Molecular Phylogenetics and Evolution 54 (2010) 10–22

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

Phylogenetic relationships and reticulation among Asian *Elymus* (Poaceae) allotetraploids: Analyses of three nuclear gene trees

Roberta J. Mason-Gamer *, Melissa M. Burns [1], Marianna Naum [2]

Department of Biological Sciences, The University of Illinois at Chicago, MC 066, Chicago, IL 60607, USA

ARTICLE INFO

Article history:
Received 19 January 2009
Revised 17 September 2009
Accepted 3 October 2009
Available online 8 October 2009

Keywords:
Elymus
Phylogeny
Polyploidy
Reticulation
Triticeae

ABSTRACT

This phylogenetic study focuses on a subset of the species in *Elymus*—specifically, the endemic Asian tetraploids presumed to combine the **St** genome from *Pseudoroegneria* with the **Y** genome from an unknown donor. The primary goals were to (1) determine whether the **St** and **Y** genomes are derived from phylogenetically distinct donors; (2) identify the closest relative, and potentially the likely donor, of the **Y** genome; and (3) interpret variation among **StStYY** species in terms of multiple origins and/or introgression. The goals were addressed using phylogenetic analyses of sequences from three low-copy nuclear genes: phosphoenolpyruvate carboxylase, β-amylase, and granule-bound starch synthase I. Data sets include 16 **StStYY** individuals representing nine species, along with a broad sample of representatives from most of the monogenomic (i.e., non-allopolyploid) genera in the tribe. To briefly summarize the results: (1) the data clearly support an allopolyploid origin for the Asian tetraploids, involving two distinct donors; (2) the **Y** genome was contributed by a single donor, or multiple closely-related donors; (3) the phylogenetic position of the *Elymus* **Y** genome varies among the three trees and its position is not strongly supported, so the identity of the donor remains a mystery; and (4) conflicts among the gene trees with regard to the **St**-genome sequences suggest introgression involving both *Elymus* and *Pseudoroegneria*.

1. Introduction

Allopolyploid species present numerous challenges to molecular phylogenetic studies. While data from the maternally-inherited chloroplast genome are generally easy to obtain from polyploids, they provide an incomplete picture of polyploid origins. Nuclear data are potentially more informative with respect to the reticulate nature of allopolyploids, but they are more expensive and time-consuming to obtain. Furthermore, data from nuclear markers can be difficult to interpret in groups where ploidy levels are not known at the outset, or when sequences are affected by recombination, copy conversion or loss, or within-genome duplication. In spite of these complications, nuclear sequence data have clarified tangled relationships among polyploids and their progenitors in many plant groups; in recent years, data from single- and low-copy markers have proven especially informative (e.g., Ainouche et al., 2004; Brysting et al., 2007; Emshwiller and Doyle, 2002; Fortune et al., 2008; Ge et al., 1999; Joly et al., 2006; Lihová et al., 2006; Mason-Gamer, 2001, 2008; Petersen et al., 2006; Popp and Oxelman, 2001; Rodríguez and Spooner, 2009; Smedmark et al., 2005; Straub et al., 2006). The present study applies phylogenetic data from three single-copy nuclear loci to a group of Asian allotetraploids in the wheat tribe, Triticeae (Poaceae).

The relationships within the Triticeae have received considerable attention, partly because of the economic importance of the group (which includes wheat, barley, and rye), and partly because the many published Triticeae phylogenetic data sets have failed to converge on a straightforward estimate of the relationships among the non-allopolyploids (Kellogg et al., 1996; Mason-Gamer, 2005; Seberg and Petersen, 2007). The tribe's numerous allopolyploids further complicate phylogenetic analyses, due to their explicitly reticulate origin (Kellogg, 1989; Kellogg et al., 1996). Although the cultivated wheats are the most economically important allopolyploids in the Triticeae, and have thus received considerable attention, the allopolyploid genus *Elymus* L. is in some ways more interesting from an evolutionary standpoint, with its many species, wide natural distribution, morphological variability, and variety of distinct genome combinations (Dewey, 1984; Löve, 1984).

The circumscription of *Elymus* has changed through time, and still varies considerably among treatments in current use (Barkworth, 2000). The practice of defining Triticeae genera by their genomic compositions, as inferred from patterns of chromosome pairing, has been critiqued for several legitimate reasons (Seberg and Petersen, 1998). However, the genomic definition of *Elymus*

* Corresponding author. Address: University of Illinois at Chicago, Department of Biological Sciences, MC 066, 845 W. Taylor St., Chicago, IL 60607, USA. Fax: +1 312 413 2435.
E-mail address: robie@uic.edu (R.J. Mason-Gamer).
[1] Present address: Exilixis, Inc., San Francisco, CA 94083, USA.
[2] Present address: Division of Microbiology, Center for Food Safety and Applied Nutrition, Food and Drug Administration, College Park, MD 20740, USA.

1055-7903/$ - see front matter © 2009 Published by Elsevier Inc.
doi:10.1016/j.ympev.2009.10.002

is more reflective of its evolutionary history than are alternative definitions based on morphology, and is the definition followed here. According to its genomic definition (Dewey, 1984; Löve, 1984), *Elymus* comprises about 150 allopolyploid species with at least one set of chromosomes derived from *Pseudoroegneria* (Nevski) Á. Löve (genome designation **St**). In *Elymus*, the **St** genome can be combined with genomes from one or more Triticeae genera, including *Hordeum* L. (genome designation **H**), *Agropyron* Gaertn. (**P**), *Australopyrum* (Tzvelev) Á. Löve (**W**), and an unknown donor (**Y**), in various allopolyploid combinations including **StStHH**, **StStYY**, **StStHHHH**, **StStStStHH**, **StStStStYY**, **StStYYYY**, **StStHHYY**, **StStYYWW**, and **StStYYPP** (Dewey, 1967, 1968, 1970, 1974, 1984; Jensen, 1990, 1993, 1996; Lu et al., 1995; Lu and von Bothmer, 1991, 1993; Salomon, 1993; Salomon and Lu, 1992, 1994a). Other **St**-containing allopolyploids include *Pascopyrum smithii* (Rydb.) Á. Löve, which combines the *Pseudoroegneria* and *Hordeum* genomes with the **Ns** genome of *Psathyrostachys* Nevski in an **StStHHNsNsNsNs** octoploid configuration (Dewey, 1975), and *Thinopyrum* Á. Löve, some species of which are hypothesized to combine the **St** genome with the **E** and/or **J** genomes usually considered characteristic of *Thinopyrum* (e.g., Chen et al., 1998; Liu and Wang, 1993; Zhang et al., 1996). Thus, the **St** genome, probably more than any other in Triticeae, plays an important role in the complex reticulate allopolyploid patterns that characterize the tribe.

The present analyses focus on the **StStYY** *Elymus* tetraploids, which comprise 30–40 species restricted to temperate Asia (Lu and Salomon, 1992). Numerous cytogenetic studies have addressed the genome content of one or more species in this group (Dewey, 1974, 1980a,b; Jensen, 1989, 1990; Jensen and Hatch, 1989; Lu et al., 1990, 1995; Lu and von Bothmer, 1989, 1990a,b, 1991; Sakamoto and Muramatsu, 1966). These studies highlight the lack of pairing between the **St**, **Y**, and/or **H** genomes, suggesting that the **Y** genome donor is unlikely to have been either *Hordeum* (**H**) or *Pseudoroegneria* (**St**). No putative **Y**-genome diploids have been identified, so the origin of the genome remains unknown. Analyses of repetitive DNA markers (Svitashev et al., 1996), and microsatellites and RAPDs (Sun et al., 1997), show a genetic distinction between the genomically-defined **StStYY** and **StStHH** *Elymus* species, and are thus consistent with the cytogenetic studies. More recently, the phylogenetic affinities of the genomes in this group were explored in two studies using DNA sequence data. Both studies included a broad sample of **StStYY** tetraploids (along with other *Elymus* allopolyploids), and a somewhat limited selection of monogenomic (i.e., non-allopolyploid) genera. One of these, based on internal transcribed spacer (ITS) sequences of the nuclear rDNA repeat (Liu et al., 2006) revealed extensive sequence variation within and among sequences from the **StStYY** individuals, but did not uncover any obvious pattern consistent with allopolyploidy (i.e., the sequences did not form distinct clades representing the presumed **St** and **Y** genomes). Based on their results, Liu et al. (2006) concluded that the **Y** genome was probably a derivative of the **St** genome. In contrast, a similar analysis based on sequences of the nuclear gene encoding the b subunit of RNA polymerase II (*RPB2*) showed intra-individual variation consistent with allotetraploidy (Sun et al., 2008): two phylogenetically divergent sequence types were recovered from within individuals, suggesting two distinct donors.

This study further explores the allopolyploid origin of the Asian tetraploid *Elymus* species using data from three low-copy nuclear genes, and a more extensive sample of representatives from most of the monogenomic genera in the tribe. The primary goals were to (1) determine whether the **St** and **Y** genomes are derived from phylogenetically distinct donors; (2) identify the closest relatives (and possible donors) of the genomes—especially the **Y** genome, whose identity has been more elusive; and (3) interpret variation among **StStYY** species in terms of multiple origins and/or introgression. In summary, the data clearly support an allopolyploid origin for the Asian tetraploids, involving donors from distinct clades; the **Y** genome was contributed from a single unknown donor (or multiple closely-related donors); the phylogenetic position of the **Y** genome is variable among trees and lacks support; and the **St** genome sequences have been subject to reticulation, possibly as a result of introgression within and between *Elymus* and *Pseudoroegneria*.

2. Materials and methods

2.1. Samples

The present analyses include 16 tetraploid individuals representing nine presumed **StStYY** tetraploid species (Table 1). This sample includes a broad representation of the geographical range of the **StStYY** group (Lu and Salomon, 1992), and includes representatives of the **StStYY** species groups based on previous analyses of morphology, hybridization, and genome pairing. These include the *E. tibeticus* group (represented here by *E. gmelinii* and *E. pendulinus*; Salomon and Lu, 1994b), the *E. semicostatus* group (*E. abolinii*, *E. nevskii*, and *E. semicostatus*; Salomon and Lu, 1994a,b), and the *E. parviglumis* group (*E. antiquus*, *E. caucasicus* and *E. longearistatus*; Lu and von Bothmer, 1993). (*Elymus ciliaris*, also included here, was not placed within one of these groups.) The present sample is not sufficient to test the groups, nor was that the intent; the proposed groups were used as a guide to ensure a diverse sample.

Three single- or low-copy nuclear genes were amplified, and multiple clones were checked with the goal of obtaining sequences representing both sets of genomes. For each gene, *Elymus* sequences were analyzed along with a reasonably broad sample of the tribe's known genomic diversity, including representatives of 15 monogenomic genera (Table 2). These include the donor of the **St** genome (*Pseudoroegneria*), and all of the genomes known to co-occur with **St** in allopolyploids: *Hordeum* (**H**), *Agropyron* (**P**), *Australopyrum* (**W**), *Psathyrostachys* (**Ns**; this genome is represented by the tetraploid *Leymus racemosus* ssp. *sabulosus* (M. Bieb.) Tzvelev in the pepC data set), and *Thinopyrum* (**J** and/or **E**). No monogenomic **Y**-genome species are known. Additional monogenomic genera included were *Aegilops*, *Crithopsis* (except for pepC), *Dasypyrum*, *Eremopyrum*, *Henrardia* (except for pepC), *Heteranthelium*, *Peridictyon*, *Secale*, *Taeniatherum*, and *Triticum*. While generic definitions within the Triticeae have been extremely variable (reviewed in Barkworth, 2000), this sample represents nearly all of the monogenomic genera accepted in genome-based classifications of the tribe. All three trees were rooted with a representative of *Bromus* L.; Bromeae and Triticeae have been shown to form a single clade, with Bromeae as either sister or paraphyletic to a monophyletic Triticeae (Davis and Soreng, 2007; Grass Phylogeny Working Group, 2001).

2.2. Molecular methods and alignment

All but four of the **StStYY** *Elymus* sequences (Table 1) are newly published here. Nearly all of the sequences from the monogenomic species, with a few exceptions as noted, were previously published in various sources (Table 2). Information about the data and taxa can be found therein, but the primary discussions about the characteristics of each marker and data set are: pepC—(Helfgott and Mason-Gamer, 2004); β-amylase—(Mason-Gamer, 2005); and GBSSI—(Mason-Gamer, 2001; Mason-Gamer et al., 1998). Based on studies of grass genomes in crop species, the three nuclear markers appear to be on three different chromosomes (more below). This is a tentative assumption, based on a small number of

Table 1
Tetraploid **StStYY** species.

	USDA PI	#	Genome	pepC	β-Amylase	Waxy
Elymus abolinii (Drobow) Tzvelev	531555	1	St	b	b	a-DQ159322[1]
			Y	a	a	b-DQ159323
Elymus abolinii	531557	2	St	b	a	k
			Y	i	—	a
			Y			e
Elymus antiquus (Nevski) Tzvelev	632564	1	St	g	a	a
			Y	a	f	b
Elymus antiquus	564958	2	St	a	a	g
			Y	c	e	a
Elymus antiquus	619528	3	St	b	a	b
			St	d	e	a
			Y	c		
Elymus antiquus	564957	4	St	c	d	b
			Y	h	a	a
Elymus caucasicus (K.Koch) Tzvelev	531573		St	a	b	—
			Y	d	a	a
			Y	—	—	b
Elymus ciliaris (Trin.) Tzvelev	531575	1	St	b	g	g-DQ159327
			Y	c	h	b-DQ159326
			Y	h	—	—
Elymus ciliaris	531577	2	St	d	f	a
			Y	a	i	b
Elymus ciliaris	531576	5	St	a	e	a
			St	—	—	b
			Y	c	a	c
Elymus gmelinii (Ledeb.) Tzvelev	499447		St	—	e	a
			Y	a	f	b
Elymus longearistatus (Boiss.) Tzvelev	401277		St	c	—	—
			Y	aa	e	d
Elymus nevskii Tzvelev	314620		St	a	k	a
			Y	b	b	b
Elymus pendulinus (Nevski) Tzvelev	499452		St	a	i	a
			Y	b	b	b
Elymus semicostatus (Nees ex Steud.) Melderis	271522	1	St	d	c	f
			Y	a	a	d
Elymus semicostatus	207453	2	St	a	—	a
			Y	—	a	e

Plant introduction (PI) numbers were assigned by the USDA. Numbers (#) identify individuals within species and letters under gene names identify cloned sequences from the **St** or **Y** genome from each individual; these correspond to the labels on the trees (Figs. 1–4). Dashes show where a genome was not recovered for a gene/individual. Genbank accession numbers for new sequences are, in order listed: pepC, GQ844927–GQ844958; β-amylase, GQ847678–GQ847706; and waxy, GQ847708–GQ847736.
[1] Mason-Gamer (2007a).

grass species, but for this study the three genes are assumed to represent independent phylogenetic estimates.

Other than the details of the primers and cycling parameters, similar molecular methods were followed for each of the three nuclear gene fragments; detailed protocols are found in the works cited above for each marker. In brief: for each *Elymus* individual, 3 PCR replicates were run per individual and combined prior to cloning, in order to counter the potential effects of PCR drift (Wagner et al., 1994). PCR products were cleaned on columns (Qiagen); cleaned products were cloned into pGEM-T Easy vectors (Promega), and transformed into *E. coli* JM109 competent cells (Promega) following Promega's protocol, except that all reactions were halved. Cloned fragments were amplified directly from white colonies using the same primers as were used for the original PCR, in 30–40 μl reactions with 0.5 U Taq polymerase (Sigma), a 1× concentration of the included 10× buffer, 45–60 nmol $MgCl_2$, 6–8 nmol of each nucleotide, and 30–40 pmol of each primer. Amplified fragments were cleaned using 1 U shrimp alkaline phosphatase (USB) and 5 U exonuclease I (USB), at 37 °C for 15 min. For each gene, 12–24 cloned PCR products from within each tetraploid individual were screened with single primers; this process yields partial sequences (about 650 bp) of one strand. Based on comparisons among these preliminary sequences, clones from within individuals that differed by more than three substitutions were fully sequenced in both directions, and the resulting sequences were added to the data set. Once it became clear that there were two main sequence types per individual in most cases, the screening process was focused toward obtaining representatives of both categories. If both types were not recovered in a sample of 12 clones, the corresponding gene from that individual was re-amplified and cloned, and 12 additional sequences were screened.

New sequences were manually added to existing aligned data sets, and alignments were inspected for chimeric sequences. In studies involving intra-individual variation (such as in allopolyploids), PCR-mediated recombination can yield chimeric products (Bradley and Hillis, 1997; Cronn et al., 2002; Judo et al., 1998). A small number of recombinants were easily identified as chimeras, either by inspection of alignments prior to phylogenetic analysis, or by closer examination of sequences that were tentatively identified as recombinants by their placement on very long branches following a short, incomplete maximum parsimony analysis. Chimeric sequences were interpreted as PCR artifacts and were discarded from the analyses.

The pepC gene is a member of a three-copy family in grasses (Lepiniec et al., 1993); the sequences used here appear to be homologous to the widely-expressed housekeeping copy. Based on the location of similar sequences in the rice genome (Genbank AP005781 and AP005802) and a comparative grass genome map (Devos and Gale, 1997), this gene copy is assumed to be on the Triticeae group 5 homoeologous chromosomes. The original Triticeae pepC data set (Helfgott and Mason-Gamer, 2004) combined two fragments designated region 1 (approximately 1 kb; Genbank AY553236–AY553269) and region 2 (approximately 600 bp; Genbank AY548399–AY548432); the present data set includes just re-

Table 2
Non-allopolyploid representatives of the Triticeae.

Species name	Accession	#	pepC	β-Amylase	GBSSI
Aegilops bicornis (Forsskål) Jaub. & Spach.	Morrison s.n.		–	[5]AY821686	–
Aegilops caudata L.	G 758		–	[5]AY821687 AY821688 AY821689	[1]AF079262
Aegilops comosa Sibth. & Smith	G 602		[3]AY553236	[5]AY821690 AY821696	–
Aegilops speltoides Tausch	Morrison s.n.		–	–	[1]AF079267
Aegilops tauschii Coss.	Morrison s.n.		–	[5]AY821695	[1]AF079268
Aegilops uniaristata Vis.	G 1297		–	[5]AY821691	[1]AF079270
Agropyron cristatum (L.) Gaertn.	PI 279802	1	[3]AY553237	[5]AY821697	[1]AF079271
Agropyron cristatum	PI 281862	2	–	–	[2]AY011002
Agropyron mongolicum Keng	D 2774		–	–	[2]AY011003
Australopyrum retrofractum (Vickery) Á. Löve	PI 533013		–	[5]AY821692	[1]AF079272
Australopyrum velutinum (Nees) B.K.Simon	D 2873-2878		[3]AY553238	[5]AY821693	[2]AY011004
Crithopsis delileana (Schult.) Rosch.	H 5562		–	[5]AY821694	[7]GQ847707
Dasypyrum 0076illosum (L.) Candargy	PI 251478	1	–	[5]AY821698	[1]AF079274
Dasypyrum villosum	PI 470279	2	–	[5]AY821699	–
Dasypyrum villosum	D 2990	3	[3]AY553240	–	–
Eremopyrum bonaepartis (Spreng.) Nevski	H 5554		–	[5]AY821700	[2]AY011005
Eremopyrum distans (C.Koch) Nevski	H 5552		–	[5]AY821701	[2]AY011006
Eremopyrum orientale (L.) Jaub. & Spach	H 5555		[3]AY553254	[5]AY821702	[2]AY011007
Henrardia persica (Boiss.) C.E.Hubb.	H 5556		–	[5]AY821703	[1]AF079276
Heteranthelium piliferum (Banks & Sol.) Hochst.	PI 402352		[3]AY553255	[5]AY821704	[1]AF079277
Hordeum bogdanii Wilensky	PI 531762	1	–	[7]GQ847675	[6]EU282316
Hordeum bogdanii	PI 531760	2	[6]EU282293	[6]EU282255	[6]EU282317
Hordeum brevisubulatum (Trin.) Link	PI 401387	1	–	[5]AY821705 AY821712	[2]AY010961
Hordeum brevisubulatum	PI 401390	2	–	[5]AY821713	[2]AY010964
Hordeum bulbosum L.	PI 440417	1	[6]EU282294 EU282295 EU282296	[5]AY821706	[2]AY010962
Hordeum californicum Covas & Stebbins	MA-138-1-40	1	[3]AY553256	[5]AY821707	[1]AF079273
Hordeum chilense Roem. & Schult.	PI 531781	1	[6]EU282297	–	[6]EU282318
Hordeum jubatum L.	RJMG 106	1	[3]AY553257	[5]AY821708 AY821709	[2]AY010963
Hordeum jubatum	RJMG 134	2	–	[5]AY821710 AY821711	–
Hordeum marinum Huds.	PI 304346	1	[3]AY553258	[6]EU282256 EU282257	[2]AY010959
Hordeum marinum	PI 304347	3	[6]EU282298	[6]EU282258	[6]EU282319
Hordeum murinum L.	PI 247054	1	[6]EU282299 EU282300	[6]EU282259	[6]EU282320
Hordeum murinum	CIho 15683	2	[3]AY553259	[6]EU282260	[2]AY010960
Hordeum pusillum Nutt.	CIho 15654		[6]EU282301	[6]EU282261	[6]EU282321
Hordeum stenostachys Godr.	PI 531791	1	[6]EU282302	[6]EU282262	[6]EU282322
Hordeum vulgare L.		1	–	–	[8]X07931
Hordeum vulgare	RJMG 107	2	[3]AY553260	[6]EU282263	–
Leymus racemosus ssp. *sabulosus* (M.Bieb.) Tzvelev	PI 531813		[3]AY553261	–	–
Peridictyon sanctum (Janka) Seberg, Fred., & Baden	KJ 248		[3]AY553262	[5]AY821714	[1]AF079278
Psathyrostachys fragilis (Boiss.) Nevski	PI 343192		–	[5]AY821715	[1]AF079279
Psathyrostachys juncea (Fisch.) Nevski	PI 206684		–	[5]AY821716	[1]AF079280
Pseudoroegneria libanotica (Hack.) D.R. Dewey	PI 228391	1	[6]EU282304	[6]EU282264	[6]EU282324
Pseudoroegneria libanotica	PI 228392	3	[6]EU282305	[6]EU282265	[6]EU282325
Pseudoroegneria spicata (Pursh) Á. Löve subsp. *spicata*	PI 232117	1	–	[5]AY821717	[1]AF079281
Pseudoroegneria spicata subsp. *inermis* (Scribn. & J.G.Smith) Á. Löve	PI 236681	2	–	[5]AY821718	[2]AY010998
Pseudoroegneria spicata subsp. *spicata*	PI 610986	3	[3]AY553263	–	[2]AY010999
Pseudoroegneria spicata subsp. *spicata*	D 2844	4	[3]AY553264	[5]AY821719	[2]AY011000
Pseudoroegneria spicata subsp. *spicata*	RJMG 112	6	–	[5]AY821720	[2]AY011001 AY010991
Pseudoroegneria stipifolia (Czern. ex Nevski) Á. Löve	PI 313960	2	[6]EU282306	[6]EU282266	–
Pseudoroegneria stipifolia	PI 531751	3	[6]EU282307 EU282308	[5]AY821721	–
Pseudoroegneria strigosa (M.Bieb.) Á. Löve	PI 499637	1	[6]EU282309 EU282310	–	[6]EU282323
Pseudoroegneria strigosa ssp. *aegilopoides* (Drobow) Á. Löve	PI 531755	2	[6]EU282311	[6]EU282267	[4]AY360823
Pseudoroegneria tauri (Boiss. & Balansa) Á. Löve	PI 380652	1	[6]EU282312	6EU282268	[6]EU282326
Pseudoroegneria tauri	PI 401319	2	[6]EU282313	–	[6]EU282327
Pseudoroegneria tauri	PI 380644	3	[6]EU282314 EU282315	–	–
Secale cereale L.	Kellogg s.n.		[3]AY553266	[5]AY821723 AY821724	[2]AY011009
Secale montanum (C.Presl.) C.Presl.	T 36554		–	[5]AY821725	–
Secale montanum	PI 440654		–	–	[1]AF079282

(*continued on next page*)

Table 2 (*continued*)

Species name	Accession	#	pepC	β-Amylase	GBSSI
Secale strictum subsp. *anatolicum* (Boiss.) K.Hammer	PI 206992		[3]AY553265	[5]AY821722	[2]AY011008
Taeniatherum caput-medusae (L.) Nevski	PI 208075	1	—	[5]AY821726	[2]AY011010
Taeniatherum caput-medusae	RJMG 189	2	[3]AY553268	[5]AY821727	[4]AY360847 AY360848
Taeniatherum caput-medusae	PI 314697	3	—	[5]AY821728	—
Taeniatherum caput-medusae	PI 317475	4	—	[5]AY821729	—
Thinopyrum bessarabicum (Savul. & Rayss) Á. Löve	PI531711		—	[5]AY821730	[1]AF079283
Thinopyrum elongatum (Host) D.R. Dewey	PI 531719	1	—	[5]AY821731	[1]AF079284
Thinopyrum elongatum	RJMG 113	2	[3]AY553269	—	—
Thinopyrum scirpeum (C.Presl) D.R. Dewey	PI 531749		—	[7]GQ847676	[2]AY011011
Triticum aestivum L.			[9]AJ007705	—	—
Triticum baeoticum Boiss.	Morrison s.n.		—	[5]AY821732	[1]AF079285
Triticum monococcum L.	PI 221413		—	[5]AY821733	—
Triticum urartu Tumanian	Morrison s.n.		—	[7]GQ847677	[1]AF079287
Bromus tectorum	Kellogg s.n.		[3]AY553239	[5]AY821734	[4]AY362757

[1] Mason-Gamer et al. (1998).
[2] Mason-Gamer (2001).
[3] Helfgott and Mason-Gamer (2004).
[4] Mason-Gamer (2004).
[5] Mason-Gamer (2005).
[6] Mason-Gamer (2008).
[7] New for this study.

gion 1 sequences. The 1100-bp PCR products obtained using primers 467F(1) and 1672R(2) (Helfgott and Mason-Gamer, 2004) include partial exons 1 and 2, along with the intervening intron, which is approximately 1000 bp long. The intron exhibits considerable length variation, including evidence of past transposon activity (Mason-Gamer, 2008). An ambiguous region of the alignment (positions 67–109), and two regions affected by transposon activity (698–779 and 1404–1488), were excluded from the analysis.

The β-amylase genes form a small family in the Triticeae, with several copies expressed in the endosperm and one that is ubiquitously expressed (Ziegler, 1999). The sequences used here appear to represent the ubiquitously-expressed copy; this gene copy has been mapped to the Triticeae group 2 homoeologous chromosomes (Sharp et al., 1988). The 1400-bp β-amylase PCR products were obtained using primers 2a-for and 5a-bac (Mason-Gamer, 2005), and include partial exons 2 and 5, complete exons 3 and 4, and introns 2–4, which are about 250, 100, and 400 bp in length, respectively. The β-amylase alignment was generally straightforward; most length differences were easy to interpret. One ambiguous simple sequence region (positions 553–570) and two regions corresponding to *Stowaway*-like transposon activity in some sequences (positions 635–765 and 1478–1644; Mason-Gamer, 2007b) were excluded from the analyses.

The GBSSI PCR products were obtained using the F-for and M-bac primers (Mason-Gamer et al., 1998), which amplify an approximately 1300-bp fragment that includes partial exons 9 and 14, exons 10–13, and introns 9–13, which are about 100 bp each. The putatively single-copy GBSSI gene maps to the Triticeae group 7 homoeologous chromosomes (Devos and Gale, 1997; Kleinhofs, 1997), or to a portion of chromosome 4 translocated from, and thus homoeologous to, the group 7 chromosomes (Devos and Gale, 1997; Korzun et al., 1997). The GBSSI alignment is generally straightforward in spite of numerous small insertions and deletions in the introns. Three ambiguously-aligned regions (positions 908–1019, 1134–1223, and 1377–1509) were excluded from the phylogenetic analyses.

2.3. Phylogenetic analyses

Prior to phylogenetic analyses, 16 nested models of sequence evolution (Frati et al., 1997; Sullivan et al., 1997; Swofford et al., 1996) were examined for each data set using preliminary maximum parsimony trees, and the resulting maximum-likelihood (ML) scores were compared using a likelihood ratio test (Felsenstein, 1981; Huelsenbeck and Crandall, 1997; Huelsenbeck and Rannala, 1997; Swofford et al., 1996). For each data set, the general time-reversible (GTR; Rodríguez et al., 1990; Tavaré, 1986) substitution model led to a large and significant increase in score compared to the Jukes–Cantor (Jukes and Cantor, 1969), Kimura two-parameter (Kimura, 1980), and Hasegawa–Kishino–Yano (Hasegawa et al., 1985) models, as did the addition of a gamma (Γ) distribution with shape parameter α to model among-site rate variation (Yang, 1993). Adding an invariable sites (I) parameter (Hasegawa et al., 1985) to the GTR + Γ model had an insignificant effect on the pepC and β-amylase scores, and caused a significant increase in the GBSSI score. Therefore, the pepC and β-amylase data were analyzed under the GTR + Γ model, and the GBSSI data under the GTR + I + Γ model. Following analyses of the three data sets, the **Y**-genome sequences from all three data sets were combined for a single analysis, with outgroups selected using information from the individual gene trees (*Dasypyrum villosum*, *Heteranthelium piliferum*, and *Secale anatolicum*). For the few individuals from which a **Y**-genome sequence was not recovered for a particular gene, the data were coded as missing in the corresponding part of the data set. The **Y**-genome analysis was analyzed under a GTR + I + Γ model.

All ML analyses were run using the Mac OS X UNIX version of GARLI v. 0.95 (Zwickl, 2006). Following the recommendations of the author, runs were set for an unlimited number of generations, and automatic termination following 10,000 generations without a meaningful (ln L increase of 0.01) change in score. For each data set, thirty analyses were run with random starting tree topologies, and the tree with the best score was used to display the gene tree. Branch support (BS) for each tree was estimated based on 100 ML bootstrap replicates in GARLI with searches as above, except that the stopping criterion was lowered to 5000 generations without a meaningful change in score. Bootstrap values $\geqslant$70% were recorded on the best ML trees.

3. Results

3.1. pepC analysis

The range of likelihood scores across 30 GARLI runs is small (−ln L 7507.42290–7507.42207), and differences between the best

tree (Fig. 1) and the others involve only nearly-identical taxa. While there is strong support for many genera, and some groups within them, most of the relationships among genera are poorly supported. The present interpretation and discussion of the pepC tree here is focused on the evolutionary history of *Elymus*; a more general discussion of relationships in the Triticeae based on pepC sequence data can be found elsewhere (Helfgott and Mason-Gamer, 2004).

The *Elymus* pepC sequences fall into two distinct groups (Fig. 1): one forms a very well-supported clade (100% BS) with *Pseudoroegneria*, and the other forms a monophyletic clade (100% BS) weakly linked (66% BS) to *Heteranthelium piliferum*. The former group is interpreted as the **St**-genome contribution from *Pseudoroegneria*, and the latter as the **Y** genome. Most of the *Elymus* individuals yield sequences in both clades. The two exceptions are *E. gmelinii*, from which only a **Y**-genome sequence was recovered, and *E. semicostatus* 2, from which only an **St**-genome sequence was obtained.

Within the *Pseudoroegneria* + *Elymus* clade, the *Elymus* **St** sequences fall into two groups. The first (100% BS) contains most of the *Elymus* sequences along with *P. strigosa*. Within it, the three *E. ciliaris* individuals and two *E. semicostatus* individuals group individually as species (both 100% BS) and together (98% BS). The rest of the *Elymus* sequences in this group form a weak (60% BS) clade containing *E. abolinii* + *E. nevskii* (95% BS), *E. pendulinus*, and three out of four *E. antiquus* individuals. The second group of *Elymus* **St** sequences includes *E. caucasicus*, *E. longearistatus*, and the

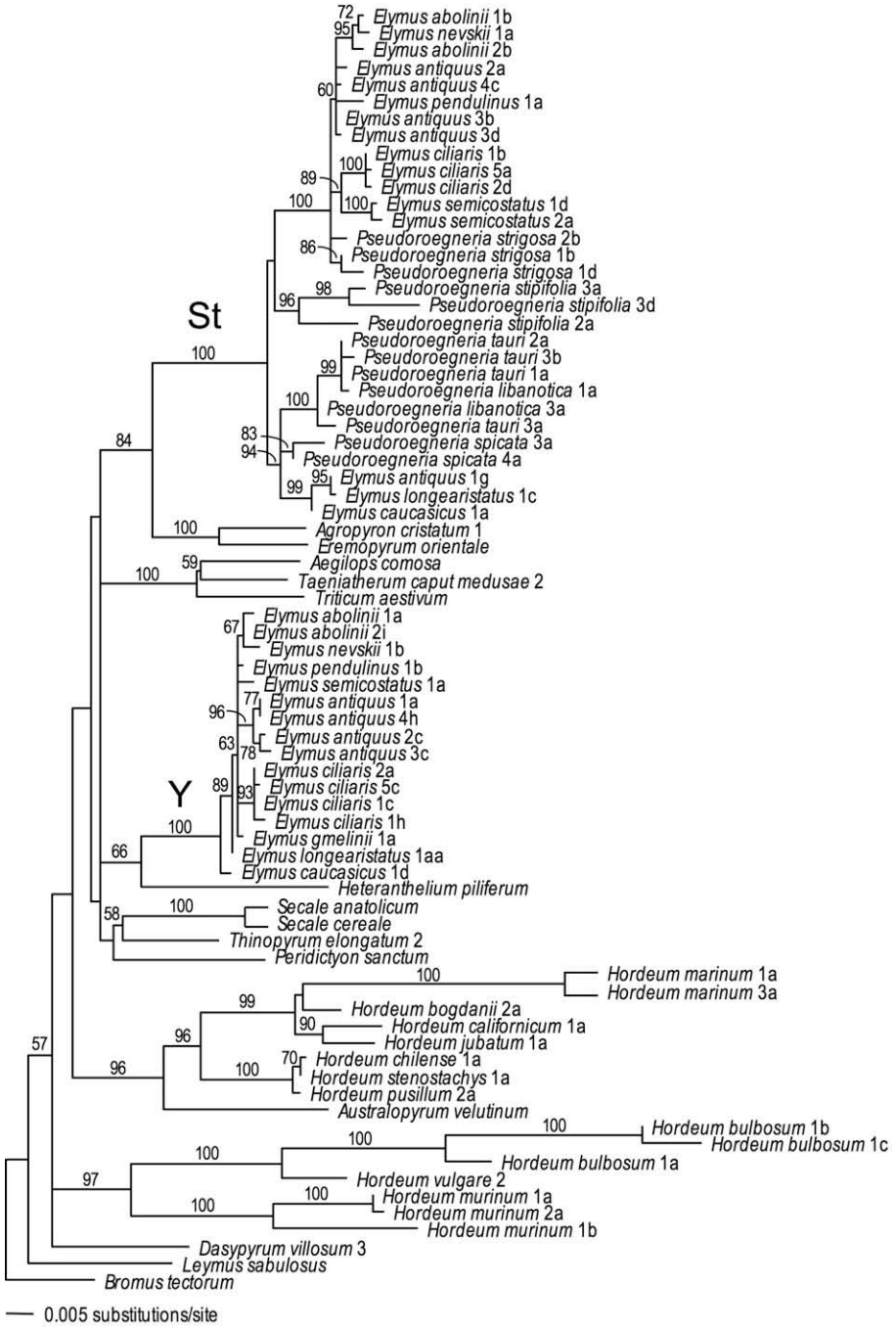

Fig. 1. The best-scoring ML tree from 30 GARLI analyses of the phosphoenolpyruvate carboxylase sequence data set under a GTR + Γ model of sequence evolution. Numbers above branches show ML bootstrap support ⩾50%. Where applicable, numbers following taxon names distinguish individuals within species, and are consistent among Figs. 1–4. Letters following these numbers designate cloned sequences from within individuals, and are specific to each gene tree.

fourth *E. antiquus* individual in a monophyletic group (99% BS) within a larger clade (94% BS) that includes *P. libanotica*, *P. spicata*, and *P. tauri*.

The presumed **Y**-genome sequences are more homogeneous than the **St**-genome sequences, as indicated by the short branch lengths separating them (Fig. 1). The *E. longearistatus* **Y**-genome sequence has a unique 361-bp insertion. BLAST searches (Altschul et al., 1990) of the Entrez nucleotide database align bases 116–361 of this sequence to a portion of the terminal repeat of a *copia*-like retrotransposon (e.g., Accession No. AM884194, 1009–1252; Di Giovanni et al., 2008). (Searches using bases 1–115 yielded nothing beyond short matches to miscellaneous sequences from a wide variety of organisms.) The limited structure within the **Y**-genome clade includes an *E. abolinii* + *E. nevskii* group and a monophyletic *E. ciliaris* group, as seen in the **St** clade. In contrast to the **St** clade, all four of the *E. antiquus* **Y**-genome sequences form a well-supported (96% BS) monophyletic group. *Elymus longearistatus* and *E. caucasicus* are not separated from the rest of *Elymus*, as they are in the **St** clade, although these species are weakly basal to the rest of the **Y** clade. The **Y**-genome's relationship to *Heteranthelium* is not very convincing (66% BS), and its broader placement within the tribe is not clarified at all.

3.2. β-Amylase analysis

The range of likelihood scores across 30 GARLI runs is larger than the range for the pepC data set (−ln L 10839.53187–10834.65461). In this case, the topology of the best β-amylase tree (Fig. 2) differs substantially from some of the other trees, including differences at some deep nodes. However, none of these nodes received bootstrap support >50%, so no conclusions were drawn from them. The description and interpretation of the β-amylase tree will focus primarily on the evolutionary history of *Elymus*; relationships among the monogenomic members of the tribe based on β-amylase sequence data have been discussed elsewhere (Mason-Gamer, 2005).

As with the pepC sequences, the *Elymus* β-amylase sequences form two main groups (Fig. 2). One is in a well-supported clade along with *Pseudoroegneria* (99% BS), and presumably represents the **St** genome. The second group includes the presumed **Y** sequences and is well-supported (100% BS), but its position within the tribe is entirely unclear. Nearly all of the *Elymus* individuals are represented in both clades, except for *E. longearistatus* and *E. semicostatus* 2, from which only **Y**-genome sequences were recovered, and *E. abolinii* 2, from which only an **St** sequence was recovered.

Within the *Elymus* + *Pseudoroegneria* clade, nearly all of the *Elymus* sequences form a clade within which *P. spicata* is nested. All four *E. antiquus* individuals form a weak (67% BS) monophyletic group (in contrast to the pepC results), and this clade is sister (88% BS) to the clade containing the four *P. spicata* individuals. This group is in turn sister (81% BS) to a group (67% BS) containing *E. ciliaris*, *E. semicostatus*, and a subclade (98% BS) of *E. abolinii* + *E. nevskii* + *E. pendulinus* + *E. gmelinii*. *Elymus caucasicus* is in a separate clade with *P. libanotica* and *P. tauri* (100% BS). This placement is partially consistent with the pepC results, in which *E. caucasicus*, *E. longearistatus*, and one *E. antiquus* individual grouped away from the rest of the *Elymus* **St** sequences, with *P. tauri*, *P. libanotica*, and *P. spicata*, but they differ in that, on the β-amylase tree, *P. spicata* is not in the group and there are no divergent *E. antiquus* sequences. The position of *E. longearistatus* on the β-amylase tree could not be addressed because no **St** sequence was recovered.

As on the pepC tree, the β-amylase **Y**-genome sequences form a well-supported group (100% BS) with little resolution within it (Fig. 2). Other similarities between the pepC and β-amylase **Y**-genome clades include the monophyly of *E. semicostatus* (66% BS) and *E. ciliaris* (98% BS), and the close relationship between *E. abolinii* and *E. nevskii* (90% BS). *Elymus caucasicus*, which is divergent from the remaining *Elymus* sequences in the **St** clade, is in an unresolved position within the **Y**-genome clade. The **Y**-genome clade has no obvious close relatives on the β-amylase tree. The one weak node (52% BS) linking the **Y**-genome to a *Secale* + *Australopyrum* + *Dasypyrum* + *Aegilops* clade is the only node with >50% BS that separates the **Y**-genome clade from the very base of the tree, so its position is essentially unresolved. Because most of the genera and some other shallow nodes are individually well-supported and exclude the **Y**-genome sequences, it appears that the donor of the β-amylase **Y**-genome sequences is not among the genera sampled here.

3.3. GBSSI analysis

The range of likelihood scores across 30 GARLI runs was −ln L 8317.02349–8300.27948. As with the β-amylase tree, the best GBSSI tree (Fig. 3) differed from some of the other trees in terms of some deeper relationships, but these nodes were also weak in terms of bootstrap support, and were not used as a basis for drawing conclusions. The relationships among the diploid genera within the Triticeae, as revealed by GBSSI, have been previously discussed (Mason-Gamer and Kellogg, 2000); the present interpretations will focus on *Elymus*.

As on the previous two trees, the *Elymus* GBSSI sequences either group with *Pseudoroegneria*, or form a separate clade of fairly homogeneous sequences (Fig. 3). One obvious difference on the GBSSI tree is that the *Elymus* + *Pseudoroegneria* sequences are diphyletic in two fairly well-supported clades (83% and 95% BS); both are interpreted as **St**-sequence clades. Although the two clades are widely separated (Fig. 3), the bootstrap support for the deep nodes throughout the tree is so weak that **St**-sequence monophyly is not convincingly ruled out; a maximum-parsimony analysis in which the **St** sequences are constrained to be monophyletic results in a tree that is only 4 steps longer than the unconstrained tree (1119 vs. 1123 steps). The **Y**-genome clade (100% BS) on the GBSSI tree is grouped with *Dasypyrum villosum* with moderate support (79% BS). Both **St**- and **Y**-genome GBSSI sequences were recovered from nearly all of the *Elymus* individuals, except for *E. caucasicus* and *E. longearistatus*, from which only **Y**-genome sequences were recovered.

The larger of the two *Pseudoroegneria* + *Elymus* clades (83% BS) comprises a heterogeneous group of sequences, within which *Elymus* is polyphyletic. A clade containing six *Elymus* species (*E. gmelinii*, *E. pendulinus*, *E. nevskii*, one of two *E. semicostatus* individuals, one of four *E. antiquus* individuals, and one of two *E. abolinii* individuals; 97% BS) is grouped (72% BS) with a paraphyletic grade including *P. strigosa*, one *P. spicata* sequence, and the second *E. abolinii* sequence. The other five *P. spicata* sequences form a separate monophyletic clade (91% BS). *Elymus ciliaris* and the second *E. semicostatus* individual are poorly resolved relative to the other sequences in this clade.

The smaller *Pseudoroegneria* + *Elymus* clade (95% BS) includes the three remaining *E. antiquus* individuals (100% BS) along with *P. tauri* and *P. libanotica*. The clade is weakly associated (60% BS) with *Peridictyon sanctum*. As stated earlier, *E. antiquus* is also polyphyletic on the pepC tree, on which one of the four individuals (#1) is separate from the rest of *E. antiquus* and from the main **St** *Elymus* group, while in this case, three of the individuals (#1, 2, and 3) are separate from the main group. The lone *E. antiquus* individual is different on each tree (individual #1 on the pepC tree and #4 on the GBSSI tree). On the β-amylase tree, the four *E. antiquus* **St** sequences form a monophyletic group.

The general heterogeneity of the GBSSI **St** sequences is reflected at the intraspecific level for three other species (Fig. 3). Not only is

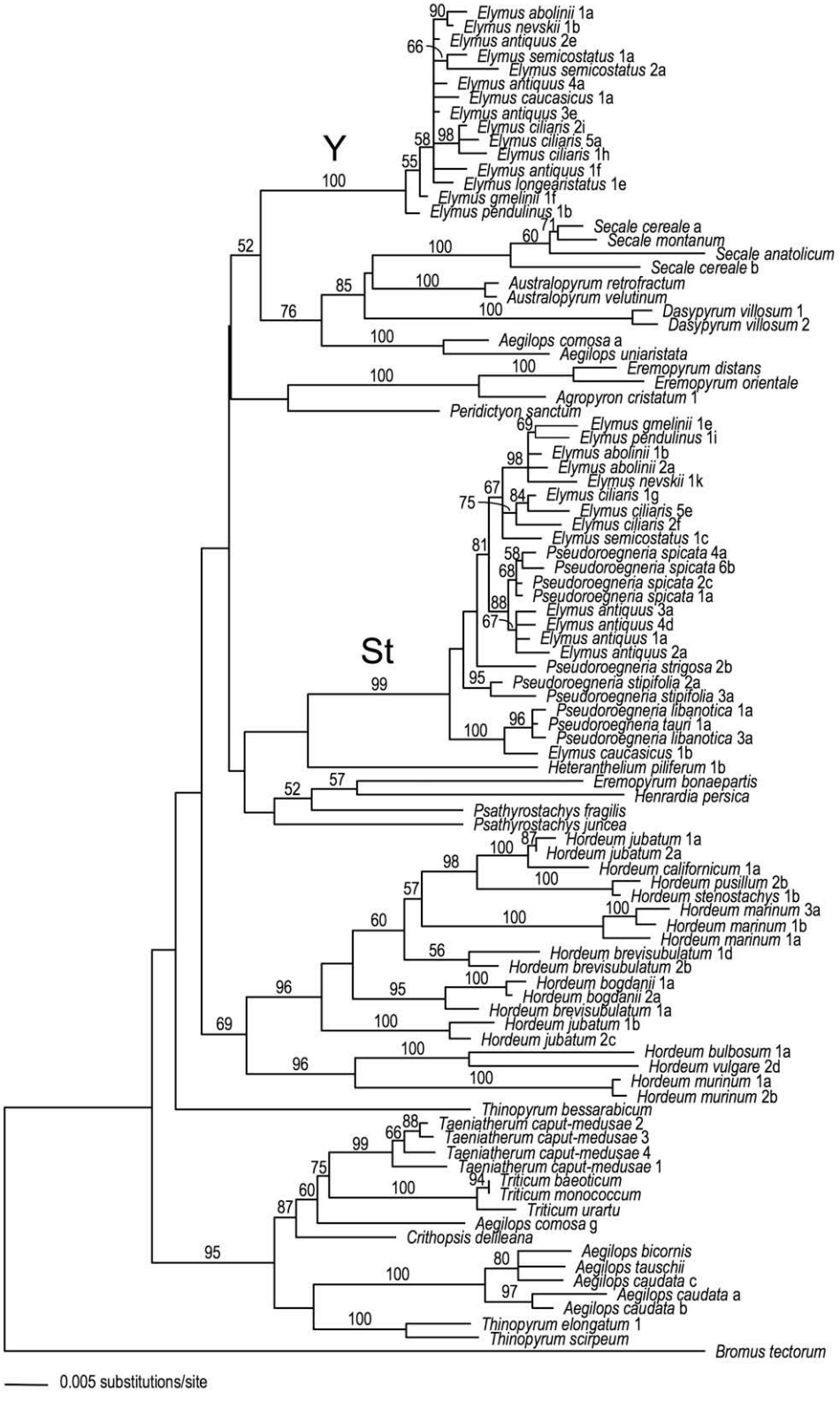

Fig. 2. The best-scoring ML tree from 30 GARLI analyses of the β-amylase sequence data set under a GTR + Γ model of sequence evolution. Numbers above branches show ML bootstrap support ⩾50%. Where applicable, numbers following taxon names distinguish individuals within species, and are consistent among Figs. 1–4. Letters following these numbers designate cloned sequences from within individuals, and are specific to each gene tree.

E. antiquus polyphyletic, as described above, but the two accessions of both *E. abolinii* (Fig. 3, clones 1a and 2k) and *E. semicostatus* (Fig. 3, clones 1a and 2k) are polyphyletic within the larger GBSSI **St** clade, as are the **St** sequences of *P. spicata* 6 (Fig. 3, clones 6a and 6c).

As on the pepC and β-amylase trees, the GBSSI **Y**-genome clade is distinct and well-supported, and the sequences within it are more homogeneous than those in the **St** clade. The little structure within the clade mainly places sequences within species together with varying levels of BS support (*E. ciliaris* 64%;

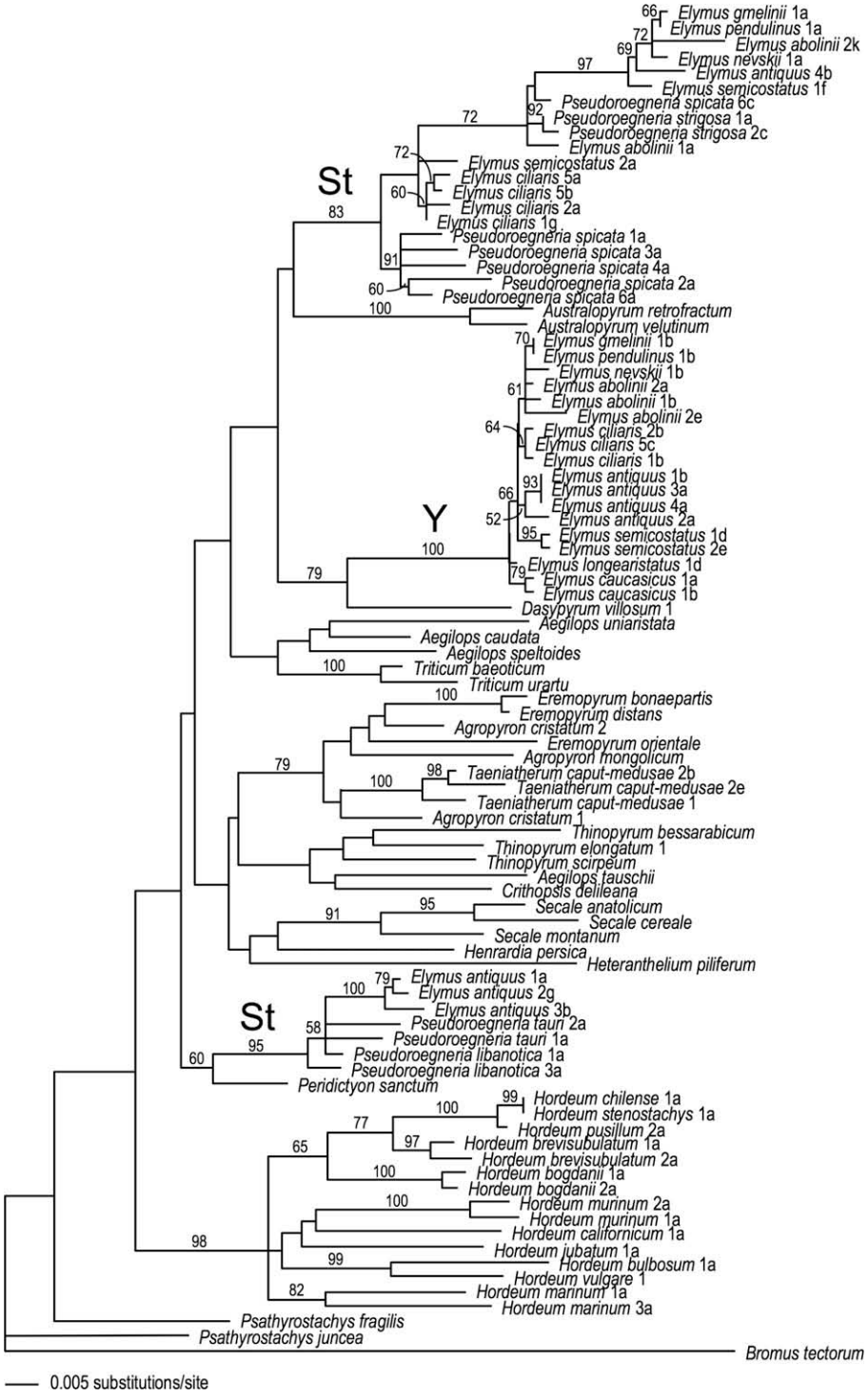

Fig. 3. The best-scoring ML tree from 30 GARLI analyses of the granule-bound starch synthase sequence data set under a GTR + I + Γ model of sequence evolution. Numbers above branches show ML bootstrap support ⩾50%. Where applicable, numbers following taxon names distinguish individuals within species, and are consistent among Figs. 1–4. Letters following these numbers designate cloned sequences from within individuals, and are specific to each gene tree.

E. antiquus 52%; *E. semicostatus* 95%; *E. caucasicus* 79%). The **Y**-genome clade is grouped with the *Dasypyrum villosum* GBSSI sequence (79% BS). The relatively long branch lengths of the **Y**-genome clade and *D. villosum* relative to one another, combined with the fact that the GBSSI tree is the only tree showing this relationship, suggest that *D. villosum* is not the immediate donor of the **Y** genome. Taken together, the three trees provide few insights into the identity of the **Y**-genome donor. It appears unlikely that any of the sampled genera, including those known to have donated genomes to other **St**-containing allopolyploids (*Agropyron*, *Australopyrum*, *Hordeum*, *Thinopyrum*, and *Psathyrostachys*) were the source of the **Y**-genome sequences. Otherwise, the **Y**-genome sequences should be similar to the sequences of their donor group, and sister to or nested within them, as is the case with the *Elymus* and *Pseudoroegneria* **St** sequences.

3.4. **Y**-genome combined analysis

Because of the differences in sampling and considerable phylogenetic conflict among the gene trees (Figs. 1–3), there was no attempt to combine the full data sets into a single analysis. Within the **Y**-genome clade alone, however, sampling was comparable for each data set, and there was no well-supported conflict among the trees (Figs. 1–3). Therefore, the **Y**-genome sequences from all three data sets were combined for a single ML analysis, in an attempt to increase resolution among the species. The selected outgroups represent relatives of the **Y**-genome clade on the pepC (Fig. 1; *Heteranthelium piliferum*), β-amylase (Fig. 2; *Secale anatolicum*), and GBSSI (Fig. 3; *Dasypyrum villosum*) trees. The resulting **Y**-genome tree (Fig. 4) shows similar or increased support for the monophyly of three of the species represented by more than one accession (*E. antiquus*, 93% BS; *E. ciliaris*, 100%; and *E. semicostatus*, 99%), and for *E. nevskii* + *E. abolinii* (93% BS), and somewhat increased support (77% BS) for the placement of *E. longearistatus* and *E. caucasicus* at the base of the **Y** clade. However, the tree does not otherwise convincingly resolve relationships among species in spite of the increased amount of data (which includes approximately 2200 basepairs of intron sequence).

4. Discussion

4.1. Allotetraploid origin of Asian Elymus

The present phylogenetic analysis of a group of Asian *Elymus* tetraploids and their diploid relatives was undertaken with the goals of (1) clarifying whether the genome donors to the tetraploids were phylogenetically distinct; (2) identifying the donor(s) of the genomes; and (3) examining the possibility of multiple tetraploid origins or introgression following polyploidy. The analyses include a reasonably broad sample of fifteen monogenomic genera from the tribe, including donors of all of the genomes known to cooccur with the **St** genome of *Pseudoroegneria* in allopolyploid species: *Hordeum* (**H**), *Agropyron* (**P**), *Australopyrum* (**W**), *Psathyrostachys* (**Ns**), and *Thinopyrum* (**J**, **E**).

The results convincingly show that the homoeologous genomes in the **StStYY** Asian *Elymus* tetraploids were derived from phylogenetically distinct donors. Nearly all of the *Elymus* individuals yield two distinct copies of all three genes, including an **St**-genome copy that groups with *Pseudoroegneria*, and a **Y**-genome copy that is not closely associated with any single Triticeae genus. In a few cases, only one copy of a gene was recovered (Table 1), but no individuals lack either the **St** or the **Y** copy of all three genes. The missing copies might result from a sampling artifact or changes in PCR primer sites, and/or they might show the occasional loss of one copy of the gene, either through homoeologous recombination or deletion. The overall results agree with many cytogenetic analyses (Dewey, 1974, 1980a,b; Jensen, 1989, 1990; Jensen and Hatch, 1989; Lu et al., 1990, 1995; Lu and von Bothmer, 1989, 1990a,b, 1991; Sakamoto and Muramatsu, 1966) that showed little pairing between the homoeologous genomes of the Asian *Elymus* species, and with the phylogenetic analysis of *RPB2* sequences (Sun et al., 2008), which placed the sequences from presumed **StStYY** tetraploid species into two distinct clades. In contrast, ITS sequences from **StStYY** *Elymus* formed a clade with *Pseudoroegneria* (Liu et al., 2006), prompting the conclusion that the **St** and **Y** genome sets were both derived from *Pseudoroegneria*. However, the weight of evidence from other studies suggests that there is some other explanation for Liu et al.'s (2006) result. Concerted evolution might explain the lack of a distinct **Y**-genome ITS sequence, but Liu et al. (2006) downplayed this possibility, pointing out that other Triticeae allopolyploids were not affected by ITS concerted evolution.

The result that one of the tetraploid genomes is derived from *Pseudoroegneria* was predicted by most of the cytogenetic and molecular phylogenetic studies cited above, but the identity of the **Y**-genome donor has been a much bigger mystery, with no obvious candidates emerging from the previous studies. The three gene trees presented here do not solve the mystery. The pepC and GBSSI trees suggest two different relatives to the **Y** clade with weak to moderate support (*Heteranthelium piliferum* with 66% BS and *Dasypyrum villosum* with 79% BS, respectively), while the **Y** clade's position on the β-amylase tree is unresolved. In contrast, support values for the *Pseudoroegneria* + *Elymus* **St** clades are 100%, 99%, and 83%/95%, respectively, on the pepC and β-amylase trees and for the two **St** clades on the GBSSI tree. We would expect to see similarly high support for the relationship between the *Elymus* **Y** genome and its donor, and conclude that the closest diploid relative to the **Y** clade has not been sampled in these (or previous) studies of the **StStYY** species. Given the intense interest in the Triticeae due to its economic importance, the wild species of the tribe have been widely collected and recorded, and the failure to discover the diploid **Y**-genome donor suggests that this entity is rare or narrowly distributed, or possibly even extinct.

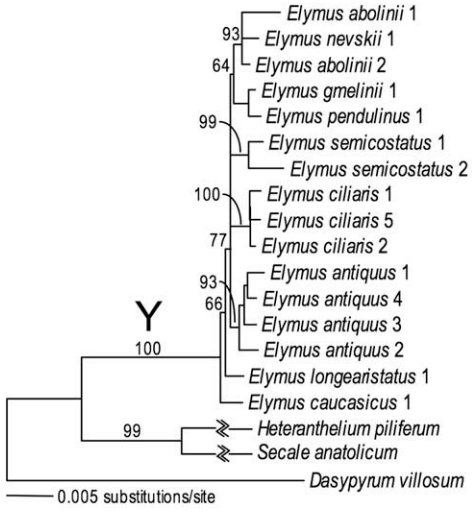

Fig. 4. The best-scoring ML tree from 10 GARLI analyses of **Y**-genome sequences from the combined phosphoenolpyruvate carboxylase, β-amylase, and granule-bound starch synthase sequence data set under a GTR + I + Γ model of sequence evolution. Numbers above branches show ML bootstrap support ⩾50%. Numbers following taxon names distinguish individuals within species where applicable, and are consistent among Figs. 1–4.

4.2. Variation within the genome-specific clades

The two genome-specific clades differ dramatically in the levels of variation among species. The lack of resolution among the **Y**-genome sequences in the separate and combined analyses, in combination with the very short branches within the clade, suggest a recent origin of the **StStYY** tetraploid combination, involving a single **Y**-genome donor. Though intraspecific sampling is very limited, the monophyly of **Y**-genome sequences from within species suggests that gene exchange involving the **Y**-genome has been limited since the origin of the **StStYY** tetraploids.

In contrast, there is considerable variation among the **St**-genome sequences, and the relationships among them are far more complicated. The conflict among the three trees with respect to their relationships suggests a history of reticulation involving *Pseudoroegneria* and *Elymus*, affecting the apparent relationships both within and between the two genera. At the interspecific level, the **St** clades on the three trees differ in terms of the relationships among *Elymus* species, among *Pseudoroegneria* species, and be-

tween *Elymus* and *Pseudoroegneria*. It is difficult to simultaneously compare and contrast the positions of all of the **St**-genome sequences on all three trees, so in the discussion below, we focus on a subset of species – *P. strigosa*, the *P. libanotica*/*P. tauri* pair, *P. spicata*, *E. antiquus*, *E. ciliaris*, and *E. caucasicus* – to illustrate the complexity of the phylogenetic relationships in the **St** clade (though examination of other combinations of species would also reveal reticulate patterns).

On the pepC tree, *P. strigosa* is a very close relative to most of the *Elymus* species, including *E. ciliaris*. *Elymus caucasicus* is in a separate clade with *P. libanotica*/*P. tauri* and *P. spicata*. *Elymus antiquus* is split between the two clades, with sequences from three individuals in the former, and one on the latter. Thus, when taken alone, the pepC tree suggests that there have been two *Pseudoroegneria* donors to *Elymus*, and that *E. ciliaris* and *E. caucasicus* acquired their pepC **St** copies from different *Pseudoroegneria* species. This could be explained by separate origins of the **StStYY** genome combination involving different *Pseudoroegneria* donors, or by introgression of pepC from *Pseudoroegneria* into *Elymus* after tetraploidization. The polyphyly of *E. antiquus* is consistent with either the retention of ancestral polymorphism within *E. antiquus*, or introgression following the species's origin. (The pattern is also consistent with two separate origins of *E. antiquus*, but the well-supported monophyly of the species in the pepC **Y**-genome clade casts doubt on this possibility.)

On the β-amylase tree, *P. spicata*, rather than *P. strigosa*, is the closest *Pseudoroegneria* relative to most of the *Elymus* species. *Elymus caucasicus* is once again separate from the majority of the *Elymus* species, and is again grouped with *P. libanotica*/*P. tauri*, although on the pepC tree, *P. spicata* was part of this clade. One possible explanation for this discrepancy between the trees is that the *Elymus* species in the larger clade have acquired one gene through vertical transmission from one *Pseudoroegneria* species, and the other through introgression from a different species. For example, if *P. strigosa* was the original **St** genome donor to *E. ciliaris* and the closely-related *Elymus* species, as suggested by the pepC tree, then the ancestor of these *Elymus* species might have acquired their *P. spicata*-like β-amylase gene through introgression. Another possible explanation is that introgression or lineage sorting has affected the phylogenetic positions of the *Pseudoroegneria* species themselves. Reticulation among *Pseudoroegneria* species would affect the apparent positions of individual *Pseudoroegneria* species with respect to *Elymus*; this could, for example, explain the discrepancy involving the placement of *P. spicata*. The *E. antiquus* **St** sequences are monophyletic in the β-amylase tree, in contrast to the pepC tree, providing additional evidence that their polyphyly on the pepC tree reflects either lineage sorting or introgression following the species origin, rather than species-level polyphyly.

On the GBSSI tree, the *Pseudoroegneria* and *Elymus* species are again separated into two groups, with one containing the majority of the *Elymus* species, though this tree is unique in its suggestion that the two **St** sequence clades might be widely diphyletic within the tribe. Given the weak support at the base of the tree, however, we provisionally interpret the apparent diphyly as a spurious artifact of poor resolution. Unfortunately, the position of *E. caucasicus* can not be discussed relative to the previous trees, because we did not recover an **St** copy of GBSSI from the species. On the GBSSI tree, the larger **St** clade is much more variable than on the other trees, and contains a further suggestion that introgression has affected the position of *P. spicata* relative to *P. strigosa*: one *P. spicata* **St** allele (6c) is far more similar to *P. strigosa* than to the other *P. spicata* copies, including a second **St** allele from the same individual (6a). Furthermore, the two polyphyletic *Elymus* species within the larger **St** clade (*E. abolinii* and *E. semicostatus*) suggest additional **St** gene reticulation involving the tetraploids in this clade, consistent with either introgression or lineage sorting. The smaller **St** clade includes *P. libanotica*/*P. tauri*, in agreement with both of the previous trees, and *E. antiquus* is split between the two main **St** clades, as it was on the pepC tree. The polyphyly of the *E. antiquus* GBSSI **St** sequences provides still more evidence for its reticulate history. As on the pepC tree, the monophyly of this species in the **Y** clade on the GBSSI tree argues against a diphyletic origin of the species, and instead suggests later reticulation involving only the **St** sequences.

While it would be possible to draw conclusions about the relationships among the **St** sequences from any one gene tree, it is the combination of differences and similarities among the trees that allow us to infer that hybridization, introgression, and/or lineage sorting have shaped the relationships among the **St** gene copies of *Elymus* and *Pseudoroegneria*. Together, the three gene trees suggest that at least two *Pseudoroegneria* species have made genetic contributions to *Elymus*; the *Elymus* **St** sequences are placed in two clades with *Pseudoroegneria* on all of the trees. The pepC and β-amylase trees both suggest that *E. caucasicus* acquired its **St** genome from a different donor than the majority of the other sampled *Elymus* species. The differences among the trees highlight reticulate evolution in *Pseudoroegneria*, possibly involving past interchange among *P. spicata*, *P. strigosa*, and *P. libanotica*/*P. tauri*. This same interchange may be responsible for the polyphyly and phylogenetic conflict characterizing *E. antiquus*, whose shifting phylogenetic placement appears to be linked to those of *P. spicata*, *P. strigosa*, and *P. libanotica*/*P. tauri*. With regard to the **St** clade in particular, however, there are too many differences among the three trees to allow us to develop a single, unifying hypothesis to describe the combination of divergence and reticulation underlying the differences. We feel that one promising approach to disentangling these processes would be to focus on just a few species – e.g., *E. caucasicus*, *E. antiquus*, *P. spicata*, *P. strigosa*, and *P. libanotica*/*P. tauri* – in a detailed study with more extensive sampling within each species.

4.3. Summary

The analyses of three independent nuclear data sets agree with earlier cytogenetic studies, and some molecular studies, that the basic genomes of the **StStYY** allopolyploids were derived from phylogenetically distinct donors. The gene trees do not identify the previously unknown **Y**-genome donor, and the **St**-genome donor is confirmed as *Pseudoroegneria*. The molecular data reveal complex relationships among the **St** genome sequences, suggesting that reticulation within and between *Elymus* and *Pseudoroegneria* has shaped the history of the **St** genome at the diploid and tetraploid levels. Further clarification of the nature of the reticulate evolutionary processes responsible for the complicated phylogenetic patterns will probably require more detailed intraspecific sampling, and our study has identified several promising species for future study.

Acknowledgments

Thanks to the USDA-ARS National Germplasm System for seeds, to Björn Salomon and Bao-Rong Lu for valuable background information on the **StStYY** species of *Elymus*, and to two anonymous reviewers for valuable feedback on the manuscript. The work was supported by National Science Foundation DEB-9974181.

References

Ainouche, M.L., Baumel, A., Salmon, A., Yannic, G., 2004. Hybridization, polyploidy and speciation in *Spartina* (Poaceae). New Phytol. 161, 165–172.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. J. Mol. Biol. 215, 402–410.

Barkworth, M.E., 2000. Changing perceptions of the Triticeae. In: Jacobs, S.W.L., Everett, J. (Eds.), Grasses: Systematics and Evolution. CSIRO, Melbourne, pp. 110–120.
Bradley, R.D., Hillis, D.M., 1997. Recombinant DNA sequences generated by PCR amplification. Mol. Biol. Evol. 14, 592–593.
Brysting, A.K., Oxelman, B., Huber, K.T., Moulton, V., Brochmann, C., 2007. Untangling complex histories of genome mergings in high polyploids. Syst. Biol. 56, 467–476.
Chen, Q., Conner, R.L., Laroche, A., Thomas, J.B., 1998. Genome analysis of *Thinopyrum intermedium* and *Thinopyrum ponticum* using genomic in situ hybridization. Genome 41, 580–586.
Cronn, R., Cedroni, M., Haselkorn, T., Grover, C., Wendel, J.F., 2002. PCR-mediated recombination in amplification products derived from polyploid cotton. Theor. Appl. Genet. 104, 482–489.
Davis, J.I., Soreng, R.J., 2007. A preliminary phylogenetic analysis of the grass subfamily Pooideae (Poaceae), with attention to structural features of the plastid and nuclear genomes, including an intron loss in GBSSI. Aliso 23, 335–348.
Devos, K.M., Gale, M.D., 1997. Comparative genetics in the grasses. Plant Mol. Biol. 35, 3–15.
Dewey, D.R., 1967. Synthetic hybrids of new world and old world Agropyrons: III. *Agropyron repens* × tetraploid *Agropyron spicatum*. Amer. J. Bot. 54, 93–98.
Dewey, D.R., 1968. Synthetic *Agropyron-Elymus* hybrids: III. *Elymus canadensis* × *Agropyron caninum*, *A. trachycaulum*, and *A. striatum*. Amer. J. Bot. 55, 1133–1139.
Dewey, D.R., 1970. Genome relations among *Elymus canadensis*, *Elymus triticoides*, *Elymus dasystachyus*, and *Agropyron smithii*. Amer. J. Bot. 57, 861–866.
Dewey, D.R., 1974. Cytogenetics of *Elymus sibiricus* and its hybrids with *Agropyron tauri*, *Elymus canadensis*, and *Agropyron caninum*. Bot. Gaz. 135, 80–87.
Dewey, D.R., 1975. The origin of *Agropyron smithii*. Amer. J. Bot. 62, 524–530.
Dewey, D.R., 1980a. Cytogenetics of *Agropyron drobovii* and five of its interspecific hybrids. Bot. Gaz. 141, 469–478.
Dewey, D.R., 1980b. Cytogenetics of *Agropyron ugamicum* and six of its interspecific hybrids. Bot. Gaz. 141, 305–312.
Dewey, D.R., 1984. The genomic system of classification as a guide to intergeneric hybridization within the perennial Triticeae. In: Gustafson, J.P. (Ed.), Gene Manipulation in Plant Improvement, Proceedings of the16th Stadler Genetics Symposium. Plenum Publishing Company, New York, pp. 209–279.
Di Giovanni, M., Cenci, A., Janni, M., D'Ovidio, R., 2008. A LTR *copia* retrotransposon and *Mutator* transposons interrupt *Pgip* genes in cultivated and wild wheats. Theor. Appl. Genet. 116, 859–867.
Emshwiller, E., Doyle, J.J., 2002. Origins of domestication and polyploidy in oca (*Oxalis tuberosa*: Oxalidaceae). 2. Chloroplast-expressed glutamine synthase data. Amer. J. Bot. 89, 1042–1056.
Felsenstein, J., 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. J. Mol. Evol. 17, 368–376.
Fortune, P.M., Portau, N., Viron, N., Ainouche, M.L., 2008. Molecular phylogeny and reticulate origins of the polyploid *Bromus* species from the section *Genea* (Poaceae). Amer. J. Bot. 95, 454–464.
Frati, F., Simon, C., Sullivan, J., Swofford, D.L., 1997. Evolution of the mitochondrial cytochrome oxidase II gene in Collembola. J. Mol. Evol. 44, 145–158.
Ge, S., Sang, T., Lu, B.-R., Hong, D.-Y., 1999. Phylogeny of rice genomes with emphasis on origins of allotetraploid species. Proc. Natl. Acad. Sci. USA 96, 14400–14405.
González, M.C., Echevarría, C., Vidal, J., Cejudo, F.J., 2002. Isolation and characterization of a wheat phosphoenolpyruvate carboxylase gene. Modelling of the encoded protein. Plant Sci. 162, 233–238.
Grass Phylogeny Working Group, 2001. Phylogeny and subfamilial classification of the grasses (Poaceae). Ann. Missouri Bot. Gard. 88, 372–457.
Hasegawa, M., Kishino, H., Yano, T., 1985. Dating the human-ape split by a molecular clock of mitochondrial DNA. J. Mol. Evol. 22, 160–170.
Helfgott, D.M., Mason-Gamer, R.J., 2004. The evolution of North American *Elymus* (Triticeae, Poaceae) allotetraploids: evidence from phosphoenolpyruvate carboxylase gene sequences. Syst. Bot. 29, 850–861.
Huelsenbeck, J.P., Crandall, K.A., 1997. Phylogeny estimation and hypothesis testing using maximum likelihood. Annu. Rev. Ecol. Syst. 28, 437–466.
Huelsenbeck, J.P., Rannala, B., 1997. Phylogenetic methods come of age: testing hypotheses in an evolutionary context. Science 276, 227–232.
Jensen, K.B., 1989. Cytology, fertility, and origin of *Elymus abolinii* (Drob.) Tzvelev and its hybrids with *Pseudoroegneria spicata*, *E. lanceolatus*, *E. dentatus* ssp. *ugamicus*, and *E. drobovii*. Genome 32, 468–474.
Jensen, K.B., 1990. Cytology and morphology of *Elymus pendulinus*, *E. pendulinus* ssp. *multiculmis*, and *E. parviglume* (Poaceae: Triticeae). Bot. Gaz. 151, 245–251.
Jensen, K.B., 1993. Cytogenetics of *Elymus magellanicus* and its intra- and intergeneric hybrids with *Pseudoroegneria spicata, Hordeum violaceum, E. trachycaulus, E. lanceolatus,* and *E. glaucus* (Poaceae: Triticeae). Genome 36, 72–76.
Jensen, K.B., 1996. Genome analysis of Eurasian *Elymus thoroldianus*, *E. melantherus*, and *E. kokonoricus* (Triticeae: Poaceae). Int. J. Plant Sci. 157, 136–141.
Jensen, K.B., Hatch, S.L., 1989. Genome analysis, morphology, and taxonomy of *Elymus gmelinii* and *E. strictus* (Poaceae: Triticeae). Bot. Gaz. 150, 84–92.
Joly, S., Starr, J.R., Lewis, W.H., Bruneau, A., 2006. Polyploid and hybrid evolution in roses east of the Rocky Mountains. Amer. J. Bot. 93, 412–425.
Judo, M.S.B., Wedel, A.B., Wilson, C., 1998. Stimulation and suppression of PCR-mediated recombination. Nucleic Acids Res. 26, 1819–1825.
Jukes, T.H., Cantor, C.R., 1969. Evolution of protein molecules. In: Munro, N.H. (Ed.), Mammalian Protein Metabolism. Academic Press, New York, pp. 21–132.
Kellogg, E.A., 1989. Comments on genomic genera in the Triticeae (Poaceae). Amer. J. Bot. 76, 796–805.
Kellogg, E.A., Appels, R., Mason-Gamer, R.J., 1996. When gene trees tell different stories: the diploid genera of Triticeae (Gramineae). Syst. Bot. 21, 321–347.
Kimura, M., 1980. A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. J. Mol. Evol. 16, 111–120.
Kleinhofs, A., 1997. Integrating barley RFLP and classical marker maps. Barley Genet. Newslett. 27, 105–112.
Korzun, V., Malyshev, S., Voymokov, A., Börner, A., 1997. RFLP-mapping of three mutant loci in rye (*Secale cereale* L.) and their relation to homoeologous loci within the Gramineae. Theor. Appl. Genet. 95, 468–473.
Lepiniec, L., Keryer, E., Philippe, H., Gadal, P., Crétin, C., 1993. *Sorghum* phosphoenolpyruvate carboxylase gene family: structure, function, and molecular evolution. Plant Mol. Biol. 21, 487–502.
Lihová, J., Shimizu, K.K., Marhold, K., 2006. Allopolyploid origin of *Cardamine asarifolia* (Brassicaceae): incongruence between plastid and nuclear ribosomal DNA sequences solved by a single-copy nuclear gene. Mol. Phylogenet. Evol. 39, 759–786.
Liu, Q., Ge, S., Tang, H., Zhang, X., Zhu, G., Lu, B.-R., 2006. Phylogenetic relationships in *Elymus* (Poaceae: Triticeae) based on nuclear ribosomal internal transcribed spacer and chloroplast *trnL–F* sequences. New Phytol. 170, 411–420.
Liu, Z.-W., Wang, R.R.-C., 1993. Genome analysis of *Elytrigia caespitosa*, *Lophopyrum nodosum*, *Pseudoroegneria geniculata* ssp. *scythia*, and *Thinopyrum intermedium* (Triticeae, Gramineae). Genome 36, 102–111.
Lu, B.-R., Salomon, B., 1992. Differentiation of the **SY** genomes in Asiatic *Elymus*. Hereditas 116, 121–126.
Lu, B.-R., Salomon, B., von Bothmer, R., 1990. Cytogenetic studies of progeny from the intergeneric crosses *Elymus* × *Hordeum* and *Elymus* × *Secale*. Genome 33, 425–432.
Lu, B.-R., Salomon, B., von Bothmer, R., 1995. Interspecific hybridizations with *Elymus confusus* and *E. dolichatherus*, and their genomic relationships (Poaceae: Triticeae). Plant Syst. Evol. 197, 1–17.
Lu, B.-R., von Bothmer, R., 1989. Cytological studies of a dihaploid and hybrid from intergeneric cross *Elymus shandogensis* × *Triticum aestivum*. Hereditas 111, 231–238.
Lu, B.-R., von Bothmer, R., 1990a. Genomic constitution of *Elymus parviglumis* and *E. pseudonutans*: Triticeae (Poaceae). Hereditas 113, 109–119.
Lu, B.-R., von Bothmer, R., 1990b. Intergeneric hybridization between *Hordeum* and Asiatic *Elymus*. Hereditas 112, 109–116.
Lu, B.-R., von Bothmer, R., 1991. Production and cytogenetic analysis of the intergeneric hybrids between nine *Elymus* species and common wheat (*Triticum aestivum* L.). Euphytica 58, 81–95.
Lu, B.-R., von Bothmer, R., 1993. Meiotic analysis of *Elymus caucasicus*, *E. longearistatus*, and their interspecific hybrids with twenty-three *Elymus* species (Triticeae, Poaceae). Plant Syst. Evol. 185, 35–53.
Löve, Á., 1984. Conspectus of the Triticeae. Feddes Repert. 95, 425–521.
Mason-Gamer, R.J., 2001. Origin of North American species of *Elymus* (Poaceae: Triticeae) allotetraploids based on granule-bound starch synthase gene sequences. Syst. Bot. 26, 757–768.
Mason-Gamer, R.J., 2004. Reticulate evolution, introgression, and intertribal gene capture in an allohexaploid grass. Syst. Biol. 53, 25–37.
Mason-Gamer, R.J., 2005. The β-amylase genes of grasses and a phylogenetic analysis of the Triticeae (Poaceae). Amer. J. Bot. 92, 1045–1058.
Mason-Gamer, R.J., 2007a. Allopolyploids of the genus *Elymus* (Triticeae, Poaceae): a phylogenetic perspective. Aliso 23, 372–379.
Mason-Gamer, R.J., 2007b. Multiple homoplasious insertions and deletions of a Triticeae (Poaceae) DNA transposon: a phylogenetic perspective. BMC Evol. Biol. 7, 92.
Mason-Gamer, R.J., 2008. Allohexaploidy, introgression, and the complex phylogenetic history of *Elymus repens* (Poaceae). Mol. Phylogenet. Evol. 47, 598–611.
Mason-Gamer, R.J., Kellogg, E.A., 2000. Phylogenetic analysis of the Triticeae using the starch synthase gene, and a preliminary analysis of some North American *Elymus* species. In: Jacobs, S.W.L., Everett, J. (Eds.), Grasses: Systematics and Evolution. CSIRO Publishing, Collingwood VIC, pp. 102–109.
Mason-Gamer, R.J., Weil, C.F., Kellogg, E.A., 1998. Granule-bound starch synthase: structure, function, and phylogenetic utility. Mol. Biol. Evol. 15, 1658–1673.
Petersen, G., Seberg, O., Yde, M., Berthelsen, K., 2006. Phylogenetic relationships of *Triticum* and *Aegilops* and evidence for the origin of the **A**, **B**, and **D** genomes of common wheat (*Triticum aestivum*). Mol. Phylogenet. Evol. 39, 70–82.
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.
Rodríguez, F., Oliver, J.L., Marín, A., Meduna, J.R., 1990. The general stochastic model of nucleotide substitution. J. Theor. Biol. 142, 485–501.
Rodríguez, F., Spooner, D.M., 2009. Nitrate reductase phylogeny of potato (*Solanum* sect. *Petota*) genomes with emphasis on the origins of the polyploid species. Syst. Bot. 34, 207–219.
Rohde, W., Becker, D., Salamini, F., 1988. Structural analysis of the waxy locus from *Hordeum vulgare*. Nucleic Acids Res. 16, 7185–7186.
Sakamoto, S., Muramatsu, M., 1966. Cytogenetic studies in the tribe Triticeae. II. Tetraploid and hexaploid hybrids of *Agropyron*. Jpn. J. Genet. 41, 155–168.

Salomon, B., 1993. Interspecific hybridizations in the *Elymus semicostatus* group (Poaceae). Genome 36, 889–905.
Salomon, B., Lu, B.-R., 1992. Genomic groups, morphology, and sectional delimitation in Eurasian *Elymus* (Poaceae, Triticeae). Plant Syst. Evol. 180, 1–13.
Salomon, B., Lu, B.-R., 1994a. Genomic relationships between species of the *Elymus semicostatus* group and *Elymus* sensu lato (Poaceae). Plant Syst. Evol. 191, 199–211.
Salomon, B., Lu, B.-R., 1994b. Interspecific hybridizations among species of the *Elymus semicostatus* and *Elymus tibeticus* groups (Poaceae). Plant Syst. Evol. 189, 1–13.
Seberg, O., Petersen, G., 1998. A critical review of concepts and methods used in classical genome analysis. Bot. Rev. 64, 372–417.
Seberg, O., Petersen, G., 2007. Phylogeny of Triticeae (Poaceae) based on three organelle genes, two single-copy nuclear genes, and morphology. Aliso 23, 362–371.
Sharp, P.J., Kreis, M., Shewry, P.R., Gale, M.D., 1988. Location of β-amylase sequences in wheat and its relatives. Theor. Appl. Genet. 75, 286–290.
Smedmark, J.E.E., Eriksson, T., Bremer, B., 2005. Allopolyploid evolution in Geinae (Colurieae: Rosaceae)—building reticulate species trees from bifurcating gene trees. Org. Div. Evol. 5, 275–283.
Straub, S.C.K., Pfeil, B.E., Doyle, J.J., 2006. Testing the polyploid past of soybean using a low-copy nuclear gene—is *Glycine* (Fabaceae: Papilionoideae) an auto – or allopolyploid? Mol. Phylogenet. Evol. 39, 580–584.
Sullivan, J., Markert, J.A., Kilpatrick, C.W., 1997. Phylogeography and molecular systematics of the *Peromyscus aztecus* species group (Rodentia: Muridae) inferred using parsimony and likelihood. Syst. Biol. 46, 426–440.
Sun, G.-L., Ni, Y., Daley, T., 2008. Molecular phylogeny of *RPB2* gene reveals multiple origin, geographic distribution of H genome, and the relationship of the Y genome to other genomes of *Elymus* species. Mol. Phylogenet. Evol. 46, 897–907.
Sun, G.-L., Salomon, B., von Bothmer, R., 1997. Analysis of tetraploid *Elymus* species using wheat microsatellite markers and RAPD markers. Genome 40, 806–814.
Svitashev, S., Salomon, B., Bryngelsson, T., von Bothmer, R., 1996. A study of 28 *Elymus* species using repetitive DNA sequences. Genome 39, 1093–1101.
Swofford, D.L., Olsen, G.J., Waddell, P.J., Hillis, D.M., 1996. Phylogenetic inference. In: Hillis, D.M., Moritz, C., Mable, B.K. (Eds.), Molecular Systematics, second ed. Sinauer, Sunderland, pp. 407–514.
Tavaré, S., 1986. Some probabilistic and statistical problems in the analysis of DNA sequences. Lec. Math. Life Sci. 17, 57–86.
Wagner, A., Blackstone, N., Cartwright, P., Dick, M., Misof, B., Snow, P., Wagner, G.P., Bartels, J., Murtha, M., Pendleton, J., 1994. Surveys of gene families using polymerase chain reaction: PCR selection and PCR drift. Syst. Biol. 43, 250–261.
Yang, Z., 1993. Maximum-likelihood estimation of phylogeny from DNA sequences when substitution rates differ over sites. Mol. Biol. Evol. 10, 1396–1401.
Zhang, X.Y., Dong, Y.S., Wang, R.R.-C., 1996. Characterization of genomes and chromosomes in partial amphiploids of the hybrid *Triticum aestivum* × *Thinopyrum ponticum* by in situ hybridization, isozyme analysis, and RAPD. Genome 39, 1062–1071.
Ziegler, P., 1999. Cereal *beta*-amylases. J. Cer. Sci. 29, 195–204.
Zwickl, D.J., 2006. Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion. Ph.D. Thesis. The University of Texas, Austin.