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

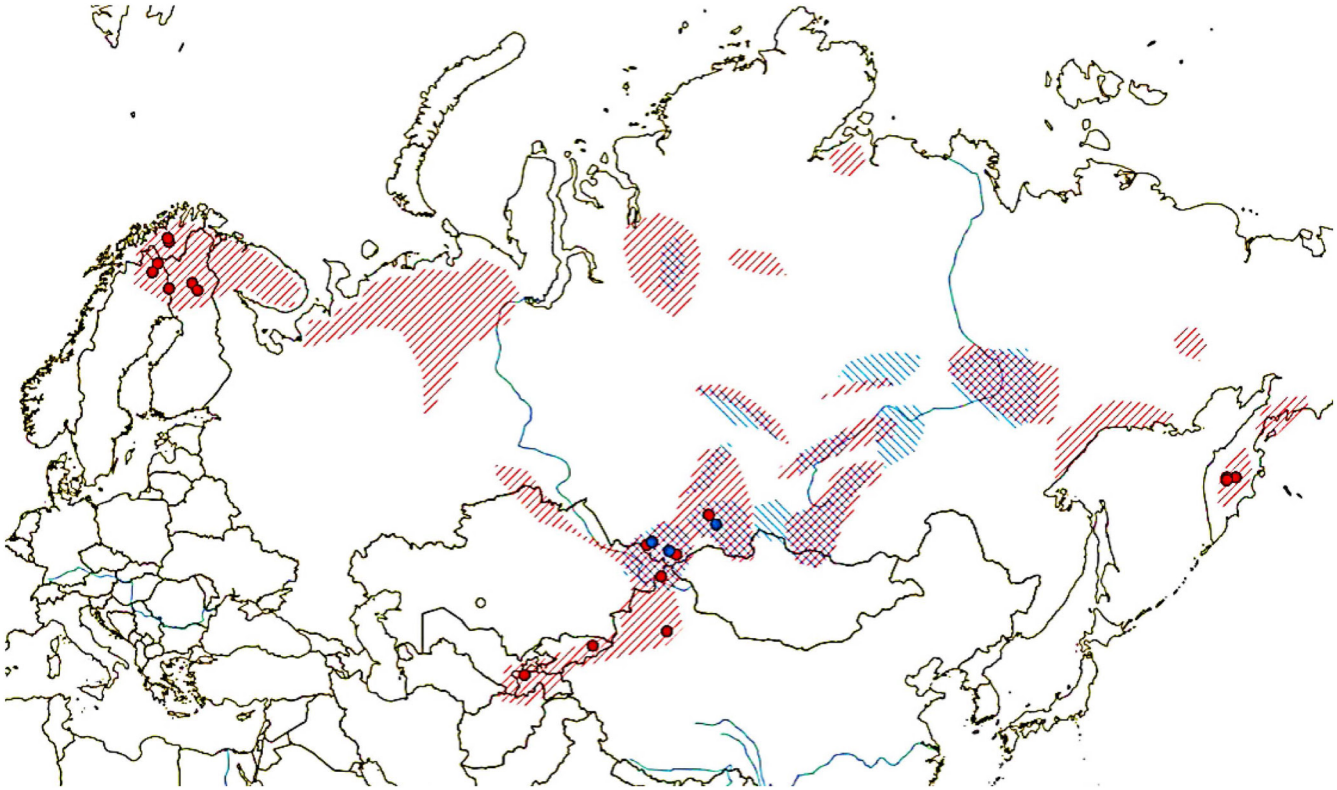

Figure 1. Distribution (lines) and sampled populations (dots) of *Elymus mutabilis* (red) and *E. transbaicalensis* (blue). Sympatric distribution in checkered pattern. Distrubution patterns are combined and modified from Peschkova (1990), Hultén (1971) and Tzvelev (1976).

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.

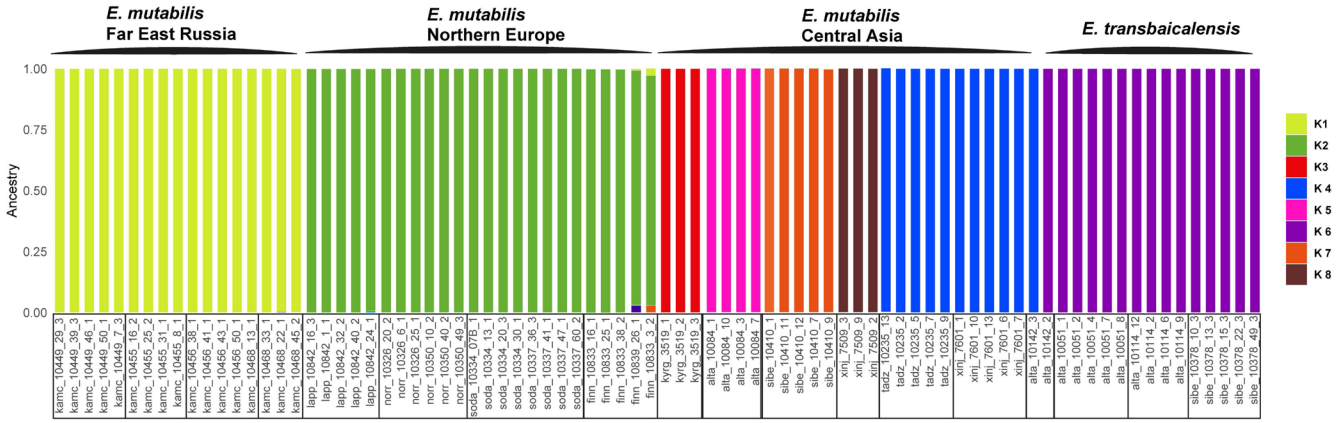

Figure 2. Population structure analysis of 18 populations of *E. mutabilis* and three populations of *E. transbaicalensis* for K = 8 using dataset c85m82 in STRUCTURE software (Pritchard et al. 2000).

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.

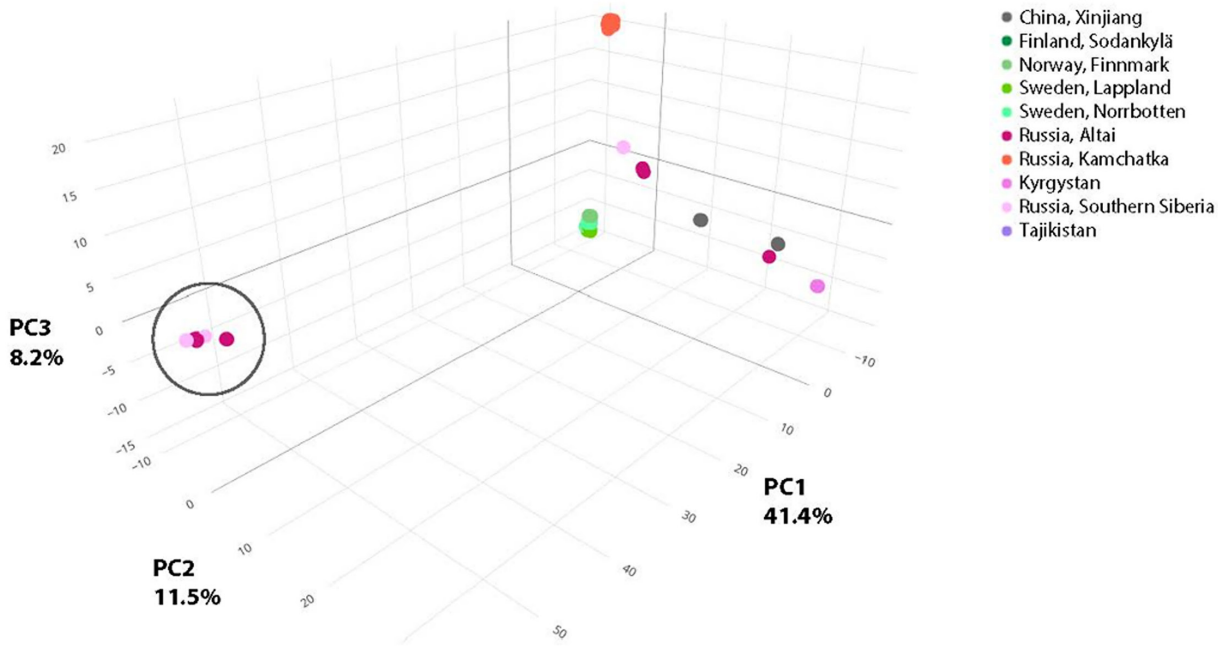

Figure 3. Genetic relationship of the 21 populations of *E. mutabilis* and *E. transbaicalensis* in the c85m82 dataset as revealed by a three dimensional principal component analysis (PCA) visualized with plotly (<https://github.com/plotly/plotly.R>). Populations are colored according to geographical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

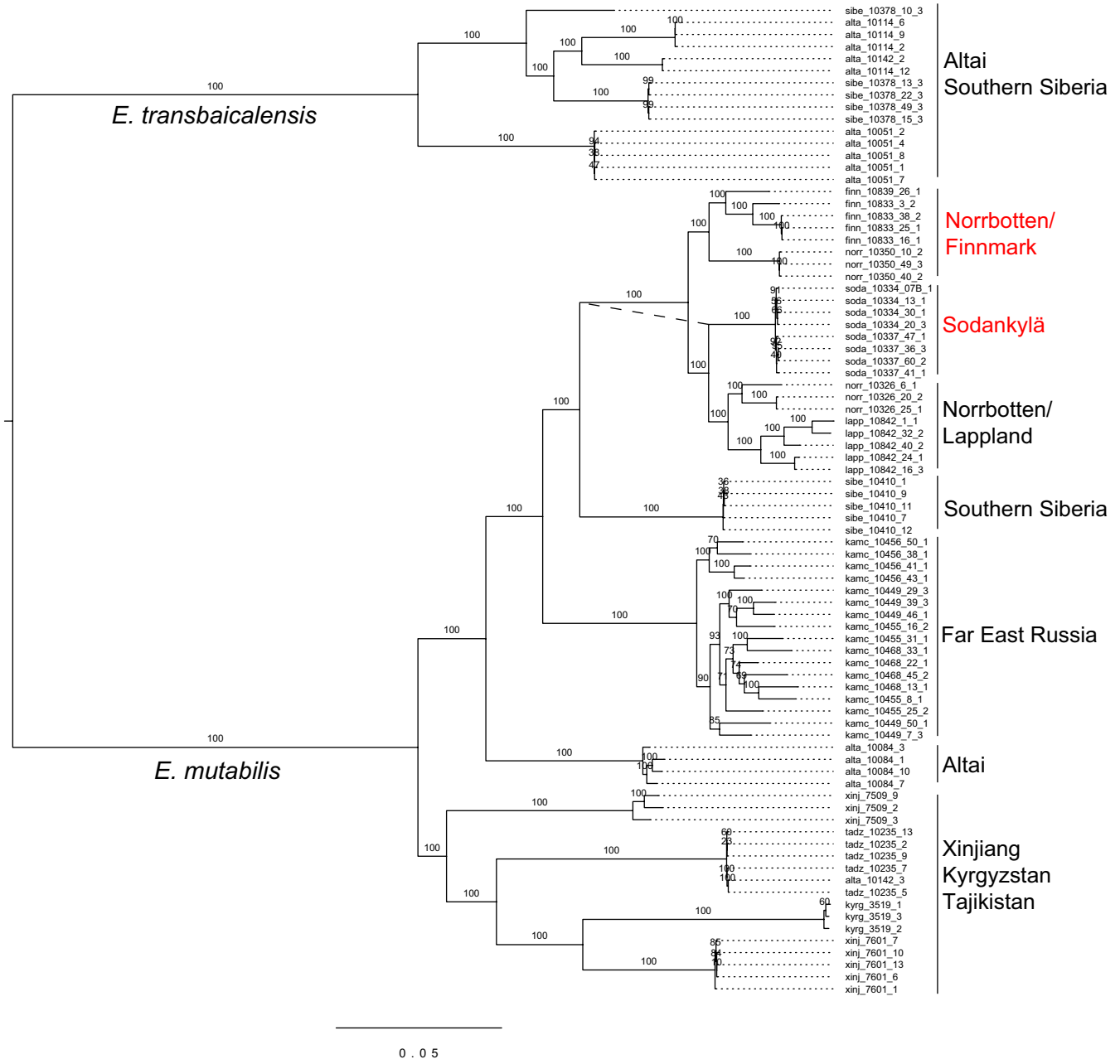

Figure 4. Phylogenetic tree of *E. mutabilis* with *E. transbaicalensis* as an outgroup using the c90m20 dataset and calculated with maximum likelihood from 152 500 SNP markers with RAxML ver. 8.0 (Stamatakis 2014) implementing the GTR + G substitution model with correction for ascertainment bias using 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

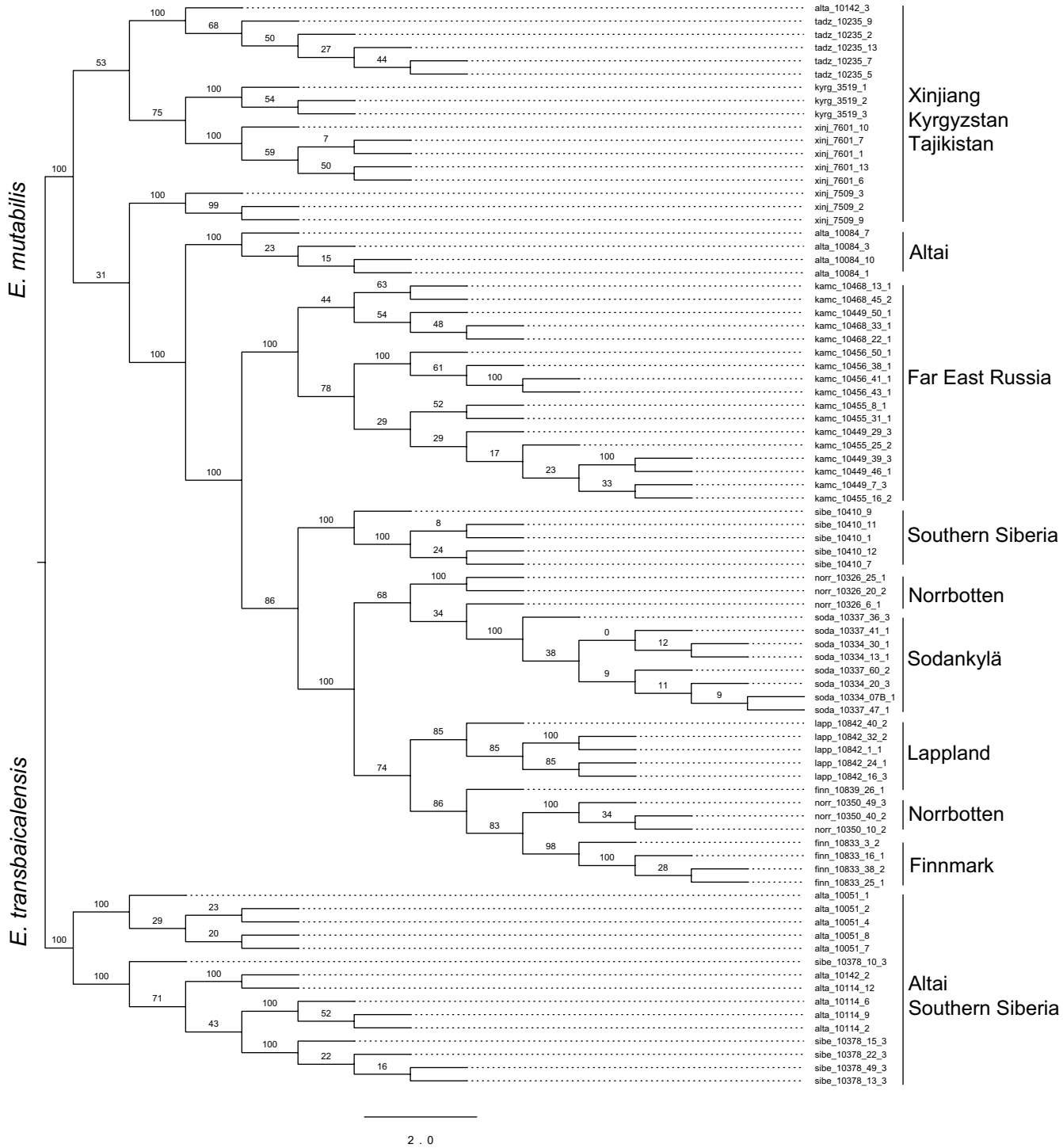

Figure 5. A coalescent phylogenetic tree of *E. mutabilis* with *E. transbaicalensis* as an outgroup. The analysis was performed using TETRAD (Chifman and Kubatko 2014, Eaton 2014) on the c90m20 dataset with all possible quartets (1 749 060) and 100 bootstrap replicates. Bootstrap values are indicated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