Soltis, P.S. (1993). Molecular data and the dynamic nature of polyploidy. *Critical Reviews in Plant Sciences*, 12 (3), 243–273. https://doi.org/10.1080/713608048

Stebbins, G.L. & Snyder, L.A. (1956). Artificial and natural hybrids in the Gramineae, tribe Hordeae, hybrids between western and eastern North American species. *American Journal of Botany*, 43 (4), 305–312. https://doi.org/10.2307/2438948

Stevens, L., Salomon, B. & Sun, G. (2007). Microsatellite variability and heterozygote excess in *Elymus trachycaulus* populations from British Columbia in Canada. *Biochemical Systematics and Ecology*, 35 (11), 725–736. https://doi.org/10.1016/j.bse.2007.05.017

Stevens, P.F. (2001). *Angiosperm Phylogeny Website*. *version 14*. http://www.mobot.org/MOBOT/research/APweb/ [2022-05-31]

Sun, D.F. & Sun, G.L. (2012). Untangling Nucleotide Diversity and Evolution of the H Genome in Polyploid *Hordeum* and *Elymus* Species Based on the Single Copy of Nuclear Gene DMC1. *Plos One*, 7 (12), 9. https://doi.org/10.1371/journal.pone.0050369

Sun, G. (2014). Molecular phylogeny revealed complex evolutionary process in *Elymus* species. *Journal of Systematics and Evolution*, 52 (6), 706–711. https://doi.org/10.1111/jse.12080

Sun, G. & Li, W. (2005). Molecular diversity of species of the *Elymus trachycaulus* species complex and their relationships to non-North American taxa. *Plant Systematics and Evolution*, 256 (1), 179–191. https://doi.org/10.1007/s00606-005-0368-7

Sun, G., Ni, Y. & Daley, T. (2008). Molecular phylogeny of RPB2 gene reveals multiple origin, geographic differentiation of H genome, and the relationship of the Y genome to other genomes in *Elymus* species. *Molecular Phylogenetics and Evolution*, 46 (3), 897–907. https://doi.org/10.1016/j.ympev.2007.12.024

Sun, G. & Salomon, B. (2003). Microsatellite variability and heterozygote deficiency in the arctic-alpine Alaskan wheatgrass (*Elymus alaskanus*) complex. *Genome*, 46 (5), 729–737. https://doi.org/10.1139/g03-052

Sun, G. & Salomon, B. (2009). Molecular evolution and origin of tetraploid *Elymus* species. *Breeding Science*, 59 (5), 487–491. https://doi.org/10.1270/jsbbs.59.487

Sun, G., Tang, H. & Salomon, B. (2006a). Molecular diversity and relationships of North American *Elymus trachycaulus* and the Eurasian *E. caninus* species. *Genetica*, 127 (1), 55–64. https://doi.org/10.1007/s10709-005-2481-5

Sun, G.L. (2007). Genetic diversity of rbcL gene in *Elymus trachycaulus* complex and their phylogenetic relationships to several Triticeae species. *Genetic Resources and Crop Evolution*, 54 (8), 1737–1746. https://doi.org/10.1007/s10722-006-9183-4

Sun, G.L., Diaz, O., Salomon, B. & von Bothmer, R. (1999). Genetic diversity in *Elymus caninus* as revealed by isozyme, RAPD, and microsatellite markers. *Genome*, 42 (3), 420–431. https://doi.org/10.1139/gen-42-3-420

Sun, G.L., Diaz, O., Salomon, B. & von Bothmer, R. (2001). Genetic diversity and structure in a natural *Elymus caninus* population from Denmark based on microsatellite and isozyme analyses. *Plant Systematics and Evolution*, 227 (3–4), 235–244. https://doi.org/10.1007/s006060170051

Sun, G.L., Diaz, O., Salomon, B. & Von Bothmer, R. (1998a). Microsatellite variation and its comparison with allozyme and RAPD variation in *Elymus fibrosus* (Schrenk) Tzvel. (Poaceae). *Hereditas*, 129 (3), 275–282. https://doi.org/10.1111/j.1601-5223.1998.00275.x

Sun, G.L. & Ma, X.Z. (2009). Nucleotide diversity and minisatellite in chloroplast Asp(GUC)-Thr(GGU) region in *Elymus trachycaulus* complex, *Elymus alaskanus* and *Elymus caninus*. *Biochemical Systematics and Ecology*, 37 (2), 67–75. https://doi.org/10.1016/j.bse.2008.12.002

Sun, G.L., Salomon, B. & von Bothmer, P. (1998b). Characterization of microsatellite loci from *Elymus alaskanus* and length polymorphism in several *Elymus* species (Triticeae : Poaceae). *Genome*, 41 (3), 455–463. https://doi.org/10.1139/gen-41-3-455

Sun, G.L., Salomon, B. & von Bothmer, R. (1997). Analysis of tetraploid *Elymus* species using wheat microsatellite markers and RAPD markers. *Genome*, 40 (6), 806–814. https://doi.org/10.1139/g97-804

Sun, G.L., Salomon, B. & von Bothmer, R. (2002). Microsatellite polymorphism and genetic differentiation in three Norwegian populations of *Elymus alaskanus* (Poaceae). *Plant Systematics and Evolution*, 234 (1–4), 101–110. https://doi.org/10.1007/s00606-002-0211-3

Sun, G.L., Shee, J. & Salomon, B. (2006b). Molecular diversity and relationships among *Elymus trachycaulus, E. subsecundus, E. virescens, E. violaceus*, and *E. hyperarcticus* (Poaceae : Triticeae) as determined by amplified fragment length polymorphism. *Genome*, 49 (9), 1160–1169. https://doi.org/10.1139/g06-062

Svitashev, S., Bryngelsson, T., Li, X.M. & Wang, R.R.C. (1998). Genome-specific repetitive DNA and RAPD markers for genome identification in *Elymus* and *Hordelymus*. *Genome*, 41 (1), 120–128. https://doi.org/10.1139/gen-41-1-120

Svitashev, S., Salomon, B., Bryngelsson, T. & VonBothmer, R. (1996). A study of 28 *Elymus* species using repetitive DNA sequences. *Genome*, 39 (6), 1093–1101. https://doi.org/10.1139/g96-137

Tomas, P.A., Gonzalez, G.E., Schrauf, G.E. & Poggio, L. (2012). Chromosomal characterization in native populations of *Elymus scabrifolius* from Argentina through classical and molecular cytogenetics (FISH-GISH). *Genome*, 55 (8), 591–598. https://doi.org/10.1139/g2012-046

Townsend, J.P., Su, Z. & Tekle, Y.I. (2012). Phylogenetic Signal and Noise: Predicting the Power of a Data Set to Resolve Phylogeny. *Systematic Biology*, 61 (5), 835. https://doi.org/10.1093/sysbio/sys036

Tzvelev, N.N. (1976). Tribe 3: Triticeae Dum. In: *Poaceae U.R.S.S.* Leningrad: Nacka Publishing House. 146–298.

Tzvelev, N.N. (1989). The system of grasses (Poaceae) and their evolution. *Botanical Review*, 55 (3), 141–204. https://doi.org/10.1007/bf02858328

Wang, Q.X., Xiang, J.S., Gao, A.N., Yang, X.M., Liu, W.H., Li, X.Q. & Li, L.H. (2010). Analysis of chromosomal structural polymorphisms in the St, P, and Y genomes of Triticeae (Poaceae). *Genome*, 53 (3), 241–249. https://doi.org/10.1139/g09-098

Wang, R.-C., von Bothmer, R., Dvorak, J., Fedak, G., Linde-Laursen, I.B. & Muramatsu, M. (1994). Genome Symbols in Triticeae. In: Wang, R.R.C., Jensen, K.B., & Jaussi, C. (eds) *Proceedings of the 2nd International Triticeae Symposium*. Logan: Utha state University Press. 29–34.

Wang, X.-L., Fan, X., Zhang, H.-Q., Zhang, C., Sha, L.-N., Kang, H.-Y., Yang, R.-W., Zhang, L., Ding, C.-B. & Zhou, Y.-H. (2011). Phylogeny of the StH genome species in Triticeae (Poaceae): Evidence from chloroplast trnL-F and mitochondria COXII intron sequences. *Biochemical Systematics and Ecology*, 39 (4), 758–765. https://doi.org/https://doi.org/10.1016/j.bse.2011.06.017

West, J.G., McIntyre, C.L. & Appels, R. (1988). Evolution and systematic relationships in the Triticeae (Poaceae). *Plant Systematics and Evolution*, 160 (1–2), 1–28. https://doi.org/10.1007/bf00936706

Whelan, S., Liò, P. & Goldman, N. (2001). Molecular phylogenetics: state-of-the-art methods for looking into the past. *Trends in genetics : TIG*, 17 (5), 262–272. https://doi.org/10.1016/s0168-9525(01)02272-7

Wu, P.P., Zuo, H.W., Sun, G.L., Wu, D.X., Salomon, B., Hu, Q.W. & Dong, Z.R. (2016). Comparison of gene flow among species that occur within the same geographic locations versus gene flow among populations within species reveals introgression among several *Elymus* species. *Journal of Systematics and Evolution*, 54 (2), 152–161. https://doi.org/10.1111/jse.12172

Xiong, Y.L., Liu, W.H., Xiong, Y., Yu, Q.Q., Ma, X., Lei, X., Zhang, X.Q. & Li, D.X. (2019). Revelation of genetic diversity and structure of wild *Elymus excelsus* (Poaceae: Triticeae) collection from western China by SSR markers. *Peerj*, 7, 18. https://doi.org/10.7717/peerj.8038

Yan, C. & Sun, G.L. (2012). Multiple origins of allopolyploid wheatgrass *Elymus caninus* revealed by RPB2, PepC and TrnD/T genes. *Molecular Phylogenetics and Evolution*, 64 (3), 441–451. https://doi.org/10.1016/j.ympev.2012.04.017

Yen, C. & Yang, J. (2022). *Biosystematics of Triticeae, Volume II. Genera:* Secale, Tritiosecale, Pseudosecale, Eremopyrum, Henrardia, Taeniantherum, Heteranthelium, Crithopsis, *and* Hordeum. Singapore: China Agriculture Press and Springer Nature Singapore Pte Ltd.

Yen, C. & Yang, J.L. (2009). Historical review and prospect of taxonomy of tribe Triticeae Dumortier (Poaceae). *Breeding Science*, 59 (5), 513–518. https://doi.org/10.1270/jsbbs.59.513

Yen, C., Yang, J.L. & Yen, Y. (2005). Hitoshi Kihara, Askell Dive and the modern genetic concept of the genera in the tribe Triticeae (Poaceae). *Acta Phytotaxonomica Sinica*, 43 (1), 82–93. https://doi.org/10.1360/aps040073

Young, A.D. & Gillung, J.P. (2020). Phylogenomics — principles, opportunities and pitfalls of big-data phylogenetics. *Systematic Entomology*, 45 (2), 225–247. https://doi.org/https://doi.org/10.1111/syen.12406

Zhang, C., Fan, X., Yu, H.Q., Zeng, J., Zhang, H.Q., Wang, X.L. & Zhou, Y.H. (2009). Phylogenetic relationships among the species of *Elymus* sensu lato in Triticeae (Poaceae) based on nuclear rDNA ITS sequences. *Russian Journal of Genetics*, 45 (6), 696–706. https://doi.org/10.1134/s102279540906009x

Zhang, X.Q., Salomon, B. & von Bothmer, R. (2002). Application of random amplified polymorphic DNA markers to evaluate intraspecific genetic variation in the *Elymus alaskanus* complex (Poaceae). *Genetic Resources and Crop Evolution*, 49 (4), 397–407

Zhang, X.Q., Salomon, B., Von Bothmer, R. & Diaz, O. (2000). Patterns and levels of genetic differentiation in North American populations of the Alaskan wheatgrass complex. *Hereditas*, 133 (2), 123–132. https://doi.org/10.1111/j.1601-5223.2000.00123.x

Zhao, Y.Q., Zhang, J.C., Zhang, Z.Y. & Xie, W.G. (2019). *Elymus nutans* genes for seed shattering and candidate gene-derived EST-SSR markers for germplasm evaluation. *Bmc Plant Biology*, 19, 102. https://doi.org/10.1186/s12870-019-1691-4

Zuo, H.W., Wu, P.P., Wu, D.X. & Sun, G.L. (2015). Origin and Reticulate Evolutionary Process of Wheatgrass *Elymus trachycaulus* (Triticeae: Poaceae). *Plos One*, 10 (5), 1–17. https://doi.org/10.1371/journal.pone.0125417

Popular science summary

Our food production is part of a system where climate, pests, politics and a number of other factors affect the plants we grow and how we breed new crops. To meet future challenges in the food supply chain, we need new crops or crops with improved properties, which we can develop through plant breeding. Grasses (Poaceae) are the most important plant group for the earth's food and feed production, and in Sweden, for example, we grow staple crops such as wheat, barley and rye. Wild relatives of our cultivated plants constitute a genetic resource where traits can be transferred in different ways from one plant to another, thus creating better-adapted crops. For example, wheat and barley, with genes from wild relatives, could become more drought-resistant, gain better resistance to pests or become more efficient at taking up a specific nutrient. These wild genetic resources must be preserved for the future.

The genus *Elymus* (elms) consists of approximately 150 species distributed in temperate regions throughout the world. They are relatives of our cultivated barley and wheat and possess a huge variety of traits, which makes them interesting for future plant breeding. However, the genus is complex. Knowledge of the overall relationships between species as well as the variation within species needs to be investigated to be able to use and preserve species and populations effectively. We need to know the variation and the relationships in order to know what is what, what to look for, and where to look to find similar characteristics. It is difficult to exploit or preserve a plant that we do not know exists. This is why plant systematics is important for both plant breeding and conservation. The task of plant systematics is to find, describe and organize plant variation.

This thesis studies the genetic relationships between species within a part of the genus *Elymus* and presents an evolutionary (phylogenetic) tree. The

results show that the American species are genetically distinct from the Eurasian ones, although there are some species that have spread between these regions. There have also been several movements of species between North and South America. It is with the knowledge of these evolutionary relationships that we organize variation. It also helps us understand how species have evolved.

The thesis also contains a study that examines the variation within the species *E. mutabilis*. It shows that the species originated in Central Asia and has spread north, then east, and west. Populations in the Nordic region are more closely related to species in eastern than central Russia. However, the populations in the Nordic region differ from other regions and are worth preserving.

Within the framework of the thesis, we have also collected populations of *Elymus* in Chile. These will be the basis for future studies on how many species there are in the country and the relationships between them. Furthermore, this will facilitate decisions on how to preserve them in a good way for use in future breeding and research projects.

Populärvetenskaplig sammanfattning

Vår matproduktion är en del av ett system där klimat, skadegörare, politik och en rad andra faktorer påverkar de växter vi odlar och hur vi får fram nya grödor. För att möta framtida utmaningar inom livsmedelsförsörjningen kommer vi behöva nya grödor eller grödor med förbättrade egenskaper vilket vi kan ta fram genom växtförädling. Gräs (Poaceae) är den viktigaste växtgruppen för jordens matproduktion och i Sverige odlar vi exempelvis stapelgrödor som vete, korn och råg. Vilda släktningar till våra odlade växter utgör en genetisk resurs där egenskaper på olika sätt kan överföras från en växt till en annan och därmed skapa bättre anpassade grödor. Exempelvis skulle vete och korn, med gener från vilda släktningar, kunna bli mer torktåliga, få bättre resistens mot skadegörare eller bli mer effektiva på att ta upp ett specifikt näringsämne. Dessa vilda genetiska resurser måste också bevaras för framtiden.

Släktet *Elymus* (elmar) utgörs av ungefär 150 arter utspridda i tempererade områden över hela världen. De är släktingar till vårt odlade korn och vete och besitter en enorm variation av egenskaper, vilket gör dem intressanta för framtida växtförädling. Släktet är dock komplext. Kunskapen om den övergripande släktskapen mellan arter så väl som variationen inom arter behöver undersökas för att kunna nyttja och bevara arter och populationer på ett effektivt sätt. Vi måste känna till variationen och släktskapen för att kunna veta vad som är vad, vad vi ska leta efter och var vi ska leta för att hitta liknande egenskaper. Det är svårt att nyttja eller bevara en växt som vi inte vet existerar. Det är därför som växtsystematik är viktigt för både växtförädling och bevarande. Växtsystematikens uppgift är att hitta, beskriva och organisera växters variation.

Den här avhandlingen studerar de genetiska släktskapen mellan arter inom en del av släktet *Elymus* och presenterar ett släktträd. Resultaten visar

att de amerikanska arterna är genetiskt skilda från de euroasiatiska, även om det finns några arter som spridit sig mellan dessa regioner. Det har också skett flera förflyttningar av arter mellan Nord- och Sydamerika.

Avhandlingen innehåller även en studie som undersöker variationen inom arten lappelm (*E. mutabilis*). Det visar sig att arten härstammar från centralasien och har spridit sig norrut, sedan österut och västerut. Populationer i Norden är närmare besläktade med arter i östra än i centrala Ryssland. Populationerna i Norden skiljer sig dock från andra regioner och är viktiga att bevara.

Inom ramen för avhandlingen har vi också samlat in populationer av *Elymus* i Chile. Dessa kommer ligga till grund för framtida studier om hur många arter som finns i landet och hur de är släkt med varandra. Detta kommer vidare att underlätta beslut om hur de ska bevaras på ett bra sätt för användning i framtida förädlings- och forskningsprojekt.

Acknowledgements

First, I want to thank my supervisors for all their support. Therése, thank you for your never-ending engagement and genuine willingness to help. You are one of the most important reasons for the completion of this work. Roland, thank you for your infectious enthusiasm for research, Triticeae, and CWR. Your deep knowledge of the subject and availability has been a true comfort throughout the project. I hope we can travel again soon! Anders, thank you for your encouragement and visionary mind. You are the steady rock in this project. I also want to thank my colleagues at the department of Plant Breeding with special thanks to the technical and bioinformatic staff.

I am also very grateful to our international collaborators Arturo Morales and Fernando Ortega at INIA, Chile, to Oscar Diaz at the University of Los Lagos, Chile, and to Frank Blattner and Jonatan Brassac at IPK, Gatersleben, Germany. Arturo, you are my partner in crime. Thank you for your hospitality and sense of humor.

The research has been financed by the Royal Physiographic Society in Lund (Kungliga Fysiografiska Sällskapet), Einar and Inga Nilssons stiftelse, and the Department of Plant Breeding at SLU (Swedish University of Agricultural Sciences).

Finally, I want to thank my family for all their love and support. You are the center of my life.

I

NORDIC JOURNAL OF
BOTANY

Research

Population structure and phylogeography of *Elymus mutabilis* and its genetic relationships with *E. transbaicalensis* (Poaceae)

Jonatan Leo, Therése Bengtsson, Anders S. Carlsson, Jonathan Brassac and Roland von Bothmer

J. Leo (https://orcid.org/0000-0002-9648-1878) ✉ (jonatan.leo@slu.se), T. Bengtsson (https://orcid.org/0000-0003-4784-1723), A. S. Carlsson (https://orcid.org/0000-0002-8525-9753) and R. von Bothmer, Dept of Plant Breeding, Swedish Univ. of Agricultural Sciences (SLU), Alnarp, Sweden. – J. Brassac (https://orcid.org/0000-0001-6476-2619), Experimental Taxonomy, Leibniz Inst. of Plant Genetics and Crop Plant Research (IPK), Gatersleben, Germany.

Nordic Journal of Botany
2022: e03520
doi: 10.1111/njb.03520
Subject Editor and
Editor-in-Chief: Torbjörn Tyler
Accepted 25 January 2022
Published 4 March 2022

Elymus mutabilis is a morphologically diverse species in the Poaceae family growing in Eurasia from northern Europe to far east Russia and southwards to central Asia. *Elymus transbaicalensis* occurs in similar habitats and is considered closely related to *E. mutabilis* and sometimes even referred to as a subspecies or synonym. Based on high similarity in morphology and habitat, molecular studies are needed to establish whether *E. mutabilis* and *E. transbaicalensis* can be considered as two distinct species. Thus, the objective of this study was to study diversity, relationships among populations and the phylogeographical structure of *E. mutabilis* and *E. transbaicalensis* using genotyping-by-sequencing (GBS). In total 68 individuals of *E. mutabilis* were sampled from 18 populations collected from northern Europe, central Asia and far east Russia, representing the central and two peripheral parts of the natural distribution of the species. The results reveal a clear distinction between *E. mutabilis* and *E. transbaicalensis* and no introgression. The phylogeographic structure of *E. mutabilis* follows the geographical distribution of the species. Populations from northern Europe, southern Siberia and far east Russia together form a clade separated from the peripheral populations in central Asia, indicating a common ancestry of the latter. Phylogenetic analyses revealed a radiation pattern among populations in northern Europe indicating a founding followed by rapid dispersal.

Keywords: Triticeae, ipyrad, phylogenetics, genotyping-by-sequencing

Introduction

Elymus s. lat. in the Triticeae tribe in Poaceae is most often circumscribed as an allopolyploid genus containing genomes from *Pseudoroegneria* (St) together with genomes from either *Hordeum* (H), an unknown donor (Y) or a combination of all three types. For most *Elymus* species, it is still uncertain which the donating species are. Apart from the widespread St, H and Y genomes, *Elymus* also contains the restricted genomes P (from *Agropyron* s. str.) and W (from *Australopyrum*) (Sun and Salomon 2009). The center of diversity for the genus *Elymus* s. lat. is in central Asia with the highest number

of species and a high degree of hybridization and introgression, making the taxonomy complicated. *Elymus* species are wild relatives to some of the major cereals, such as wheat and barley, hence a potential genetic resource for future cereal and forage crop breeding (von Bothmer and Salomon 1994). Thus, one major aim for studies of this group is to collect data of value for breeding and conservation purposes.

Elymus mutabilis (Drob.) Tzvel. is a perennial, caespitose species with a scattered distribution in boreal Eurasia, from northern Europe and eastwards to far east Russia where it occupies forest clearings, riverbanks and meadows (Hultén 1971, Peschkova 1990). The species is morphologically variable, also depending on growing conditions, and with naturally occurring intermediates (Agafonov 2004). It is sometimes referred to as a species complex, including a number of closely related taxa concentrated to central and eastern Asia such as *E. mutabilis* s.s., *E. transbaicalensis* (Nevski) Tzvelev, *E. praecaespitosus* (Nevski) Tzvelev, *E. viridiglumis* (Nevski) Czerep., *E. charkeviczii* (Prob.) Czerep. and *E. subfibrosus* (Nevski) Czerep. (Agafonov et al. 1998, 2005, 2019, Agafonov and Salomon 2002). The latter two have been treated as morphological variants of *E. mutabilis* (Agafonov and Salomon 2002, Agafonov et al. 2005). *Elymus viridiglumis* is probably not a monophyletic distinct taxon but has a complex origin where populations have been derived from both *E. caninus* and *E. mutabilis* (Agafonov 2004, Emtseva and Agafonov 2018, Shabanova (Kobozeva) et al. 2020). Flora of China (eFloras 2008) considers *E. praecaespitosus* as a variety of *E. mutabilis* (*E. mutabilis* var. *praecaespitosus* (Nevski) S. L. Chen) that differs in lacking rhizomes and having glaucous spikelets, but the taxon has also been treated as a subspecies (*E. mutabilis* subsp. *praecaespitosus* (Nevski) Tzvelev). *Elymus transbaicalensis* is considered closely related to *E. mutabilis*, and often described as a subspecies or a synonym (Tzvelev 1973, POWO 2019). The two taxa have sympatric distributions and occur in similar habitats, though *E. transbaicalensis* is restricted to western and southern Siberia, Mongolia and northwestern China and may grow at lower altitudes compared to *E. mutabilis* (Agafonov 2004). They are morphologically similar but still considered as two distinct species in the Flora of Siberia based on differences in glume characteristics and anther size (Peschkova 1990, Agafonov 2004). *Elymus mutabilis* has a more or less hairy internal glume surface and 1.5–2.5 mm long anthers, while *E. transbaicalensis* has glabrous or scabrous internal glume surface and 1.0–1.5 mm long anthers. They have partly sympatric but scattered distribution areas though *E. transbaicalensis* is restricted to western and southern Siberia, Mongolia and northwestern China. The distributions are disjunct for both taxa, probably mainly due to the fact that suitable habitats are missing in many areas (Fig. 1).

Both *E. mutabilis* and *E. transbaicalensis* are allotetraploids (2n = 4x = 28) with a StStHH genome combination (Löve 1984, Salomon et al. 1988). The two taxa belong to a larger

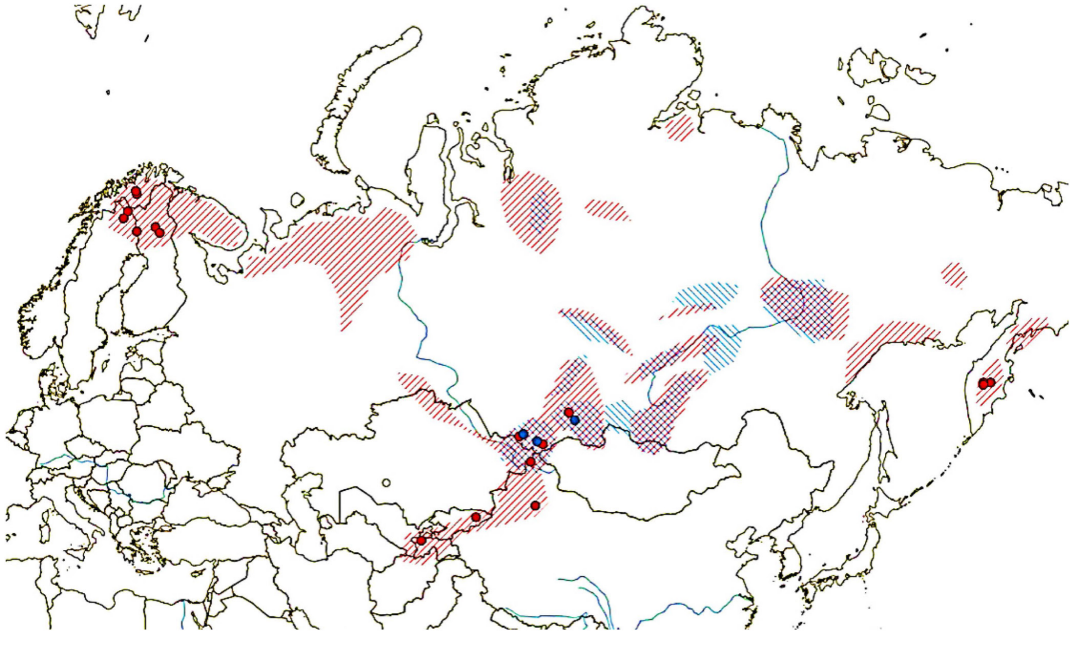

Figure 1. Distribution (lines) and sampled populations (dots) of *Elymus mutabilis* (red) and *E. transbaicalensis* (blue). Sympatric distribution in checkered pattern. Distrubution patterns are combined and modified from Peschkova (1990), Hultén (1971) and Tzvelev (1976).

species group with a circumpolar distribution with StStHH genome combinations including, for example, the widespread *E. trachycaulus*, *E. alaskanus*, *E. fibrosus* and *E. caninus* in Eurasia and North America together with a number of taxa with more restricted distributions.

Agafonov (2004) expressed a need to verify the taxonomical status of *E. mutabilis* and *E. transbaicalensis* and used morphology, biosystematics and genetic data for this purpose. He found that the two taxa are morphologically discrete and differ in electrophoretic patterns of endosperm storage proteins. The species are predominately self-pollinating and thus reproductively isolated, but crossing experiments show a variation in sexual compatibility among populations (biotypes). Offspring is semi-fertile and the taxa belong to the same introgressive gene pool (Agafonov et al. 1998, Agafonov 2004). Based on these findings, Agafonov argued that *E. mutabilis* and *E. transbaicalensis* should be considered as distinct species.

Diaz et al. (1999a) used isozyme and allozyme data from 17 loci to study the genetic diversity within and among ten populations of *E. mutabilis* in northern Europe. They found no genetic variation and a low morphological variation in this region. The techniques are, however, not as accurate as modern DNA marker technology. An earlier study using isozyme markers comparing populations from northern Europe, China, Russia and Pakistan showed a higher genetic diversity at the species level over the whole species distribution area (Diaz et al. 1998).

Using single nucleotide polymorphisms (SNPs) obtained from genotyping by sequencing (GBS) has proven to be a powerful approach in the characterization of genetic diversity, breeding and phylogenetic inferences (Geibel and Hohlfeld 2003, Lu et al. 2013, Xiong et al. 2016, Chung et al. 2017, Wagner et al. 2020), also in young lineages, within species and complexes (Anderson et al. 2017, Pérez-Escobar et al. 2020). The benefit of creating assemblies also without the requirement of a reference genome has made the technique popular in non-model taxa. Contrary to whole-genome sequencing, GBS uses a combination of building short read restriction site associated libraries together with fragment barcoding and subsequent DNA sequencing in a high-throughput platform (Elshire et al. 2011). The major benefits are the low cost, high-density genotyping and restricted sequencing to reduce genome complexity and facilitate building assays. The main drawbacks are relatively high proportion of missing data, uneven genome coverage and potential issues related to polyploidy. Diversity studies using GBS in *Elymus* have previously been done by Li et al. (2018), analyzing *E. lanceolatus* ssp. *lanceolatus*.

The objectives of this study are to 1) find molecular evidence for a taxonomic differentiation between *E. mutabilis* and *E. transbaicalensis*, 2) find molecular evidence of eventual differentiation between Nordic and Asian *E. mutabilis* populations, 3) study the diversity, gene flow and phylogenetic relationships among populations and 4) study the phylogeographic structure of *E. mutabilis*. Most studies on *Elymus* species diversity focus on limited geographical areas, but this study covers a large area capturing a wide range of the species distribution. Sympatrically grown individuals of *E. mutabilis* and *E. transbaicalensis* are compared with *E. mutabilis* populations from northern Europe and far east Russia. The present study is part of a larger investigation studying genome relationships, genetic diversity and phylogenetic pathways in Eurasian and American species of *Elymus* with the StH genome combination.

Material and methods

Plant materials

In the present study, 68 individuals of *Elymus mutabilis* were sampled from a total of 18 populations (accessions) from northern Europe, central Asia and far east Russia, representing the central and two peripheral parts of the natural distribution of the species (Fig. 1). In addition, 14 individuals of *E. transbaicalensis* from three populations were included for genetic comparisons and as outgroup for phylogenetic analyses of *E. mutabilis*. Seed material was collected in wild native stands on several collection expeditions between 1986 and 2003. Seeds were stored in −18°C freezers at the Triticeae germplasm collection at the Swedish University of Agricultural Sciences (SLU) in Alnarp, Sweden. Determination of polyploidy was previously conducted for some of the populations, indicated in Table 1. Since all taxa within this group are known to be tetraploid the remaining individuals in this study were assumed tetraploid (Löve 1984, Salomon et al. 1988, Diaz et al. 1999a). The number of individuals per population and accompanying data is presented in Table 1. Seeds were germinated in a greenhouse and samples of leaf tissue were taken from one leaf for each individual seedling. The plants were transplanted and put outside for morphological confirmation using the diagnostic traits from Peschkova (1990) and Agafonov (2004). Herbarium voucher specimens are kept at SLU, Alnarp.

Genotyping-by-sequencing

DNA extraction

The leaf samples were freeze-dried overnight, ground in a shaker (Retsch MM400) at 13 200 rpm for 2 min with two 4 mm size glass beads and finally stored in a −80°C freezer prior to DNA extraction. The samples were pretreated using the CTAB (cetyl trimetheyl ammonium bromide) method, as described by Åhman and Bengtsson (2019) and genomic DNA was extracted using DNeasy Plant DNA Extraction Kit (Qiagen) with a Qiacube DNA extraction robotic workstation (Qiagen, Hilden, Germany), according to the protocol of the manufacturer. DNA quality was controlled on a 1% agarose gel relative to a DNA standard and quantifications were performed using both Nanodrop and QubitFluorometer. Samples were diluted to a final concentration of 20 ng μl^{-1}.

Restriction enzyme and sequencing

For each sample, 200 ng of genomic DNA were used to prepare a GBS library (Elshire et al. 2011). Library preparations were

Table 1. List of included species, population (accession) numbers, the number of individuals per population, ploidy level (when known) and region, country and location of origin.

Population number	Species	No. of individuals	2n*	Region	Country	Location
H7509	*Elymus mutabilis*	3	4x	Central Asia	China	Xinjiang
H7601	*Elymus mutabilis*	5	4x	Central Asia	China	Xinjiang
H3519	*Elymus mutabilis*	3		Central Asia	Kyrgyzstan	Kyrgyzstan
H10084	*Elymus mutabilis*	4	4x	Central Asia	Russia	Altai
H10142	*Elymus mutabilis*	2	4x	Central Asia	Russia	Altai
H10410	*Elymus mutabilis*	5		Central Asia	Russia	Southern Siberia
H10235	*Elymus mutabilis*	5	4x	Central Asia	Tajikistan	Tajikistan
H10449	*Elymus mutabilis*	5		Far East Russia	Russia	Kamchatka
H10455	*Elymus mutabilis*	5		Far East Russia	Russia	Kamchatka
H10456	*Elymus mutabilis*	3		Far East Russia	Russia	Kamchatka
H10468	*Elymus mutabilis*	4		Far East Russia	Russia	Kamchatka
H10334	*Elymus mutabilis*	4		Northern Europe	Finland	Sodankylä
H10337	*Elymus mutabilis*	4		Northern Europe	Finland	Sodankylä
H10833	*Elymus mutabilis*	4		Northern Europe	Norway	Finnmark
H10839	*Elymus mutabilis*	1		Northern Europe	Norway	Finnmark
H10842	*Elymus mutabilis*	5		Northern Europe	Sweden	Lappland
H10326	*Elymus mutabilis*	3		Northern Europe	Sweden	Norrbotten
H10350	*Elymus mutabilis*	3		Northern Europe	Sweden	Norrbotten
H10051	*Elymus transbaicalensis*	5	4x	Central Asia	Russia	Altai
H10114	*Elymus transbaicalensis*	4	4x	Central Asia	Russia	Altai
H10378	*Elymus transbaicalensis*	5		Central Asia	Russia	Southern Siberia

conducted at the Institute for Plant Genetics and Crop Plant Research (IPK), Gatersleben, Germany and the procedure followed Wendler et al. (2014) using the two restriction enzymes PstI-HF (CTGCAG, NEB Inc., Ipswich, UK) and MspI (CCGG, NEB Inc.). An Illumina HiSeq 2000/2500 (100 bp single-end reads) was used to sequence the genomic libraries. Barcoded reads were de-multiplexed using the CASAVA pipeline 1.8 (Illumina, Inc.). Adapter trimming was performed with CUTADAPT (Martin 2011) and reads shorter than 60 bp, after removal of the adapter, were discarded. All demultiplexed reads are available through the NCBI Short Read Archive (SRA) under BioProject PRJNA770613.

Bioinformatic and genetic diversity analysis

Assembly

De novo assemblies of loci were performed using the ipyrad 0.9.19 pipeline (Eaton 2014, Eaton and Overcast 2020) with a few modifications to the recommended default parameter settings for single-end read GBS data. The polyploidy level was appointed as tetraploid with filtering of putative paralogs set to a maximum of four alleles and eight heterozygous positions per consensus sequence (the number of maximal shared heterozygotic sites and indels). To assess the impact of missing data and the effects of different sequence similarity thresholds on tree topologies and population structures, 10 sets of assembled GBS loci data were created, compared and evaluated. To determine the effect of clustering strategies (which combine de novo assembly of short reads into clusters of alleles representing different loci) on the subsequent analysis, two contrasting thresholds of clustering values were employed (for within-sample and across sample sequence clustering (c)), $c = 0.85$ and $c = 0.90$, i.e. the threshold when two sequences are identified as being homologous. Clustering threshold settings are known to influence the number of alleles per locus and the number of divergent alleles. An overly stringent clustering threshold, as well as highly variable species, may result in loss of divergent alleles or that orthologous sequences are divided into separate loci, so called 'over-splitting' (Harvey et al. 2015). On the other hand, a too liberal clustering threshold may result in so called 'under-splitting' with paralogous loci being combined into a single locus. As a further filter of putative paralogs, five assemblies for each clustering value were produced with different thresholds for the minimum number of samples that must have shared data at a given locus for it to be retained in the dataset (m): 4, 20, 40, 60 and 82, with the last being the full dataset.

Population genetic analyses

Maximum likelihood (ML) trees were inferred from the ten assemblies using RAxML ver. 8.0 (Stamatakis 2014). After removal of invariant sites with raxml_ascbias (<https://github.com/btmartin721/raxml_ascbias>), the GTR + G substitution model was implemented with correction for ascertainment bias using the Lewis method, and with 100 bootstrap replicates. The trees were visualized with FigTree ver. 1.4.4 (Rambaut 2018). Analyses using STRUCTURE ver. 2.3.4 (Pritchard et al. 2000), within the ipyrad-analysis toolkit module, were performed to investigate relationships and estimated number of populations (K) among individuals using unlinked SNPs including all 82 individuals and clustering set to 0.85. This dataset was selected to avoid bias due to too much missing data (1.7%). The range of K was set from two to 15, a burn-in of 9999 and MCMC of 9999 replicates, with ten replicates for each value of K. STRUCTURE HARVESTER (Earl and von Holdt 2012) was used to estimate the optimal value of K using delta K (Evanno et al. 2005). Principal component analyses (PCA) was further used for studying population structures and were

performed for all datasets within the ipyrad-analysis toolkit module, and 3D plots of the first three resulting principal components were generated using the plotly package in R (<https://github.com/plotly/plotly.R>). To assess genetic variability and compare genetic distances among populations and clusters from the STRUCTURE analyses (c85m82 dataset with 22 941 SNP markers), genetic variation between and within all populations was quantified using a standard AMOVA (analysis of molecular variance) under default settings using pairwise distances based on haplotypes in Arlequin ver. 3.5.2 (Excoffier and Lischer 2010). Three genetic groups were tested: 1) K = 3 + admixed genotypes (n = 4), 2) K = 8 (n = 8) and 3) populations (n = 21). The fixation index, a measure of deviation from the Hardy–Weinberg equilibrium (HWE) in total population (F_{IT}) and within sub-populations (F_{IS}), and of genetic differentiation among sub-populations (F_{ST}), was calculated. The significance of the fixation index was tested using a non-parametric permutation approach described in Excoffier et al. (1992). SplitsTree4 (Huson and Bryant 2006) was used in order to reconstruct possible network-like evolutionary relationships among populations. The analyses were performed on dataset c90m20 and c85m60 using nexus files converted from ipyrad u.snps files as input files. Default settings, implementing neighbor-net analysis with variance of ordinary least squares, were used. Missing data were treated as unknown. Bootstrapping to test for statistical branch support was conducted with 1000 replicates. For comparison with the ML approach, TETRAD within the ipyrad-analysis toolkit module was used to conduct coalescent phylogenetic analyses based on SVDquartets (Chifman and Kubatko 2014, Eaton 2014). The analyses were performed on dataset c85m60 and c90m20 with all possible quartets (1 749 060) and 100 bootstrap replicates.

Results

GBS sequencing

Illumina sequencing provided in total 485 993 257 reads with an average of 5 926 747 raw reads per sample with a standard deviation of 996 114. After quality filtering 474 750 846 reads remained to be used in the ipyrad assembly pipeline. The assemblies with a parameter series of differing minimum sample coverage and clustering thresholds are shown in Table 2. Changing the clustering parameter from 0.90 to 0.85 increased the percentage of missing data and the number of variable sites, while the number of retained loci and the number of variable sites decreased. Changing the minimum sample coverage from 82, representing the full set, to four resulted in an increase of missing data, number of loci and number of variable sites. For both clustering parameters, the average number of variable sites per locus increased when reducing the sample parameter to reach a maximum for 20 samples (m20). The drop in the average number of variable sites per locus for the lowest minimum number of sample coverage (m04) shows a lower gain of variable sites in proportion to the increase of loci retained (Table 2).

Genetic diversity and structure analysis

The Bayesian clustering analyses with STRUCTURE were performed with no prior population information to elucidate the origin of the populations and patterns of admixture between individuals. The delta K method revealed eight optimal clusters (K = 8) and additionally three, ten and 12 suboptimal clusters (Fig. 2). Grouping into K = 10 and K = 12 did not provide any further information while K = 3 was considered informative due to investigations for potential division into subspecies (Supporting information). All cluster models make it evident that the 14 *E. transbaicalensis* individuals from three populations (H10051, H10114 and H10378), as well as one individual from population H10142 are distinct from the rest. The remaining 67 *E. mutabilis* individuals in the K = 3 group are divided into two clusters, one including populations from northern Europe and far east Russian and the other populations from China, Kyrgyzstan and Tajikistan, with two intermediate admixed populations in Altai and Xinjang. The eight-cluster grouping K = 8 showed a more detailed pattern, where the populations from northern Europe and far east Russia formed two evident groups, while

Table 2. Output from ten GBS datasets assembled with ipyrad 0.9.19 pipeline (Eaton 2014, Eaton and Overcast 2020).

	c90m82	c90m60	c90m40	c90m20	c90m04	c85m82	c85m60	c85m40	c85m20	c85m04
Clustering threshold	0.90	0.90	0.90	0.90	0.90	0.85	0.85	0.85	0.85	0.85
Minimum sample coverage	82	60	40	20	4	82	60	40	20	4
No. of loci	11 971	60 933	93 886	139 740	291 559	9039	53 235	84 197	122 759	250 945
Concatenated length (bp)	1 243 951	5 725 736	8 598 496	12 814 565	26 486 031	941 625	4 997 886	7 687 933	11 208 157	22 692 584
Missing data	1.5%	12.6%	23.8%	39.2%	60.4%	1.7%	13.2%	24.9%	39.9%	60.4%
No. of variable sites (1)	28 952	160 964	256 300	395 150	693 240	22 941	152 500	254 973	385 653	667 967
Average no. of variable sites per locus (2)	2.42	2.64	2.73	2.83	2.38	2.54	2.86	3.03	3.14	2.66

[1] Parsimony-informative sites plus autapomorphies.
[2] Total no. of variable sites divided by the number of loci.

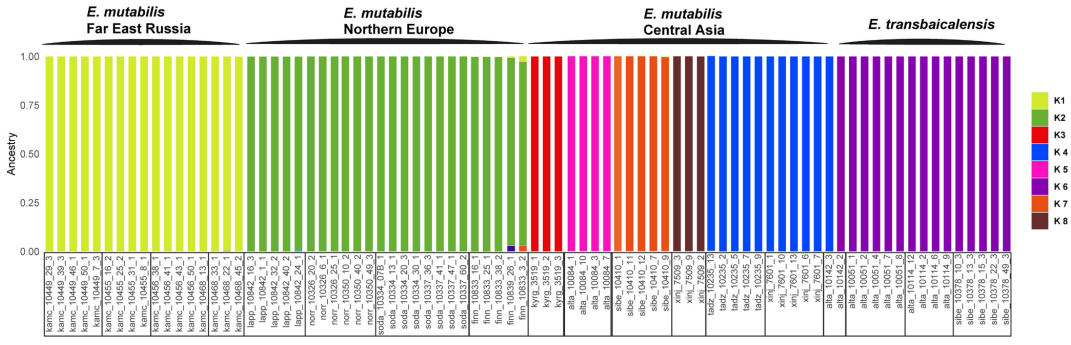

Figure 2. Population structure analysis of 18 populations of *E. mutabilis* and three populations of *E. transbaicalensis* for K = 8 using dataset c85m82 in STRUCTURE software (Pritchard et al. 2000).

the central Asian populations were distributed over six well-defined clusters without admixtures.

AMOVA analyses were performed in Arlequin using the c85m82 dataset, for both populations and population structure groups: K3 plus admixed genotypes (here after referred to as K3), K8, K10 and K12. The total molecular variance explained within the K3 group was 58.5% ($F_{ST} = 0.585$), and 20.0% ($F_{IS} = 0.493$) among sub clusters in K3 (Table 3). The corresponding molecular variance within and between clusters of K8, K10 and K12 groups was 69.0% ($F_{ST} = 0.680$) and 7.4% ($F_{IS} = 0.237$), respectively. This further suggests a redundancy of describing the genetic structure as more than eight clusters. The highest molecular variance can be found among the 21 original populations (77.7%, $F_{ST} = 0.777$).

Principal component analyses (PCA) showed clear and similar patterns of genetic structure across all datasets (Fig. 3). The results correspond to the STRUCTURE plots and the phylogenetic networks revealing the same genetic relationships with eight distinct clusters. The individuals within populations grouped closely together indicating a low genetic diversity within populations. The main axis of variation (PC1 41.4%) distinctly differentiated *E. transbaicalensis* from *E. mutabilis*. The second (PC2 11.5%) and third axis (PC3 8.5%) additionally revealed a pattern corresponding to the geographical distributions of the *E. mutabilis* populations.

ML phylogenetic analyses were conducted across all ten datasets resulting in consistently supported and distinct clades for the majority of the populations (Fig. 4, Supporting information). However, the different assembly settings showed two conflicting topologies for the northern European populations. Six datasets (c85m04, c85m40, c90m40, c90m20, c90m04 and c90m60) place Finnmark/Norrbotten populations as a sister group to Sodankylä and Lappland/Norrbotten populations, while four datasets (c85m20 c85m60,

Table 3. Results of analysis of molecular variance (AMOVA) performed in Arlequin ver. 3.5.2 (Excoffier and Lischer 2010) using pairwise distances based on haplotypes and dataset c85m82 including all 21 populations of *E. mutabilis* and *E. transbaicalensis*. Three genetic groups were tested: 1) K = 3 + admixed genotypes (n = 4), 2) K = 8 (n = 8) and 3) populations (n = 21). The fixation index is a measure of deviation from the Hardy–Weinberg equilibrium (HWE) in total population (F_{IT}) and within sub-populations (F_{IS}), and of genetic differentiation among sub-populations (F_{ST}). The significance of the fixation index is tested using a non-parametric permutation approach described in Excoffier et al. (1992).

Source of variation	Degrees of freedom	Variance components	Percentage (%) of molecular variance explained	Fixation index
K3 + admixed				
Among clusters	3	1074.2	58.5	$F_{ST} = 0.585$*
Among individuals within clusters	78	375.6	20.0	$F_{IS} = 0.493$*
Within individuals	82	386.0	21.0	$F_{IT} = 0.790$*
Total	163	1835.8	100	
K8				
Among clusters	7	1126.0	69.0	$F_{ST} = 0.690$*
Among individuals within clusters	74	120.2	7.4	$F_{IS} = 0.237$*
Within individuals	82	386.0	23.7	$F_{IT} = 0.763$*
Total	163	1632.2	100	
Populations				
Among clusters	20	1146.7	77.7	$F_{ST} = 0.777$*
Among individuals within clusters	61	0	0	
Within individuals	82	386.0	26.2	$F_{IT} = 0.739$*
Total	163	1476.4	100	

* $p \leq 0.01$, using 1023 permutations.

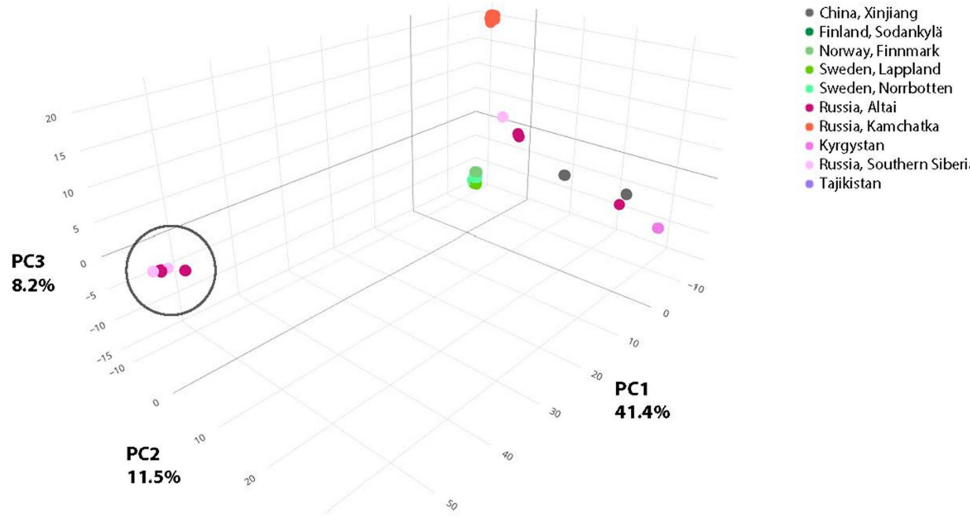

Figure 3. Genetic relationship of the 21 populations of *E. mutabilis* and *E. transbaicalensis* in the c85m82 dataset as revealed by a three dimensional principal component analysis (PCA) visualized with plotly (<https://github.com/plotly/plotly.R>). Populations are colored according to geographical origin with *E. transbaicalensis* encircled.

c85m82 and c90m82) place the Sodankylä population as a sister group to the Finnmark/Norrbotten and Lappland/Norrbotten populations. The topology of the TETRAD trees complied with the ML analyses except for the relationships within the monophyletic European group (Fig. 5). The two datasets (c85m60 and c90m20) place one of the Norrbotten populations (norr_10326) as a sister group to the populations from Sodankylä (soda_10337 and soda_10334) and the Lappland population (lapp_10842) as a sister group to the two Finnmark populations (finn_10839 and 10822) and the other population from Norrbotten (norr_10350). In addition, branch support showed low values within several populations.

The same pattern of genetic relationships as revealed with STRUCTURE and PCA analyses was obtained from the ML and TETRAD phylogenetic analyses with a clear differentiation between *E. transbaicalensis* forming a monophyletic clade separated from the monophyletic *E. mutabilis* clade (Fig. 4). The topology suggests that individuals belonging to a population mostly form monophyletic clades with the notable exception of individuals from population H10142. The *E. mutabilis* populations from Tajikistan, Kyrgyzstan and China formed a monophyletic clade and were resolved as a sister group to the rest of the populations with populations from Altai and far east Russia forming a grade. Northern Europe and southern Siberia populations are sister groups.

The overall result from SplitsTree correspond to the ML and TETRAD analyses and show a low degree of reticulate relationships. A close up on the populations from northern Europe show similar radiation patterns in the c85m60 and c90m20 datasets (Supporting information).

Discussion

Division of *E. mutabilis* and *E. transbaicalensis*

Species identification, delimitation and description are means to understand and describe biodiversity and are important for conservation planning, utilization and further biological research (Bickford et al. 2007, Heath et al. 2008). Multiple polyploidization events from the same progenitor taxa may lead to cryptic speciation forming distinct lineages that are not accompanied by clear morphological differentiation (Soltis et al. 2010). Recurrent formations of polyploid species from the same diploid parental taxa are considered to be more common than single events and may be an important source of increased genetic variation in polyploids depending on the contribution from diploid progenitors (Symonds et al. 2010, McAllister and Miller 2016, Welles and Ellstrand 2016). The case of *E. mutabilis* and *E. transbaicalensis* is a good example of a cryptic relationship. They are morphologically similar, but in this study both the PCA and STRUCTURE analysis show a clear distinction between the two taxa. No hybrids or introgression were detected by the phylogenetic analyses, however, the two species are referred to as belonging to the same introgressive gene pool (Agafonov 2004). The only population with individuals not grouping together was H10142 in the Altai Region in Central Asia with one individual placed in the *E. transbaicalensis* clade and the other in the *E. mutabilis* clade. A careful examination showed a faint morphological differentiation between the two individuals. One of them had more or less scabrous internal surface of the glumes with a

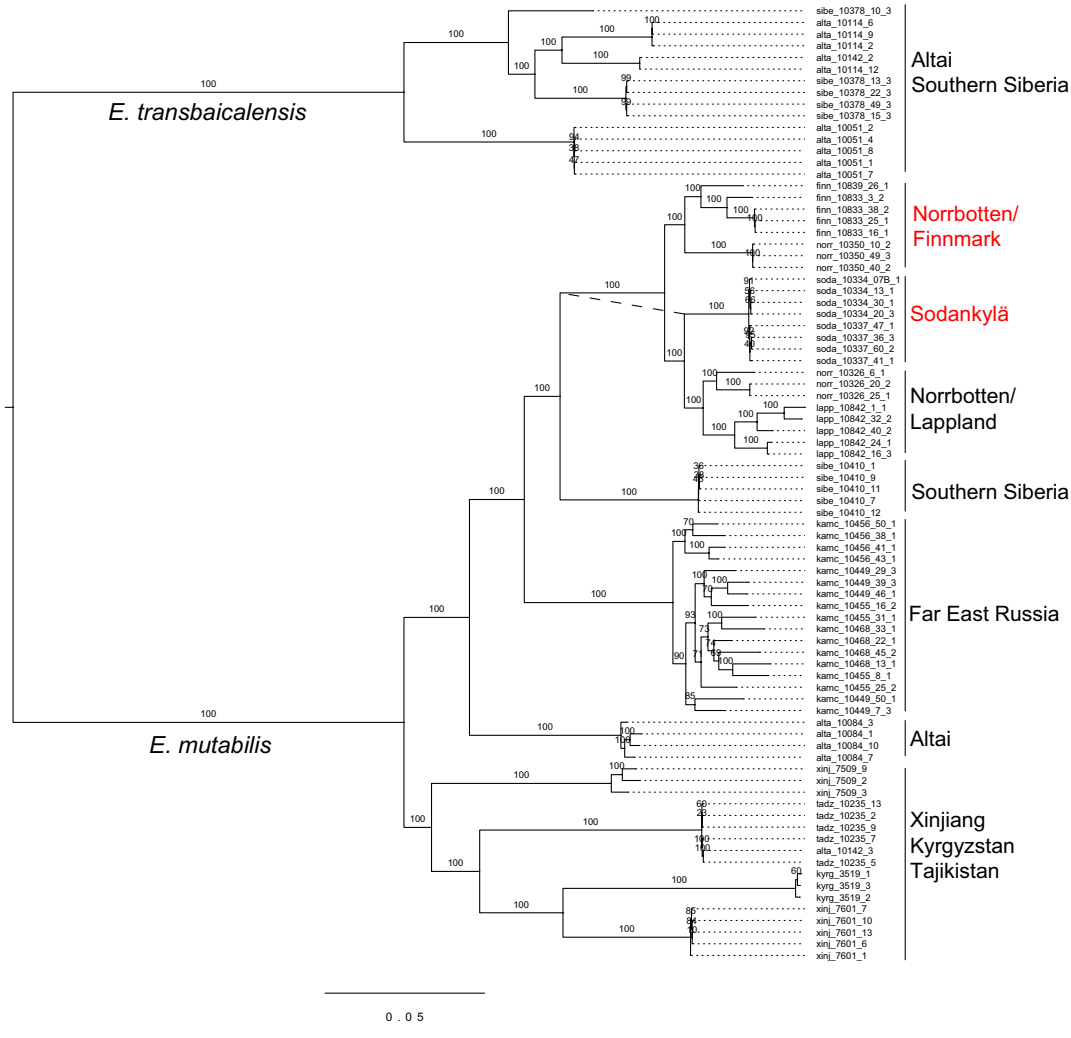

Figure 4. Phylogenetic tree of *E. mutabilis* with *E. transbaicalensis* as an outgroup using the c90m20 dataset and calculated with maximum likelihood from 152 500 SNP markers with RAxML ver. 8.0 (Stamatakis 2014) implementing the GTR + G substitution model with correction for ascertainment bias using the Lewis method. Bootstrap values are indicated above branches. The tree represents the topology of the majority of the datasets (c85m04, c85m40, c90m40, c90m20, c90m04 and c90m60), and the dashed line and red text indicating where the conflicting branches occur for the other datasets (c85m20 c85m60, c85m82 and c90m82).

tendency of hairiness, intermediate between the two species, whereas the other was glabrous, typical for *E. transbaicalensis*. This morphological ambiguity might be an indication of introgressive hybridization but none of them show genetic admixture in the STRUCTURE analysis.

It is evident that the two taxa should be considered as distinct species based on the data from this study together with the morphological and biosystematic analyses of Agafonov (2004). However, it is not clear if *E. mutabilis* and *E. transbaicalensis* are derived from a split in a single lineage or have originated from recurring hybridization events.

Elymus with its high diversity and multigenome constitution is an appropriate model genus to investigate formation and diversification of polyploid lineages (Kellogg 2016). The population structures suggest that *E. fibrosus* originated from a single event while *E. trachycaulus*, *E. caninus* and *E. alaskanus* have multiple independent origins, even though ecology and other environmental factors cannot be excluded (Yan

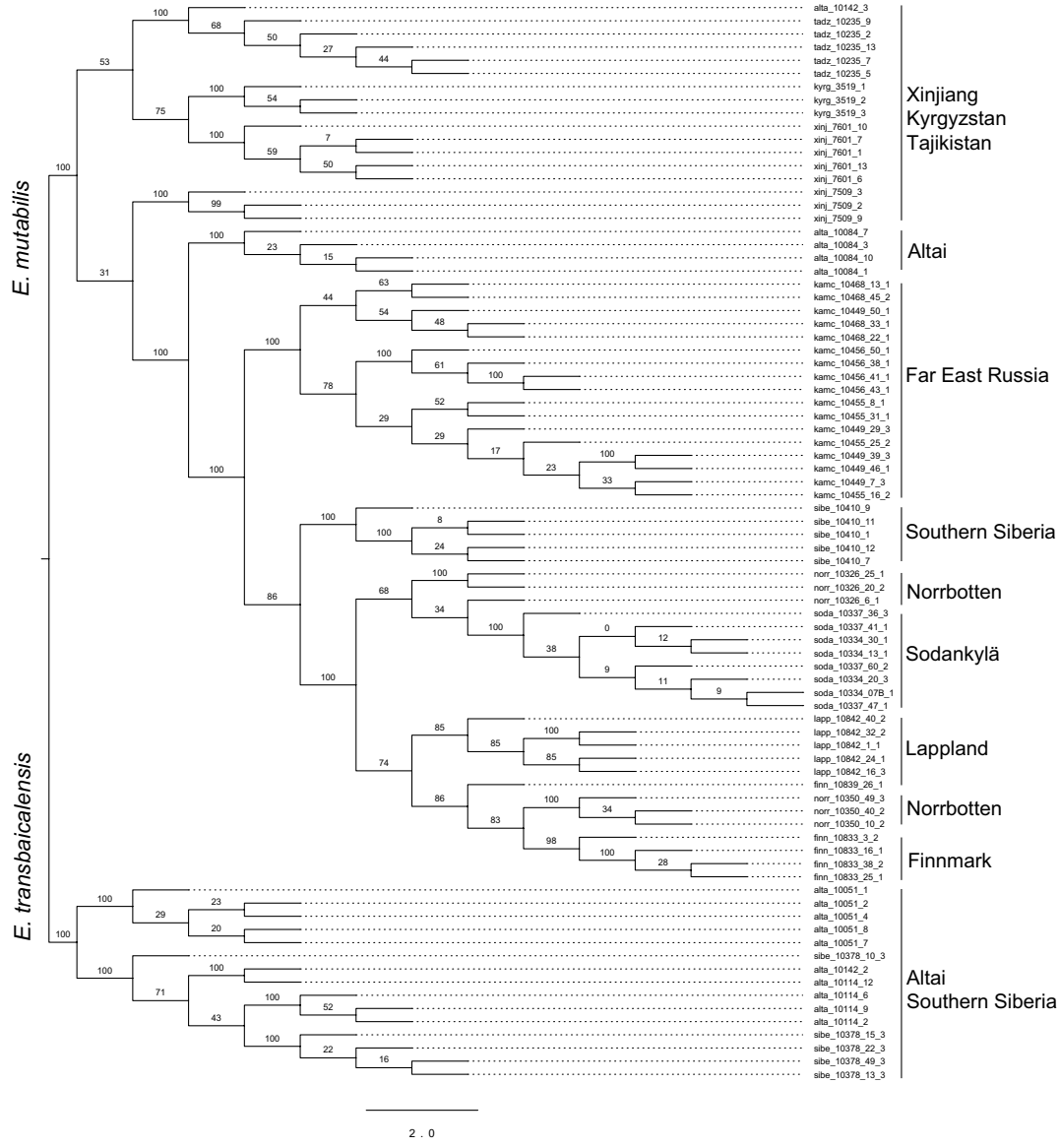

Figure 5. A coalescent phylogenetic tree of *E. mutabilis* with *E. transbaicalensis* as an outgroup. The analysis was performed using TETRAD (Chifman and Kubatko 2014, Eaton 2014) on the c90m20 dataset with all possible quartets (1 749 060) and 100 bootstrap replicates. Bootstrap values are indicated above branches.

and Sun 2012, Wu et al. 2016a). A polymorphic H genome in *E. trachycaulus* indicates a polyphyletic origin either from multiple independent formation events or introgression by subsequent hybridization (Zuo et al. 2015). The genetic variation in the *E. mutabilis* clade, as seen in the PCA analysis, follows the geographical distribution of the species. This suggests a diverse and dispersed, but still coherent species that has probably spread by a step-by-step expansion from a single evolutionary lineage rather than multiple hybridization events. However, in a study by Agafonov et al. (2019),

accessions of *E. mutabilis* fell into different clades with different H genomes, indicating multiple origins. They studied three populations from Altai (H_1), Ural and China (H_2) and have likely captured variation not seen in this study.

Further division within *E. mutabilis*

Divergence among *E. mutabilis* populations could be driven by the joint effect of geographic distance and climatic heterogeneity. The STRUCTURE K=3 data suggests a potential split in at least a southern taxon, including the populations from Tajikistan, Kyrgyzstan and China, a northern taxon including populations from southern Siberia, far east Russia and northern Europe, and a potential hybrid zone or a center of origin in Altai and north-western China (Supporting information). This differentiation could potentially represent the subdivision of *E. mutabilis* ssp. *mutabilis* (Drobow) Tzvelev and *E. mutabilis* ssp. *praecaespitosus* (Nevski) Tzvelev. According to Flora of China (eFloras 2008), *E. mutabilis* var. *praecaespitosus* is without rhizomes and spikelets are usually glaucous or purplish glaucous while other varieties are more or less rhizomatous with green or purple spikelets. Examination of available herbarium material from the entire distribution area shows a tendency to a geographical differentiation (in e.g. stature, spike and awn length) but due to the large variation and morphological overlap, for most areas it does not seem justified with a subspecific taxonomic recognition. However, the material from central Asia (China, Tajikistan and Kyrgyzstan) has longer and denser spikes and is conspicuously more hairy and scabrid on the lemmas than material from other areas, but for a formal taxonomic recognition further material should be studied.

There is a genetic similarity between populations from northern Europe and far east Russia seen in the PCA plot. A distinct split is visible first with the third axes of variation, which indicates a closer relationship between the peripheral populations compared to the populations in Tajikistan, Kyrgyzstan and China. One explanation could be that the environment is more important than geography as an isolation factor. The northern populations have probably experienced range reductions during the last glacial period and cold-tolerant populations have followed the ice as it retreated as pioneers recolonized northern Eurasia. The analyses suggest that *E. mutabilis* most likely originated in central Asia and then spread northwards and southwards. Even though there is a close relationship, the molecular analyses in this study show that the populations in northern Europe are distinct from populations in Asia, including far east Russia. Morphological and phenological observations on wild-collected material and plants in cultivation support this differentiation. When cultivating the plants in southern Sweden for the present study, individuals from northern Europe showed a more compact growth habit with shorter internodes and were less keen to flower compared to the Asian populations. This could be an extreme on a gradient of variation or evidence for a distinct group. This observed differentiation makes it unlikely that more recent long-distance dispersal have shaped the species genetic structure. Populations from far east Russia are more morphologically similar to populations from central Asia. Nevertheless, this shows the importance of conserving peripheral populations.

The topological disparity in the ML and TETRAD analyses and the radiation pattern in the Splitstree networks for the northern European populations suggest a founding followed by a rapid dispersal and incomplete lineage sorting making it difficult to draw phylogenetic and biogeographical conclusions. The difference in ML tree topology correlates with the number of variable sites in the data. Four out of the five assemblies with the most variable sites place Finnmark/Norrbotten populations as a sister group to Sodankylä and Lappland/Norrbotten populations. Allowing sites with missing data may increase the proportion of parsimony-informative sites and in some cases also branch support for the phylogenetic hypothesis (Huang and Knowles 2016, Crotti et al. 2019, Pérez-Escobar et al. 2020). The results show that even large datasets can have difficulties entangle relationships from rapid radiation.

Introgression between *E. mutabilis* and *E. transbaicalensis*

Introgressive hybridization between species is an additional factor affecting genetic diversity and population structures. More or less sterile interspecific hybrids are found where *E. mutabilis* and *E. caninus* grow neighboring or at the same site (Meledris 1955, Diaz et al. 1999a). Diaz et al. (1999b) found that populations of *E. caninus* have a higher genetic diversity when growing sympatrically with *E. mutabilis* than growing alone or together with *E. fibrosus*, which suggests some degree of interspecific gene flow. Wu et al. (2016b) used microsatellite markers to investigate the amount of gene flow between *E. mutabilis*, *E. alaskanus* and *E. fibrosus* in northern Europe. Their results show that gene flow is higher between sympatrically grown species than between spatially separated populations of the same species. They observed asymmetrical rates of gene flow among the studied species, and the highest was from *E. fibrosus* to *E. mutabilis*. They also found a weak correlation between genetic distance and geographic distance in *E. mutabilis*. In the present study, the STRUCTURE plot suggests no gene flow between *E. mutabilis* and *E. transbaicalensis*. This is in line with the findings of Agafonov (2004) of reproductive barriers, even though he suggests that the two taxa belong to the same introgressive gene pool where hybrid formation is possible. High seed fertility in hybrids have been confirmed in crosses between *E. mutabilis* s.s., *E. charkeviczii* and *E. subfibrosus* (Agafonov et al. 2005). The lack of hybrids and introgression is another argument for separating both taxa.

Species variation

Genetic variation and population structures are affected by several abiotic and biotic factors such as ecological forces, population sizes, the spatial distribution of populations,

breeding systems and dispersal (Loveless and Hamrick 1984). Spatially well-separated populations are likely to show at least some variation due to differences in fixed alleles and/or allele frequencies caused by mutations, natural selection and/or genetic drift (Diaz 1999). Members of Poaceae show great variability in genetic composition and population structure and the main influencing factors are in general the breeding system and the geographical distribution range (Godt and Hamrick 1998). *Elymus trachycaulus*, *E. fibrosus*, *E. alaskanus* and *E. caninus* are similar to *E. mutabilis* and *E. transbaicalensis* in their habit (perennial and caespitose), ecology (grow in forest clearings, riverbanks and meadows), reproduction system (predominantly self-pollinating) and genome composition (allotetraploid with a StStHH combination) (Sun et al. 1998a, b, Sun and Salomon 2003). However, molecular studies indicate that the five species significantly differ in genetic variability and population structure, which suggests that other factors play an important role in shaping population structures and species variation. Altogether, *E. trachycaulus* and *E. caninus* have high genetic variation both within and between populations, *E. alaskanus* has high genetic variation among populations but not within populations, and *E. fibrosus* has low genetic variation both among and within populations. Wu et al. (2016b) also found that *E. mutabilis* was less variable than, in ascending order, *E. fibrosus*, *E. alaskanus* and *E. caninus*. In the present study, a higher genetic variation in northern Europe populations is shown as expected from high throughput genome-wide markers compared to older marker methods. Isozymes and allozymes used by Diaz et al. (1999a) are less sensitive to detect genetic variation than modern molecular techniques and could be the reason why Diaz et al. did not find variation in the Nordic *E. mutabilis* populations. Still, the population sizes sampled in this study are too small to gain reliable unbiased population statistics and further data is needed to accurately assess the population genetics.

Geographical variation in diversity in *E. mutabilis*

Studies of geographical variation in population genetic structure in the temperate zone of the Northern Hemisphere show in general a decline in within-population diversity and an increase in the differentiation among populations from the center of the species distribution range towards the periphery (Eckert et al. 2008). The 'abundant center' model explains this as a result of a decline in population size and increasing in isolation towards the range limit and both historical and contemporary ongoing evolutionary factors are potential causes (Sagarin and Gaines 2002, Vucetich and Waite 2003, Samis and Eckert 2007). Even though the sampled number of individuals per population is low, the AMOVA analyses show a higher variation between populations than within, indicating a substantial inbreeding behavior and a low gene flow between populations within *E. mutabilis*. The neighbor-net analysis further shows a low degree of evolutionary reticulation within *E. mutabilis*. The main genetic variation of *E. mutabilis* is found in central Asia, which is in accordance with the average pattern of distribution of genetic diversity. It is necessary to study intraspecific variation in order to assess conservation values and predict the consequences of habitat losses due to global climate change and other reasons (Eckert et al. 2008, Pauls et al. 2013). In this study, peripheral populations at the edge of the species range differentiate from the central populations, which should be considered in future conservation programs. However, Volis et al. (2016) concluded that the extent of variation in molecular markers cannot predict plant performance in novel environments while extent of variation in quantitative traits can. Even though there is less diversity in the peripheral northern populations, there is most likely still diversity and adaptive potential not reflected in the diverse central populations.

Conclusion

Genotyping-by-sequencing is a well-suited method for the purpose of both phylogenetic and population studies. The current study provides molecular evidence for considering *E. mutabilis* and *E. transbaicalensis* as two distinct species. Additionally, a clear phylogeographic structure with a pattern of variation corresponding to the geographical distribution of *E. mutabilis* with the main genetic diversity in central Asia is evident. Both geography and climate are potential drivers for the divergence of the species. The genetic variation of populations in northern Europe are distinct from that of Asian populations, but shows a high similarity to populations from far east Russia and southern Siberia. Thus, the populations in the peripheral regions most likely originated in Altai and southern Siberia from where it spread to other areas. A phylogenetic radiation pattern among populations in northern Europe indicates a founding followed by rapid dispersal. These findings give further understanding of the complexity of *Elymus* and pose new questions of polyploid formation and diversification. Continued taxonomic work in *Elymus* is important to explore the large variation over wide distribution areas.

Acknowledgements – We want to thank IPK Gatersleben and the Experimental Taxonomy Group for help with the analyses. We specifically thank Dr. Frank Blattner and Dr. Mulatu Geleta Dida for their valuable comments. This work was supported by the Department of Plant Breeding SLU and the Udda server.
Funding – The research was funded by The Royal Physiographic Society of Lund (38937 för Nilsson-Ehle-donationerna – 2017-09-28) and the Swedish University of Agricultural Sciences, Department of Plant Breeding. Genotyping-by-sequencing was financed by basic funds from IPK Gatersleben.

Author contributions

Jonatan Leo: Conceptualization (equal); Data curation (equal); Formal analysis (equal); Funding acquisition (equal); Investigation (equal); Methodology (equal); Project administration (equal); Visualization (equal); Writing – original draft (lead). **Therése Bengtsson**: Conceptualization (equal); Data curation (equal); Formal analysis (equal); Funding acquisition

(equal); Investigation (equal); Methodology (equal); Project administration (equal); Supervision (equal); Visualization (equal); Writing – original draft (supporting); Writing – review and editing (equal). **Anders S. Carlsson**: Conceptualization (equal); Formal analysis (supporting); Funding acquisition (equal); Investigation (supporting); Methodology (equal); Project administration (supporting); Supervision (equal); Writing – original draft (supporting); Writing – review and editing (equal). **Jonathan Brassac**: Conceptualization (supporting); Data curation (equal); Formal analysis (supporting); Investigation (supporting); Methodology (equal); Supervision (supporting); Visualization (supporting); Writing – original draft (supporting); Writing – review and editing (equal). **Roland von Bothmer**: Conceptualization (equal); Formal analysis (supporting); Funding acquisition (equal); Investigation (equal); Methodology (supporting); Project administration (equal); Supervision (equal); Writing – original draft (supporting); Writing – review and editing (equal).

Data availability statement

Data are available from the NCBI Short Read Archive (SRA) under BioProject PRJNA770613.

Supporting information

The supporting information associated with this article is available from the online version.

References

Agafonov, A. V. 2004. Intraspecific structure and reproductive relationships between *Elymus mutabilis* and *E. transbaicalensis* (Poaceae) in southern Siberia from the viewpoint of taxonomical genetics. – Russ. J. Genet. 40: 1229–1238.

Agafonov, A. V. and Salomon, B. 2002. Genepools among SH genome *Elymus* species in boreal Eurasia. – In: Hernándes, P. et al. (eds), Triticeae IV. Consejeria de Agricultura y Pesca, pp. 37–41.

Agafonov, A. V. et al. 1998. Biosystematic relationships between *Elymus komarovii* (Nevski) Tzvel. and related species (Jaradat, A. A. Ed.). – Science Publishers Inc.

Agafonov, A. V. et al. 2005. A biosystematic study of *Elymus charkeviczii*, *E. mutabilis* and *E. subfibrosus* from the Peninsula Kamchatka. – Czech J. Genet. Plant Breed. 41: 79–85.

Agafonov, A. V. et al. 2019. Genome constitution and differentiation of subgenomes in Siberian and far eastern endemic species of the genus *Elymus* (Poaceae) according to the sequencing of the nuclear gene waxy. – Vavilovskii Zhurnal Genet. Selektsii 23: 817–826.

Åhman, I. and Bengtsson, T. 2019. Introgression of resistance to *Rhopalosiphum padi* L. from wild barley into cultivated barley facilitated by doubled haploid and molecular marker techniques. – Theor. Appl. Genet. 132: 1397–1408.

Anderson, B. M. et al. 2017. Genotyping-by-sequencing in a species complex of Australian hummock grasses (*Triodia*): methodological insights and phylogenetic resolution. – PLoS One 12: e0171053.

Bickford, D. et al. 2007. Cryptic species as a window on diversity and conservation. – Trends Ecol. Evol. 22: 148–155.

Chifman, J. and Kubatko, L. 2014. Quartet inference from SNP data under the coalescent model. – Bioinformatics 30: 3317–3324.

Chung, Y. S. et al. 2017. Genotyping-by-sequencing: a promising tool for plant genetics research and breeding. – Hortic. Environ. Biotechnol. 58: 425–431.

Crotti, M. et al. 2019. Causes and analytical impacts of missing data in RADseq phylogenetics: insights from an African frog (*Afrixalus*). – Zool. Scr. 48: 157–167.

Diaz, O. 1999. Genetic diversity in *Elymus* species (Triticeae) with emphasis on the Nordic region. – PhD thesis, Dept of Plant Breed. Res. Svalöv, Sweden.

Diaz, O. et al. 1998. Description of isozyme polymorphisms in *Elymus* species using starch gel electrophoresis (Jaradat, A. A., ed.). – Science Publishers Inc., pp. 199–208.

Diaz, O. et al. 1999a. Absence of genetic variation in Scandinavian populations of *Elymus mutabilis* (Drob.) Tzvel. (Poaceae). – Hereditas 131: 83–86.

Diaz, O. et al. 1999b. Genetic diversity and structure in populations of *Elymus caninus* (L.) L. (Poaceae). – Hereditas 131: 63–74.

Earl, D. A. and von Holdt, B. M. 2012. Structure Harvester: a website and program for visualizing STRUCTURE output and implementing the Evanno method. – Conserv. Genet. Resour. 4: 359–361.

Eaton, D. A. R. 2014. PyRAD: assembly of de novo RADseq loci for phylogenetic analyses. – Bioinformatics 30: 1844–1849.

Eaton, D. A. R. and Overcast, I. 2020. ipyrad: interactive assembly and analysis of RADseq datasets. – Bioinformatics 36: 2592–2594.

Eckert, C. G. et al. 2008. Genetic variation across species' geographical ranges: the central–marginal hypothesis and beyond. – Mol. Ecol. 17: 1170–1188.

eFloras 2008. Flora of China. – Miss. Bot. Gard., MO Harvard Univ. Herb.

Elshire, R. J. et al. 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. – PLoS One 6: e19379.

Emtseva, M. and Agafonov, A. 2018. Identification of artificial interspecific hybrids in the genus *Elymus* (Poaceae) by using ISSRmarkers. – BIO Web Conf. 11 00011: 1–4.

Evanno, G. et al. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. – Mol. Ecol. 14: 2611–2620.

Excoffier, L. and Lischer, H. E. L. 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. – Mol. Ecol. Resour. 10: 564–567.

Excoffier, L. et al. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. – Genetics 131: 479–491.

Geibel, M. and Hohlfeld, B. 2003. *Malus* germplasm from Asia and its evaluation at the German fruit genebank. – In: Forsline, P. L. et al. (eds), Plant genetic resources: the fabric of horticulture's future. Acta Horticult. 63: 275–282.

Godt, M. J. W. and Hamrick, J. L. 1998. Allozyme diversity in the grasses. – In: Cheplick, G. P. (ed.), Population biology if grasses. Cambridge Univ. Press, pp. 11–29.

Harvey, M. G. et al. 2015. Similarity thresholds used in DNA sequence assembly from short reads can reduce the comparability of population histories across species. – PeerJ 3: e895.

Heath, T. A. et al. 2008. Taxon sampling and the accuracy of phylogenetic analyses. – J. Syst. Evol. 46: 239–257.

Huang, H. and Knowles, L. L. 2016. Unforeseen consequences of excluding missing data from next-generation sequences: simulation study of RAD sequences. – Syst. Biol. 65: 357–365.

Hultén, E. 1971. Atlas över växternas utbredning i Norden- 2:a uppl. – Generalstabens Litografiska Anstalts Förlag.
Huson, D. H. and Bryant, D. 2006. Application of phylogenetic networks in evolutionary studies. – Mol. Biol. Evol. 23: 254–267.
Kellogg, E. A. 2016. Has the connection between polyploidy and diversification actually been tested? – Curr. Opin. Plant Biol. 30: 25–32.
Li, P. C. et al. 2018. Genetic diversity of northern wheatgrass (*Elymus lanceolatus* ssp *lanceolatus*) as revealed by genotyping-by-sequencing. – Diversity 10: 14.
Löve, A. 1984. Conspectus of the Triticeae. – Feddes Repert. 95: 425–521.
Loveless, M. D. and Hamrick, J. L. 1984. Ecological determinants of genetic structure in plant populations. – Annu. Rev. Ecol. Syst. 15: 65–95.
Lu, F. et al. 2013. Switchgrass genomic diversity, ploidy and evolution: novel insights from a network-based SNP discovery protocol. – PLoS Genet. 9: e1003215.
Martin, M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. – EMBnet J. 17: 10–12.
McAllister, C. A. and Miller, A. J. 2016. Single nucleotide polymorphism discovery via genotyping by sequencing to assess population genetic structure and recurrent polyploidization in *Andropogon gerardii*. – Am. J. Bot. 103: 1314–1325.
Meledris, A. 1955. Species problems in recent Scandinavian works on grasses. – In: Lousley, J. E. (ed.), Species studies in the British flora. Botanical Society of the British Isles, pp. 140–159.
Pauls, S. U. et al. 2013. The impact of global climate change on genetic diversity within populations and species. – Mol. Ecol. 22: 925–946.
Pérez-Escobar, O. A. et al. 2020. Resolving relationships in an exceedingly young Neotropical orchid lineage using genotyping-by-sequencing data. – Mol. Phylogenet. Evol. 144: 106672.
Peschkova, G. A. 1990. *Elymus*, L. – In: Malyschev, L. I. and Peschkova, G. A. (eds), Flora of Siberia, vol. 2, Poaceae (Gramineae). Scientific Publishers, Inc., pp. 361.
POWO 2019. *Elymus mutabilis* (Drobow) Tzvelev. – Plants of the World Online. – Facilitated by the Royal Botanic Gardens, Kew. <www.plantsoftheworldonline.org/>.
Pritchard, J. K. et al. 2000. Inference of population structure using multilocus genotype data. – Genetics 155: 945–959.
Rambaut, A. 2018. Figtree v1.4.4, a graphical viewer of phylogenetic trees. – <http://tree.bio.ed.ac.uk/software/figtree/>.
Sagarin, R. D. and Gaines, S. D. 2002. The 'abundant centre' distribution: to what extent is it a biogeographical rule? – Ecol. Lett. 5: 137–147.
Salomon, B. et al. 1988. Notes on perennial Triticeae – species in northern Pakistan. – Bot. Jahrbücher Syst. Pflanzengesch. Pflanzengeogr. 110: 7–15.
Samis, K. E. and Eckert, C. R. G. 2007. Testing the abundant center model using range-wide demographic surveys of two coastal dune plants. – Ecology 88: 1747–1758.
Shabanova (Kobozeva), E. et al. 2020. Assessment of the variability of *Elymus caninus* (Poaceae) and closely related taxa from Russia and Kazakhstan using ISSR markers. – In: Banaev, E. V. et al. (eds), Int. Conf. Plant Divers. status, trends, Conserv. Concept, 24. EDP Science, pp. 1–5.
Soltis, D. E. et al. 2010. What we still don't know about polyploidy. – Taxon 59: 1387–1403.
Stamatakis, A. 2014. RAxML ver. 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. – Bioinformatics 30: 1312–1313.
Sun, G. and Salomon, B. 2003. Microsatellite variability and heterozygote deficiency in the arctic-alpine Alaskan wheatgrass (*Elymus alaskanus*) complex. – Genome 46: 729–737.
Sun, G. and Salomon, B. 2009. Molecular evolution and origin of tetraploid *Elymus* species. – Breed. Sci. 59: 487–491.
Sun, G. L. et al. 1998a. Microsatellite variation and its comparison with allozyme and RAPD variation in *Elymus fibrosus* (Schrenk) Tzvel. (Poaceae). – Hereditas 129: 275–282.
Sun, G. L. et al. 1998b. Characterization of microsatellite loci from *Elymus alaskanus* and length polymorphism in several *Elymus* species (Triticeae: Poaceae). – Genome 41: 455–463.
Symonds, V. V et al. 2010. Dynamics of polyploid formation in *Tragopogon* (Asteraceae): recurrent formation, gene flow and population structure. – Evolution 64: 1984–2003.
Tzvelev, N. N. 1973. A review of the Tribe Triticeae Dum. from the family Poaceae of the Flora of the Soviet Union. – Nov. Sist. Sosud. Rast. 10: 19–59.
Tzvelev, N. N. 1976. Tribe 3: Triticeae Dum. – In: Fedorov, A. A. (ed.), Poaceae U.R.S.S. Nacka Publishing House, pp. 146–298.
Volis, S. et al. 2016. The conservation value of peripheral populations and a relationship between quantitative trait and molecular variation. – Evol. Biol. 43: 26–36.
von Bothmer, R. and Salomon, B. 1994. Triticeae: a tribe for food, feed and fun. – In: Wang, R. R. C. et al. (eds), Proc. 2nd Int. Triticeae Symp. Utah State Univ. Press, pp. 1–12.
Vucetich, J. A. and Waite, T. A. 2003. Spatial patterns of demography and genetic processes across the species' range: null hypotheses for landscape conservation genetics. – Conserv. Genet. 4: 639–645.
Wagner, N. D. et al. 2020. Phylogenomic relationships and evolution of polyploid *Salix* species revealed by RAD sequencing data. – Front. Plant Sci. 11: 1077.
Welles, S. R. and Ellstrand, N. C. 2016. Genetic structure reveals a history of multiple independent origins followed by admixture in the allopolyploid weed *Salsola ryanii*. – Evol. Appl. 9: 871–878.
Wendler, N. et al. 2014. Unlocking the secondary gene-pool of barley with next-generation sequencing. – Plant Biotechnol. J. 12: 1122–1131.
Wu, D. C. et al. 2016a. Origin and evolution of allopolyploid wheatgrass *Elymus fibrosus* (Schrenk) Tzvelev (Poaceae: Triticeae) reveals the effect of its origination on genetic diversity. – PLoS One 11: e0167795.
Wu, P. P. et al. 2016b. Comparison of gene flow among species that occur within the same geographic locations versus gene flow among populations within species reveals introgression among several *Elymus* species. – J. Syst. Evol. 54: 152–161.
Xiong, H. Z. et al. 2016. Genetic diversity and population structure of cowpea (*Vigna unguiculata* L. Walp). – PLoS One 11: e0160941.
Yan, C. and Sun, G. L. 2012. Multiple origins of allopolyploid wheatgrass *Elymus caninus* revealed by RPB2, PepC and TrnD/T genes. – Mol. Phylogenet. Evol. 64: 441–451.
Zuo, H. W. et al. 2015. Origin and reticulate evolutionary process of wheatgrass *Elymus trachycaulus* (Triticeae: Poaceae). – PLoS One 10: e0125417.

Acta Universitatis agriculturae Sueciae
Doctoral Thesis No. 2022:68

Elymus (wild-rye) constitutes wild relatives to some of the most important cereal crops and forages. The genus is a genetic resource and contains many valuable traits that could be used in future crop breeding. In order to explore, exploit and conserve the genetic diversity of *Elymus,* basic knowledge of the systematics of the genus is needed. This thesis investigates the phylogeny of *Elymus s.s.* on different taxonomical levels, and gives insights into the evolutionary history of the genus.

Jonatan Leo received his graduate education at the Department of Plant Breeding, Swedish University of Agricultural Sciences (SLU), Alnarp Sweden. He received his MSc, in Biology from Lund University, Lund, and his degree of Master of Science in Horticulture from SLU, Alnarp.

Acta Universitatis agriculturae Sueciae presents doctoral theses from the Swedish University of Agricultural Sciences (SLU).

SLU generates knowledge for the sustainable use of biological natural resources. Research, education, extension, as well as environmental monitoring and assessment are used to achieve this goal.

ISSN 1652-6880
ISBN (print version) 978-91-8046-012-5
ISBN (electronic version) 978-91-8046-013-2