ELSEVIER

Available online at www.sciencedirect.com

ScienceDirect

Molecular Phylogenetics and Evolution xxx (2008) xxx–xxx

MOLECULAR PHYLOGENETICS AND EVOLUTION

www.elsevier.com/locate/ympev

Allohexaploidy, introgression, and the complex phylogenetic history of *Elymus repens* (Poaceae)

Roberta J. Mason-Gamer *

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

Received 28 June 2007; revised 11 January 2008; accepted 13 February 2008

Abstract

The phylogenetic position of hexaploid *Elymus repens* within the tribe Triticeae (Poaceae) was examined using cloned sequences from the low-copy nuclear genes encoding phosphoenolpyruvate carboxylase (pepC) and β-amylase. A previous analysis of *E. repens* using data from the nuclear granule-bound starch synthase I (GBSSI) gene had yielded five phylogenetically distinct gene copies, two more than expected from hexaploidy alone. The three gene trees share three distinct *E. repens* clades, suggesting that *E. repens* contains three phylogenetically divergent genomes, contributed by *Hordeum*, *Pseudoroegneria*, and an unknown donor. The two additional GBSSI sequences, including one that was apparently derived from outside of the tribe, appear to reflect past introgression of GBSSI sequences into the *E. repens* genome. On all three trees, the *Hordeum*-like *E. repens* sequences are polyphyletic within *Hordeum*, and the trees are in conflict with regard to the placement of these sequences within *Hordeum*, highlighting multiple contributions from *Hordeum* to *E. repens*.
© 2008 Elsevier Inc. All rights reserved.

Keywords: *Elymus*; Triticeae; Polyploidy; Introgression; Phylogeny

1. Introduction

Over the last two decades, phylogenetic analyses of chloroplast and nuclear DNA data have clarified the relationships between many polyploid species and their diploid relatives. In some of the earliest of these, chloroplast DNA (cpDNA) restriction site data were successfully used to clarify polyploid origins (Doyle et al., 1990; Soltis and Soltis, 1989; Soltis et al., 1989a,b), and cpDNA trees remain important components of phylogenetic analyses of polyploids (Guggisberg et al., 2006; Johnson and Johnson, 2006; Sharbel and Mitchell-Olds, 2001; Tate and Simpson, 2003). The nuclear ribosomal rDNA repeat, including the internal transcribed spacer (ITS) region, has the added potential to identify multiple parental genome donors (Guggisberg et al., 2006; Hughes et al., 2002; Rauscher et al., 2002; Sang et al., 1995; Soltis and Soltis, 1991; Tate and Simpson, 2003; Widmer and Baltisberger, 1999), although the widespread phenomenon of ITS homogenization following polyploid formation (reviewed in Álvarez and Wendel, 2003) can limit the utility of ITS sequences in studies of polyploid origins. Single- and low-copy nuclear gene trees are now common in studies of polyploid phylogeny, and have been used to verify the existence of suspected polyploids (e.g., Smith et al., 2006; Winkworth and Donoghue, 2004), identify genome donors (e.g., Emshwiller and Doyle, 2002; Fortune et al., 2007; Ge et al., 1999; Helfgott and Mason-Gamer, 2004; Johnson and Johnson, 2006; Mason-Gamer, 2001; Petersen et al., 2006; Popp et al., 2005; Slotte et al., 2006), demonstrate multiple polyploid origins (Doyle et al., 2002), clarify combinations of polyploidy and hybridization or introgression (e.g., Cronn et al., 2003; Ferguson and Sang, 2001; Lihová et al., 2006; Mason-Gamer, 2004), and examine gene evolution or gene silencing in polyploids (Barrier et al., 2001; Cronn et al., 1999; Ford and Gottlieb, 2002; Hughes et al., 2002).

The present study clarifies the evolutionary history of the hexaploid species *Elymus repens* (L.) Gould, a

* Fax: +1 312 413 2435.
E-mail address: robie@uic.edu

1055-7903/$ - see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.ympev.2008.02.008

widespread, morphologically variable species in the wheat tribe (Triticeae), using sequence data from low-copy nuclear genes. This native of Europe and Asia has become a problematic weed throughout much of northern North America since its introduction about a century ago (Batcher, 2002). The species is allohexaploid ($2n = 42$), with generally normal, diploidized meioses in which the chromosomes form 21 bivalents (Cauderon, 1958; Dewey, 1961). Cytogenetic studies suggest that two of the three haploid chromosome sets are closely related, with the third set distinct from the other two (Cauderon, 1958; Dewey, 1961). The identities of the genome donors have been investigated using cytogenetic (Assadi and Runemark, 1995; Cauderon and Saigne, 1961; Dewey, 1970b, 1976) and GISH (Ørgaard and Anamthawat-Jónsson, 2001) techniques, but the results are not in complete agreement. Cytogenetic studies of hybrids between *E. repens* and two different *Pseudoroegneria* species (Dewey, 1970b, 1976) suggested that the two similar genome sets of *E. repens* were derived from *Pseudoroegneria* (genome designation **St**), while in a more recent GISH-based study (Ørgaard and Anamthawat-Jónsson, 2001), *P. spicata* probes did not hybridize strongly to the *E. repens* genome, leading the authors to question whether *Pseudoroegneria* is a genome donor. In the same GISH study, *Hordeum californicum* probes hybridized to the *E. repens* genome, indicating the presence of an **H** genome, in agreement with some cytogenetic studies (Cauderon and Saigne, 1961), but Dewey (1984) doubted the presence of the **H** genome on morphological grounds. The cytogenetic analysis of Assadi and Runemark (1995) was one of few that simultaneously indicated both **St** and **H** genomes in *E. repens*: based on chromosome pairing in hybrids between *E. repens* and the presumed **StStStStHH** hexaploid *E. transhyrcanus*, they concluded that *E. repens* shared the **StStStStHH** genome complement. Although the analyses of the *E. repens* genome content are not in complete agreement, they collectively suggest that the two similar genome sets were derived from *Pseudoroegneria* (genome designation **St**) and the third from *Hordeum* (genome **H**). However, studies attempting to ascertain polyploid genome content through observations of meiotic chromosome behavior in hybrids do have some shortcomings. First, they are based on the assumption that chromosome pairing in such hybrids is a direct indicator of relatedness among chromosomes, an assumption that has long been questioned (e.g., Darlington, 1932; de Wet and Harlan, 1972; Seberg and Petersen, 1998). Second, studies of hybrid meioses typically include only a limited number of hybrid combinations. In these cases, patterns of chromosome pairing in hybrids between polyploids and diploids can potentially identify candidate genome donors, but they cannot necessarily provide positive identifications if there are other, related taxa that have not been tested.

The hypothesis that *E. repens* has an **StStStStHH** genome complement was recently examined for six individuals using molecular data from the chloroplast genome and from the single-copy nuclear gene encoding granule-bound starch synthase I (GBSSI or *waxy*; Mason-Gamer, 2004). The chloroplast data identified *Pseudoroegneria* as the maternal genome donor, and the GBSSI data revealed **St** and **H** genome contributions from *Pseudoroegneria* and *Hordeum*, respectively. There were also three additional, phylogenetically distinct GBSSI gene copies within *E. repens*, representing apparent contributions from *Taeniatherum* (genome designation **Ta**), from an unknown donor within the Triticeae, and from a second unknown-donor outside of the tribe. Based on these results, a combination of allohexaploidy and introgression was hypothesized, but it remained unclear which, if any, of the three unexpected gene copies represented an entire third genome set, or which were acquired through introgression. In the present study, phylogenetic analyses of *E. repens* sequences from two additional, unlinked single-copy nuclear loci are used to further clarify the evolutionary history of *E. repens*. Under the assumption that *Pseudoroegneria* and *Hordeum* were donors to *E. repens*, three hypotheses are considered: (1) the basic *E. repens* genome complement is **StStStStHH**, as suggested by the cytogenetic studies, and all three additional GBSSI gene copies have been acquired through introgression; (2) *Taeniatherum* is a genome donor (giving *E. repens* an **StStHHTaTa** complement), and the GBSSI gene copies from both of the unknown donors were acquired through introgression; or (3) the Triticeae unknown-donor sequence represents a complete genome (giving an **StStHHXX** complement, where the origin of **X** is unknown), and the GBSSI gene copies from both *Taeniatherum* and the more distant unknown donor were acquired through introgression. While other hypotheses are conceivable, the three listed here seemed the most reasonable. The ability to distinguish among the hypotheses using comparisons among gene trees is based on the assumption that gene copies present on all three gene trees represent contributions of entire genomes, while copies that are unique to just one gene tree probably represent introgression.

2. Materials and methods

2.1. Molecular data sets

Separate analyses were run for the three nuclear gene data sets, each of which includes sequences from between five and six *E. repens* individuals (Table 1). The GBSSI *E. repens* sequences were previously published (Mason-Gamer, 2004), and pepC and β-amylase *E. repens* sequences were obtained for this analysis. These were analyzed with a broadly representative sample of monogenomic, mostly diploid species from throughout the Triticeae. Most of the sequences from the monogenomic species were previously published, but all three data sets were augmented with new sequences from *Hordeum* and *Pseudoroegneria*, the two presumed genetic contributors to *E. repens* (Table 2). Based on genomic studies of grass crop species

ARTICLE IN PRESS

Table 1
New (boldface) and previously published (plain font) sequences from *E. repens*

Individual	Collection	Locality	GBSSI[a]	pepC	β-Amylase
1	RJMG119	Bonner Co., ID	a1,c,e,f*,q	**a,b,g,q**	**c,d,q,r**
2	RJMG122	Boundary Co., ID	aa,bb,cc,g*,hh	**a,b,h**	**a,i,k**
3	RJMG131	Payette Co., ID	a,aa,dd,f*,h*	**d,f,h,i**	**a,c,d,l**
4	RJMG159	Dane Co., WI	a,aa,b*,dd,i*,t	**c,e,m,q**	**a,b,x**
5	RJMG166	Washington Co., ME	cc,dd,hh,j*,k	**a,b,c,e,i,l**	**a,b,c**
6	RJMG167	Washington Co., ME	a*,a2,dd,g*,hh	**c,d,e**	—

Numbers 1–6 identify individual plants and letters identify cloned sequences from within individuals; these correspond to the labels on the trees (Figs. 1–3).
[a] From Mason-Gamer (2004). Asterisks represent shorter fragments from an alternative primer pair used to offset PCR bias.

(more details are provided below), the three nuclear markers appear to be on three different chromosomes. This is a tentative assumption, based on a small number of grass species, but for this study the three genes are provisionally 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 nuclear gene fragments (more details for each gene can be found in the corresponding publications, cited below). For *E. repens*, 3–4 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). For *Hordeum* and *Pseudoroegneria*, single PCRs were cloned. In all cases, PCR products were cleaned on columns (Qiagen), cloned into pGEM-T Easy vectors (Promega), and transformed into JM109 competent cells (Promega) following the manufacturer'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

Table 2
New (boldface) and previously published (plain font) sequences from *Hordeum* and *Pseudoroegneria*

Taxon (**Genome**)	Identifier[a]	Ploidy	GBSSI	pepC	β-Amylase
Hordeum bogdanii 1 (**H**)	PI 531762	2X	**a**	—	—
Hordeum bogdanii 2 (**H**)	PI 531760	2X	**a**	**a**	**a**
Hordeum brevisubulatum 1 (**H**)	PI 401387	4X	a	—	a,d
Hordeum brevisubulatum 2 (**H**)	PI 401390	2X	a	—	b
Hordeum bulbosum 1 (**I**)	PI 440417	4X	a	**a,b,c**	a
Hordeum californicum 1 (**H**)	MA-138-1-40	2X	a	a	a
Hordeum chilense 1 (**H**)	PI 531781	2X	**a**	**a**	—
Hordeum jubatum 1 (**H**)	RJMG 106	4X	a	a	a,b
Hordeum jubatum 2 (**H**)	RJMG 134	4X	—	—	a,c
Hordeum marinum 1 (**Xa**)	PI 304346	4X	a	a	**a,b**
Hordeum marinum 3 (**Xa**)	PI 304347	2X	**a**	**a**	**a**
Hordeum murinum 1 (**Xu**)	PI 247054	—	**a**	**a,b**	**a**
Hordeum murinum 2 (**Xu**)	CIho 15683	2X	a	a	**b**
Hordeum pusillum 2 (**H**)	CIho 15654	2X	**a**	**a**	**b**
Hordeum stenostachys 1 (**H**)	PI 531791	2X	**a**	**a**	**b**
Hordeum vulgare 1 (**I**)	X07931	2X	n.a.	—	—
*Hordeum vulgare 2 (**I**)*	RJMG 107	2X	—	**a**	**d**
Pseudoroegneria libanotica 1	PI 228391	2X	**a**	**a**	**a**
Pseudoroegneria libanotica 3	PI 228392	2X	**a**	**a**	**a**
Pseudoroegneria spicata 1	PI 232117	2X	a	—	a
Pseudoroegneria spicata 2	PI 236681	2X	a	—	c
Pseudoroegneria spicata 3	PI 610986	2X	a	a	—
Pseudoroegneria spicata 4	D 2844	2X	a	a	a
Pseudoroegneria spicata 6	RJMG 112	—	a,c	—	b
Pseudoroegneria stipifolia 2	PI 313960	2X	—	**a**	**a**
Pseudoroegneria stipifolia 3	PI 531751	4X	—	**a,d**	a
Pseudoroegneria strigosa 1	PI 499637	—	**a**	**b,d**	—
Pseudoroegneria strigosa 2	PI 531755	2X	c	**d**	**b**
Pseudoroegneria tauri 1	PI 380652	2X	**a**	**a**	**a**
Pseudoroegneria tauri 2	PI 401319	2X	**a**	**a**	—
Pseudoroegneria tauri 3	PI 380644	2X	—	**a,b**	—

Numbers after taxon names identify individual plants and letters identify cloned sequences from within individuals; these correspond to the labels on the trees (Figs. 1–3).
[a] RJMG identifiers are author collection numbers; PI, MA, CIho, and D identifiers were assigned by USDA; X07931 is from GenBank (Rohde et al., 1988).

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). Colony PCR products from within each individual were given letter designations, and were screened for sequence variation with single primers. For *E. repens*, 25–38 GBSSI and 15–20 pepC and β-amylase products were screened; for *Hordeum* and *Pseudoroegneria*, 2–5 products were screened for all three genes. Products from within individuals were compared, and any products that differed by more than three base pair substitutions were fully sequenced and added to the data set. While the sampling of *Hordeum* and *Pseudoroegneria* does uncover intra-individual variation in some cases, the sampling strategy was not designed to thoroughly sample allelic variation within individuals.

The GBSSI data set serves as a basis of comparison to the other two nuclear data sets. A phylogenetic analysis from a Triticeae GBSSI data set, including 30 sequences from 6 *E. repens* individuals (Table 1), has been previously published (Mason-Gamer, 2004). The analyses have been rerun here with 7 new sequences from 7 *Hordeum* accessions (Table 2; GenBank EU282316–EU282322) and 5 new sequences from 5 *Pseudoroegneria* accessions (Table 2; GenBank EU282323–EU282327). 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 the five intervening introns 9–13, which are approximately 100 bp each. The original study of *E. repens* (Mason-Gamer, 2004) also included an analysis of a small but broad sample of grass GBSSI exon-only sequences, in order to place a divergent group of *E. repens* sequences that falls outside of the Triticeae. Present discussion of that sequence group is based entirely on the original analysis. 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 present phosphoenolpyruvate carboxylase (pepC) data set includes 24 new sequences from 6 *E. repens* individuals (Table 1; GenBank EU282269–EU282292), 11 new sequences from 8 *Hordeum* accessions (Table 2; GenBank EU282293–EU282303), and 12 new sequences from 9 *Pseudoroegneria* accessions (Table 2; GenBank EU282304–EU282315). These were added to a previously published data set (Helfgott and Mason-Gamer, 2004). The original data set combined two fragments labeled region 1 (approximately 1 kb; GenBank AY553236–AY553269) and region 2 (approximately 600 bp; GenBank AY548399–548432), but the present data set includes just region 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. Based on the location of similar sequences in the rice genome (GenBank AP005781 and AP0058802) and a comparative grass genome map (Devos and Gale, 1997), this member of a small gene family is tentatively assumed to be on the Triticeae group 5 homoeologous chromosomes.

The present β-amylase data set includes 17 new sequences from 5 *E. repens* individuals (Table 1; GenBank EU282238–EU282254), 9 new sequences from 8 *Hordeum* accessions (Table 2; Genbank EU282255–EU282263) and 5 new sequences from 5 *Pseudoroegneria* accessions (Table 2; GenBank EU282264–EU282268). These were added to a published β-amylase data set (Mason-Gamer, 2005), which included monogenomic, mostly diploid individuals. The 1400-bp PCR products obtained using primers 2a-for and 5a-bac (Mason-Gamer, 2005) 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 sequences are from one member of a small gene family; this copy has been mapped to the Triticeae group 2 homoeologous chromosomes (Sharp et al., 1988).

2.2. Alignment and phylogenetic analyses

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 polyploids), PCR-mediated recombination can yield chimeric products (Bradley and Hillis, 1997; Cronn et al., 2002; Judo et al., 1998), and a few were recovered from *E. repens*. Some recombinants were not detected in the initial clone-screening process, when the recombination point fell beyond the read of the screening primer. After being fully sequenced, however, recombinants were easily identified as chimeras in the alignments, either prior to phylogenetic analysis, or by closer inspection of sequences that were tentatively identified as recombinants by their placement on long branches following a quick maximum parsimony analysis. These sequence artifacts were discarded from the analyses.

The GBSSI alignment is generally straightforward in spite of numerous small insertions and deletions in the introns. There were three regions of length variation (aligned positions 916–1033, 1154–1243, and 1397–1529) that were difficult to resolve with the addition of insertions or deletions; these sites were excluded from the phylogenetic analyses. The pepC alignment is a modification of the original alignment (Helfgott and Mason-Gamer, 2004), primarily to better account for apparent transposon activity. This data set was the most difficult to align due to extensive length differences, but upon examination, most of these could be aligned with the addition of gaps. The high level of length variation among the pepC sequences may be related to the fact that the sequence mostly comprises one long intron (approximately 900–1500 bp), in contrast to the five in the GBSSI sequence (about 100 bp each) or the three

in the β-amylase sequence (about 250, 100, and 400 bp). In addition, some *Hordeum* sequences have experienced insertion and/or excision of *Stowaway*-like transposable elements (Bureau and Wessler, 1994) involving at least two sites in the intron, further contributing to length variation within *Hordeum*. At one site, *H. marinum* has an element while *H. murinum* has an apparent excision footprint (Yang et al., 2006) at the same site. At another site, *H. marinum* has an element, most of the remaining *Hordeum* and *Hordeum*-like *E. repens* sequences have one to two apparent excision footprints immediately adjacent to the site, and *H. murinum* has neither an element nor any obvious footprints. Following the final adjustment and addition of gaps, a repetitive simple-sequence region was excluded due to alignment ambiguity (aligned positions 67–111), as were the two regions corresponding to *Stowaway*-like insertions and apparent excision footprints (positions 699–780 and 1045–1138). The β-amylase alignment was the most straightforward of the three data sets; there are length differences in the introns, but they are generally easy to interpret. Excluded from the analysis due to alignment ambiguity were one simple-sequence region (aligned positions 553–570) and two regions corresponding to *Stowaway*-like element insertions and/or excision footprints in some sequences (positions 635–765 and 1478–1674; Mason-Gamer, 2007).

Prior to phylogenetic analysis, 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 trees, and the resulting 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 led to a large, significant improvement in the GBSSI score, a small but significant increase in the β-amylase score, and a non-significant increase in the pepC score. Thus, the GTR+I+Γ model was used for ML analyses of all the GBSSI and β-amylase data sets, and the GTR+Γ model was used for the analysis of the pepC data.

Maximum likelihood (ML) analyses were run using the Mac OS X GUI 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 significant ($\ln L$ increase of 0.01) topology change. Initially, 10 analyses were run with random starting tree topologies. The range of likelihood scores for the pepC data sets across 10 runs was very small (0.0037); thus, the best of the 10 trees was used as the phylogenetic estimate. The ranges of scores for the GBSSI and β-amylase data were larger (8.2626 and 10.2004, respectively), so an additional 20 independent runs were performed for each. Branch support 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 significant topology change. Bootstrap values $\geqslant 70\%$ were recorded on the best ML trees. Bayesian posterior probabilities were estimated in MrBayes (Huelsenbeck and Ronquist, 2001) under the same models of sequence evolution. The analyses were run for 3,000,000 generations, using four chains and uniform priors. Trees were sampled at 1000-generation intervals; the first 300 sampled trees (300,000 generations) were discarded, and Bayesian posterior probability values $\geqslant 0.95$ were obtained from a 95% majority-rule consensus tree generated from the remaining 2700 sampled trees.

3. Results

3.1. GBSSI analysis

The range of likelihood scores across 30 GARLI runs was $-\ln L$ 8433.1948–8442.7291. The tree scores fell into more-or-less discrete categories, in which the four best trees had very similar scores ($-\ln L$ 8433.1948–8433.1952) and differed by almost 0.5 from the next best set of five trees ($-\ln L$ 8433.6242–8433.6253). Topological differences among the nine best trees involve similar sequences but are identical at all deeper nodes; Fig. 1 shows the best of the trees. The complete set of 30 GARLI trees differed in terms of some deeper relationships, but these nodes were also weak in terms of both bootstrap and Bayesian support, and were not used as a basis for drawing conclusions.

The present GBSSI analysis includes more samples of *Pseudoroegneria* and *Hordeum* than the earlier analysis (Mason-Gamer, 2004), but the general results regarding *E. repens* are similar: the *E. repens* sequences fall into four positions within the Triticeae (Fig. 1), and a fifth group falls outside of the tribe (not shown; see Mason-Gamer, 2004). Of the four groups within the tribe, one is closely related to *Pseudoroegneria*, one to *Hordeum*, one to *Taeniatherum*, and one is not closely related to any one of the sampled genera. It was more difficult to sample the *E. repens* GBSSI gene copies than it was for either pepC or β-amylase, because PCR bias strongly favored the *Taeniatherum*-like group at the expense of the others (Mason-Gamer, 2004). Primers were successfully designed to exclude this sequence type, but they, in turn, preferentially amplified sequences from the unknown-donor outside of the tribe, though that bias was not as extreme. As a result, the *Pseudoroegneria*-like and *Hordeum*-like sequences have been undersampled, and their absence from certain individuals is probably an artifact.

The *Pseudoroegneria*-like *E. repens* sequences form a clade with *P. strigosa* (and one of the two *P. spicata* 6

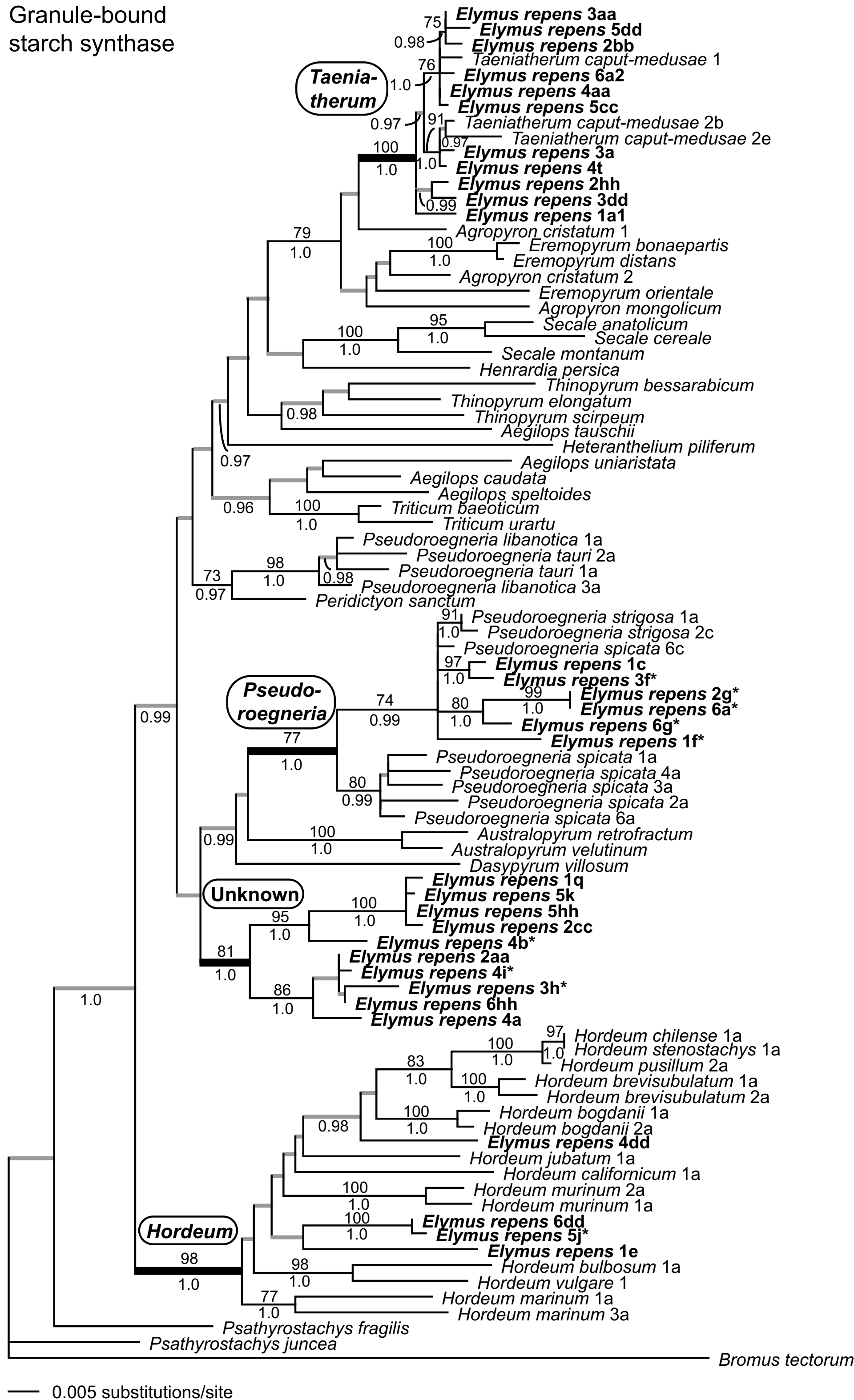

Fig. 1. The best-scoring ML tree from 30 GARLI analyses of the granule-bound starch synthase data set under a GTR+I+Γ model of sequence evolution. Numbers above branches show ML bootstrap support ⩾70%; numbers below branches show Bayesian support ⩾0.95. Gray nodes have bootstrap support <70% and/or Bayesian support <0.95. Where applicable, numbers following taxon names distinguish individuals within species, and letters following these numbers distinguish cloned sequences from within an individual. Asterisks following some *E. repens* sequences indicate shorter sequences obtained using an alternative set of primers designed to counter PCR bias. Four of the five recovered *E. repens* sequence groups are shown; the fifth group falls outside of the tribe (not shown; see Fig. 3 of Mason-Gamer, 2004).

sequences). The *Hordeum*-like sequences from *E. repens* are polyphyletic within *Hordeum*, but none are closely associated with any of the sampled *Hordeum* species. The third *E. repens* sequence group is unassociated with any of the sampled genera, and thus represents an unknown donor. The fourth sequence group is closely related, and very similar, to sequences from *Taeniatherum*, and shares with *Taeniatherum* two frameshifting deletions. The remaining *E. repens* GBSSI sequences (not shown in Fig. 1) form a group outside of the Triticeae, probably within the tribe Poeae (Mason-Gamer, 2004). These have multiple frameshifting length changes that are not shared with Poeae sequences obtained elsewhere (Davis and Soreng, 2007). Thus, this may represent a loss of function that occurred within *E. repens*. In addition, the Poeae-like sequences from *E. repens* lack intron 10, a loss they share with Poeae s. lat. (Davis and Soreng, 2007), a large Pooid clade that includes Aveneae and Poeae s. str., along with the small tribes Hainardieae and Phleeae (Soreng et al., 2007); this character further suggests that the divergent *E. repens* sequence originated outside of the Triticeae.

3.2. pepC analysis

The range of likelihood scores across all 10 GARLI runs was small ($-\ln L$ 6508.4524–6508.4561), and differences between the best tree (Fig. 2) and the other nine involved only nearly-identical taxa. The six *E. repens* individuals combine three phylogenetically distinct pepC gene copies: a *Pseudoroegneria*-like copy, a *Hordeum*-like copy, and a third copy that is not closely associated with any one genus. *Taeniatherum*-like and Poeae-like copies were not recovered for pepC. The *Pseudoroegneria*-like sequences are similar to one another, as indicated by their short branch lengths, and, as on the GBSSI tree, they form a well-supported clade with *P. strigosa*. Five of the eight *E. repens* copies in this clade are in a well-supported subclade (copies 1g, 3d, 4q, 5i, and 6d). Three individuals (3, 4, and 5) show intra-individual polymorphism, suggesting allelic variation. Only one *Pseudoroegneria*-like copy was recovered from individuals 1 and 6; a copy may have been missed in the sampling, or these individuals may be homozygous at the locus. No *Pseudoroegneria*-like copy was obtained from individual 2, out of 18 clones screened; this is the only individual for which one of the three pepC lineages is missing. The *Hordeum*-like *E. repens* pepC sequences are polyphyletic within *Hordeum*, as are the *Hordeum*-like GBSSI sequences. Unlike GBSSI, however, the pepC *Hordeum*-like sequences are closely related to some of the sampled *Hordeum* taxa: sequences from individuals 1, 2, 3, 5, and 6 are closely related to *H. californicum* and the one sequence representing tetraploid *H. jubatum*, and sequences from individuals 1, 4, and 5 are closely related to *H. chilense*, *H. stenostachys*, and *H. pusillum*. Sequence polymorphism within individuals 1 and 5 might represent allelic variation. Individuals 2, 3, 4, and 6 might be homozygous, or one or the other variant may have been missed as a sampling artifact. The third set of *E. repens* pepC copies, recovered from all six individuals, is not closely related to any single genus; thus, a monogenomic donor cannot by hypothesized based on the current tree. The sequences are in a clade with *Heteranthelium*, *Peridictyon*, *Secale*, and *Thinopyrum* with strong Bayesian but very weak bootstrap support; thus, this is not interpreted as a strong hypothesis of the placement of the unknown-donor sequences.

3.3. β-Amylase analysis

The range of likelihood scores across 30 GARLI runs was $-\ln L$ 10403.9933–10414.2014; as with the GBSSI trees, the 30 β-amylase trees differed at many of the deep nodes. None of these nodes received bootstrap support >50% or Bayesian posterior probabilities >0.95; thus, they are weak by any measure and no interpretations were based upon them. The tree scores fell into more-or-less discrete categories, with the very similar scores of the five best trees ($-\ln L$ 10403.9933–10404.0010) separated from the scores of the next six trees by nearly a point ($-\ln L$ 10404.8728–10404.9316), and from the third best group of two trees by about 2.5 points ($-\ln L$ 10406.4198–10406.5067). The five best trees all had the same branching order (Fig. 3).

As in the GBSSI and pepC trees (Figs. 1 and 2), some *E. repens* β-amylase sequences group with *Pseudoroegneria*, some with *Hordeum*, and some form a clade that is not closely related to any one genus. *Taeniatherum*-like and Poeae-like β-amylase sequences were not recovered. The *Pseudoroegneria*-like β-amylase sequences are very similar to one another, and form a distinct clade within *Pseudoroegneria*. They are most closely related to the *P. libanotica*/*P. tauri* clade, which is at odds with their placement with *P. strigosa* on the GBSSI and pepC trees. The *Hordeum*-like β-amylase sequences from *E. repens* are polyphyletic within *Hordeum*. As on the pepC tree, several individuals (1, 2, and 3) have sequences closely associated with *H. pusillum* and *H. stenostachys*. In contrast to pepC, however, no *E. repens* β-amylase sequences were closely related to the *H. californicum*/*H. jubatum* clade, and instead, a gene copy from one individual (3) is closely related to *H. bogdanii*, and the sequences from two individuals (4 and 5) fall outside the main *Hordeum* **H**-genome group. The third set of *E. repens* β-amylase sequences is not associated with any one of the sampled monogenomic genera, but it is in a fairly well-supported clade with *Secale*, *Australopyrum*, and *Dasypyrum*. This is the only one of the three gene trees on which the position of the unknown clade was supported in bootstrap (80%) and Bayesian (1.0) analyses.

4. Discussion

4.1. Genome diversity within E. repens

The analyses of nuclear pepC and β-amylase gene sequences are interpreted in light of a previous study of

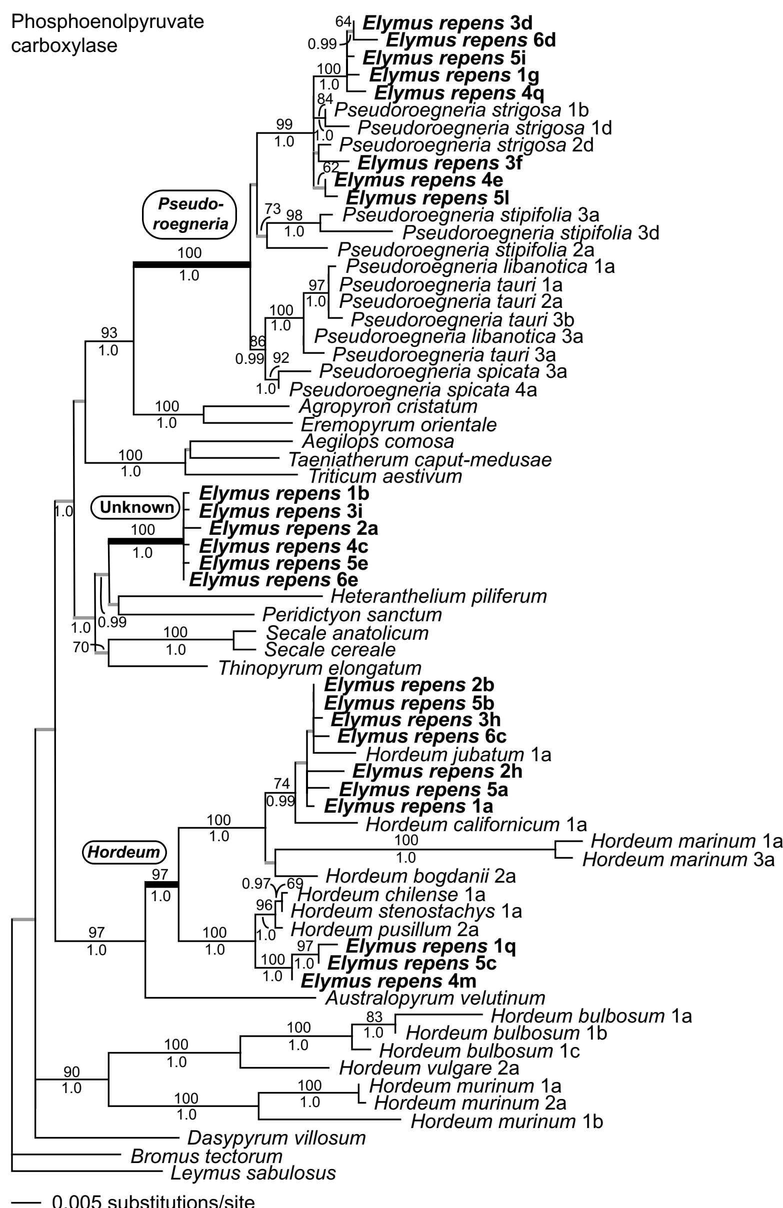

Fig. 2. The best-scoring ML tree from 10 GARLI analyses of the phosphoenolpyruvate carboxylase data set under a GTR+Γ model of sequence evolution. Numbers above branches show ML bootstrap support ⩾70%; numbers below branches show Bayesian support ⩾0.95. Gray nodes have bootstrap support <70% and/or Bayesian support <0.95. Where applicable, numbers following taxon names distinguish individuals within species, and letters following these numbers distinguish cloned sequences from within an individual.

Elymus repens based on cpDNA and nuclear GBSSI data (Mason-Gamer, 2004). While the uniparentally-inherited chloroplast genome cannot identify multiple parental donors, the cpDNA tree did indicate *Pseudoroegneria* as the likely maternal genome donor to *E. repens*. In contrast, the GBSSI tree revealed more potential donors than were expected, including *Pseudoroegneria* (**St**), *Hordeum* (**H**), *Taeniatherum* (**Ta**), an unknown donor from within the tribe, and an unknown donor from outside of the tribe. The *Pseudoroegneria* and *Hordeum* copies are consistent with previous hypotheses about the origin of the hexaploid, but it was not clear which of the additional GBSSI gene copies, if any, represented an entire genome from a third donor, or whether all three were acquired through introgression. Results from two additional nuclear genes, when compared to an updated GBSSI gene tree, allow a clearer understanding of the origin of *E. repens*, based on the assumption that a complete genome will be evident on all three gene trees, whereas gene copies acquired through introgression will only appear on one tree.

The GBSSI, pepC, and β-amylase trees together support the third *a priori* hypothesis proposed in Section 1. On all three trees, *E. repens* sequences group into a *Pseudoroegneria* + *E. repens* clade, a *Hordeum* + *E. repens* clade, and an *E. repens* clade that is distinct from the monogenomic Triticeae genera included in the analysis. Thus, the trees

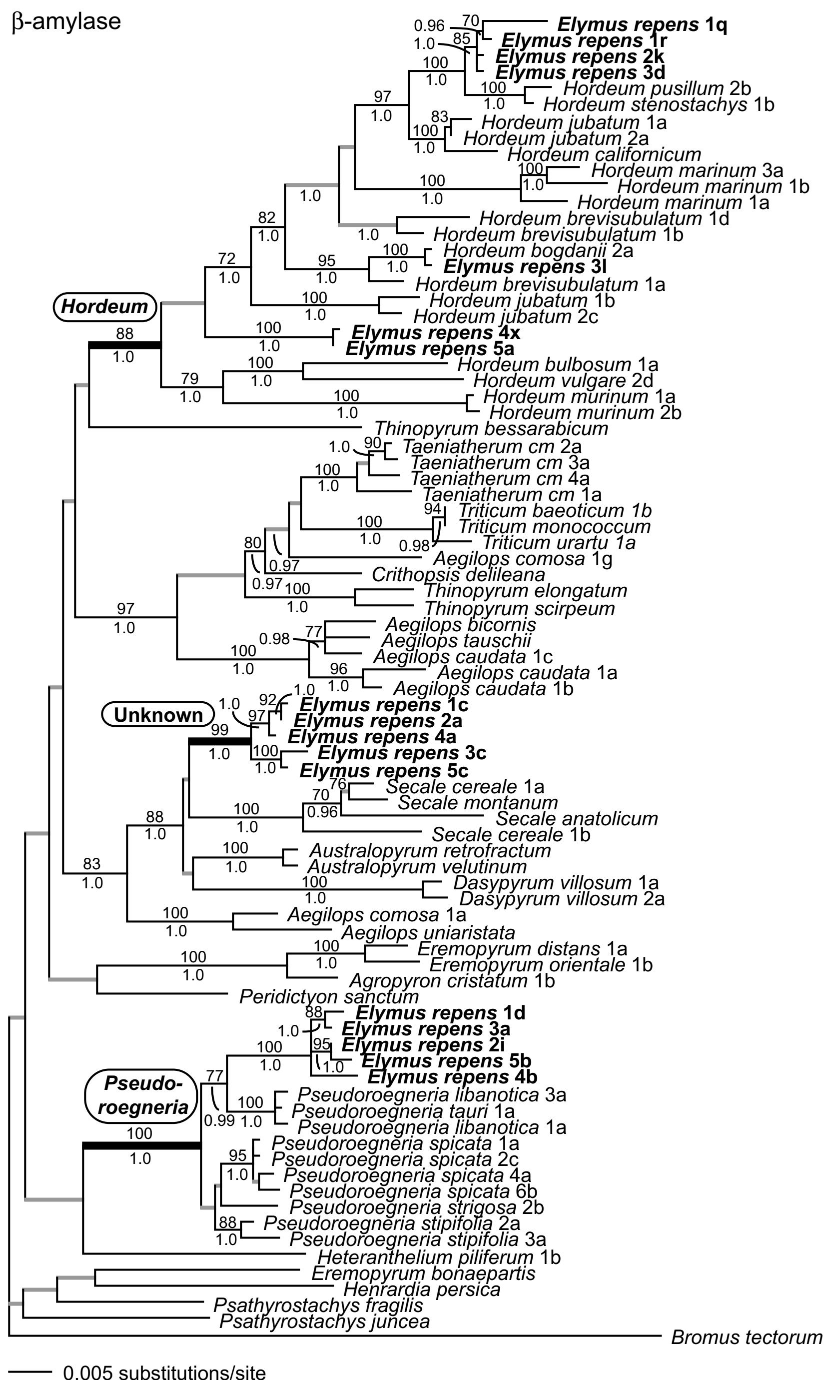

Fig. 3. The best-scoring ML tree from 30 GARLI analyses of the β-amylase data set under a GTR+I+Γ model of sequence evolution. Numbers above branches show ML bootstrap support ⩾70%; numbers below branches show Bayesian support ⩾0.95. Gray nodes have bootstrap support <70% and/or Bayesian support <0.95. Where applicable, numbers following taxon names distinguish individuals within species, and letters following these numbers distinguish cloned sequences from within an individual.

together reveal three phylogenetically distinct genome donors to *E. repens*, including *Pseudoroegneria*, *Hordeum*, and an unknown donor, and suggest that the two additional GBSSI gene copies, one closely related to *Taeniatherum* and the other placed outside of the tribe, are products of introgression. However, as discussed in more detail below, the unusual placement of *Taeniatherum* on the GBSSI tree has led to an additional, *a posteriori* hypothesis that the *Taeniatherum*-like GBSSI sequences of *E. repens* might represent the same genome as the unknown *E. repens* clades on the β-amylase and pepC trees. If this is true, both of the unknown GBSSI sequences would appear to be products of introgression. In either case, it is difficult to speculate about whether successful hybridization events that could lead to introgression have occurred repeatedly, or continue to occur. Meiotic pairing has been examined

in some *E. repens* hybrids (Assadi and Runemark, 1995; Cauderon and Saigne, 1961; Dewey, 1970a, 1976), but *E. repens* × *Taeniatherum* hybrids have not been examined. Naturally-occurring hybrids between *Taeniatherum* and *E. repens* have not been reported, but *E. repens* has been shown to hybridize with other Triticeae species in areas of overlap (Mahelka et al., 2005, 2007), demonstrating at least some potential to acquire genetic material through introgression.

Duplication of the GBSSI gene is another mechanism that would lead to extra GBSSI copies. The gene is clearly duplicated in some taxa, including Rosaceae (Evans et al., 2000) and *Viburnum* (Winkworth and Donoghue, 2004), and possibly in Araliaceae (Mitchell and Wen, 2004). Recently, complex patterns of GBSSI copy diversity in the grass genus *Spartina* were interpreted as a combination of allopolyploidy, duplication, and copy loss (Fortune et al., 2007). While GBSSI duplication might potentially complicate phylogenetic analyses in grasses, the placement of the extra *E. repens* GBSSI copies is not consistent with gene duplication. For example, two duplication events, one during the history of the tribe, and one prior to the origin of the tribe, would be required to explain the two additional GBSSI sequences. Each duplication event should have resulted in two clades that are more-or-less similar to one another, with possible content differences due to copy loss. There is no evidence of such a pattern on the GBSSI tree, without assuming rigidly non-random gene loss involving precisely non-overlapping sets of taxa in each clade (i.e., with all genera losing one or the other copy, and with no taxa except *E. repens* retaining both). More recent duplications within *E. repens* would also lead to additional gene copies, but these would group within existing *E. repens*-containing clades, rather than in distinct clades.

Similarly, lineage sorting of ancestral polymorphisms can potentially explain divergent gene copies within an individual, but it does not appear to be a likely explanation for the GBSSI sequence diversity with *E. repens*, primarily because the polymorphisms are seen only in *E. repens* and not in any of the related diploids. Because none of the distinct *E. repens* GBSSI sequence copies form sister clades, an ancestral polymorphism involving any two of them would have to have been maintained through numerous speciation events in order to persist in one of *E. repens*'s immediate diploid progenitors. The polymorphism would have to be passed through two events of polyploidy into hexaploid *E. repens*, and lost from all of the related diploid Triticeae species sampled here. Thus, at present, introgression appears to be a more parsimonious explanation for the additional GBSSI gene copies than either gene duplication or lineage sorting.

4.2. The Pseudoroegneria, Hordeum, and unknown-donor sequences of E. repens

The acquisition of **St**-genome nuclear sequences from *Pseudoroegneria* is consistent with the chloroplast DNA tree (Mason-Gamer, 2004), on which all six *E. repens* individuals have a *Pseudoroegneria*-like chloroplast genome. On both the GBSSI and pepC trees, the *E. repens* **St**-like sequences are related to and phylogenetically indistinguishable from *P. strigosa*, indicating that species as a potential genome donor. In contrast, the β-amylase tree places the *E. repens* **St**-like sequences in a distinct monophyletic group sister to the *P. libanotica*/*P. tauri* clade. The very close relationship among the *E. repens* **St**-like sequences on each individual tree is consistent with a limited number of *Pseudoroegneria* gene donors. On the other hand, the differences among the trees with respect to the placement of these sequences within *Pseudoroegneria* suggest that the *E. repens* β-amylase gene might have been acquired from a different *Pseudoroegneria* species than were the GBSSI and pepC genes, possibly through introgression following polyploidy. It is also possible that the *Pseudoroegneria* species themselves have undergone introgression or lineage sorting before or after the origin of *E. repens*, which could affect their placement relative to *E. repens* in various ways. This hypothesis of β-amylase lineage sorting or introgression could be strengthened with a more extensive sample of β-amylase sequences from *Pseudoroegneria*, which would provide a better estimate of the relationships among *Pseudoroegneria* species, and a more detailed picture of the placement of the **St**-like *E. repens* sequences.

The sampling from *Hordeum* is not extensive, and so does not allow relationships within the genus to be addressed in detail, but it does span much of the phylogenetic diversity revealed in other molecular analyses (e.g., Blattner, 2004, 2006; Petersen and Seberg, 2003). The sample includes all of the designated genome groups in *Hordeum* (**H**, **I**, **Xa**, **Xu**; Table 2) and a reasonable representation of diversity within the **H**-genome group (Blattner, 2006). The pepC tree in particular shows some unexpected relationships, especially the non-monophyly of the genus. The monophyly of *Hordeum* is well supported by many lines of evidence, so this is probably an artifact. This might reflect the overall pepC sequence instability within *Hordeum*, as revealed by numerous insertions and deletions, some of which involve gain or loss of *Stowaway*-like transposable elements. Large gaps are required to align the *Hordeum* pepC sequences, and thus, the phylogenetic information contained in these length differences is excluded from the phylogenetic analysis. As a result, relationships among clades that differ by numerous length differences may be poorly resolved.

All of the gene trees reveal *Hordeum* as a genome donor to *E. repens*. In contrast to the *Pseudoroegneria*-like clade, in which the *E. repens* sequences form a tight monophyletic group, the *Hordeum*-like sequences of *E. repens* are polyphyletic on all three trees. Thus, there have been multiple contributions from *Hordeum* to *E. repens*, whether due to multiple origins of the hexaploid, or to subsequent hybridization. The trees are incongruent with regard to how the sequences are placed within *Hordeum*, possibly suggesting further introgression from *Hordeum* into *E. repens* follow-

ing the hexaploid's origin. On the pepC tree, the sequences are found within two separate, well-defined **H**-genome clades, and are very similar to their closest *Hordeum* relatives. On the β-amylase tree, some *E. repens* sequences are closely related to *Hordeum* sequences, though two (4x and 5a) are divergent enough that it is not clear which *Hordeum* genome group they are derived from. On the GBSSI tree, *Hordeum*-like sequences were recovered from just four of the six *E. repens* individuals, probably because of the PCR bias described earlier, and none of these are very similar to any of the sampled *Hordeum* sequences.

All three gene trees have an *E. repens* clade that is not closely associated with any one potential donor. This is consistent with a third parental genome donor to *E. repens*, in addition to *Hordeum* and *Pseudoroegneria,* but the identity of the hypothetical third donor is not clear. The GBSSI tree is the least informative with regard to the position of the unknown clade; the relevant nodes on the best of the GARLI trees (Fig. 1) lack both Bayesian and bootstrap support and do not merit mention. On the pepC tree (Fig. 2), the sequences are weakly associated with *Heteranthelium*, *Peridictyon*, *Secale*, and *Thinopyrum elongatum*; these relationships show either bootstrap or Bayesian support, but not both, and are not considered to be strong hypotheses. On the β-amylase tree, the *E. repens* copies are part of a larger, fairly well-supported clade with *Secale*, *Australopyrum*, and *Dasypyrum* (Fig. 3), but the tree does not point to any one potential donor. The inability to identify the donor of the third clade of *E. repens* sequences reflects a sampling artifact if the closest relative is extant and simply has not been included in the analysis. However, extinction of the donor following the origin of *E. repens* is also a possibility. Readers familiar with Triticeae polyploids might wonder whether the donor of this sequence is the same unknown, possibly extinct, donor that contributed the enigmatic **Y** genome to many Asian *Elymus* polyploids. Published analyses of **Y**-genome tetraploids will appear elsewhere, but on all three trees, the **Y**-genome clade is entirely distinct from the *E. repens* unknown-donor clade (unpublished data).

4.3. The "Taeniatherum-like" GBSSI sequence in E. repens

The GBSSI tree is the only one that links *Taeniatherum* to *E. repens*. Based on the GBSSI tree alone (Fig. 1), the *Taeniatherum*-like sequences were initially interpreted as clear evidence of a genetic contribution from *Taeniatherum* (Mason-Gamer, 2004), representing either a complete set of *Taeniatherum* genomes, or introgression of a portion of the *Taeniatherum* genome into *E. repens*. The pepC and β-amylase trees effectively rule out the presence of an entire *Taeniatherum* genome, leaving introgression as the better hypothesis. The difficulty with introgression, however, is that *E. repens* × *Taeniatherum* hybrids have not been reported from areas of overlap, so it is not clear whether, or how often, they hybridize. Analyses of experimental hybrids have not been reported, but in general, *Taeniatherum* shows little tendency to form hybrids with other members of the tribe (Frederiksen and von Bothmer, 1989).

A direct contribution from *Taeniatherum* to *E. repens* through introgression remains a possibility, but the conflicting placement of *Taeniatherum* on different gene trees suggests an alternative to this seemingly obvious conclusion. Specifically, the moderately-supported placement of *T. caput-medusae* with *Agropyron* and *Eremopyrum* on the GBSSI tree (Fig. 1) conflicts with its placement on several other gene trees, including those from pepC and β-amylase, and suggests that *T. caput-medusae* itself may have acquired its GBSSI gene copy from elsewhere. On both the pepC (Fig. 2) and β-amylase (Fig. 3) trees, *T. caput-medusae* groups in or near the wild wheats, *Triticum* and *Aegilops.* This result is further supported by analyses of chloroplast DNA data (Mason-Gamer et al., 2002; Petersen and Seberg, 1997), 5S rDNA sequences (Kellogg and Appels, 1995), and some analyses of ITS sequences (Hsiao et al., 1995). In contrast, the nuclear locus *DMC1* (Petersen and Seberg, 2002) places *T. caput-medusae* well outside of the *Triticum/Aegilops* group, but nowhere near *Agropyron* or *Eremopyrum*. If *Taeniatherum* is, in fact, related to *Triticum* and *Aegilops*, then both *T. caput-medusae* and *E. repens* must have acquired GBSSI gene copies through introgression from an as-yet unknown donor. If so, it remains unclear whether one of the two species acquired the copy and then passed it to the other, or whether they acquired it independently. Perhaps this copy, with its frameshifting deletions, initially became a pseudogene in *E. repens* as a result of relaxed selection following polyploidization (Wendel, 2000). Given the possibility of introgression involving *Taeniatherum*, it is possible that the *E. repens/Taeniatherum* GBSSI copy is the true ortholog of the unknown-donor sequences on the pepC and β-amylase trees, in which case this copy would represent the third complete genome, while the "unknown-donor" clade on the GBSSI tree would represent a case of introgression rather than an entire genome. This generally confusing set of possibilities arises, in part, because of the lack of genus-level congruence among the gene trees.

4.4. Summary

The molecular phylogenetic analyses of *E. repens* agree in part with the earlier cytogenetic studies, but uncover additional genome-level complexity. The combination of three nuclear gene trees has clarified the hexaploid origin of *E. repens*, confirming the involvement of three distinct genome donors, and the likely introgression of two additional GBSSI copies. All of the *E. repens* specimens included in the study were introduced individuals, collected in the United States. It is not yet clear whether the complex genetic makeup demonstrated here is specific to *E. repens* in North America, and work is underway to determine whether Eurasian individuals show a similar genetic composition.

Together, the three data sets show that at least three levels of reticulate evolution have shaped the genome of *E. repens*. The allohexaploid genome complement of *E. repens*, a confluence of three phylogenetically distinct lineages, represents the broadest level of reticulation. The non-monophyly of the *E. repens* sequences within *Hordeum* on all three trees represents a second level, involving multiple genetic contributions from *Hordeum* to *E. repens*. Finally, the two additional GBSSI gene copies show a third level of reticulation, involving introgression into the allohexaploid.

Acknowledgments

Sincere thanks to the USDA-ARS National Germplasm System for seed material, and to Megan Helfgott and Melissa Burns for generating preliminary sequence data from pepC and β-amylase. Barbara Whitlock provided a valuable critique of an earlier version of the manuscript, and two anonymous reviewers provided thoughtful comments on the initial submission. The work was supported by National Science Foundation DEB-9974181 and DEB-0426194.

References

Álvarez, I., Wendel, J.F., 2003. Ribosomal ITS sequences and plant phylogenetic inference. Mol. Phylogenet. Evol. 29, 417–434.

Assadi, M., Runemark, H., 1995. Hybridisation, genomic constitution and generic delimitation in *Elymus* s.l. (Poaceae: Triticeae). Plant Syst. Evol. 194, 189–205.

Barrier, M., Robichaux, R.H., Purugganan, M.D., 2001. Accelerated regulatory gene evolution in an adaptive radiation. Proc. Natl. Acad. Sci. USA 98, 10208–10213.

Batcher, M.S., 2002. Element Stewardship Abstract for *Elytrigia repens* (L.) Desv. ex B.D. Jackson. The Nature Conservancy, Arlington, VA.

Blattner, F.R., 2004. Phylogenetic analysis of *Hordeum* (Poaceae) as inferred by nuclear rDNA ITS sequences. Mol. Phylogenet. Evol. 33, 289–299.

Blattner, F.R., 2006. Multiple intercontinental dispersals shaped the distribution area of *Hordeum* (Poaceae). New Phytol. 169, 603–614.

Bradley, R.D., Hillis, D.M., 1997. Recombinant DNA sequences generated by PCR amplification. Mol. Biol. Evol. 14, 592–593.

Bureau, T.E., Wessler, S.R., 1994. *Stowaway*: a new family of inverted repeat elements associated with the genes of both monocotyledonous and dicotyledonous plants. Plant Cell 6, 907–916.

Cauderon, Y., 1958. Étude cytogénétique des *Agropyrum* français et de leurs hybrides avec les blés. Ann. Amélior. Plant. 8, 389–567.

Cauderon, Y., Saigne, B., 1961. New interspecific and intergeneric hybrids involving *Agropyron*. Wheat Inf. Serv. 12, 13–14.

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.

Cronn, R.C., Small, R.L., Haselkorn, T., Wendel, J.F., 2003. Cryptic repeated genomic recombination during speciation in *Gossypium gossypioides*. Evolution 57, 2475–2489.

Cronn, R.C., Small, R.L., Wendel, J.F., 1999. Duplicated genes evolve independently after polyploid formation in cotton. Proc. Natl. Acad. Sci. USA 96, 14406–14411.

Darlington, C.D., 1932. The control of the chromosomes by the genotype and its bearing on some evolutionary problems. Am. Nat. 66, 25–51.

Davis, J.L., 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.

de Wet, J.M.J., Harlan, J.R., 1972. Chromosome pairing and phylogenetic affinities. Taxon 21, 67–70.

Devos, K.M., Gale, M.D., 1997. Comparative genetics in the grasses. Plant Mol. Biol. 35, 3–15.

Dewey, D.R., 1961. Hybrids between *Agropyron repens* and *Agropyron desertorum*. J. Hered. 52, 13–21.

Dewey, D.R., 1970a. A cytogenetic study of *Agropyron stipifolium* and its hybrids with *Agropyron repens*. Bull. Torrey Bot. Club 97, 315–320.

Dewey, D.R., 1970b. Genome relations among *Elymus canadensis*, *Elymus triticoides*, *Elymus dasystachyus*, and *Agropyron smithii*. Am. J. Bot. 57, 861–866.

Dewey, D.R., 1976. Derivation of a new forage grass from *Agropyron repens* × *Agropyron spicatum* hybrids. Crop Sci. 16, 175–180.

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 16th Stadler Genetics Symposium. Plenum Publishing Company, New York, pp. 209–279.

Doyle, J.J., Doyle, J.L., Brown, A.H.D., Palmer, R.G., 2002. Genomes, multiple origins, and lineage recombination in the *Glycine tomentella* (Leguminosae) polyploid complex: histone H3-D gene sequences. Evolution 56, 1388–1402.

Doyle, J.J., Grace, J.P., Brown, A.H.D., 1990. Reproductively isolated polyploid races of *Glycine tabacina* (Leguminosae) had different chloroplast genome donors. Syst. Bot. 15, 173–181.

Emshwiller, E., Doyle, J.J., 2002. Origins of domestication and polyploidy in oca (*Oxalis tuberosa*: Oxalidaceae). 2. Chloroplast-expressed glutamine synthase data. Am. J. Bot. 89, 1042–1056.

Evans, R.C., Alice, L.A., Campbell, C.S., Kellogg, E.A., Dickinson, T.A., 2000. The granule-bound starch synthase (GBSSI) gene in the Rosaceae: multiple loci and phylogenetic utility. Mol. Phylogenet. Evol. 17, 388–400.

Felsenstein, J., 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. J. Mol. Evol. 17, 368–376.

Ferguson, D., Sang, T., 2001. Speciation through homoploid hybridization between allotetraploids in peonies (*Paeonia*). Proc. Natl. Acad. Sci. USA 98, 3915–3919.

Ford, V.S., Gottlieb, L.D., 2002. Single mutations silence *PgiC2* genes in two very recent allotetraploid species of *Clarkia*. Evolution 56, 699–707.

Fortune, P.M., Schierenbach, K.A., Ainouche, A.K., Jacquemin, J., Wendel, J.F., Ainouche, M.L., 2007. Evolutionary dynamics of *Waxy* and the origin of hexaploid *Spartina* species. Mol. Phylogenet. Evol. 42, 1040–1055.

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.

Frederiksen, S., von Bothmer, R., 1989. Intergeneric hybridization between *Taeniatherum* and different genera of Triticeae, Poaceae. Nord. J. Bot. 9, 229–240.

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.

Guggisberg, A., Bretagnolle, F., Mansion, G., 2006. Allopolyploid origin of the Mediterranean endemic, *Centaurium bianoris* (Gentianaceae), inferred by molecular markers. Syst. Bot. 31, 368–379.

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.

Hsiao, C., Chatterton, N.J., Asay, K.H., Jensen, K.B., 1995. Phylogenetic relationships of the monogenomic species of the wheat tribe, Triticeae (Poaceae), inferred from nuclear rDNA (internal transcribed spacer) sequences. Genome 38, 211–223.

ARTICLE IN PRESS

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.

Huelsenbeck, J.P., Ronquist, F., 2001. MrBayes: Bayesian inference of phylogenetic trees. Bioinformatics 17, 754–755.

Hughes, C.E., Bailey, C.D., Harris, S.A., 2002. Divergent and reticulate species relationships in *Leucaena* (Fabaceae) inferred from multiple data sources: insights into polyploid origins and nrDNA polymorphism. Am. J. Bot. 89, 1057–1073.

Johnson, L.A., Johnson, R.L., 2006. Morphological delimitation and molecular evidence for allopolyploidy in *Collomia wilkenii* (Polemoniaceae), a new species from northern Nevada. Syst. Bot. 31, 349–360.

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., Appels, R., 1995. Intraspecific and interspecific variation in 5S RNA genes are decoupled in diploid wheat relatives. Genetics 140, 325–343.

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. Newsl. 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.

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.

Mahelka, V., Suda, J., Jarolímová, V., Trávnícek, P., Krahulek, F., 2005. Genome size discriminates between closely related taxa *Elytrigia repens* and *E. intermedia* (Poaceae: Triticeae) and their hybrid. Folia Geobot. 40, 367–384.

Mahelka, V., Fehrer, J., Krahulek, F., Jarolímová, V., 2007. Recent natural hybridization between two allopolyploid wheatgrasses (*Elytrigia*, Poaceae): ecological and evolutionary implications. Ann. Bot. 100, 249–260.

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). Am. J. Bot. 92, 1045–1058.

Mason-Gamer, R.J., 2007. Multiple homoplasious insertions and deletions of a Triticeae (Poaceae) DNA transposon: a phylogenetic perspective. BMC Evol. Biol. 7, 92.

Mason-Gamer, R.J., Orme, N.L., Anderson, C.M., 2002. Phylogenetic analysis of North American *Elymus* and the monogenomic Triticeae (Poaceae) using three chloroplast DNA data sets. Genome 45, 991–1002.

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.

Mitchell, A., Wen, J., 2004. Phylogenetic utility and evidence for multiple copies of granule-bound starch synthase I (GBSSI) in Araliaceae. Taxon 53, 29–41.

Ørgaard, M., Anamthawat-Jónsson, K., 2001. Genome discrimination by *in situ* hybridization in Icelandic species of *Elymus* and *Elytrigia* (Poaceae:Triticeae). Genome 44, 275–283.

Petersen, G., Seberg, O., 1997. Phylogenetic analysis of the Triticeae (Poaceae) based on *rpo*A sequence data. Mol. Phylogenet. Evol. 7, 217–230.

Petersen, G., Seberg, O., 2002. Molecular evolution and phylogenetic application of *DMC1*. Mol. Phylogenet. Evol. 22, 43–50.

Petersen, G., Seberg, O., 2003. Phylogenetic analyses of the diploid species of *Hordeum* (Poaceae) and a revised classification of the genus. Syst. Bot. 28, 293–306.

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., Erixon, P., Eggens, F., Oxelman, B., 2005. Origin and evolution of a circumpolar polyploid species complex in *Silene* (Caryophyllaceae) inferred from low copy nuclear RNA polymerase introns, rDNA, and chloroplast DNA. Syst. Bot. 30, 302–313.

Rauscher, J.T., Doyle, J.J., Brown, A.H.D., 2002. Internal transcribed spacer repeat-specific primers and the analysis of hybridization in the *Glycine tomentella* (Leguminosae) polyploid complex. Mol. Ecol. 11, 2691–2702.

Rodríguez, F., Oliver, J.L., Marín, A., Meduna, J.R., 1990. The general stochastic model of nucleotide substitution. J. Theor. Biol. 142, 485–501.

Rohde, W., Becker, D., Salamini, F., 1988. Structural analysis of the waxy locus from *Hordeum vulgare*. Nucleic Acids Res. 16, 7185–7186.

Sang, T., Crawford, D.J., Stuessy, T.F., 1995. Documentation of reticulate evolution in peonies (*Paeonia*) using internal transcribed spacer sequences of nuclear ribosomal DNA: implications for biogeography and concerted evolution. Proc. Natl. Acad. Sci. USA 92, 6813–6817.

Seberg, O., Petersen, G., 1998. A critical review of concepts and methods used in classical genome analysis. Bot. Rev. 64, 372–417.

Sharbel, T.F., Mitchell-Olds, T., 2001. Recurrent polyploid origins and chloroplast phylogeography in the *Arabis holboellii* complex (Brassicaceae). Heredity 87, 59–68.

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.

Slotte, T., Ceplitis, A., Neuffer, B., Hurka, H., Lascoux, M., 2006. Intrageneric phylogeny of *Capsella* (Brassicaceae) and the origin of the tetraploid *C. bursa-pastoris* based on chloroplast and nuclear DNA sequences. Am. J. Bot. 93, 1714–1724.

Smith, J.F., Funke, M.M., Woo, V.L., 2006. A duplication of g*cyc* predates divergence within tribe Coronanthereae (Gesneriaceae): phylogenetic analysis and evolution. Plant Syst. Evol. 261, 245–256.

Soltis, D.E., Soltis, P.S., 1989. Allopolyploid speciation in *Tragopogon*: insights from chloroplast DNA. Am. J. Bot. 76, 1119–1124.

Soltis, D.E., Soltis, P.S., Ness, B.D., 1989a. Chloroplast DNA variation and multiple origins of autopolyploidy in *Heuchera micrantha* (Saxifragaceae). Evolution 43, 650–656.

Soltis, D.E., Soltis, P.S., Ranker, T.A., Ness, B.D., 1989b. Chloroplast DNA variation in a wild plant, *Tolmiea menziesii*. Genetics 121, 819–826.

Soltis, P.S., Soltis, D.E., 1991. Multiple origins of the allotetraploid *Tragopogon mirus* (Compositae): rDNA evidence. Syst. Bot. 16, 407–413.

Soreng, R.J., Davis, J.L., Voionmaa, M.A., 2007. A phylogenetic analysis of Poaceae tribe Poeae sensu lato based on morphological characters and sequence data from three plasmid-encoded genes: evidence for reticulation, and a new classification for the tribe. Kew Bull. 62, 425–454.

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.

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.

Tate, J.A., Simpson, B.B., 2003. Paraphyly of *Tarasa* (Malvaceae) and diverse origins of the polyploid species. Syst. Bot. 28, 723–737.

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.

Wendel, J.F., 2000. Genome evolution in polyploids. Plant Mol. Biol. 42, 225–249.

Widmer, A., Baltisberger, M., 1999. Molecular evidence for allopolyploid speciation and a single origin of the narrow endemic *Draba ladina* (Brassicaceae). Am. J. Bot. 86, 1282–1289.

Winkworth, R.C., Donoghue, M.J., 2004. *Viburnum* phylogeny: evidence from the duplicated nuclear gene GBSSI. Mol. Phylogenet. Evol. 33, 109–126.

Yang, G., Weil, C.F., Wessler, S.R., 2006. A rice *Tc1/mariner*-like element transposes in yeast. Plant Cell 18, 2469–2478.

Yang, Z., 1993. Maximum-likelihood estimation of phylogeny from DNA sequences when substitution rates differ over sites. Mol. Biol. Evol. 10, 1396–1401.

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.

Please cite this article in press as: Mason-Gamer, R.J., Allohexaploidy, introgression, and the complex phylogenetic ..., Mol. Phylogenet. Evol. (2008), doi:10.1016/j.ympev.2008.02.008