Mol Genet Genomics (2015) 290:329–342
DOI 10.1007/s00438-014-0907-4

ORIGINAL PAPER

Diversity within the genus *Elymus* (Poaceae: Triticeae) as investigated by the analysis of the nr5S rDNA variation in species with St and H haplomes

B. R. Baum · T. Edwards · D. A. Johnson

Received: 27 May 2014 / Accepted: 23 August 2014 / Published online: 24 September 2014

Abstract The genus *Elymus* ("Ryegrass") is a repository for a range of species with a variety of haplome contents; hence the pejorative name "dustbin" genus. We have analyzed 1,059 sequences from 128 accessions representing 24 species to investigate the relationships among the StH haplomes-containing species described by Yen and Yang (Genus *Elymus* Beijing 5:58–362, 2013). Sequences were assigned to "unit classes" of orthologous sequences and subjected to a suite of analyses including BLAST (Basic Local Alignment Search Tool) searches, phylogenetic analysis and population genetic analysis to estimate species diversity. Our results support the genome analyses in Yen and Yang (Genus *Elymus* Beijing 5:58–362, 2013), i.e., genomic constitution StStHH including variants restricted to *Elymus*. Population genetic analysis of the 5S nrDNA sequence data revealed that the within-species variance component is roughly ±89 %; thus, we were unable to identify molecular markers capable to separate the 24 species analyzed. Separate phylogenetic analyses of the two unit classes and of all the data exhibit a trend only of the species to cluster on the phylograms. Finally, the analysis provides evidence for the multiple origins of American and Eurasian species.

Keywords Unit classes · Haplome · Analysis of molecular variance · Phylogenetic analysis

Communicated by S. Hohmann.

Electronic supplementary material The online version of this article (doi:10.1007/s00438-014-0907-4) contains supplementary material, which is available to authorized users.

B. R. Baum (✉) · T. Edwards
Eastern Cereal and Oilseed Research Centre, Agriculture and Agri-Food Canada, Ottawa, ON K1A OC6, Canada
e-mail: bernard.baum@agr.gc.ca

Present Address:
B. R. Baum
Department of Biology, University of Ottawa, 30 Marie Curie, Ottawa, ON K1N 6N5, Canada
e-mail: bbaum@uottawa.ca

D. A. Johnson
Ottawa-Carleton Institute of Biology, University of Ottawa, P.O. Box 450, Station A, Ottawa, ON K1N 6N5, Canada

Introduction

Elymus species grow on all habitable continents and are commonly known as ryegrass or wheatgrass. Since its establishment by Linnaeus (1753), the genus *Elymus* L. has grown to include a large number of morphologically defined species. In fact, *Elymus* has become the largest genus in the Triticeae Tribe of the Grass family and one of the largest in the family. The reason for this asymptotic growth in the number of species likely was due to the reliance on morphological characters for discrimination among species that demonstrate high levels of polymorphism among individual plants and in different habitats. Thus, taxonomists described an endless number of species by exclusion, leading to *Elymus* becoming a repository of miscellaneous species or a "dustbin" genus, by default. The concept of "dustbin" taxon was invented by Davis and Heywood (1963) who stated that "the situation may arise particularly easily in genera in which good taxonomic characters are hard to find or in which distinctive characters are found only in the majority of species". In addition, several hundreds of species names became transferred due to synonymy (see for instance International Plant Names Index [IPNI, http://www.ipni.org/index.html] and specifically Löve's Conspectus of the Triticeae (1984)). The number of

new species keeps on growing. IPNI currently lists 1,336 records under the genus *Elymus*, including generic synonyms and several sectional and sub-sectional names, but with more than a thousand specific epithets.

As with other disciplines, botany is constantly undergoing progressive phases of development by incorporating new investigative technologies along with novel analytical methodologies (Constance 1964). New approaches lead to new evaluations and re-evaluations. During the "morphological phase" Nevski (1934) separated *Roegneria* C. Koch from *Agropyron* Gaertn., a very large genus with at least 1,050 names (IPNI). But then in (1976) Tsvelev increased the size of *Elymus* to a very big genus by incorporating the species of *Roegneria* into *Elymus*, again strictly on the basis of morphology. During the "cytogenetic phase" the establishment of a new criterion to determine generic status, i.e., a genus as a class or as a taxon is primarily based on karyology and chromosome pairing led to new findings in the Triticeae. Major contributions from Dewey (1984), Löve (1984) were key to the establishment of the genome constitutions of perennial diploid and polyploid genera as being combinations of genomes or a combination of multiples: *Pseudoroegneria* (**StSt**), *Agropyron* (**PP**), *Elymus* (**StStHH**), *Pascopyrum* (**HHJJNNStSt**), *Thinopyrum* (**JJ** or multiples such as **JJJJ**, etc.), to name a few. Löve followed Tsvelev's morphological concept of *Elymus* while believing that its genomic constitution is entirely **StStHH**, but Dewey (1984) changed it to **StStHHYY** by introducing the **Y** and **X** genomic symbols "to indicate unspecified genomes of unknown origin".

As the new areas of morphology, cytogenetics and more recently DNA data were analyzed, new genomically based genera were established or resurrected in the Triticeae, such as *Kengyilia* Yen et Yang (Yen and Yang 1990) with genomic constitution **StStPPYY**, *Douglasdeweya* Yen, Yang & Baum (**StStPP**) (Yen et al. 2005), *Roegneria* with **StStYY** and **StStStStYY** genomes (Yang et al. 2008), and *Campeiostachys* Drobov (**StStHHYY**) (Baum et al. 2011). Several of these genera had been consequently segregated out of *Elymus* thus confining it to the **StH** haplome combinations, namely to the **StStHH**, **StStStStHH**, **StStHHHH** and **StStStStHHHH** genomic constitutions (Yen and Yang 2013).Yen and Yang summarized their many publications in a multi-volume entitled Biosystematics of the Triticeae of which the genus *Elymus* is one of the genera treated in their fifth volume, pages 58–362. In all of their publications, Yen and Yang were staunch followers of the genomic system of classification established by Löve (1984), (Dewey 1984).

We are interested in the genetic diversity within and among species of *Elymus* with the focus on those species for which their genome constitution is known based principally on the work of Yen and Yang (2013). The present investigation employs the 5S nrDNA units as a tool to study variation and relationships. In the nucleus of higher plants, 5S nrDNA is organized into arrays of tandem repeats consisting of a highly conserved transcribed genes separated by a non-transcribed spacer (NTS) which varies considerably in length and pattern. The 5S NTSs especially can be assembled into orthologous sequence groups mainly based upon conserved patterns of nucleotide substitution (Baum et al. 2001), which we called "unit classes", that are useful in studies of variation including population genetics and phylogenetic analysis (Appels et al. 1992; Baum and Johnson 2002, 2004). For example, the 5S nrDNA multigene family has proven to be a useful tool to characterize genome composition in genomes that have previously determined by cytogenetics, based upon the combinations of unit classes that were detected (Baum et al. 2001; Baum and Johnson 2008). Hence, the genome constitutions for species that had not previously been determined can in some cases be predicted. The NTS has also been used as a tool to differentiate between very closely related species belonging to the same genome (Baum et al. 2012a) and to assess biodiversity of a taxon for example Einkorn wheat (*Triticum urartu*) a bread wheat progenitor (Baum and Bailey 2012).

Thus, the aims of the present 5S nrDNA investigation are (1) to check for presence of the **StH** haplomes in 24 of the 30 cytogenetically documented **StH** species studied by Yen and Yang (2013); (2) from sequence similarity to infer the putative **St** and **H** donor species of each of the 24 species for further hypotheses testing; (3) to perform a population genetic analysis and calculate diversity statistics within and among species; (4) to examine whether 5S rDNA can be used to distinguish among species, among genomic constitutions, and among geographic distribution of the species (Phylogenetic analysis); and (5) to examine and discuss the hypothesis that the American species have evolved independently from the Eurasian species previously based on genome (Wang and Hsiao 1986), isozyme (Jaaska 1992) and karyotype (Linde-Laursen et al. 1994) analyses.

Materials and methods

In previous publications, we have described in detail the methodology that is used to collect cloned 5S nrDNA, to create data sets of 5S nrDNA sequences, to assign them to unit classes (Baum and Johnson 1994, 1996, 1998), to carry out population genetic analyses (Baum and Bailey 2012, Baum et al. 2012a) and to preform phylogenetic analyses (Baum et al. 2001;2012b). In this paper, we supply a short summary. Table S1 provides a summary of the sample locations, GenBank DNA accessions and other details.

DNA material was obtained from plants grown in the greenhouse or growth cabinets from seed or from leaf tissue

from herbarium specimen. Seed samples were obtained from the National Plant Germplasm System (NPGS), Agricultural Research Service (ARS), United States Department of Agriculture (USDA), National Small Grains Collection after selection from the database Germplasm Resources Information Network (GRIN) at http://www.ars-grin.gov/npgs/acc/acc_queries.html.

DNA data acquisition

DNA isolation, PCR amplification and cloning were carried out as outlined in Baum et al. (2012a). The primers in Appels et al. (1992) were designed to use the BamHI sites found in tandem arrays 5S nrDNA unit repeats for amplification, thus avoiding organelle DNA.

Sequencing reactions were carried out in both directions (McGill University and Genome Quebec Innovation Centre, Montreal, Quebec, Canada or Plant Biotechnology Institute, Saskatoon, Saskatchewan, Canada). The sequences were checked to ensure removal of vector sequences using the VecScreen program at NCBI (National Center for Biotechnology Information, USA) and then submitted to NCBI (Table S1, in column entitled GenBank accession #).

Identification of unit classes in each taxon and in all the data

The following procedures were carried out separately in each species as originally identified for each accession. Initially synonyms were retained, and then upon verification were amalgamated under the correct species name. For instance, *Elymus agropyroides* is a synonym of *E. angulatus*, often known by the synonym in the area where it is collected (Table S1).

The cloned PCR amplicon sequences were arranged with the NTS followed by the 5S nrDNA gene of approximately 120 bp. Clustal Omega (Sievers et al. 2011) was used to generate an alignment that was further improved with the aid of SeaView Version 4.4.3 (Gouy et al. 2010), followed by visual examination and editing with GeneDoc© Version 2.6.002 (Nicholas and Nicholas HB Jr 1997). One important option from among several in Clustal Omega is an option that changes the order of the input sequences into the descending order of the similarity among them. SeaView then refines further the alignment without changing the input order of the output from Clustal Omega. Then, we used GeneDoc© to identify groups of sequences that contain putative orthologs based upon sequence similarity and pattern of the nucleotide sequence, clearly rendered in color and visualized. For each species including synonyms, sequences representative of each orthologous group were used to search GenBank© databases using the NCBI Web-based BLAST service (Altschul et al. 1990, 1997). BLAST searches were done in two rounds. The first using just the default, i.e., no option, was to ascertain that the sequence is similar to an *Elymus* sequence in the database. The second was to establish that a particular representative of a group was either an **St** or an **H** haplome from a diploid 5S nrDNA sequence present in NCBI database submitted earlier by us (Baum et al. 2008) as part of publications on *Hordeum* and *Pseudoroegneria*. The second round has been completed using the option "Exclude" that is excluding *Elymus*. This has been purposely carried out to identify the closest *Pseudoroegneria* (**StSt**) or *Hordeum* (**HH**) species as potential genome donors of each of the *Elymus* species in this study.

To ascertain that all the sequences together fall into the three expected unit classes belonging to the Long H1 (including the Long H2 that following BLAST analysis was separated out, but their sequences are very close), Long S1 and Short S1, all the aligned sequences were subjected to a phylogenetic analysis using PhyloBayes (Lartillot et al. 2009) with the option -CAT –GTR and to MrBayes (Ronquist et al. 2011) using the substitution parameters obtained from jModelTest 2 (see below under section Phylogenetic analyses). An outcome of three clear clusters would support the three unit classes as separate orthologous groups.

Population genetic analyses

To understand the diversity of the species of *Elymus* through the diversity of the NTS over all the sequences, we investigated the apportionment of the interspecific and intraspecific variance under different group structures: (1) One region: the total distribution; (2) Two regions: Eurasia–circumpolar distribution versus North–South Americas; (3) Three regions: Eurasia–Circumpolar, North America, South America; (4) Four regions: Eurasia, Circumpolar, North America, South America; (5) Five regions: strict Eurasia, Asia, Circumpolar, North America, South America; (6) Two regions based upon genomes: tetraploid **StStHH** species and hexaploid **StStHHHH** species and their variants including **XXXXHH** (X stands for unknown); (7) Three regions: i.e., same as the former except that the hexaploids were divided into the South American and the Asian. To obtain the results, the NTS sequences were subjected to a series of population genetic analysis, especially AMOVA (Analysis of molecular variance), using two complimentary programs: Arlequin (Excoffier and Lischer 2010) and GenAlEx 6.501 (Peakall and Smouse 2006, 2012) where in addition Principal Coordinates Analyses (Gower 1966) were carried out. It is noteworthy that these analyses were carried out as a guide only where the species were treated as populations.

Statistical parsimony networks

Since the divergence within *Elymus* as defined in this study is low, and since we already regarded its species as populations, it would be sensible to investigate the pattern of relationship among the species in network form. This can be accomplished in Statistical Parsimony using TCS (the method of Templeton et al. 1992), a program designed "to estimate gene genealogies including multifurcations and/or reticulations (i.e., networks)".The DNA sequences from each unit class were subjected to separate TCS (programmed by Clement et al. 2000) runs, as well as the alignment of all the sequence data. Furthermore, the visual display of the network resulting from each TCS run was made possible by TempNet (Prost and Anderson 2011). TempNet is a script that runs on the R environment (Venables et al. 2014). In this study, TempNet was executed by assigning each species to a "layer".

Phylogenetic analyses

At first, we assessed the evolutionary models of each unit class drawn from the DNA sequence data, using jModelTest2 (Darriba et al. 2012; Guindon and Gascuel 2003). Then, we carried out a separate analysis of the DNA sequence alignment of each unit class using MrBayes (Ronquist et al. 2011) with input of the parameters obtained beforehand with ModelTest2, with the purpose to examine whether the individual sequences cluster by species on the obtained gene trees, i.e., the unit class trees. PhyloBayes was used in a similar fashion as MrBayes above with the–CAT–GTR option, and for the same purpose. In PhyloBayes, the substitution models cannot be used, however, the option used is the most generalized one.

Results

We generated our basic dataset of 1,059 sequences using material from 128 accessions derived from 24 of the available 30 species (Table S1) for an average of 8.3 sequences per accession and 44.1 sequences per species.

Identification of unit classes

Almost all sequences from each taxon as received were readily assigned to putative unit classes based upon the alignment. In some cases, this assignment was more difficult as illustrated by the following example, available online. One sequence (EGAY002) has been identified to belong to one of three putative unit classes identified (Fig. S1). However, when applicable taxa that were synonyms were each pulled together with sequences of the species with the correct name (following the nomenclature of Yen and Yang 2013) and realigned, we could confirm a single putative unit class status for it.

The result of BLAST analyses (Table 1) shows that the majority of the several sequences selected per species resulted in the identification of the Long S1 and/or Short S1 unit classes, both characteristically present in *Pseudoroegneria* **St**, and the Long H1 and/or the Long H2 unit class present in *Hordeum*. The Long H1 unit class is common in Eurasia and found in a few species in the Americas whereas the Long H2 is most common in the Americas (Baum and Johnson 2004). Table 1 also shows several BLAST results in which alternative unit classes were identified (see Table heading "Unit classes captured").For instance, in *E. alaskanus,* the three unit classes were identified, Long S1, Short S1 and Long H1, but a few sequences were similar to the Short J1 and Long J1 (from *Thinopyrum*). The Long Y2 and Long H2 unit classes, both significant variants of the H haplome (Baum and Johnson 2002, 2003), were found exclusively among three hexaploid South American species, whereas the Long H2 unit class was found in seven, probably nine, American species, and not in Eurasian ones. The Long S1 sequences and the Short S1 sequences were most similar to a variety of *Pseudoroegneria* species. Possible interpretation of these results with respect to the origin and evolution of *Elymus* is explored in the "Discussion".

Finally, a phylogenetic analysis of all the 1,059 sequences (unrooted PhyloBayes tree in Fig. 1) demonstrates that the three putative unit classes, Long H1, Long S1 and Short S1 did indeed fall into three major clades—orthologous groups of sequences, thus confirming the status of unit classes. A different rendition of the relationships within the same tree can be seen in the expanded PhyloBayes tree (Fig. S2), where the individual sequences are more clearly identified. Indeed, for a clearer understanding of the relationships among the accessions, these two unrooted trees should be considered at the same time.

Population genetic analyses

Several different approaches were used to help assess and display genetic diversity. We re-emphasize here that the statistics reported are to be used only as guides, not as absolutes. As part of the genetic diversity, the expected heterozygosity over 567 loci, which are the positions in the alignment, shows that there is a tendency of the variation to be grouped over three strata, but definitely less in the area from the position 420 to the end (data not shown) which is the transcribed part. Among the 24 species, the least variable is *E. sierrae*; probably due to the small sample available for this study only 10 clones all from the Long S1 unit class were recovered. Interestingly, the expected heterozygosity did not correlate with sample size (data not shown). *E.*

Table 1 Species used in the *Elymus* study with synonyms relevant to the study

Species number in Yen and Yang 2013	Species number in some analyses	Material available:1=Yes 2=No	Species nomenclature, includes synonyms pertinent to Table2	Genome/ Alternate genome	Chromosome number	Continent	Unit classes captured LongH2: +same E values but higher total score than LongH1 or vice versa; =equivocal; ^: very few sequences found among Long and Short S1						
							Long S1	Short S1	LongH1	Long Y2	Long H2	Short J1	Long J1
01	01	1	*_E. alaskanus_ (Scribn. Merr.) A. Love includes vars. _alaskanus_ and _hyperarcticus_	**StStHH**	2n=4x=28	Circ	Long S1	Short S1	Long H1			^Short J1	^Long J1
04	03	1	*_E. breviaristatus_ (Keng) Keng ex Keng f.	**StStHH**	2n=4x=28	A	Long S1	Short S1	Long H1				
05	04	1	*_E. canadensis_ L.	**StStHH**	2n=4x=28	NA	Long S1	Short S1	Long H1				
06	05	1	*_E. caninus_ (L.)L.	**StStHH**	2n=4x=28	A	Long S1	Short S1	Long H1				
08	06	1	*_E. elymoides_ ((Raf.) Sweezey includes var. _brevifolius_	**StStHH**	2n=4x=28	NA	Long S1	Short S1	Long H1				
09	07	1	*_E. glaucus_ Buckley, includes _E. parishii_	**StStHH**	2n=4x=28	NA	Long S1	Short S1	Long H1				
12	08	1	*_E. hystrix_ L.	**StStHH**	2n=4x=28	NA	Long S1	Short S1				Short J1	
13	09	1	*_E. lanceolatus_ (Scribn. et JG Smith) Gould	**StStHH**	2n=4x=28	NA	Long S1	Short S1	Long H1				
14	10	1	*_E. magellanicus_ (Desvaux) A. Love includes _E. glaucescens_	**StStHH**	2n=4x=28	SA	Long S1	Short S1	Long H1		=Long H2		
15	11	1	*_E. mutabilis_ (Drob.) Tzvel.	**StStHH**	2n=4x=28	A	Long S1	Short S1	Long H1				
18	14	1	*_E. riparius_ Wiegand	**StStHH**	2n=4x=28	NA	Long S1	Short S1	Long H1		+Long H2		
20	16	1	*_E. scribneri_ (Vasey) M.F. Jones	**StStHH**	2n=4x=28	NA	Long S1	Short S1	Long H1		+Long H2		
21	17	1	*_E. sibiricus_ L.	**StStHH**	2n=4x=28	Circ	Long S1	Short S1	Long H1				
22	18	1	*_E. sierrae_ Gould	**StStHH**	2n=4x=28	NA	Long S1						
24	19	1	*_E. tilcarensis_ (J.H. Hunziker) A. Love	**StStHH**	2n=4x=28	SA	Long S1	Short S1			+Long H2		
25	20	1	*_E. trachycaulus_ (Link) Gould ex Shinners	**StStHH**	2n=4x=28	Circ	Long S1	Short S1	Long H1				
27	22	1	*_E. vaillantianus_ (Wulfen et Schreb.) K.B. Jensen	**StStHH**	2n=4x=28	NA	Long S1	Short S1			Long H2		
29	23	1	*_E. virginicus_ L.	**StStHH**	2n=4x=28	NA	Long S1	Short S1			Long H2		
30	24	1	*_E. wawawaiensis_ J. Carlsen et Barkworth	**StStHH**	2n=4x=28	NA	Long S1	Short S1	Long H1				
03	02	1	*_E. angulatus_ J. Presl includes _E. agropyroides, andinus, antarcticus, gayanus_	**StStH1H1H2H2**	2n=6x=42	SA	Long S1	Short S1	+Long H1	LongY2	Long H2		
16	12	1	*_E. patagonicus_ Speg.	**StStH1H1H2H2 StStHHHH**	2n=6x=42	SA	Long S1	Short S1	Long H1	Long Y2	Long H2		
19	15	1	*_E. scabriglumis_ (Hack.) A. Love	**StStHHHH**	2n=6x=42	SA		Short S1	LongH1	Long Y2	+Long H2		
26	21	1	*_E. transhyrcanicus_ (Nevski) Tzvel.	**StStSt2St2HH StStStxStxHH**	2n=6x=42	A	Long S1	Short S1	Long H1				
17	13	1	*_E. repens_ (L.) Gould includes vars. _elongatiformis_ and _repens_	**StStStStHH XXXXHH**	2n=6x=42	A, NA	Long S1	Short S1	Long H1				^Long J1
02		2	*_E. albicans_ (Scribn. Et J.G. Smith) A. Love	**StStHH**	2n=4x=28	NA							
07		2	*_E. dentatus_ (Hook.f.) T.A. Cope	**StStHH**	2n=4x=28	A							
11		2	*_E. hordeoides_ (Suksdorf) Barkworth et D.R. Dewey	**StStHH**	2n=4x=28								
23		2	*_E. subsecundus_ (Link) A. Love et D. Love	**StStHH**	2n=4x=28	NA							
28		2	*_E. violaceus_ (Hornem.) J.F. Feiberg	**StStHH**	2n=4x=28	Circ							
10		2	*_E. hoffmanni_ K.B. Jensen et K.H. Asay	**StStStStHH**	2n=6x=42	A							

Species listed by their genomic constitution

canadensis and *E. hystrix*, with fairly large sampling, were found to be low compared to the rest of the species. Similarly to the expected heterozygosity, the molecular diversity index as calculated by Theta (S) (Fig. S3, θ_s the blue line) is lowest in *E. sierrae* and in *E. hystrix*, but not comparatively so in *E. canadensis*.

The matrix of pairwise F_{st}'s depicted in Fig. S4 shows that several species appear to be genetically remote from the majority of species, such as *E. alaskanus*, *E. angulatus*, *E. elymoides*, *E. lanceolatus*, etc. *E. sierrae* appears to be closer to many *Elymus* species than observed for the majority of species' pairwise comparisons, perhaps stemming from the small sampling available for this species in this study.

The average number of pairwise differences between species calculated using Nei and Li's (1979) π between populations is depicted on Fig.S5. On the diagonal showing π within species, within *E. sierrae* there are no differences, for the same reason as above; in *E. trachycaulus* and E. *canadensis,* the within differences are higher (40–50) and so forth. In the

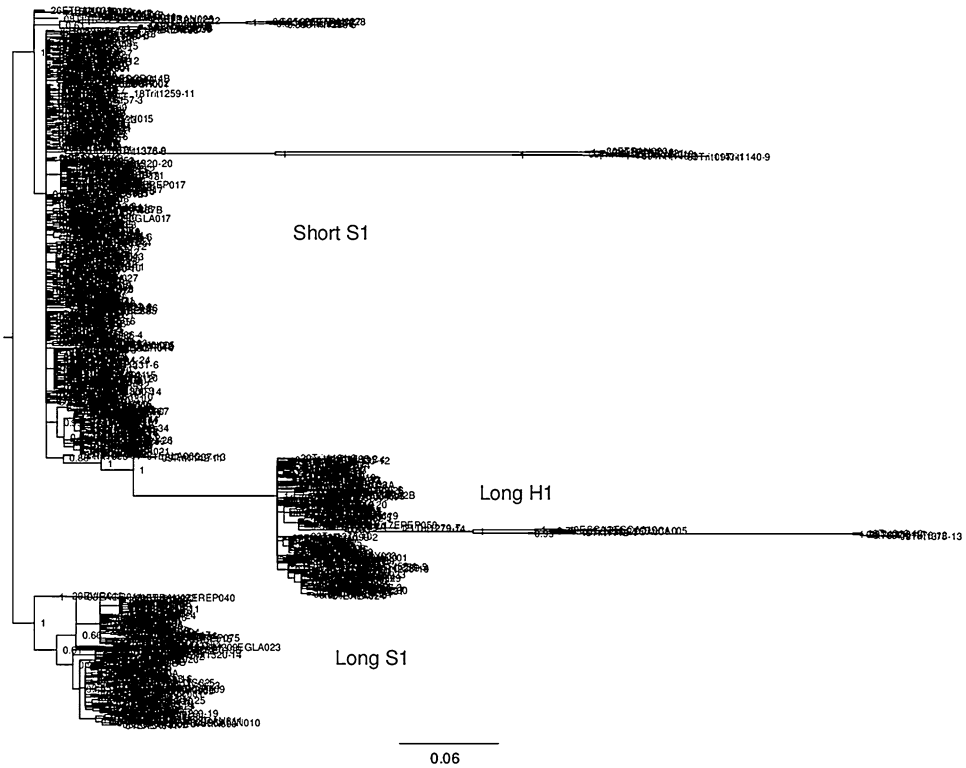

Fig. 1 An unrooted PhyloBayes tree resulting from input of all the 1,059 5S DNA sequences in the *Elymus* study displayed on a single page using the program Fig Tree. A "1" indicates 100 % bootstrap support. The identities of the three groups are, from *top*: Short S1, Long H1, and Long S1. Branches are proportional to their actual length. *Scale bar* indicates relative distance on a 0–1 scale. See Supplementary Fig. S1 for more details

plot area above the diagonal the between pairwise differences between species, π_{xy}, are depicted mostly as high with some exceptions, notably *E. trachycaulus* with *E. canadensis* and with *E. hystrix*, and between *E. canadensis* and *E. hystrix*. Overall the number of average pairwise differences between species is high. In the area below the diagonal the net number of nucleotide differences between species, Nei and Li's D_A (1979 between populations), is generally low except for most pairwise comparisons with *E. sierrae*.

AMOVA analyses were conducted on these structures as mentioned in the "Methods" section (Table 2). Essentially most of the variance is apportioned within populations (87–89 %) with less among populations (10–12 %) and only 2–3 % among regions/genomes (Fig. 2). Thus, negligible differences were found within the different genetic structures. However, the highest amount of variation among regions (3 %) was found when the regions were determined by the five different genomes in Table 1.

Principal coordinate analyses of the 24 species (Fig. 3) demonstrated that in general the plots of species on the first three principal coordinate axes were similar when the various genetic structures were implemented. However, in the case of a genetic structure of all the species in one region, the disposition of the species points is much more spread out on the 3D space than when the genetic structure is enlarged to include more than one region. In this case, the majority of species are more concentrated. When the genetic structure is based on the genome constitution (Table 1) the plot of species in 3D space is similar to the genetic structure of all species in a single region. In all different principal coordinate plots, *E. canadensis* and *E. sierrae* are more dispersed outward along with *E. magellanicus* and *E. caninus* and *E. hystrix* and *E. trachycaulus*, the latter two close to *E. canadensis*.

Statistical parsimony networks

Discerning relationships among genera in the grasses are rendered more difficult by the reticulate nature of their evolution. To help understand these relationships, we turned to network analysis (Baum et al. 2012b). The TCS networks representation by TempNet shows great similarities among the 24 species. When the 1,059 sequences belonging to the 24 species were analyzed together they yielded a large and overcrowded network of connection patterns of very little practical use (data not presented). Instead, we broke the rendition of the 24 species (the layers in the graph) into six parts of four. We show here the first of six (Fig. 4). The pattern of all six analyses is similar among the species and the connections among species have almost been obliterated in the rendition, but we can discern them connecting the centers with highest density of points among the four in the graph (Fig. 4) and among the 24 species (not shown).

The TCS networks obtained for the species within each unit class exhibited pattern similarities among species. Long H1 unit class showed as above a similar core among

Table 2 AMOVA analyses

Source of variation	d. f.	Sum of squares	Variance components	Percentage of variation
1 Region				
Among populations	23	5,118.334	4.362	11.34
Within populations	1,035	35,309.738	34.11569	88.66
Total	1,058			
Fixation index		FST: 0.113366		
5 Regions				
Among groups	4	1,665.115	0.87558	2.26
Among populations within groups	19	3,452.219	3.69172	9.54
Within populations	1,035	35,309.738	34.11569	88.19
Total	1,058	40,428.072	36.68298	
Fixation index	FSC :0.09765	FST: 0.11807		
3 Regions as genomes				
Among groups	2	1,175.381	0.85496	2.2
Among populations within groups	21	3,942.953	3.93514	10.11
Within populations	1,035	35,309.738	34.11569	87.69
Total	1,058	40,428.072	38.90579	
Fixation index	FSC: 0.10342	FST: 0.12312		
2 Regions (North America + South America and Eurasia + Circumboreal)				
Among groups	1	622.455	0.61584	1.59
Among populations within groups	22	4,495.878	4.04808	10.44
Within populations	1,035	35,309.738	34.11569	87.97
Total	1,058	40,428.072	38.77961	
Fixation index	FSC: 0.10607	FST: 0.12027		
2 Regions as genomes				
Among groups	1	266.507	0.29169	−0.76
Among populations within groups	22	4,851.827	4.49336	11.73
Within populations	1,035	35,309.738	34.11569	89.03
Total	1,058	40,428.72	38.31736	
Fixation index	FSC: 0.11638	FST: 0.10965		

For the five regions genetic structure *E. caninus* and *E. mutabilis* were considered Eurasian (EA), separate from Asian

Refer to Table 1 for the regions

species all with an extended variation of sequences such as in Fig. 5. However, the patterns generated by the analysis of the individual unit classes showed some differences, such as between the Long H1 (Fig. 5) and the Long S1 (Fig. 6) and the Short S1 (Fig. 7). All the different networks have in common a core with expanded network edges made of long chains of sequence variants.

Phylogenetic analyses

First nucleotide substitution models were selected for "all the data" and separately for the Long H1, Long S1 and Short S1 unit classes (Table 3). The evolutionary model for all the data together (not broken into unit classes) and for the Long H1 unit class was HKY + G. It is TPM1uf + G for the Long S1 and Short S1 unit classes data. The unrooted tree with all the data (Fig. S2) supported the distinction between the three unit classes. In each of the three gene trees, the Long H1 (Fig. S7), Long S1 (Fig. S8) and Short S1 (Fig. S9), the cloned sequences fell largely together by species and this tendency was general. However, there were sequences intermixed among species. The trees generated by MrBayes with the specific substitution parameters (Table 3) were similar to the trees generated by PhyloBayes when using the "–CAT–GTR" options.

Discussion

This project was initiated with the aim of achieving five goals: assign sequences to unit classes and use them to identify haplome content in *Elymus*, using sequence data to

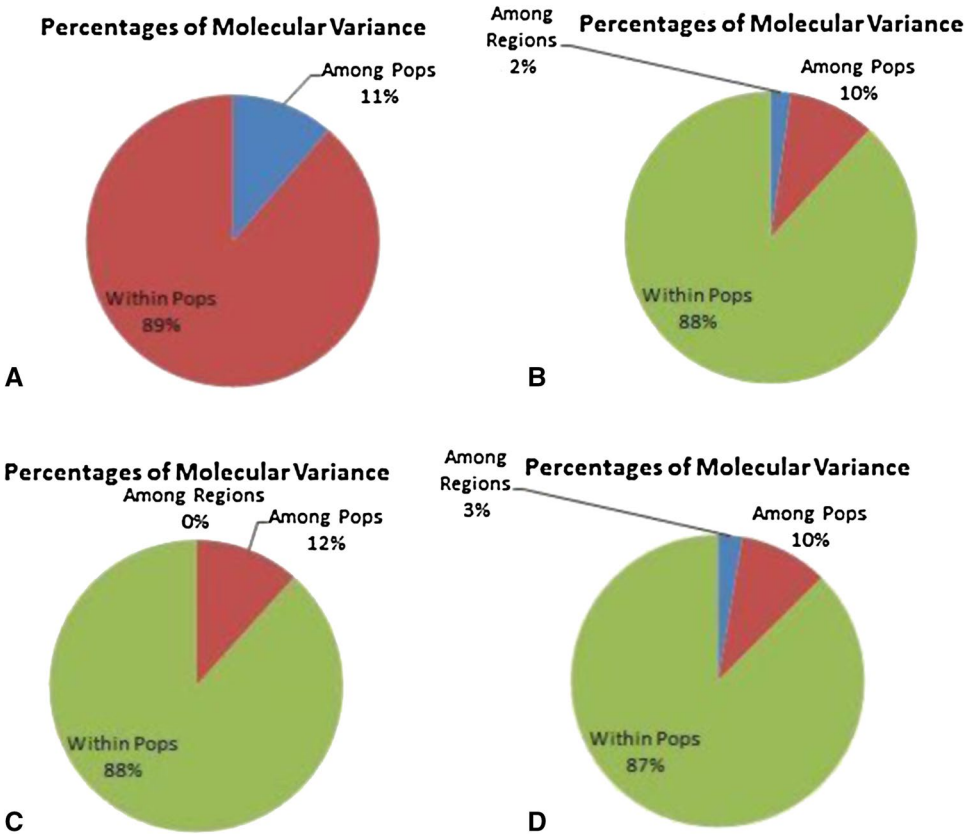

Fig. 2 AMOVA of the *Elymus* sequence data divided as described in the text. **a** One region with all the 24 species; **b** Five geographic regions; **c** Two regions, i.e., the tetraploid in one and hexaploids in the other; **d** Five regions, i.e., the five different genomes

estimate infra and interspecific diversity, to infer potential haplome donor species and to distinguish among species, genomic constitutions and geographic distributions, and finally testing the hypothesis that American and Eurasian species evolved independently. Analyzing 1,059 sequences in *Elymus* has allowed us to gain insight into these goals.

Haplome identification and utility

We have shown here and in several previous publications that BLAST followed by phylogenetic analysis can be used to assign the 1,059 sequences to unit classes that are paralogous. The unit classes can be regarded as separate genes belonging to a multigene family. In *Elymus,* we identified the Long S1 and Short S1 unit classes typically present in *Pseudoroegneria* **St**, and the Long H1 and Long H2 unit class present in *Hordeum*. We also found in the Triticeae that each unit class represents different genomes (see for instance Baum et al. 2001; Baum and Johnson 2008). Based upon the identified unit classes, all species of *Elymus* as defined in this work have the same genome or include multiples of some haplomes or include some haplome variants (Table 1); but for practical purposes, they all were found to contain the same genome components as cytogenetically defined. Our results support the genome analysis of Yen and Yang (2013).

The BLAST findings also allude to a possible dual (or multiple) origin of the species of *Elymus*. While the BLAST results for every clone cannot be documented due to space limitations, several examples support this view. Sequences identified as belonging to the Long H2 unit class were identified (Table 1, Table S1). The Long H2 unit class was found to be prevalent among the American species of *Hordeum* (Baum and Johnson 2002, 2004), and is not present among the Eurasian *Hordeum* species. The Long Y2 unit class was identified in *E. angulatus* (Table 1); the BLAST results for clone TRIT436-5 (Table S1) identified it as a Short S1, specifically identical to the North American *Pseudoroegneria spicata* (Pursh) Á Löve clone PSEU044 5S ribosomal RNA gene (GenBank Accession EU093363) and with two additional *P. spicata* clones, hence Short S1 unit class (Baum et al. 2008). In the same fashion clone Trit435-2 had been found 97 % identical to *Hordeum stenostachys* Godr. clone HSTE031 (GenBank AY544565) which is a Long H2 unit class member (and which in this work has been treated as a Long H1 unit in the phylogenetic analyses because of the close relationship between the two). In another example *E. caninus*, a Eurasian species, clone TRIT1309-1 has been identified with the Asian *H. brevisubulatum* (Trin.) Link clone HNEV001 being a typical member of the Long H1 unit class. Several cases involving the Long Y2 unit class (Table 1) found

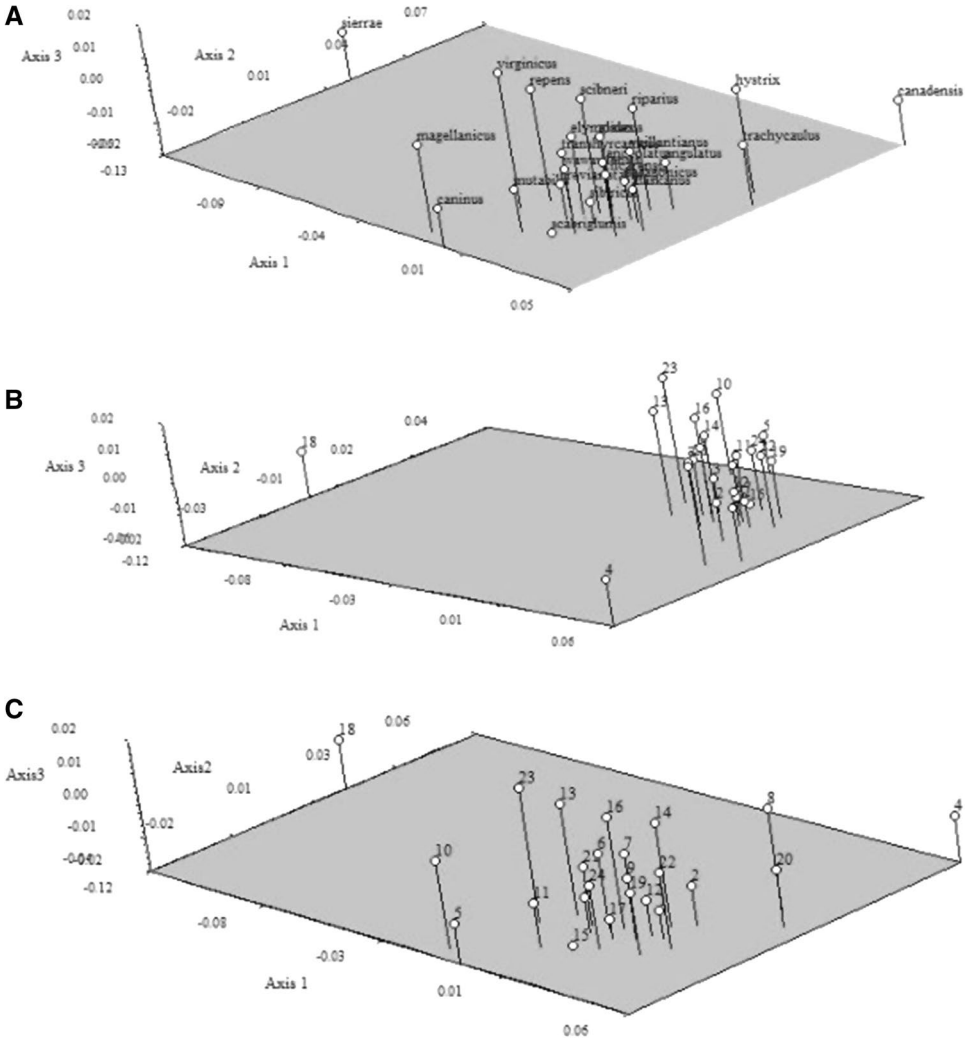

Fig. 3 Principal coordinate analysis of the *Elymus* sequence data. Plots of the first three axes with different genetic structures. **a** One region, variation explained by *axis 1* 57.03 %, *axis 2* 26.80 %, 4.02 %; **b** Five regions, *axis 1* 67.94 %, *axis 2* 26.74 %, *axis 3* 3.65 %; **c** Two and five regions as different genomes *axis 1* 67.94 %, *axis 2* 26.74 % Axis3 3.65 %. In Figs. **b** and **c** the species labels are replaced by numbers which one can extrapolate from **a**, but refer to Table 1, second column from left for the species name equivalent

three South American species need to be mentioned. Thus, BLAST results alone point to separate origins of *Elymus* from different diploid donor species of *Hordeum* and *Pseudoroegneria* in different continents.

Our analyses are in agreement with Mason-Gamer et al. (2010) with respect to the donor genera, but not with respect to the specific species in different continents. For instance, in *E. hystrix* a North American species, sequences of the Long S1 were similar to the comparable ones in *P. stipifolia*an Asian species and the Short S1 sequences had higher similarity with sequences of *P. strigosa*, also an Asian species, but not exclusively with *P. spicata*. Another example is *E. magellanicus* where sequences of the Short S1 unit class were most similar to the comparable ones found in *P. strigosa* and *P. gracillima* both Asian species. In fact, our results suggest the notion of multiple origins for *Elymus* in support of Sun and Salomon (2009). Multiple sources for genome donors in different Swedish populations of *E. repens* had been proposed (Fahleson et al. 2008) based upon AFLP-derived evidence for the donor of **J** haplomes, as we found with the 5S nrDNA (Table 1). However, it could instead be argued that the resulting sequence similarities detected originated as a result of introgression events and not strictly of a onetime multiple diploid donors of *Elymus* species. To make things more complicated, both types of events were possible in the course of *Elymus* evolution.

Diversity within and among *Elymus* species

When taken as populations the species of *Elymus* in this study were each found to be highly variable, so much in fact that the variance component within species was roughly 89 % and the variance component among them was roughly 11 %.When the species were divided into regions in different ways the variance component of the region was a very small 1–2 % and reached 3 % when ploidy was used as a regional criterion (Fig. 2d). A similar breakdown of variance components had been found in more detailed population studies of several single species, such as *E. caninus*

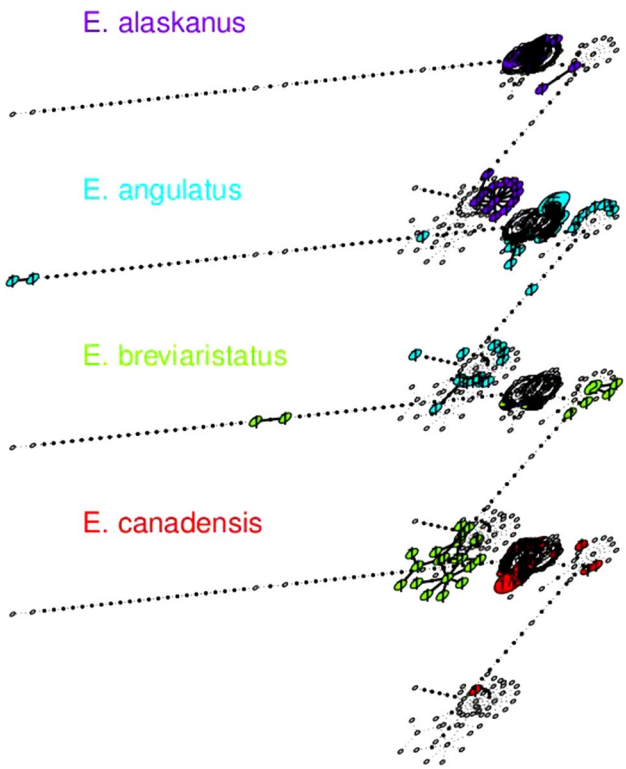

Fig. 4 Statistical parsimony of all the sequence data. TCS analysis rendition by the TempNet script in R, example of the first four species number 01–04 of 24 species; see Table 1 second column from the left. See text

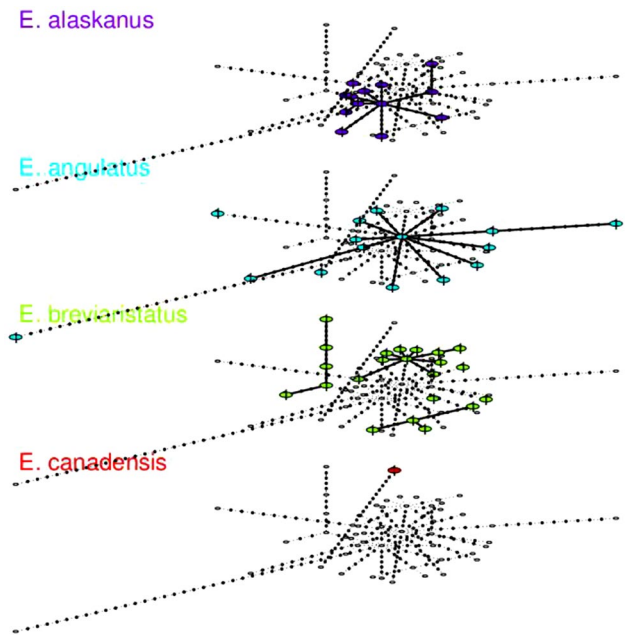

Fig. 5 Statistical parsimony of the Long H1 unit class sequences data. TCS analysis rendition by the TempNet script in R, example of the first four species number 01–04 of 24 species; see Table 1 second column from the left. See text

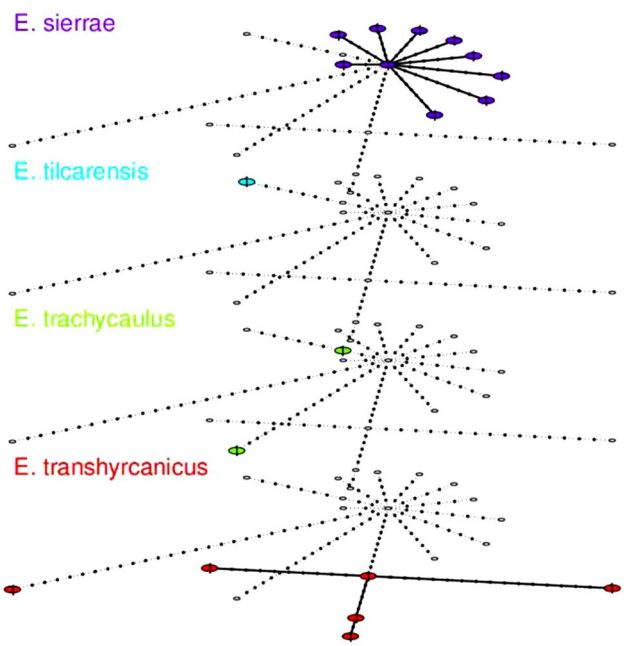

Fig. 6 Statistical parsimony of the Long S1 unit class sequences data. TCS analysis rendition by the TempNet script in R, example of the species number 18–21; see Table 1 second column from the left. See text

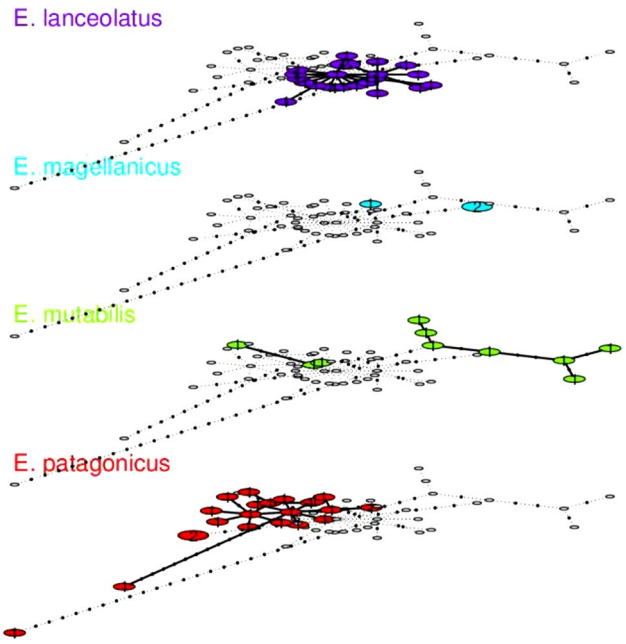

Fig. 7 Statistical parsimony of the Short S1 unit class sequences data. TCS analysis rendition by the TempNet script in R, example of the species number 09–12; see Table 1 second column from the left. See text

(Diaz et al. 1999), *E. alaskanus* (Sun and Salomon 2003), *E. sibiricus* (Ma et al. 2008), *E. canadensis*, *E. sibiricus* and their hybrids (Baum et al. 2012a).

Table 3 Several results of selection of substitution models using jModelTest2, of all the nucleotide data and the three unit classes' data decoupled from the former

	All data	Long H1	Long S1	Short S1
Number of sequences	1,059	198	190	671
Number of sites	1,043	611	577	859
Model selected	HKY + G	HKY + G	TPM1uf + G	TPM3uf + G
lnL	29,588.7301	7,916.307	6,923.1099	15,580.9871
K	2,103	399	384	1,346
FreqA	0.2723	0.2767	0.2829	0.261
FreqC	0.2013	0.1769	0.1787	0.2142
FreqG	0.2227	0.2325	0.2314	0.2348
FreqT	0.3038	0.3139	0.307	0.29
Kappa	2.6689	2.4971		
Gamma shape	4.109	4.388	2.447	4.208

The Principal coordinate plots may indicate that some species exhibit various amounts of localized genetic isolation due to distribution in different eco-geographic environments. These are the species in the periphery of the 3D plots (Fig. 3). For instance *E. canadensis* is known to be distributed from Northern Canada South to Northern Mexico. Apparently, there is less to no genetic interaction and gene exchange, especially south of the boundary of the Wisconsin Glacial which could explain the low genetic polymorphism (Sanders et al. 1979) among regions. This we found is the general case when species are akin to populations as in the present study, i.e., the genetic polymorphism is greatest within species and smallest among them. However, the amount of polymorphism alone has no bearing on species differentiation.

The results of TCS analyses supported the picture obtained from the population genetic analyses. Overall, similar central cores (of sequences) surrounded with long lines of network edges (interconnected sequences) resulted (Figs. 4, 5, 6, 7). These long edges unveiled the existence of the great diversity of sequences within each species. The cores of each species in these figures are separated artificially in the graphs. This is because each species is represented as a different layer in the graphic rendition. However, in reality the cores are linked by varying edge lengths and not in alphabetic order of the species names. Statistical parsimony can at least in this case be regarded as another view of the result obtained with the population genetic analyses where the greatest amount of polymorphism occurs within species. The separate TCS analyses of the data within each unit class revealed only slightly different patterns of variation among species (see for instance Fig. 5, 6, 7). The TCS graphic results with this amount of sequences (1,059) were difficult to render in a meaningful way, and TempNet has been a very helpful addition for obtaining a better rendition to visualize these results. An example showing the minute, but significant molecular differences between *E. canadensis* and *E. sibiricus*, two morphologically very closely similar species, obtained by TCS with the present data is available online (Fig. S6).

Phylogenetic analyses

As indicated above, phylogenetic analysis supports the placement of sequences into unit classes where a unit class can represent a different haplome (genome) in the genome constitution of a taxon. In the present study, each gene tree was unable to conclusively cluster together all the sequences by species. However, within a unit class in each gene tree the individual sequences belonging to the same species largely clustered together (Figs. S7, S8, S9).Even in the tree generated from all the data together the sequences largely clustered together by species (Fig. S2). We had not expected to be able to cluster all 5S DNA units at the species level, i.e., below the level of genome. Thus, the expectation of the resulting phylogenetic analyses has been stretched too far for the potential capacity of 5S nrDNA gene as a marker to distinguish between species.

We pooled together sequences of the Long H2 and Long Y2 unit classes found with the Long H1 in the phylogenetic analyses because of their close similarity and because they were found in some species together as in the present study and certainly in previous studies of American *Hordeum* sp. (Baum and Johnson 2002, 2003, 2004; Baum et al. 2005). However, sequences of the Long H2 and Long Y2 unit classes were never found in the Eurasian *Hordeum* species. Their presence intermixed with the Long H1 sequences in the same accessions in several American *Elymus* (Table 1) and *Hordeum* species clearly supports the view of multiple origins for *Elymus* species even within continents. As above these results do not preclude gene exchange and introgression among species, or both processes, over time. These two processes, and a third-loss of species, could have impacted species distribution even at the continental level.

Concluding general comments

In the present study, we again demonstrate the utility of the 5S nrDNA to identify haplomes and thus confirm genome constitution. We first discovered the potential genomic assignment of the rDNA unit classes in *Kengyilia* a genus with genomic constitution **StStPPYY** (Baum and Bailey 1997) that had been segregated out from *Elymus* (Yen and Yang 1990). The utility of the 5S NTS for genomic differentiation was also found in the genus *Silene*

(Popp and Oxelman 2001). We have also investigated highly repetitive sequences such as the ITS1, ITS2, and IGS of the rDNA repeat (Baum et al. 2001) and found them potentially useful as haplome markers in *Triticum*. We recognize that for this kind of investigation, e.g., assignment of NTS and variable repetitive sequences to unit classes is laborious as it requires the sampling and sequencing of a large number of DNA clones, but this approach avoids problems that may arise from comparing paralogues.

As expected from the outset, we were unable to use this approach to differentiate among species with the same genome constitution. For species differentiation other genes have been investigated and many more are required. The genes used to date in *Elymus* such as phosphoenolpyruvate carboxylase (PEP carboxylase; Helfgott and Mason-Gamer 2004; Mason-Gamer et al. 2010), β-amylase and granule-bound starch synthase 1 (GBSS1; Mason-Gamer et al. 2010), RNA polymerase II (RPB2; Sun and Daley 2007; also studied in *Hordeum,* one of the haplome contributors to *Elymus* Sun et al. 2009) although not aiming at differentiating among species, have not yet shown any promise in this regard because the species remained unresolved on the phylograms.

Elymus species are self-fertilizing, but can outcross with surprising frequency to other self-fertilizing species (Dewey 1984). Although the species are genetically close to each other and have the potential to interbreed (Church 1958), many have adapted to specific eco-geographic conditions and have developed specific genetic characteristics (Sun and Salomon 2003; Diaz et al. 1999).When two or more species meet in adjacent habitats interspecific hybrids may be partially fertile permitting introgression (Barkworth et al. 2007). For these reasons (and others), each species is still very variable in morphological characters making visual or morphological identification somewhat difficult. Some species that are difficult to separate morphologically can be distinguished genetically, or their status as species assessed on molecular basis such we found for *E. canadensis* and *E. sibiricus* (data from Baum et al. 2012a rerun here with TempNet) and their hybrids (Fig. S10).

For various purposes, such as biodiversity and habitat conservation or for a variety of utilitarian purposes, it is important to understand and recognize the inventory and the components of the organisms in different geographical locations. For many organisms, the species category is sufficient, for others it is necessary to classify the organisms of interest below the species level, such as varieties, strains, etc. This is very relevant because *Elymus* is an important pasture and forage grass and some species have also been improved for cultivation. If the "species" are difficult to separate are they truly species? We mentioned beforehand that introgression is not uncommon, and that many names subsequently became synonyms. *Elymus charkeviczii* Probat., *E. mutabilis* (Drob.) Tzvel. and *E. subfibrosus* (Tzvel.) Tzvel are variants of the same species, i.e., *E. mutabilis,* but introgression with *E. kamczadalorum* (a synonym of *E. trachycaulus*) is common (Agafonov and DE Gerus 2008). Identification based on morphological features, such as in Barkworth et al. (2007), is difficult and the attribution of specimens to species is qualified in many entries by terms such as "most", "usually", "in part" and by multiple entries for the same species. This raises the question: on what basis can we define a species within a genus defined genomically that is to date established by morphological circumscription, i.e., keys based on morphology? Identification by morphology is not sufficient because it is not precise enough, as mentioned above, and requires subjectivity. Thus, there is a need to introduce molecular markers into species circumscriptions and identification keys.

Acknowledgments The authors would like to thank Harold Bockelman (U.S.D.A) for supplying seed, Glen Newton and Christine Lowe, Biodiversity Bioinformatics Programmers at this institution, respectively, for running jModelTest2 and PhyloBayes on the Linux clusters. We also thank D. Laczko, and R. Kang for technical assistance. This research was funded in part by grants from the Natural Sciences and Engineering Research Council of Canada and the University of Ottawa to D.A.J.

References

Agafonov AV, DE Gerus (2008) Study of the polymorphic complex Elymus charkeviczii Probat.S.L. (Triticeae: Poaceae) on the Kamchatka peninsula from the viewpoint of biosystematics and taxonomic genetics. Растительный мир Азиатской России 1:58–70

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403–410

Altschul SF, Madden TL, Schoffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389–3402

Appels R, Baum BR, Clarke BC (1992) The 5S DNA units of bread wheat (*Triticumaestivum* L.). Pl Syst Evol 183:183–194

Barkworth ME, Campbell JJN, Salomon B (2007) Named hybrids. In: Barkworth ME, Capels KM, Long S, Anderton LK, Piep MB (eds) Flora North America, vol 24. Oxford Univ Press, Oxford, pp 338–343

Baum BR, Bailey LG (1997) The molecular diversity of the 5S rRNA gene in *Kengyilia alatavica* (Drobov) J.L. Yang, Yen & Baum (Poaceae: triticeae): potential genomic assignment of different rDNA units. Genome 40:215–228

Baum BR, Bailey LG (2012) Genetic diversity in the Red wild einkorn: *T. urartu* Gandilyan (Poaceae: Triticeae). Genet Res Crop Evol 60:77–87. doi:10.1007/s10722-012-9817-7

Baum BR, Johnson DA (1994) The molecular diversity of the 5S rRNA gene in barley (*Hordeum vulgare*). Genome 37:992–998

Baum BR, Johnson DA (1996) The 5 s rRNA gene units in ancestral two rowed barley (*Hordeum spontaneum* C. Koch) and bulbous barley (*H. bulbosum* L.): sequence analysis and phylogenetic relationships with the 5 s rDNA units of cultivated barley (*H. vulgare* L.). Genome 39:140–149

Baum BR, Johnson DA (1998) The 5S rRNA gene in sea barley (*Hordeum marinum* Hudson sensu lato): sequence variation among repeat units and relationship to the X haplome in barley (*Hordeum*). Genome 41:652–661

Baum BR, Johnson DA (2002) A comparison of the 5S rDNA diversity in the *Hordeumbrachyantherum–californicum* complex with those of the eastern Asiatic *Hordeum roshevitzii* and the South American *Hordeum cordobense* (Triticeae: Poaceae). Can J Bot 80:752–762. doi:10.1139/B02-057

Baum BR, Johnson DA (2003) The South African *Hordeum capense* is more closely related to some American Hordeum species than to the European *H. secalinum*—a perspective based on the 5S DNA units (Triticeae: Poaceae). Can J Bot 81:1–11

Baum BR, Johnson DA (2004) Differences between South American **H** haplome diploids and **I** haplome diploids from the perspective of the 5S rDNA gene in the genus *Hordeum*. Czech J. Plant Breed 40:45–50

Baum BR, Johnson DA (2008) Molecular confirmation of the genomic constitution of Douglasdeweya (Triticeae: Poaceae). Demonstration of the utility of the 5S DNA sequence as a genomic tool. Mol Genet Genom 279:621–628. doi:10.1007/s00438-008-0338-1

Baum BR, Johnson DA, Bailey LG (2001) Defining orthologous groups among multicopy genes prior to inferring phylogeny, with special emphasis on the Triticeae (Poaceae). Hereditas 135:123–138

Baum BR, Johnson DA, Bailey LG (2005) Ancient differentiation of the H and I haplomes in diploid Hordeum species based on 5S rDNA. Genome 48:610–618

Baum BR, Edwards T, Johnson DA (2008) Loss of 5S rDNA units in the evolution of *Agropyron*, *Pseudoroegneria*, and *Douglasdeweya*. Genome 51:589–598

Baum BR, Yang JL, Yen C, Agafonov AV (2011) A taxonomic synopsis of the genus *Campeiostachys* Drobov. J Syst Evol 49:146–159. doi:10.1111/j.1759-6831.2010.00106.x

Baum BR, Edwards T, Ponomareva E, Johnson DA (2012a) Are the great plains Wildrye (*Elymus canadensis*) and the Siberian Wildrye (*E. sibiricus*) conspecific? Botany 90:407–421

Baum BR, Edwards T, Mamuti M, Johnson DA (2012b) Phylogenetic relationships among the polyploid and diploid *Aegilops* species inferred from 5S rDNA units (Triticeae: Poaceae). Genome 55:1–17

Church GL (1958) Artificial hybrids of *Elymus virginicus* with *E. canadensis, interruptus, riparius*, and *wiegandii*. Amer J Bot 45:410–417

Clement M, Posada D, Crandall K (2000) TCS: a computer program to estimate gene genealogies. Mol Ecol 9(10): 1657–1660. http://darwin.uvigo.es/software/tcs.html. Accessed in 2005. Clement M, Derington J, Woolley S, Posada D 2005. TCS 1.21

Constance L (1964) Systematic botany—an unending synthesis. Taxon 13:257–273

Darriba D, Taboada GL, Doallo R, Posada D (2012) jModelTest 2: more models, new heuristics and parallel computing. Nature Methods 9: 772, https://code.google.com/p/jmodeltest2/wiki/. Accessed 2013

Davis PH, Heywood VH (1963) Principles of Angiosperm Taxonomy. D. Van Nostrand Company, Princeton

Dewey DR (1984) The genomic system of classification as a guide to intergeneric hybridization with the perennial Triticeae. In: Gustafson JP (ed) Gene manipulation in plant improvement. Plenum Publ. Corp, New York

Diaz O, Salomon B, von Bothmer R (1999) Genetic diversity and structure in populations of *Elymuscaninus* (L.) L. (Poaceae). Hereditas 131:63–74

Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. Mol Ecol Res 10:564–567 (Ver. 3.5.1.3. Accessed 2013)

Fahleson J, Okori P, Ákerblom-Expeby L, Dixelius C (2008) Genetic variability and genomic divergence of *Elymus repens* and related species. Pl Syst Evol 271:143–156. doi:10.1007/s00606-007-0623-1

Gouy M, Guindon S, Gascuel O (2010) SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. Mol Biol Evol 27(2):221–224

Gower JC (1966) Some distance properties of latent root and vector methods used in multivariate analysis. Biometrika 53:325–338

Guindon S, Gascuel O (2003) A simple, fast and accurate method to estimate large phylogenies by maximum-likelihood. Syst Biol 52:696–704

Helfgott DM, Mason-Gamer RJ (2004) The evolution of North American *Elymus* (Triticeae, Poaceae) Allotetraploids: evidence from phosphoenolpyruvate carboxylase gene sequences. Syst Bot 29:850–861. doi:10.1600/0363644042451017

Jaaska V (1992) Isoenzyme variation in the grass genus *Elymus* (Poaceae). Hereditas 117:11–22

Lartillot N, Lepage T, Blanquart S (2009) PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. Bioinformatics 25:2286–2288

Linde-Laursen I, 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). Pl Syst Evol 192:165–176

Linnaeus, C (1753) Species plantarum. Stockholm

Löve Á (1984) Conspectus of the Triticeae. Feddes Repertorium 95:425–521

Ma X, Zhang XQ, Zhou YH, Bai SQ, Liu W (2008) Assessing genetic diversity of Elymus sibiricus (Poaceae: Triticeae) populations from Qinghai-Tibet plateau by ISSR markers. Biochem Syst Ecol 36:514–522. doi:10.1016/j.bse.2008.o3.003

Mason-Gamer RJ, Burns MM, Naum M (2010) Reticulate evolutionary history of a complex group of grasses: phylogeny of *Elymus* StStHH allotetraploids based on three nuclear genes. PLoS One 5(6):e10989. doi:10.1371/journal.pone.0010989

Nei M, Li WH (1979) Mathematical model for studying genetic variation in terms of restriction endonucleases. Proc Nat Acad Sci USA 76:5269–5273

Nevski SA (1934) Tribe XIV. Hordeae Benth. In: Komarov VL, Rozhevits RYu, Shishkin BK (eds) Flora SSSR, vol 2 Grasses. Leningrad [Translated from Russian by the Israel Program for Scientific Translations, National Science Foundation and Smithsonian Institution], pp 590–728

Nicholas KB, Nicholas HB Jr (1997) GeneDoc©: a tool for editing and annotating multiple sequence alignments distributed by the authors Available from http://www.psc.edu/biomed/genedoc/ or http://www.nrbsc.org/downloads/. Accessed 2008

Peakall R, Smouse PE (2006) GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. Mol Ecol Notes 6:288–295

Peakall R, Smouse PE (2012) GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research–an update. Bioinformatics 28: 2537–2539. http://bioinformatics.oxfordjournals.org/content/28/19/2537. Accessed 2013

Popp M, Oxelman B (2001) Inferring the history of the polyploid *Silene aegaea* (Caryophyllaceae) using plastid and homoeologous nuclear DNA sequences. Mol Phylogenet Evol 20:474–481

Prost S, Anderson CNK (2011) TempNet: a method to display statistical parsimony networks for heterochronous DNA sequence data. Methods Ecol Evol 2:663–667. doi:10.1111/j.2041-210X.2011.00129.x

Ronquist F, Huelsenbeck J, Teslenko M (2011) Draft MrBayes version 3.2 Manual: tutorials and model summaries. Version 3.2.2, http://mrbayes.sourceforge.net/index.php. Accessed 2013

Sanders TB, Hamrick JL, Holden LR (1979) Allozyme variation in *Elymus canadensis* from the Tallgrass Prairie region: geographic variation. Amer Midl Naturalist 101:1–12

Sievers F, Wilm A, Dineen DG, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Söding J, Thompson JD, Higgins DG (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol 7:539. doi:10.1038/msb.2011.75

Sun GL, Daley T (2007) Molecular evolution and genome divergence at *RPB2* gene of the St and H genome in *Elymus* species. Pl Mol Biol 64:645–655. doi:10.1007/s11103-007-9183-6

Sun G, Salomon B (2003) Microsatellite variability and heterozygote deficiency in the arctic-alpine Alaskan wheatgrass (*Elymus alaskanus*) complex. Genome 46:729–773

Sun GL, Salomon B (2009) Molecular evolution and origin of tetraploid Elymus species. Breeding Sci 59:487–491

Sun GL, Pourkheirandish M, Komatsuda T (2009) Molecular evolution and phylogeny of the *RPB2* gene in the genus *Hordeum*. Annals Bot 103:975–983. doi:10.1093/aob/mcp020

Templeton AR, Crandall KA, Sing CF (1992) A cladistics analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data III. Cladogram estimation. Genetics 132:619–633

Tsvelev NN (1976) Zlaki SSSR. Leningrad, AkademiaNauk [Translated from Russian, Smithsonian Institution and National Foundation 1983]

Venables WN, Smith DM, and the R Core Team (2014) An introduction to R. Notes on R: a programming environment for data analysis and graphics. Version 3.0.3 (2014-03-06)

Wang RR-C, Hsiao C (1986) Differentiation of the **H** genome of the genus *Critesion* and one natural hybrid of *C. violaceum* and *C. bogdanii*. Can J Genet Cytol 28:947–953

Yang JL, Baum BR, Yen C (2008) A revision of the genus Roegneria C. Koch (Triticeae: Poaceae). J Sichuan AgricUniv 26:311–381

Yen C, Yang JL (1990) *Kengyliagobicola*, a new taxon from West China. Can J Bot 68:1894–1897

Yen C, Yang JL (2013) Biosystematics of the Triticeae. Genus *Elymus* 5:58–362 [Chinese]

Yen C, Yang JL, Baum BR (2005) *Douglasdeweya*: a new genus, with a new species and a new combination (Triticeae: Poaceae). Can J Bot 83:413–419