Allopolyploids of the Genus Elymus (Triticeae, Poaceae): a Phylogenetic Perspective

Roberta J. Mason-Gamer

University of Illinois, Chicago

Follow this and additional works at: http://scholarship.claremont.edu/aliso

Part of the Botany Commons, and the Ecology and Evolutionary Biology Commons

Recommended Citation


Available at: http://scholarship.claremont.edu/aliso/vol23/iss1/30
ALLOPOLYPLOIDS OF THE GENUS ELYMUS (TRITICAEA, POACEAE): A PHYLOGENETIC PERSPECTIVE

ROBERTA J. MASON-GAMER

University of Illinois at Chicago, Department of Biological Sciences, MC 066, 845 West Taylor Street, Chicago, Illinois 60607, USA
(robie@uic.edu)

ABSTRACT

The wheat tribe, Triticeae, includes many genomically distinct polyploid taxa. Elymus is an entirely allopolyploid genus, with all species containing the \( \text{St} \) genome of Pseudoroegneria. The \( \text{St} \) genome may be combined with one or more distinct genomes representing multiple, diverse diploid donors from throughout the tribe. This study includes a simultaneous phylogenetic analysis of new and previously published data from several distinct Elymus groups, including North American and Eurasian \( \text{StStHH} \) tetraploids, in which the \( \text{H} \) genome is derived from Hordeum, Eurasian \( \text{StStYY} \) tetraploids, in which the \( \text{Y} \) genome is derived from an unknown donor, and a putative \( \text{StStStStHH} \) hexaploid. Elymus species were analyzed with a broad sample of diploid genera from within the tribe using a combination of molecular data from the chloroplast and the nuclear genomes. The data confirm the genomic constitution of the \( \text{StStHH} \) and \( \text{StStYY} \) tetraploids, but do not provide additional information on the identity of the \( \text{Y} \)-genome donor. The genomic diversity in the hexaploid is greater than expected, inconsistent with the hypothesis of an \( \text{StStStHH} \) genome complement.

Key words: Elymus, Poaceae, polyploidy, reticulation, Triticeae, wheat tribe.

INTRODUCTION

The wheat tribe, Triticeae (Pooidae, Poaceae), is widely known for its economic importance. This monophyletic tribe includes three major grain crops—wheat, barley, and rye—along with several important forage grasses and numerous weedy, invasive species. From an evolutionary standpoint, the tribe’s complex reticulate history has been of interest for many years. At the diploid level, conflict among gene trees, especially among those based on chloroplast and nuclear DNA data (Mason-Gamer and Kellogg 1996a), suggests a history of gene exchange, lineage sorting, or a combination of both (Kellogg et al. 1996). Furthermore, polyploidy is common in the tribe; about 75% of the species with known chromosome numbers are polyploid (Löve 1984).

Allopolyploidy represents a reticulate process by which distinct genomes are united in a single nucleus. While there are problems with placing reticulate taxa within the bifurcating trees that are obtained using many methods of analysis (e.g., Hull 1979; Cronquist 1987; McDade 1992, 1998), studies of individual gene trees often allow these problems to be circumvented. Thus, molecular phylogenetic data have characterizing trees that are obtained using many methods of analysis (e.g., Hull 1979; Cronquist 1987; McDade 1992, 1998), studies of individual gene trees often allow these problems to be circumvented. Thus, molecular phylogenetic data have recently revealed the reticulate histories of several polyploid species or groups, for example: Gossypium L. (e.g., Cronn et al. 1996, 2003; Seelanan et al. 1997; Small and Wendel 2000), Geum L. (Smedmark et al. 2003), Glycine Wild. (e.g., Doyle et al. 2002; Rauscher et al. 2002), Oxalis L. (Emshwiller and Doyle 1998, 2002), Oryza L. (Ge et al. 1999), and Paeonia L. (e.g., Sang and Zhang 1999). In Triticeae, Elymus L. is a genomically heterogeneous polyploid group of about 140 species (Löve 1984). According to its genomic definition, Elymus comprises allopolyploid species with at least one set of chromosomes derived from Pseudoroegneria (Nevski) A. Löve (genome designation \( \text{St} \)). The \( \text{St} \) genome can also be found in both diploid and autotetraploid species, which are classified as Pseudoroegneria (Dewey 1984). In Elymus, the \( \text{St} \) genome can be combined with genomes from one or more Triticeae genera, including Hordeum L. (genome designation \( \text{H} \)), Agropyron Gaertn. (P), Australopyrum (Tzvelev) A. Löve (W), and an unknown donor (Y), in various allopolyploid combinations including \( \text{StStHH} \), \( \text{StStYY} \), \( \text{StStHHH} \), \( \text{StStS} \), \( \text{StStYY} \), \( \text{StStYYY} \), \( \text{StStHH} \), \( \text{StStYY} \), and \( \text{StStYYY} \) (e.g., Dewey 1967, 1968, 1970b, 1974, 1984; Jensen 1990, 1993, 1996; Salomon and Lu 1992, 1994; Lu and von Bothmer 1993; Lu et al. 1995). Other \( \text{St} \)-containing allopolyploids include Pascopyrum smithii (Rydib.) Barkworth & D. R. Dewey, which combines the Pseudoroegneria and Hordeum genomes with the \( \text{Ns} \) genome of Psathyrostachys Nevski in an \( \text{StStHHNsNsNsNs} \) octoploid configuration (Dewey 1975), and Thinopyrum A. Löve, some species of which are hypothesized to combine the \( \text{St} \) genome with the \( \text{E} \) and/or \( \text{J} \) genomes usually considered characteristic of Thinopyrum (e.g., Liu and Wang 1993; Zhang et al. 1996; Chen et al. 1998). Thus, the \( \text{St} \) genome, probably more than any other in Triticeae, plays an important role in the complex reticulate allopolyploid patterns that characterize the tribe.

This overview focuses on a subset of species from the genomically heterogeneous Elymus, and presents molecular phylogenetic data that address their relationships to the diploid members of the tribe. These species include (1) North American tetraploid species with presumed \( \text{StStHH} \) genome configurations (e.g., Dewey 1982, 1983a, b, 1984; Mason-Gamer 2001; Mason-Gamer et al. 2002; Helfgott and Mason-Gamer 2004), (2) Eurasian tetraploid species, with presumed \( \text{StStHH} \) or \( \text{StStYY} \) genome configurations (e.g., Lu and von Bothmer 1990, 1993; Salomon and Lu 1992, 1994; Lu 1993; Lu et al. 1995), and (3) a Eurasian hexaploid spe-
cies (E. repens L.), with a presumed StStStHHH genomic configuration (Cauderon and Saigné 1961; Dewey 1970a, 1976; Ørgaard and Anamthawat-Jönsson 2001). These genomically diverse allopolyploids are analyzed together with a broad sample of diploid genera from throughout the tribe, using three sources of chloroplast DNA (cpDNA) data, and sequences from the single-copy nuclear gene encoding granule-bound starch synthase (GBSSI).

MATERIALS AND METHODS

Both of the molecular data sets include a combination of new and previously published data. Specimens corresponding to the new data are currently in the possession of the author, and herbarium vouchers will be permanently deposited upon the completion of the ongoing study. Information about the previously published sequence data and the plants from which they were derived can be found in the corresponding publications, cited below.

The cpDNA data set includes restriction sites, sequences from trnT, trnL, and trnF genes and their intergenic spacers, and sequences of the rpoA gene. The existing restriction site data set (Mason-Gamer and Kellogg 1996b) includes most of the monogenomic genera and three species of Elymus (E. lanceolatus, E. trachycaulus, E. virginicus, and E. wawawaiensis; Mason-Gamer 2001) and from hexaploid E. repens (Mason-Gamer 2004). New data were obtained from E. abolini (China; PI531555), E. caninus L. (Uzbekistan; PI314205), E. ciliaris (China; PI531575), E. dentatus (Russia; PI628702), and E. mutabilis (Russia; PI628704). A 1.3 kilobase portion of the GBSSI gene was amplified, cloned, and sequenced as in Mason-Gamer (2001) using primers F-for, H-for, J-bac, L1-for, L2-bac, and M-bac (Mason-Gamer et al. 1998). Because multiple, distinct nuclear sequences are expected to coexist in allopolyploid individuals, multiple clones per individual (4–12 from each tetraploid, 20+ from E. repens) were partially sequenced and compared in an attempt to obtain, and then fully sequence, all variants within each individual. The ten new tetraploid sequences included here have been assigned GenBank accession numbers DQ159322–DQ159331.

The GBSSI data set was analyzed using PAUP* vers. 4.0b10 (Swofford 2002). Initial cladistic parsimony analyses were carried out with all characters equally weighted, except that those in ambiguously aligned regions were excluded from the analysis. Nucleotide frequencies, probabilities of different substitution types, and rate variation among sites were estimated on the strict consensus tree from the cladistic analysis using maximum likelihood (ML). Estimated model parameters were used as settings in a subsequent ML search under the general time reversible (GTR) model of sequence evolution (Yang 1994), with an ML-estimated proportion of sites presumed to be invariable (I) and with the remaining sites following a gamma (Γ) distribution (Gu et al. 1995; Waddell and Penny 1996). The initial search settings were: nucleotide frequencies A = 0.240954, C = 0.271580, G = 0.300879, and T = 0.186587; relative character state changes AC = 0.01382, AG = 3.49960, AT = 0.92930, CG = 1.54439, CT = 5.45252, GT = 1.00000; gamma shape parameter describing rate heterogeneity among sites = 0.728590; and proportion of invariable sites = 0.320805. Branch support was estimated based on a bootstrap analysis under a minimum-evolution criterion, with the distances estimated using ML under the same GTR + I + Γ model described above, and using the starting parameters listed above.

RESULTS

Chloroplast DNA Data

The restriction site data set consists of 129 variable sites, and is discussed in detail elsewhere (Mason-Gamer and Kel-
logg 1996b). The tRNA gene/spacer PCR products range from 1643 to 1763 base pairs (bp) in length, and the length of the aligned data set (including insertions/deletions [indels]) is 2157 bp. The rpOA PCR products range from 1345 to 1371 bp in length, and the length of the aligned data set is 1393 bp. In both cases, sequence alignment was straightforward and all data were included in the analysis.

The St-containing allopolyploids are grouped with Pseudoroegneria, one of the two hypothesized genome donors (Fig. 1). The Elymus–Pseudoroegneria group includes two other genera, Dasyphyrum (Coss. & Durieu) T. Durand (genome designation V) and Thinopyrum (genomes J and/or E), which therefore represent additional potential chloroplast genome donors to Elymus. Although bootstrap support for the Elymus–Pseudoroegneria–Dasyphyrum–Thinopyrum relationship is very weak, the branch lengths indicate that the Elymus chloroplast genomes are very similar to those from Pseudoroegneria. There is little resolution within the Elymus–Pseudoroegneria–Dasyphyrum–Thinopyrum clade other than the well-supported Dasyphyrum and Thinopyrum clades themselves, and the branches leading to most of the Elymus and Pseudoroegneria individuals are very short.

### Starch Synthase Data and Tetraploid Elymus

The starch synthase PCR products range from 1193 to 1360 bp in length, and the length of the aligned data set, including indels, is 1598 bp. A total of 114 bp (two regions, 65 and 49 bp) were judged difficult to align such that the positional homology was questionable; these positions were excluded from the analysis.

All of the tetraploid Elymus species include a Pseudoroegneria-like (St) sequence (Fig. 2). The pattern of diversity in the Pseudoroegneria–Elymus clade does not fully correspond to either geography or genomic constitution. For example, there are two main subclades within the clade, and each includes both Eurasian and North American Elymus tetraploids. Furthermore, some individuals have copies of gene sequences from both of the subclades (E. lanceolatus 1a and 1d, and E. wawawaiensis 3a and 3b). On the other hand, the three Eurasian SisHH species (E. caninus, E. dentatus, and E. mutabilis) do form a clade within one of the subclades (89% bootstrap support). Elymus ciliaris does not fall cleanly into either subclade; preliminary analyses suggest that this is a naturally occurring recombinant sequence, having been recovered from multiple independent PCR reactions.

All of the presumed SisHH tetraploids yield Hordeum-like (H) sequences. The H-genome sequences from the North American tetraploids are very similar to one another (0.0036–0.0091 uncorrected pairwise distance) and, along with the sequence from Eurasian E. dentatus, form a well-supported clade with H. californicum Covas & Stebbins. The three Eurasian SisHH Hordeum-like sequences are more divergent from one another (0.0437–0.0538 uncorrected pairwise distance), and do not form a clade.

A third clade of tetraploid Elymus sequences apparently represents the Y-genome sequences from the SisHY tetraploids E. abolinii and E. ciliaris. There is no strong evidence to indicate which of the monogenic genera are most closely related to the Y-genome clade. When the very poorly supported nodes (<50%; unmarked nodes in Fig. 2) are disregarded, many of the basal nodes on the tree collapse, leaving the Y-genome clade in an unresolved position relative to the monogenic genera.

### Starch Synthase Data and Hexaploid Elymus repens

Like the SisHH tetraploids, the two E. repens individuals include both Pseudoroegneria-like (1c and 6g) and Hordeum-like sequences (1e and 6dd) (Fig. 2). The positions of the E. repens St- and H-genome sequences within their respective clades are not resolved. Sequences 1c and 6g do not form a monophyletic group within the Pseudoroegneria–Elymus clade, and 1e and 6dd form only a poorly supported (<50% bootstrap) group within the Hordeum–Elymus clade.

Elymus repens sequences appear in two other clades within Triticeae. The third E. repens clade groups sequences 6a2 and 1a1 with Taeniatherum Nevski with strong support (100% bootstrap). The Taeniatherum–Elymus clade is part of a larger clade with Agropyron s.s. and Eremopyrum (Ledb.) Jaub. & Spach. The fourth Elymus repens clade includes sequences 1q and 6th and does not appear to be closely related to any of the included monogenic genera. This clade superficially appears to group with Dasyphyrum and the Y-genome clade on the best maximum-likelihood tree, but these relationships lack support (<50% bootstrap; Fig. 2); thus, the donor of the GBSSI sequences in the fourth E. repens clade is currently unknown. Earlier analyses of GBSSI exon data within a taxonomically broader sample of grasses (Mason-Gamer 2004) disclosed a fifth E. repens sequence type that, while related to sequences in Triticeae, appeared to fall outside of Triticeae (Fig. 3). While this sequence clearly appears to be pooid in origin, the identity of its donor is not revealed by the data currently available.

### DISCUSSION

The wheat tribe has been the subject of extensive studies of meiotic behavior in hybrids. Based on these, the monogenic genera have been delimited according to whether their chromosomes pair at meiosis. The generic definition of the allopolyploid genus Elymus is somewhat more complicated. Members of all species of the genus have a Pseudoroegneria (St) genome, but the other Elymus genomes represent diverse sources. Thus, Elymus is genomically heterogeneous, as has been further demonstrated in phylogenetic analyses of DNA sequence data.

Chloroplast DNA data are consistent with three possible maternal genome donors: Pseudoroegneria, Dasyphyrum, or Thinopyrum (Fig. 1). As previously suggested (Mason-Gamer et al. 2002), Pseudoroegneria is the most likely donor. First, it is consistent with the existing cytogenetic data that formed the basis of the initial hypothesis that the St genome is a component of all Elymus species. Second, although some studies have suggested that Thinopyrum may have been a genome contributor to E. repens (e.g., Ørgaard and Ananthawat-Jónsson 2001), molecular data from nuclear genes encoding GBSSI (Fig. 2), phosphoenolpyruvate carboxylase (PepC; Hellgott and Mason-Gamer 2004), and the tissue-ubiquitous form of β-amylase (β-amy; R. J. Mason-Gamer unpubl. data) provide no evidence for the presence of Thinopyrum-like or Dasyphyrum-like sequences in any Elymus...
Fig. 1.—Consensus of 63,465 most-parsimonious cpDNA trees based on a combination of data from restriction sites, sequences of the \textit{trnT}, \textit{trnL}, and \textit{trnF} genes and their associated noncoding regions, and sequences of the \textit{rpoA} gene. \textit{Elymus} polyploids are shown in boldface. Labels following the \textit{Elymus} species names indicate whether each species is native to North America (NA) or Eurasia (E), and provide the presumed genomic constitution of each (\textit{StStHH}, \textit{StStYY}, or \textit{StStStStHH}). Numbers following some species names correspond to specific individual plants within a species, and match the numbers used in previous studies (Mason-Gamer et al. 2002; Mason-Gamer 2004), in which specimen information can be found. Maximum-parsimony bootstrap support values $\geq 50\%$ are shown above the nodes.
Fig. 2.—Maximum-likelihood tree based on GBSSI sequence data, analyzed under a GTR + I + Γ model of sequence evolution. *Elymus* polyploids are shown in boldface. Numbers following some species names correspond to specific individual plants within a species, and match the numbers used in previous studies (Mason-Gamer 2001, 2004), in which specimen information can be found. The character(s) after each numerical identifier is (are) specific to a single cloned sequence from within that individual. Brackets on the right and the corresponding thick nodes mark all of the clades in which *Elymus* sequences are found; boldface labels between brackets indicate whether each species is native to North America (NA) or Eurasia (E), and provide the presumed genomic constitution of each. Support values ≥50% are shown above the nodes, and were based on a minimum-evolution criterion with distances estimated using maximum likelihood under a GTR + I + Γ model of sequence evolution.
species examined to date. Third, among the available molecular data sets, a *Pseudoroegneria–Dasypyrum–Thinopyrum* group is unique to cpDNA trees, and the level of cpDNA divergence among them is low (Fig. 1). Thus, their cpDNA similarity may reflect introgression of the chloroplast genome from *Pseudoroegneria* into *Thinopyrum* and *Dasypyrum* (Kellogg et al. 1996). The cpDNA data thus highlight the consistent involvement of *Pseudoroegneria* as the maternal genome donor to *Elymus* species of varying polyploid genomic content. This appears to be a general pattern for all St-containing allopolyploids examined to date (Redinbaugh et al. 2000; McMillan and Sun 2004).

The nuclear starch synthase gene data show, as expected, multiple genome donors to *Elymus* (Fig. 2). The data support the hypothesized origins of the StStHH tetraploid from both North America (*E. elymoides*, *E. glaucus*, *E. lanceolatus*, *E. trachycladus*, *E. virginicus*, and *E. wawawaiensis*) and Eurasia (*E. caninus*, *E. denticus*, and *E. mutabilis*). Both Hordeum-like and *Pseudoroegneria*-like starch synthase gene sequences have been recovered from each presumed StStHH tetraploid. The St-genome sequences from the North American StStHH species are scattered throughout the *Pseudoroegneria–Elymus* clade, and are represented in both of two main subclades that divide the group. The two subclades do not appear to correspond to either geographical or taxonomic boundaries (Mason-Gamer 2001), and some North American individuals contain representatives of both subclades. The St-genome sequences from the three Eurasian StStHH species, in contrast to the North American species, form a clade. The pattern in the Hordeum–Elymus clade is nearly the reverse: the six North American H-genome sequences are very similar to one another (and to *H. californicum*), while the H-genome sequences from the three Eurasian StStHH species are much more diverse. The pattern suggests a single H-genome donor to the North American species, while several Hordeum species may have been involved in multiple origins of the Eurasian species.

Hybridization and cytogenetic studies have yet to reveal a potential diploid Y-genome donor (e.g., Dewey 1984 and references therein). The GBSSI gene sequence data from the two StStYY tetraploids, *E. abolithii* and *E. ciliaris*, provide little additional information to help solve the mystery of the Y-genome’s origin. The Y-genome clade on the GBSSI tree (Fig. 2) does not show a close relationship to any of the monogenic genera on the tree, suggesting that none of the included genera are very closely related to the donor of this genome. Preliminary data from genes encoding PEPc and β-amylase (*R. J. Mason-Gamer unpubl. data) also leave the position of the Y-genome sequences unresolved.

The GBSSI data from *E. repens* reveal several distinct starch synthase gene copies (Fig. 2). Both of the *E. repens* individuals have a *Pseudoroegneria*-like sequence, consistent with earlier cytogenetic data (Dewey 1970a, 1976), and a Hordeum-like sequence, consistent with cytogenetic analyses (Cauderon and Saigne 1961) and genomic in-situ hybridization results (Ørgaard and Anamthawat-Jónsson 2001). Furthermore, both of the individuals included here have two additional, phylogenetically distinct, gene copies. One of these forms a well-supported clade with *Taeniatherum* (100% bootstrap). This clade is grouped with *Agropyron* and *Eremopyrum* with moderate (84%) support. Both *Elymus repens* and *Taeniatherum* are widespread introduced weeds in the USA, but they are morphologically very distinct, so the discovery of a *Taeniatherum*-like gene copy in *E. repens* was unexpected. Preliminary analyses of three Eurasian accessions of *E. repens* (*R. J. Mason-Gamer unpubl. data) reveal *Taeniatherum*-like sequences in each, suggesting that this gene copy was acquired prior to the introduction of *E. repens* to North America. The analysis also reveals a fourth GBSSI gene copy within *E. repens* (1q and 6h), which does not group strongly with any other members of the tribe, so its origin remains unknown. While these sequences superficially appear to be related to the *Dasypyrum* and/or Y-genome sequences, there is no appreciable support for these relationships (bootstrap <50%).

Previous analyses of *E. repens* GBSSI exon sequences within a broader phylogenetic context (Mason-Gamer 2004) revealed the presence of a fifth, distinct *E. repens* gene copy—one that falls outside of Triticeae (Fig. 3). While the copy is apparently pooid, the sparse sampling within the subfamily does not allow speculation about the identity of its donor. The closest relative within the sample is *Cutandia Willk.*, but the related *E. repens* sequence is more divergent from the *Cutandia* sequence than any of the other *E. repens* sequences are from their apparent donors (Mason-Gamer 2004). Thus, a more precise hypothesis regarding the identity of the donor of the fifth *E. repens* sequence awaits a broader sample of GBSSI sequences from within Pooidae, an undertaking being pursued elsewhere (*J. I Davis pers. comm.*). Another intriguing question about this distinct sequence is whether it was acquired before or after the introduction of *E. repens* to North America over a century ago. With additional *E. repens* samples representing Eurasia, it may be possible to address this interesting issue.

Both of the *E. repens* individuals included here (1 and 6) have all five of the GBSSI sequence variants described above. The variation within this species is thus more com-
plex than expected in an allohexaploid, in which no more than three distinct copies would be expected. Therefore, the origin of the species may have involved other processes, such as introgression, in addition to polyploidy. It remains to be seen whether results based on other genes will be as complex as those from the analyses of the GBSSI data. Data from genes encoding PepC and β-amy will be added to those from GBSSI, and are expected to show whether or not the complex pattern of GBSSI diversity in *E. repens* represents a genome-wide pattern.

Within Triticeae, the St genome plays a major role in polyploid evolution—it is a component of many distinct allopolyploid combinations at tetraploid, hexaploid, and octoploid levels. This paper presents a simultaneous analysis of three such genomic combinations, including North American and Eurasian *Sisymbrium* tetraploids, Eurasian *Sisymbrium* tetraploids, and North American representatives of presumed *Sisymbrium* *E. repens*, a native of Eurasia. However, these groups represent only a portion of the diversity found among the St-containing allopolyploids. A fuller understanding of the role of this genome in the evolution of the tribe will require the use of multiple molecular markers, and the investigation of additional St-containing allopolyploids, including: (1) more representatives within the *Elymus* groups already included here; (2) representatives of additional genome combinations at higher ploidy levels, including the StIIHY and StIIYYP hexaploids from Eurasia, the StIYYWW hexaploids of Australasia, and the StIIHHImSnSnSn octoploid *Pascopyrum smithii*; (3) samples of the putative St-containing members of *Thinopyrum*; and (4) St-genome autopolyploid representatives of *Pseudoroegneria*. The gradual untangling of the history of polyploidy and hybridization within Triticeae will further enhance our understanding of the tribe’s complex and fascinating evolutionary history.

**ACKNOWLEDGMENTS**

Sincerest thanks to D. Megan Helfgott and Mary Barkworth for ongoing discussions of the Triticeae, to Travis Columbus for reviewing the manuscript, to the USDA for seeds, to the Organizing Committee of the Monocots III/Grasses IV Symposium, and to Pilar Catalán for organizing the Pooidae symposium session. The work was supported by NSF-DEB 9774181.

**LITERATURE CITED**


—. 1996. Genome analysis of Eurasian *Elymus thoroldianus*,


