

## THE $\beta$ -AMYLASE GENES OF GRASSES AND A PHYLOGENETIC ANALYSIS OF THE TRITICEAE (POACEAE)<sup>1</sup>

ROBERTA J. MASON-GAMER<sup>2</sup>

University of Illinois at Chicago, Department of Biological Sciences, MC 066, 845 West Taylor Street, Chicago, Illinois 60607 USA

There are two forms of  $\beta$ -amylase in the Triticeae crop plants wheat, barley, and rye: an endosperm-specific form encoded by two or three closely linked genes, and a tissue-ubiquitous form encoded by a single gene. Both rice and corn have one ubiquitously expressed form encoded by a single gene. This study focuses on two phylogenetic analyses of  $\beta$ -amylase gene sequences. First, a phylogenetic analysis of coding sequences from wheat, barley, rye, rice, and corn was expected to clarify the relationship between the endosperm-specific and tissue-ubiquitous forms of the protein. Instead, it illustrates possible effects of distant outgroups, based on conflicting patterns of character state variation consistent with different root positions. Next, a broad sample of the monogenomic Triticeae was included in a phylogenetic analysis based on sequences from a portion of the tissue-ubiquitous  $\beta$ -amylase gene. The results were compared to existing Triticeae gene trees, among which extensive conflict had been noted in the past. One additional gene tree has not completely clarified the complexity of the group, but has shed additional light on reticulate phylogenetic patterns within the tribe, including relationships involving *Eremopyrum*, *Thinopyrum*, and the *Triticum/Aegilops* group.

**Key words:**  $\beta$ -amylase; character conflict; Gramineae; outgroups; phylogeny; Poaceae; tree comparisons; Triticeae.

In grasses,  $\beta$ -amylase (1,4- $\alpha$ -glucan maltohydraz; E.C. 3.2.1.2) has been characterized from five major cereal crops: wheat, barley, rye, rice, and corn. The representatives of the Triticeae (wheat, barley, and rye) have two distinct forms of  $\beta$ -amylase, which differ in their expression patterns: one form is specific to the endosperm, while the other has a tissue-ubiquitous pattern of expression (e.g., Ziegler, 1999). The endosperm-specific form has been most extensively studied in barley because of its importance to the brewing industry: of the four enzymes that contribute to the diastatic power of barley malt (Dunn, 1974), i.e., its ability to convert starch into fermentable sugars,  $\beta$ -amylase is considered to be the principle contributor (Gibson et al., 1995). The endosperm-specific enzyme is characterized by a glycine-rich region at the 3' end that is subject to extensive post-translation modification (Lundgard and Svensson, 1987; Bureau et al., 1989). In barley, it is encoded by two to three closely linked genes on chromosome 4 (Kreis et al., 1987; Li et al., 2002).

The ubiquitously expressed form of  $\beta$ -amylase lacks the 3' glycine-rich tail seen on the endosperm-specific form. Wheat, barley, and rye all express a tissue-ubiquitous form of  $\beta$ -amylase that appears, based on studies in barley, to be encoded by a single gene on chromosome 2 (Kreis et al., 1987; Li et al., 2002). A second, highly divergent and possibly paralogous gene is transcribed ubiquitously in wheat tissues, but a corresponding protein has not been detected (Wagner et al., 1999). Unlike wheat, barley, and rye, both rice and corn have just a single, ubiquitously expressed form of  $\beta$ -amylase, which ap-

pears to be encoded by a single gene in corn (Wang et al., 1997) and rice (Yamaguchi et al., 1999).

The genes for  $\beta$ -amylase have not been extensively explored as phylogenetic markers, although a few papers have discussed broader-scale relationships among grass  $\beta$ -amylase sequences (e.g., Daussant et al., 1994; Wang et al., 1997; Ziegler, 1999), and a portion of the gene has been used in a lower-level phylogenetic analysis of *Ipomoea* series *Batatas* (Rajakpase et al., 2004). The many advantages of using single- and low-copy nuclear genes for plant phylogenetic studies, in addition to the more commonly used chloroplast DNA and highly repetitive nuclear genes, have recently been reviewed (Small et al., 2004). In the case of the wheat tribe, Triticeae Dumort., the reconstruction of its complex reticulate history requires data from several genes, representing different portions of the genome.

In the past 10 years, the Triticeae have been the focus of numerous molecular phylogenetic analyses based on data from chloroplast DNA markers, high-copy nuclear genes, and low- or single-copy nuclear genes. Some of the earlier data sets (Hsiao et al., 1995b; Kellogg and Appels, 1995; Mason-Gamer and Kellogg, 1996b) showed appreciable conflict with one another (Mason-Gamer and Kellogg, 1996a), a possible result of past hybridization among the genera, lineage sorting of ancestral variation, introgression involving polyploid intermediates, or a combination of these (Kellogg et al., 1996). Although the phylogenetic relationships among Triticeae genera are not consistent with a single bifurcating tree, it should be possible to clarify reticulate patterns with the use of multiple gene trees, based on molecular markers from throughout the genome.

The goals of this paper are twofold. The first is to examine the relationships among the endosperm-specific and tissue-ubiquitous forms of grass  $\beta$ -amylase using coding sequences from Genbank. This analysis will be used to demonstrate a case of character conflict within the data set that simultaneously supports two alternative hypotheses about the placement of the root of the tree. The second goal is to present a  $\beta$ -amylase

<sup>1</sup> Manuscript received 16 August 2004; revision accepted 21 March 2005.

The author thanks Mary E. Barkworth and D. Megan Helfgott for discussion of the Triticeae; the USDA Germplasm Resource Information Network for seeds; two anonymous reviewers for thoughtful and useful suggestions, including comments on the effects of distant outgroups; and Jack Sullivan for discussion of the data analysis and for providing a copy of an in-press manuscript. The work was funded by the National Science Foundation (DEB-9974181 and DEB-0426194).

<sup>2</sup> E-mail: robie@uic.edu

TABLE 1. Sequences for analysis of endosperm-specific and ubiquitous  $\beta$ -amylase.

Taxon	Expression	GenBank	Reference
<i>Hordeum vulgare</i>	Endosperm	AF414082	Clark et al., 2003
<i>Hordeum vulgare</i>	Endosperm	D49999	Yoshigi et al., 1995
<i>Hordeum vulgare</i>	Endosperm	X52321	Kreis et al., 1987
<i>Secale cereale</i> <sup>a</sup>	Endosperm	X56785	Rorat et al., 1991
<i>Triticum aestivum</i> <sup>b</sup>	Endosperm	AF470353	J.R. Li et al., unpublished data
<i>Hordeum vulgare</i>	Ubiquitous	AF012343	Jung et al., 2001
<i>Secale cereale</i>	Ubiquitous	Z11772	Sadowski et al., 1993
<i>Triticum aestivum</i> <sup>c</sup>	Ubiquitous	X98505	Wagner et al., 1999
<i>Triticum aestivum</i> <sup>d</sup>	Ubiquitous	Y16242	Wagner et al., 1999
<i>Oryza sativa</i>	Ubiquitous	L10345	J. Chen, unpublished data
<i>Oryza sativa</i>	Ubiquitous	L10346	J. Chen, unpublished data
<i>Zea mays</i>	Ubiquitous	AF068119	Y.-N. Lin and S.-M. Wang, unpublished data
<i>Zea mays</i>	Ubiquitous	Z25871	Wang et al., 1997
<i>Castanea crenata</i>	—	AF353207	K. Nomura et al., unpublished data
<i>Ipomoea batatas</i>	—	D12882	Yoshida et al., 1992
<i>Trifolium repens</i>	—	AF049098	Gana et al., 1998

<sup>a</sup> Partial sequence corresponding to amino acids 305–end of accession AF414082.

<sup>b</sup> Partial sequence corresponding to amino acids 183–443 of accession AF414082.

<sup>c</sup> Typical Triticeae ubiquitous form.

<sup>d</sup> Divergent, atypical ubiquitous form.

gene tree for the monogenomic Triticeae, using partial sequences of the single-copy, ubiquitously expressed  $\beta$ -amylase gene. This will be followed by a discussion of other published phylogenetic analyses of the Triticeae, with emphasis on how they compare to selected groups on the  $\beta$ -amylase gene tree.

## MATERIALS AND METHODS

**Relationships among Poaceae sequences**—Selected  $\beta$ -amylase coding sequences from Genbank were analyzed with the goal of clarifying the relationships among the genes encoding endosperm and ubiquitous  $\beta$ -amylase in grasses. The analysis included sequences encoding endosperm and ubiquitous  $\beta$ -amylase from *Hordeum vulgare* L., *Secale cereale* L., and *Triticum aestivum* L.; the single, ubiquitously expressed proteins from *Zea mays* L. and *Oryza sativa* L.; and as outgroups, sequences from the dicotyledons *Castanea crenata* Sieb. & Zucc., *Ipomoea batatas* L. Lam., and *Trifolium repens* L. (Table 1).

Sequences were aligned using Clustal V (Higgins et al., 1992), and alignments were manually adjusted in MacClade version 4.06 (Maddison and Maddison, 2000). Amino acid translations were used to guide the nucleotide alignments. A short portion of the beginning of the gene was difficult to align, as was a longer region at the end of the gene, over half of which coincides with the 3' glycine-rich region characteristic of the Triticeae endosperm form of the protein. The final DNA sequence alignment used in the analysis corresponds to the exons encoding amino acids 8–441 of the 531 amino acid protein from *Hordeum* L. endosperm. The *S. cereale* and *T. aestivum* endosperm sequences are lacking the codons for the first 304 and 182 amino acids, respectively. Unweighted maximum parsimony (MP) analyses of this data set were carried out using PAUP\* 4.0b10 (Swofford, 2002).

The data set was tested for nucleotide stationarity using PAUP\* 4.0b10 and was shown to deviate significantly from stationarity ( $P < 0.0001$ ), violating an assumption that underlies the use of maximum likelihood (ML) models that incorporate nucleotide frequencies. Three sequences were removed to achieve stationarity ( $P = 0.976$ ): *Castanea*, *Trifolium*, and the highly divergent sequence from *Triticum*. Maximum-likelihood analyses were carried out on the remaining taxa using a general-time reversible (GTR) model of sequence evolution (Rodríguez et al., 1990), with some proportion of sites assumed to be invariable (I; Hasegawa et al., 1985), and variation among the remaining sites assumed to follow a gamma ( $\Gamma$ ) distribution (Yang, 1993; Gu et al., 1995; Waddell and Penny, 1996) with shape parameter  $\alpha$ . Parameters for the tree search under the GTR + I +  $\Gamma$  model were first estimated on the shortest MP trees using maximum likelihood and were fixed for an initial ML

tree search. Parameters were further optimized using a successive approximations approach (e.g., Sullivan et al., 1996; Swofford et al., 1996; Frati et al., 1997), in which parameters are re-estimated on the resulting ML tree, fixed for a new search, and re-estimated on the resulting tree, until the same tree is found in two successive searches. Recent empirical tests indicate that this approach is robust to starting tree topology and suggest that it should, under most circumstances, yield results indistinguishable from those obtained using full ML optimization searches (Sullivan et al., in press). For this data set, the second ML tree had the same score as the first. Estimated model parameters based on this tree were: nucleotide frequencies A = 0.26706, C = 0.24645, G = 0.27549, T = 0.21101; relative nucleotide substitution rates AC = 1.66349, AG = 3.90720, AT = 1.05719, CG = 2.17106, CT = 5.83171, GT = 1.00000; I = 0.2488; and  $\alpha = 1.3476$ . Support for this tree was estimated using 1000 heuristic ML bootstrap replicates, under the GTR + I +  $\Gamma$  model, with the same parameters as above. Bayesian posterior probability values were obtained using MrBayes version 3.0 (Huelsenbeck and Ronquist, 2001). Markov chain Monte Carlo analyses were run with random starting trees and four simultaneous chains, one cold and three incrementally heated. Analyses were run for 5 000 000 generations, with flat prior distributions and with a burn-in of 100 000 generations.

Two a priori hypotheses of relationships among the grass  $\beta$ -amylase sequences (Fig. 1A and B), which differ in the placement of the root of the tree, were compared using an MP-based Wilcoxon signed-ranks (WSR; Templeton, 1983) test and ML-based Kishino-Hasegawa (KH; Kishino and Hasegawa, 1989) and Shimodaira-Hasegawa (SH; Shimodaira and Hasegawa, 1999) tests. Hypothesis 1 (Fig. 1A) is that the grass sequences group by expression pattern, with the sequences encoding the Triticeae endosperm sequences in one clade and the ubiquitously expressed sequences from corn, rice, and the Triticeae in another. This was suggested by an earlier distance-based analysis of several grass and dicotyledon  $\beta$ -amylase sequences (Wang et al., 1997) and would imply that the gene duplication leading to the dichotomy between the ubiquitously expressed and endosperm-specific forms of the enzyme arose relatively early during grass evolution (or before the origin of grasses), and that the ortholog corresponding to the endosperm-specific form in the Triticeae was lost from rice and corn. Hypothesis 2 (Fig. 1B) is that all of the Triticeae sequences, endosperm-specific and tissue-ubiquitous, form a single clade relative to rice, corn, and the dicotyledonous outgroups, i.e., that the Triticeae endosperm-specific sequences arose via a later gene duplication closer to the origin of the Triticeae (Daussant et al., 1994; Ziegler, 1999). This would require no subsequent losses in rice and corn.

The WSR tests were carried out with PAUP\* 4.0b10 (Swofford, 2002) using two constraint trees (Fig. 1C and D) corresponding to hypotheses 1 and

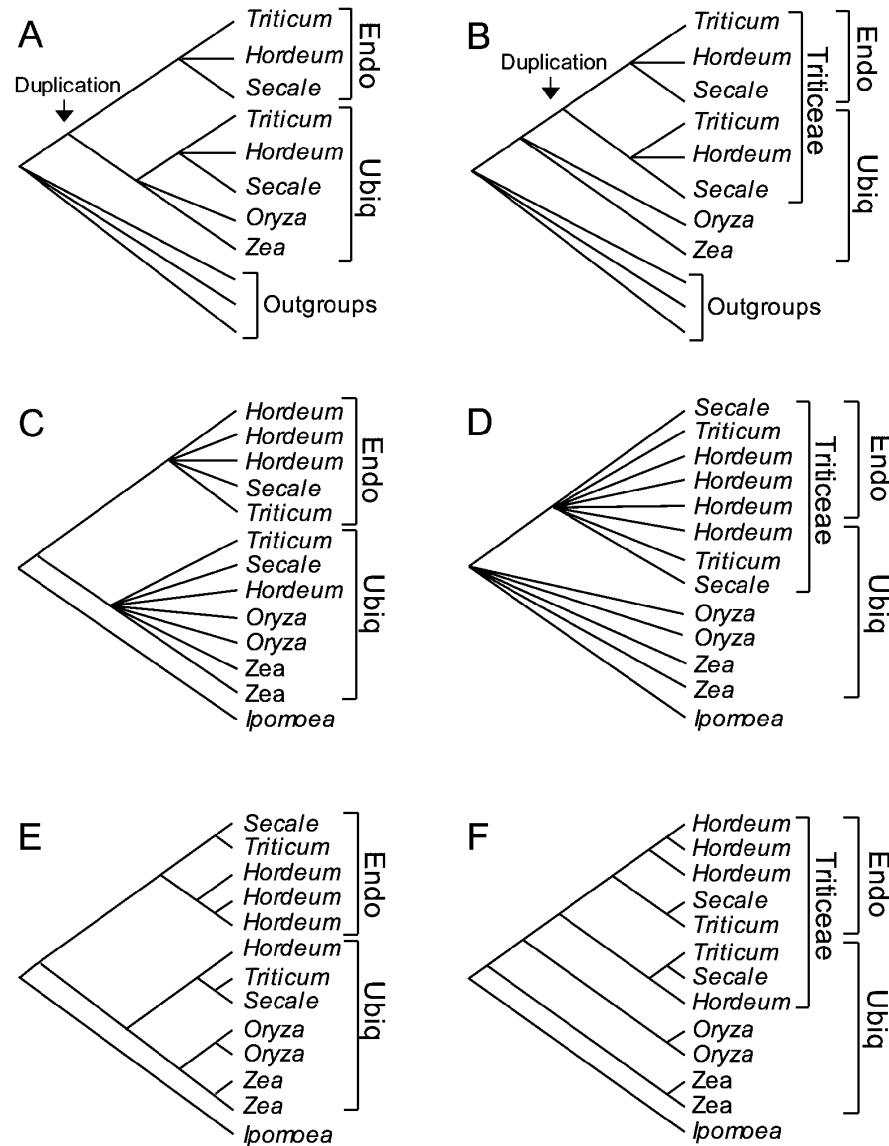


Fig. 1. (A, B) Two hypotheses of relationships among grass  $\beta$ -amylase gene sequences. (A) Two clades correspond to expression pattern, tissue-ubiquitous (Ubiq) vs. endosperm-specific (Endo). According to this hypothesis, the divergence between the tissue-ubiquitous and endosperm-specific forms occurred relatively early, and the endosperm-specific form was subsequently lost from some grasses. (B) All Triticeae sequences form a clade. According to this hypothesis, the tissue-ubiquitous and endosperm-specific forms diverged in a more recent common ancestor of the Triticeae. (C, D) Constraint trees used in Wilcoxon signed-ranks tests comparing the hypotheses shown in (A) and (B), respectively. (E, F) Constraint trees used in Kishino-Hasegawa and Shimodaira-Hasegawa tests comparing the hypotheses shown in (A) and (B), respectively. These topologies were obtained in maximum-likelihood searches constrained with trees (C) and (D), under the GTR + I +  $\Gamma$  model of sequence evolution (model parameters provided in the Materials and Methods, Relationships among Poaceae sequences).

2 (Fig. 1A and B), respectively. Because the placement of the root of the tree is important for distinguishing between the two hypotheses, the WSR tests were repeated with *Castanea* and with *Trifolium* as outgroups and with all three dicotyledons included simultaneously. (The divergent *Triticum* sequence, Y16242, was excluded from all tests.) Characters requiring different numbers of steps on the two constraints were listed, and their character-state distributions among taxa were examined using MacClade version 4.06 (Madison and Maddison, 2000). The KH and SH tests were used to compare the two best ML trees (Fig. 1E and F) obtained in ML tree searches using the same constraint trees (Fig. 1C and D) as were used for the WSR test. In addition, the likelihood of each character on each tree was obtained using PAUP\* 4.0b10 (Swofford, 2002), and the difference in each character's likelihood on ML Tree 2 (Fig. 1F) vs. ML Tree 1 (Fig. 1E) was calculated. The

differences were plotted in order of increasing magnitude using Microsoft Excel 10.0.0 (Microsoft, Redmond, Washington, USA).

**Phylogenetic analysis of the monogenomic Triticeae**—Phylogenetic analyses were done on a broad sample of monogenomic, mostly diploid members of the Triticeae (Table 2). The phylogenetic estimate is based on a 1400-base pair (bp) portion of the ubiquitously expressed  $\beta$ -amylase gene (Fig. 2), amplified using a forward primer in exon 2 and a backward primer in exon 5 (Fig. 2, Table 3). Amplification reactions were carried out in a 10- $\mu$ L volume using 0.5 units *Taq* DNA polymerase (Invitrogen, Carlsbad, California, USA), a 1 $\times$  concentration of the included *Taq* buffer, 15 nmol  $MgCl_2$ , 2 nmol of each nucleotide, and 10 pmol of each primer. Amplification products were cloned into Promega pGem-T Easy vectors (Promega, Madison, Wisconsin,

TABLE 2. Taxa used for the phylogenetic analysis of the monogenomic Triticeae.

Taxon	Authorities	Accession number	Voucher
<i>Aegilops bicornis</i>	Forsskål (Jaub. & Spach)	Morrison s.n.	GH
<i>Aegilops caudata</i> 1a	L.	G 758	GH
<i>Aegilops caudata</i> 1b	—	—	—
<i>Aegilops caudata</i> 1c	—	—	—
<i>Aegilops comosa</i> 1a	Sibth. & Smith	G 602	GH
<i>Aegilops comosa</i> 1g	—	—	—
<i>Aegilops tauschii</i>	Coss.	Morrison s.n.	GH
<i>Aegilops uniaristata</i>	Vis.	G 1297	GH
<i>Agropyron cristatum</i>	(L.) Gaertn.	PI 279802	GH
<i>Australopyrum retrofractum</i>	(Vickery) Á.Löve	PI 533013	GH
<i>Australopyrum velutinum</i>	(Nees) B.K.Simon	D 2873–2878	GH
<i>Bromus tectorum</i>	L.	Kellogg s.n.	GH
<i>Crithopsis delileana</i>	(Schult.) Rosch.	H 5562	GH
<i>Dasypyrum villosum</i> 1a	(L.) Candargy	PI 251478	GH
<i>Dasypyrum villosum</i> 2a	—	PI 470279	GH
<i>Eremopyrum bonaepartis</i>	(Spreng.) Nevski	H 5554	GH
<i>Eremopyrum distans</i>	(C.Koch) Nevski	H 5552	GH
<i>Eremopyrum orientale</i>	(L.) Jaub. & Spach	H 5555	GH
<i>Henrardia persica</i>	(Boiss.) C.E.Hubb.	H 5556	GH
<i>Heterantherium piliferum</i>	(Banks & Sol.) Hochst.	PI 402352	GH
<i>Hordeum brevisubulatum</i> 1a	(Trin.) Link	PI 401387	ID
<i>Hordeum brevisubulatum</i> 1d	—	—	—
<i>Hordeum bulbosum</i>	L.	PI 440417	GH
<i>Hordeum californicum</i>	Covas & Stebbins	MA-138-1–40	GH
<i>Hordeum jubatum</i> 1a	L.	RJMG 106	ID
<i>Hordeum jubatum</i> 1b	—	—	—
<i>Hordeum jubatum</i> 2a	—	RJMG 134	ID
<i>Hordeum jubatum</i> 2c	—	—	—
<i>Hordeum violaceum</i>	Boiss. & Hohen	PI 401390	GH
<i>Peridictyon sanctum</i>	(Janka) Seberg, Fred., & Baden	KJ 248	GH
<i>Psathyrostachys fragilis</i>	(Boiss.) Nevski	PI 343192	GH
<i>Psathyrostachys juncea</i>	(Fisch.) Nevski	PI 206684	GH
<i>Pseudoroegneria spicata</i> 1a	(Pursch.) Á.Löve	PI 232117	GH
<i>Pseudoroegneria spicata</i> 2c	—	PI 236681	GH
<i>Pseudoroegneria spicata</i> 4a	—	D 2844	GH
<i>Pseudoroegneria spicata</i> 6b	—	RJMG 112	ID
<i>Pseudoroegneria stipifolia</i>	(Czern. Ex Nevski) Á.Löve	PI 531751	GH
<i>Secale cereale</i> 1a	L.	Kellogg s.n.	GH
<i>Secale cereale</i> 1b	—	—	—
<i>Secale montanum</i>	Guss.	T 36554	GH
<i>Secale montanum</i> ssp. <i>anatolicum</i>	(Boiss.) Tzelev	PI 206992	GH
<i>Taeniatherum caput-medusae</i> 1a	(L.) Nevski	PI 208075	GH
<i>Taeniatherum caput-medusae</i> 2a	—	RJMG 189	ID
<i>Taeniatherum caput-medusae</i> 3a	—	PI 314697	GH
<i>Taeniatherum caput-medusae</i> 4a	—	PI 317475	GH
<i>Thinopyrum bessarabicum</i>	(Savul. & Rayss) Á.Löve	PI 431711	GH
<i>Thinopyrum elongatum</i>	(Host) D.R.Dewey	PI 531719	GH
<i>Triticum baeoticum</i>	Boiss.	Morrison s.n.	GH
<i>Triticum monococcum</i>	L.	PI 221413	GH

USA) using the Promega PCR cloning kit according to instructions, except that reaction volumes were halved. Cloned products were amplified directly from colonies in 40- $\mu$ L PCR reactions with 0.5 unit *Taq* DNA polymerase, a 1 $\times$  concentration of *Taq* buffer, 60 nmol MgCl<sub>2</sub>, 8 nmol of each nucleotide, and 40 pmol of each primer. Amplified clones were cleaned using 1 unit shrimp alkaline phosphatase (USB, Cleveland, Ohio, USA) and 5 unit exonuclease I (USB, Cleveland, Ohio, USA); the mixture was heated to 37°C for

15 min to allow the reactions to occur and 75°C for 15 min to denature the enzymes.

After cleaning, 2  $\mu$ L of each amplified, cloned fragment was sequenced using 2  $\mu$ L BigDye Terminator version 3.1 (Applied Biosystems, Foster City, California, USA) and 3.2 pmol of sequencing primer in a 10- $\mu$ L reaction volume. Both strands of each product were sequenced using six primers (Fig. 2, Table 3), and sequence fragments were assembled in Sequencher version



Fig. 2. A schematic drawing of the tissue-ubiquitous  $\beta$ -amylase gene from barley showing the relative locations of exons (rectangles), introns (lines), and amplification and sequencing primers (arrows) used in the analysis of the diploid Triticeae.

TABLE 3. Triticeae  $\beta$ -amylase amplification and sequencing primers corresponding to Fig. 2.

Primer	Sequence	Used for
2a-for	GCCATCATGTCRTTCCACCA	Amplification, sequencing
3a-bac	ATGAATTCTCCRAYGCCTGG	Sequencing
3a-for	CCAGGCRTNGGAGAATTCAT	Sequencing
4a-bac	CTGCTGCTGCTTTGAARTCTG	Sequencing
4b-for	TACCTRSAAAGCAGACTTCAAAG	Sequencing
5a-bac	TCRGCTGCATGGTTTGGAAC	Amplification, sequencing

4.1 (Genecodes, Ann Arbor, Michigan, USA). Sequence alignments were done using Clustal V (Higgins et al., 1992), with some manual adjustments. From the aligned data set of 1926 bp, positions 1–23 and 1905–1926, corresponding to the amplification primers, were excluded from the phylogenetic analysis. In addition, positions 1518–1654 were excluded; seven taxa have transposable elements of varying length within this region, and these were difficult to align unambiguously. Their characteristic terminal repeats identify these as *Stow-away* elements (Bureau and Wessler, 1994), which are relatively common in grass genes (e.g., Feschotte et al., 2002). Sequences from the Triticeae data set have been accessioned in GenBank under the numbers AY821686–AY821734, and the aligned data set is available for download at the website [http://www.uic.edu/depts/bios/ecoevo/masongamer/Diploids\\_AJB.txt](http://www.uic.edu/depts/bios/ecoevo/masongamer/Diploids_AJB.txt).

Tree searches were done using ML under the GTR + I +  $\Gamma$  model of sequence evolution. First, because of the size of the data set and the time required to run analyses under this relatively complex model, the GTR + I +  $\Gamma$  model was compared to 15 other models (e.g., Swofford et al., 1996; Frati et al., 1997; Sullivan et al., 1997) to determine whether a simpler, less computationally intensive model would be sufficient. Four nucleotide substitution models were examined: Jukes-Cantor (Jukes and Cantor, 1969), Kimura two-parameter (Kimura, 1980), Hasegawa-Kishino-Yano (HKY; Hasegawa et al., 1985), and GTR (Rodríguez et al., 1990). Each substitution model was paired with each of four models of among-site rate variation: (1) no rate heterogeneity; (2) some sites invariable (I; Hasegawa et al., 1985) with equal rates of change among the remaining sites; (3) rate heterogeneity among sites following a gamma distribution ( $\Gamma$ ; Yang, 1993); and (4) some sites invariable, with gamma-distributed variation among the remaining sites (I +  $\Gamma$ ; Gu et

al., 1995; Waddell and Penny, 1996). Models were compared based on ML scores estimated for all 14 trees obtained in an equally weighted MP analysis. The best score (GTR + I +  $\Gamma$ ) was compared to the two next-best scores (HKY + I +  $\Gamma$  and GTR +  $\Gamma$ ) using a likelihood ratio test (Felsenstein, 1981; Huelsenbeck and Crandall, 1997; Huelsenbeck and Rannala, 1997; Sanderson, 1998), which showed the GTR + I +  $\Gamma$  model to have the best fit to the data ( $P < 0.05$ , Bonferroni-corrected for two non-independent comparisons).

The GTR + I +  $\Gamma$  parameters estimated on the best MP tree were fixed for an ML tree search, with the starting tree obtained by stepwise addition with 10 trees held at each step. Further optimization was done using a successive approximations approach as described for the exon data set. The second ML tree topology was identical to the first and was used as the phylogenetic estimate for the tribe. The estimated model parameters were: nucleotide frequencies A = 0.31903, C = 0.21553, G = 0.20887, T = 0.25657; relative substitution rates AC = 1.35197, AG = 4.37860, AT = 1.26551, CG = 2.42244, CT = 5.44094, GT = 1.00000; I = 0.3063; and  $\alpha = 1.8357$ . Support for the tree was estimated using 100 ML bootstrap replicates under the GTR + I +  $\Gamma$  model, with the same fixed model parameters as were estimated from the tree. Bayesian posterior probability values were obtained using MrBayes version 3.0 (Huelsenbeck and Ronquist, 2001) following the same procedure as used for the exon sequences.

## RESULTS

**Relationships among Poaceae sequences**—The MP analysis (Fig. 3A) resulted in three trees of length 1661, with consistency index excluding uninformative characters of 0.658, retention index of 0.706, and rescaled consistency index of 0.505. The grass sequences form a monophyletic group relative to the three dicotyledon sequences. The strongly divergent, ubiquitously transcribed wheat sequence (Y16242) falls at the base of the grasses, well outside of all of the other sequences from the Triticeae. The ML tree (Fig. 3B), which excludes *Castanea*, *Trifolium*, and the divergent *Triticum* sequence, is consistent with the MP tree. On both trees, the Triticeae endosperm-specific and tissue-ubiquitous sequences form a well-supported clade relative to the tissue-ubiquitous sequences of *Oryza* and *Zea*.

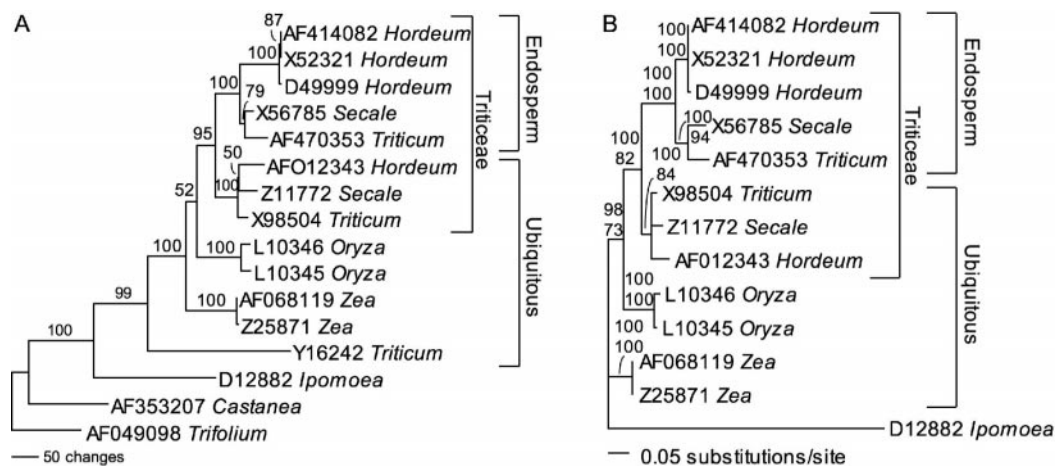


Fig. 3. Analysis of exon sequences encoding both endosperm-specific and tissue-ubiquitous  $\beta$ -amylase. Numbers preceding taxon names are GenBank accession numbers. (A) One of three most-parsimonious trees; the nodes with lowest support (50% and 52%) collapse in the strict consensus tree. Tree includes three dicotyledon outgroups and a divergent, apparently paralogous,  $\beta$ -amylase gene copy (*Triticum*, Y16242). Bootstrap support values (above nodes) are from 1000 heuristic replicates. (B) Maximum likelihood tree derived from a search under the GTR + I +  $\Gamma$  model of sequence evolution (parameters are provided in the Materials and Methods, Relationships among Poaceae sequences). Support values: Bayesian posterior probabilities under the GTR + I +  $\Gamma$  model (top, above nodes) and bootstrap support values from heuristic ML searches of 1000 replicates under the same model (bottom, above nodes). Relative to the maximum-parsimony analysis (A), three sequences were excluded in order to achieve stationarity of nucleotide frequencies.

Although there is reasonable support for the Triticeae clade on both trees, the WSR test (with *Ipomoea* as an outgroup and the divergent *Triticum* sequence excluded) suggests a more complex underlying pattern. The shortest tree consistent with the all-Triticeae clade in Hypothesis 2 (Fig. 1B) is equivalent to the shortest unconstrained trees (Fig. 3A), at 955 steps long. The shortest tree consistent with the tissue-ubiquitous clade in Hypothesis 1 (Fig. 1A), is 964 steps, an increase of nine steps. However, underlying the nine-step increase are 93 characters that require different numbers of steps on the two trees: 51 characters require more steps on the tissue-ubiquitous constraint, while 42 characters simultaneously require more steps on the monophyletic Triticeae tree. Among these characters, several distinct patterns of variation are obvious. Among the characters that are more consistent with Triticeae monophyly are 17 uncontradicted characters that support the Triticeae clade and 14 that support a Triticeae + *Oryza* clade. Among the characters that are more consistent with the tissue-ubiquitous clade are 13 uncontradicted characters that support that clade, 10 that support a *Zea* + *Oryza* clade, and two that support an *Oryza* + Triticeae-tissue-ubiquitous clade. The 56 characters listed above lack homoplasy and are uniform within the clades they support. (The remaining characters are variable in their patterns and don't fall into obvious categories: for example, a state might be shared by all of Triticeae and one of the rice sequences or shared by all of Triticeae except for one member with an autapomorphy, etc.) Because of the ambiguous signal in the data set, the net change of nine steps on the constraint tree was not significant in the WSR test ( $P = 0.366$ ;  $N = 93$ ; TS (test statistic) = 1978.5). Unweighted MP analyses of amino acid sequences with *Ipomoea* as the outgroup (results not shown) were similarly ambivalent: the tree consistent with Hypothesis 1 was, in this case, slightly shorter than the tree consistent with Hypothesis 2, with lengths of 341 and 342 steps, respectively. Ten amino acid characters require one additional step on the Triticeae monophyly constraint, and nine simultaneously require one additional step on the tissue-ubiquitous constraint. Bootstrap support for the conflicting ubiquitously expressed and Triticeae clades on the amino acid tree were both low (45% and 35%, respectively).

Because character state variation among the outgroups could change the interpretation of character state distribution within the grasses, WSR tests were rerun on the nucleotide data with each of the other outgroups and with all three. With *Castanea* as the outgroup, the tree constrained by Hypothesis 1 required 952 steps, 13 steps more than the tree constrained by Hypothesis 2. There were 77 characters that differed in the number of changes on the two trees, with 45 requiring more steps on the tissue-ubiquitous monophyly constraint, and 32 requiring more on the Triticeae monophyly constraint. The difference in length was not significant ( $P = 0.139$ ;  $N = 77$ ; TS = 1248). With *Trifolium* as the outgroup, trees constrained by Hypothesis 1 required just five steps more than those constrained by Hypothesis 2 (948 vs. 943 steps). Of the 83 characters that differed in number of changes, 44 required more steps on the tissue-ubiquitous monophyly constraint, and 39 required more on the Triticeae monophyly constraint. The difference in tree length was not significant ( $P = 0.583$ ;  $N = 83$ ; TS = 1638). Finally, with all three outgroups, the tree constrained by Hypothesis 1 was 16 steps longer than the tree constrained by Hypothesis 2 (1394 vs. 1378 steps), with 81 characters differing in number of changes on the two trees. Of these, 48 required more steps on the tissue-ubiquitous monophyly con-

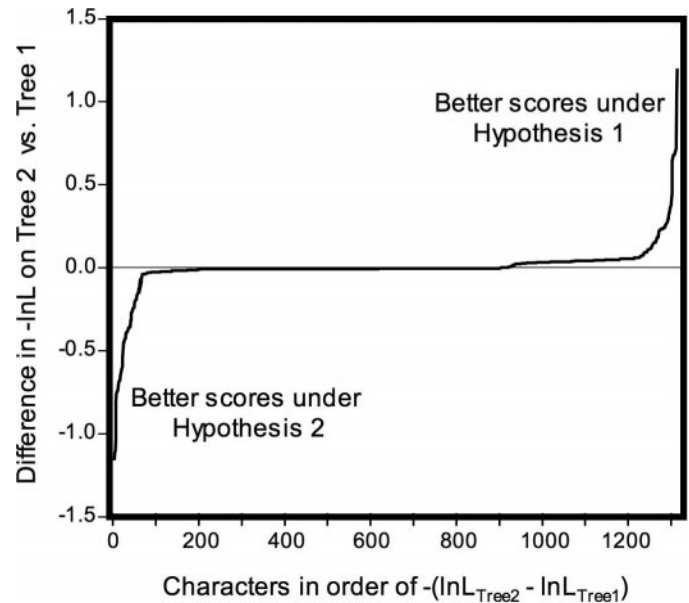


Fig. 4. Differences in likelihood scores of all characters in increasing order. Likelihood scores for each character were compared on the two trees used in the Kishino-Hasegawa and Shimodaira-Hasegawa tests. Tree 1 (Fig. 1E) and Tree 2 (Fig. 1F) correspond to hypotheses of tissue-ubiquitous monophyly and Triticeae monophyly, respectively. Negative differences (left side of diagram) involve characters with better (lower) scores on trees constrained for Triticeae monophyly, while positive differences involve characters with better scores on trees constrained for tissue-ubiquitous monophyly.

straint, while 33 required more steps on the Triticeae monophyly constraint. The 16-step increase in length was not significant ( $P = 0.081$ ;  $N = 81$ ; TS = 1336.5).

The inability of the  $\beta$ -amylase exon characters to distinguish between the two hypotheses is further illustrated by the results of both the SH and KH tests ( $P = 0.075$  and  $P = 0.152$ , respectively). Furthermore, comparisons of likelihood scores of each character on both trees showed a pattern of character conflict within the data set similar to that seen in the WSR test; some characters have better scores in analyses constrained by Triticeae monophyly (Fig. 4, left side), while others have better scores when constrained by tissue-ubiquitous monophyly (Fig. 4, right side).

**Phylogenetic analysis of the monogenomic Triticeae**—The ML analysis resulted in a single tree with score  $-\ln L = 8846.175$  (Fig. 5). Although the sampling was not specifically designed to address the monophyly of the genera, several are supported as monophyletic groups, including *Australopyrum*, *Hordeum*, *Pseudoroegneria*, *Secale*, and *Taeniatherum*. Others are represented by only a single species (*Dasypyrum villosum*) or individual (*Agropyron cristatum*, *Crithopsis delileana*, *Henrardia persica*, *Heteranthelium piliferum*, and *Peridictyon sanctum*). Even with the small numbers of species sampled per genus, however, several genera are non-monophyletic on the tree; these will be addressed in more detail in the Discussion: (1) Two species of *Psathyrostachys* form a paraphyletic grade with *Henrardia* and *Eremopyrum bonaepartis*. This is supported with high ( $\geq 95\%$ ) posterior probability, but very low ( $< 50\%$ ) ML bootstrap values. (2) *Eremopyrum* is polyphyletic, with *E. distans* and *E. orientale* grouped with *Agropyron cristatum* and *E. bonaepartis* with *Henrardia*. (3)

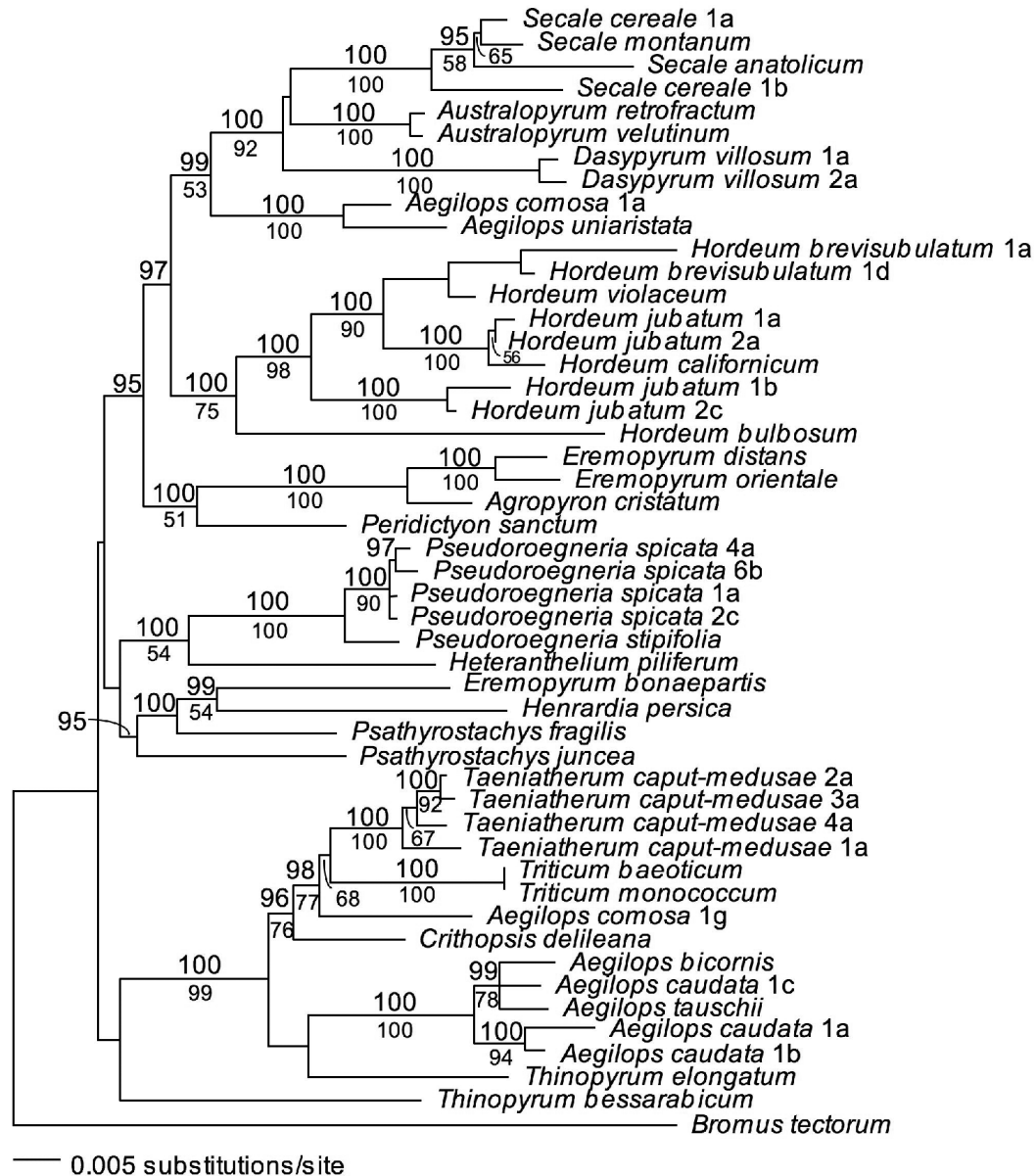


Fig. 5. Maximum-likelihood (ML) estimate of relationships among the monogenomic Triticeae, under a GTR + I +  $\Gamma$  model of evolution (model parameters are provided in the Materials and Methods, Relationships among Poaceae sequences). Numbers above nodes are Bayesian posterior probability values >95%; numbers below nodes are bootstrap values >50% based on 100 full heuristic ML replicates under a GTR + I +  $\Gamma$  model, using the same parameters as those used in the initial tree search. Number/letter labels after taxon names distinguish individuals within species and clones within individuals, respectively, and are applied only when two or more individuals per species and/or two or more clones per individual are presented.

*Thinopyrum bessarabicum* and *Thinopyrum elongatum*, representing the only two diploid species within an evolutionarily complex genus, do not form a clade. *Thinopyrum elongatum* is part of a well-supported clade with *Aegilops*, *Triticum*, *Crithopsis*, and *Taeniatherum*, while *Thinopyrum bessarabicum* falls outside of that clade. (4) Sequences from the *Triticum/Aegilops* (wild wheat) group are not monophyletic. Most fall within a well-supported clade with *Thinopyrum elongatum*, *Crithopsis* and *Taeniatherum*. Within that clade, sequences from *A. bicornis*, *A. caudata*, and *A. tauschii* form a weak clade with *Thinopyrum elongatum*, while *Aegilops comosa*, *Triticum monococcum*, and *T. baeticum* are in a well-supported clade with *Taeniatherum*. *Aegilops uniaristata* and a

second sequence type from *A. comosa* form a clade that is distant from the other *Triticum/Aegilops* sequences. An additional round of PCR and cloning of  $\beta$ -amylase from *Aegilops* failed to yield the more common *Aegilops* sequence type from *A. uniaristata* or the less common type from *A. bicornis*, *A. caudata*, or *A. tauschii*.

The tree supports several intergeneric relationships. Most are mentioned above in the context of nonmonophyletic genera, including: *Aegilops* + *Triticum* + *Thinopyrum* + *Crithopsis* + *Taeniatherum*; its two subclasses *Crithopsis* + *Triticum* + *Aegilops* + *Taeniatherum* and *Triticum* + *Aegilops* + *Taeniatherum*; and the two *Eremopyrum* clades *Eremopyrum* + *Agropyron* and *Eremopyrum* + *Henrardia*. One additional

intergeneric clade with strong bootstrap and posterior probability support is *Australopyrum* + *Dasyphyrum* + *Secale*. Numerous other intergeneric clades have high ( $\leq 95\%$ ) posterior probability values, but low ( $< 50\%$ ) ML bootstrap support; most of these will not be discussed in detail.

## DISCUSSION

**Relationships among Poaceae sequences**—*Triticum* is unique (so far) among grasses in having two very distinct ubiquitously transcribed  $\beta$ -amylase genes. The function of one of the two genes is not clear, because its protein product has not been detected (Wagner et al., 1999). Based on its high level of amino acid sequence divergence relative to the more typical *Triticum* ubiquitous form (Wagner et al., 1999) and its basal placement relative to other grass sequences in the present analysis, it appears to represent a distinct lineage. However, until the exact function (if any) of this gene is elucidated and/or its ortholog is detected in additional grass species, the details of its evolutionary history and significance will not be clear.

Among the remaining grass  $\beta$ -amylase sequences, one hypothesis of relationships, derived from a dendrogram of  $\beta$ -amylase amino acid sequences (Wang et al., 1997), suggests that the ubiquitous and endosperm-specific sequences form separate evolutionary lineages. This would be consistent with a gene duplication early in the evolutionary history of grasses, before the deep divergence that separates the Panicoideae (which includes *Zea*) from the Ehrhartoideae and Pooideae (which include *Oryza* and the Triticeae, respectively). This scenario would require the loss of the endosperm-specific ortholog from rice and corn. In fact, given the likely relationship among the three subfamilies [(Panicoideae, (Ehrhartoideae, Pooideae)); GPWG, 2001], two independent losses would be needed to explain the pattern. A second hypothesis is that the two forms in the Triticeae arose via a more recent gene duplication, closer in time to the origin of the tribe (Daussant et al., 1994; Ziegler, 1999), in which case all of the Triticeae sequences would form a monophyletic group relative to the rice and corn sequences. This is more intuitively appealing because it does not require the subsequent loss from rice and corn. Both the MP and ML trees are more consistent with the latter hypothesis.

Although the nucleotide data are more consistent with the second hypothesis above, they are unable to reject the first hypothesis in WSR, KH, or SH tests. The patterns of character state variation, apparent in both MP and ML analyses of the fit of the characters to the two hypotheses, clearly illustrate ambiguous signal with regard to the two hypotheses. Because the difference between the two hypotheses involves the placement of the root of the tree, it may be that the observed ambiguity results from the use of a too-distant outgroup. Enough time may have passed since the monocot–dicot divergence such that informative signal in the dicot outgroup sequences is effectively random relative to grasses. Too-distant outgroups can be problematic if enough time has passed that phylogenetic signal is obliterated by homoplasious character states shared by the long branch leading to the outgroup and any nonbasal ingroup lineage (Miyamoto and Boyle, 1989; Wheeler, 1990), leading to the topological phenomenon of long branch attraction (e.g., Felsenstein, 1978; Henny and Penny, 1989; Kim, 1996; Huelsenbeck, 1998). Topological ambiguities due to distant outgroups have been explored in several

studies of plants (e.g., Qiu and Palmer, 1999; Qiu et al., 2001; Graham et al., 2002; Xiang et al., 2002). In the present case, the topology is the same for each outgroup or combination of outgroups, and, unlike some of the previous studies, the support for the placement of the root is relatively well supported. It is the pattern of character conflict that, in this case, reveals a potential problem with the outgroup. Character conflict is demonstrated on a broad scale with the WSR, SH, and KH tests, and in more detail with MP and ML character scores across the two main competing hypotheses. Thus, the unavailability of an appropriate  $\beta$ -amylase outgroup sequence has confounded the initial goal of clarifying the relationship between the tissue-ubiquitous and endosperm-specific  $\beta$ -amylases in grasses. The data do, however, provide a convincing demonstration of character-level conflict in a situation where it might easily have been overlooked: where a tree with moderate to high bootstrap support and high posterior probability is in agreement with an intuitively reasonable hypothesis.

**Phylogenetic analysis of the monogenomic Triticeae**—A detailed description of relationships among Triticeae genera has been confounded for two major reasons. First, in some of the studies that sample widely throughout the tribe, the relationships among the genera are unresolved or poorly supported. Second, even in cases where well-supported intergeneric relationships have been recovered, there are numerous conflicts among published phylogenetic trees. Because the conflicts involve many taxa, and because there are no data sets in total agreement, it is difficult to develop even a simplified core phylogeny by removing problematic taxa or by disregarding one or two especially odd data sets. For some genera, the combination of conflict and poor resolution means that no strong conclusions can be drawn about their relationships. Other relationships, with the addition of more gene trees, have appeared repeatedly. In these cases, we can begin to draw phylogenetic conclusions about certain taxa or at least portions of their genomes.

The discussion that follows is not an attempt to compare all published Triticeae trees in terms of the placement of every taxon. Instead, the intent is to use clades from the  $\beta$ -amylase tree as examples to illustrate varying amounts of conflict when multiple trees are considered. The trees used for comparison with the  $\beta$ -amylase tree were based on (1) combined analysis of cpDNA data from the *trnT*, *trnL*, and *trnF* noncoding spacers, the RNA polymerase  $\alpha$ -subunit (*rpoA*) gene, and restriction sites (Mason-Gamer et al., 2002), or individual analyses of restriction sites (Mason-Gamer and Kellogg, 1996b) and/or *rpoA* (Petersen and Seberg, 1997); (2) integrated and/or individual analyses of three highly repetitive nuclear DNA loci (Kellogg et al., 1996), including two 5S rDNA spacer loci (long spacers and short spacers; Kellogg and Appels, 1995) and the internal transcribed spacer of the nuclear DNA repeat (ITS; Hsiao et al., 1995b); (3) sequences from a single-copy nuclear granule-bound starch synthase I gene (GBSSI; Mason-Gamer, 2001); (4) sequences of the single-copy disrupted meiotic cDNA gene (*DMC1*; Petersen and Seberg, 2002); (5) sequences from one member of the small phosphoenolpyruvate carboxylase gene family (PEPC; Helfgott and Mason-Gamer, 2004); and (6) a comprehensive morphological analysis (Seberg and Frederiksen, 2001). The sampling in the different studies is not identical, so discussions of some  $\beta$ -amylase clades do not refer to each previous study.

TABLE 4. A  $\beta$ -amylase clade and its members' closest relatives in other molecular data sets cited in text. Clades with moderate to strong support are shown in boldface type. BS = bootstrap support; JK = jackknife support.

$\beta$ -amylase clade (BS = 92)	cpDNA	BS	Integrated repetitive nuclear	BS	GBSSI	BS	DMC1	JK	PEPC	BS
<i>Secale</i>	<i>Triticum</i> <i>Aegilops</i> <i>Taeniatherum</i>	<b>99</b>	<i>Henrardia</i>	56	<i>Heterantheium</i>	54	<i>Dasypyrum</i>	<50	<i>Aegilops</i> <i>Taeniatherum</i>	<50
<i>Australopyrum</i>	<i>Agropyron</i> <i>Eremopyrum</i> <i>Henrardia</i> <i>Thinopyrum</i>	<b>99</b>	<i>Agropyron</i> <i>Pseudoroegneria</i>	<b>79</b>	<i>Pseudoroegneria</i>	<50	<i>Taeniatherum</i>	<b>95</b>	<i>Dasypyrum</i>	<50
<i>Dasypyrum</i>	<i>Thinopyrum</i> <i>Pseudoroegneria</i>	<b>88</b>	Unresolved	NA	<i>Triticum</i> <i>Aegilops</i> <i>Thinopyrum</i>	<50	<i>Secale</i>	<50	<i>Australopyrum</i>	<50

*Close relationships among Secale, Australopyrum, and Dasypyrum*—This clade, though well-supported on the  $\beta$ -amylase tree, is not recovered on any other published Triticeae trees. The placement of each of the three taxa on the cpDNA (Mason-Gamer et al., 2002), integrated highly repetitive gene (Kellogg et al., 1996), GBSSI (Mason-Gamer, 2001), DMC1 (Petersen and Seberg, 2002), and PEPC (Helfgott and Mason-Gamer, 2004) trees is summarized in Table 4. In many cases, the support for differing placements is very low and thus may not represent meaningful conflict. However, it is notable that, regardless of the level of support, no other molecular data sets support a *Secale* + *Australopyrum* + *Dasypyrum* clade; in general, there is little agreement with regard to the placement of any of these taxa. Furthermore, the morphological data (Seberg and Frederiksen, 2001) are not in full agreement with any of the molecular trees: *Secale* is sister to a *Dasypyrum* + *Triticum* clade, while *Australopyrum* forms a paraphyletic grade at the base of a large clade containing over half of the remaining taxa. This group of genera illustrates the most complicated kind of scenario within the Triticeae. On the few trees on which genera are resolved with reasonable support, their positions conflict. Other trees are poorly supported with regard to these three genera and thus provide no clues to help interpret the conflict. While *Secale*, *Australopyrum*, and *Dasypyrum* remain frustratingly intractable, the remaining discussion will focus on taxa whose evolutionary history has been clarified by the acquisition of multiple, though sometimes conflicting, gene trees.

*Basal relationships within the tribe and paraphyly of Psathyrostachys*—Most of the published Triticeae trees were rooted using one or more species of *Bromus* L. from the tribe Bromeae Dumort., the likely sister group to the Triticeae (e.g., MacFarlane and Watson, 1982; Soreng et al., 1990; Hsiao et al., 1995a; Soreng and Davis, 2000), and most place *Psathyrostachys* at or near the base of the tree (Table 5). With regard to basal relationships, the  $\beta$ -amylase gene tree is one of the least informative trees to date. The two basal clades, separating *Triticum* + *Aegilops* + *Crithopsis* + *Taeniatherum* + *Thinopyrum* from the rest of the Triticeae, are very weak. Furthermore, *Psathyrostachys* is nested well within the  $\beta$ -amylase tree and is resolved as paraphyletic. The only other trees on which *Psathyrostachys* is not monophyletic are a small subset of the GBSSI MP trees (Mason-Gamer, 2001), but these did not otherwise resemble the  $\beta$ -amylase tree.

Most of the other molecular trees show considerable agreement with regard to the earliest-diverging taxa (Table 5). The ITS data (Hsiao et al., 1995b) placed *Hordeum* at the base of

the tree, followed by *Psathyrostachys*. All cpDNA analyses (Mason-Gamer and Kellogg, 1996b; Petersen and Seberg, 1997; Mason-Gamer et al., 2002) placed *Psathyrostachys* at the base of the tree, followed by *Hordeum*, and most of the GBSSI MP trees revealed the same basal relationships, although with very weak support (Mason-Gamer, 2001). The PEPC tree did not include *Psathyrostachys*, but two polyploids thought to include the *Psathyrostachys* genome (Ns) were at the base of the tree, followed by *Hordeum* with weak support (Helfgott and Mason-Gamer, 2004). Finally, the 5S long spacer data set placed both *Psathyrostachys* and *Hordeum* at the base of the tree, unresolved relative to one another (Kellogg and Appels, 1995). Not all of the published tree are in agreement about the earliest-diverging taxa. For example, the morphology (Seberg and Frederiksen, 2001) and DMC1 (Petersen and Seberg, 2002) data sets place *Thinopyrum bessarabicum* and *Henrardia persica* at the bases of their respective trees. Although there is not a complete consensus among trees with regard to basal relationships in the Triticeae, the accumulated trees indicate that (most of the genome of) either *Psathyrostachys* or *Hordeum* is basal to the rest of the tribe.

*Polyphyletic placement of Eremopyrum with Agropyron and Henrardia*—*Eremopyrum* is a small genus of four species, including two diploids (*E. triticeum* and *E. distans*), one tetraploid (*E. orientale*), and one species with both diploid and tetraploid cytotypes (*E. bonaepartis*) (Frederiksen, 1991). A dual placement of *Eremopyrum* with *Agropyron* and with *Henrardia*, as on the  $\beta$ -amylase tree, has also been supported by other data sets (Table 5). The accession of *E. bonaepartis* that groups with *Henrardia* on the  $\beta$ -amylase tree (accession H5554, tetraploid) was grouped with *E. distans*, *E. orientale*, and *Agropyron* on the chloroplast DNA tree, while a different accession of *E. bonaepartis* (H5569, diploid) was strongly grouped with *Henrardia* (Mason-Gamer et al., 2002). In the GBSSI tree, only one accession of *E. bonaepartis* (H5554) was included and was placed in a weak clade with *E. distans*, *E. orientale*, and *Agropyron* (Mason-Gamer, 2001). On the ITS tree, on the other hand, *E. bonaepartis*, the sole representative of *Eremopyrum* on that tree (and a different accession altogether), was grouped with *Henrardia* (Hsiao et al., 1995b). The DMC1 (Petersen and Seberg, 2002) and PEPC (Helfgott and Mason-Gamer, 2004) data sets both yield well-supported *Eremopyrum* + *Agropyron* clades; neither analysis included *E. bonaepartis*.

When taken together, the molecular data appear to support two distinct *Eremopyrum* relationships, and in fact, nonmolecular data have suggested both before. A close relationship be-

TABLE 5. Summary of among-tree comparisons of relationships discussed in the text. Aegi = *Aegilops*; Agro = *Agropyron*; Crit = *Crithopsis*; Erem = *Eremopyrum*; Henr = *Henrardia*; Hord = *Hordeum*; Psat = *Psathyrostachys*; Taen = *Taenitherum*; Thin = *Thinopyrum*; Trit = *Triticum*.

Relationship	β-amylase		cpDNA		Integrated repetitive nuclear		5S-Long		5S-Short		ITS		GBSSI		DMCI		PEPC		Morphology	
	Unresolved	Yes	Psat, Hord	Hord	Outgroup too distant	Psat or Hord	Hord, Psat	On most trees	Hord	Yes	Agro	Yes	Agro	Yes	Agro	Yes	Agro	Yes	Ns-genome polyploids	Thin bess
Basal relationships	No	Yes	Yes	N. A. <sup>a</sup>	N. A. <sup>a</sup>	N. A. <sup>a</sup>	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	N. A. <sup>b</sup>	Thin bess	
Monophyletic	No	Yes	Yes	N. A. <sup>a</sup>	N. A. <sup>a</sup>	N. A. <sup>a</sup>	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	N. A. <sup>b</sup>	Thin bess	
Psat?	No	Yes	Yes	N. A. <sup>a</sup>	N. A. <sup>a</sup>	N. A. <sup>a</sup>	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	N. A. <sup>b</sup>	Thin bess	
Erem sister(s)	Agro, Henr	Yes	Agro, Henr	N. A. <sup>a</sup>	N. A. <sup>a</sup>	N. A. <sup>a</sup>	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	N. A. <sup>b</sup>	Thin bess	
Trit near Aegi?	Yes	Yes	Yes	Conflict	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	N. A. <sup>b</sup>	Thin bess	
Taen near Aegi?	Yes	Yes	Yes	Yes	N. A. <sup>a</sup>	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	N. A. <sup>b</sup>	Thin bess	
Crit near Aegi?	Yes	Yes	Yes	Yes	N. A. <sup>a</sup>	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	N. A. <sup>b</sup>	Thin bess	
Crit near Taen?	Yes	Yes	Yes	Yes	N. A. <sup>a</sup>	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	N. A. <sup>b</sup>	Thin bess	
Monophyletic	No	Yes	Yes	No	N. A. <sup>b</sup>	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	N. A. <sup>b</sup>	Thin bess	
Thin?	No	Yes	Yes	No	N. A. <sup>b</sup>	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	N. A. <sup>b</sup>	Thin bess	
Thin near Aegi?	Yes	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	N. A. <sup>b</sup>	Thin bess	

<sup>a</sup> Not included in data set.  
<sup>b</sup> Single individual included; monophyly not addressed.

tween *Eremopyrum* and *Agropyron* seems reasonable on morphological grounds; *Eremopyrum* has even been described as looking “like an annual crested wheatgrass [*Agropyron*]” (Barkworth, 1998). The relationship was suggested more formally based on overall morphological similarity (Clayton and Renvoise, 1986) and on specific features such as, for example, one-keeled glumes (Frederiksen, 1991) and caryopsis morphology (Terrell and Peterson, 1993). Chromosomal features, on the other hand, point to a relationship between *Eremopyrum* and *Henrardia*. Even though these two genera are morphologically dissimilar, both are unusual within the Triticeae for their predominantly telocentric or subtelocentric chromosomes (Frederiksen, 1991). Given the results from the molecular analyses, the apparently disparate relationships suggested by the morphological vs. the cytological data may reflect a complex origin of *Eremopyrum*, involving both *Agropyron* and *Henrardia*. Broader sampling from within the genus would help to define this potentially interesting pattern more clearly and might also clarify what, if any, role polyploidy plays in its apparently reticulate history.

*Nonmonophyly of Triticum/Aegilops (the wild wheats) and their close relatives*—The *Triticum/Aegilops* group includes some of the most familiar members of the tribe, including the polyploid cultivated wheats and their diploid progenitors. Recent taxonomic treatments of the wild wheats (reviewed in van Slageren, 1994) have been varied, with the number of proposed diploid genera ranging from one to eight. This study follows van Slageren (1994), who placed the A-genome diploids in *Triticum* and the remaining diploid species in *Aegilops* except for *A. tripsacoides* Jaub. & Spach (not included here), which was recognized as *Amblyopyrum muticum* (Boiss.) Eig. The most common points of disagreement among the recent phylogenetic data sets with regard to *Triticum* and *Aegilops* sensu van Slageren are (1) whether or not the two genera form a single monophyletic group and (2) the monophyly vs. paraphyly of *Aegilops* relative to other Triticeae genera. On the β-amylase tree, most of the *Triticum* and *Aegilops* sequences are found together, but they form a paraphyletic group with *Crithopsis*, *Taenitherum*, and *Thinopyrum*. Similarly, *Triticum* is grouped with *Aegilops* on the cpDNA (Mason-Gamer and Kellogg, 1996b; Petersen and Seberg, 1997; Mason-Gamer et al., 2002), 5S rDNA short spacer (Kellogg and Appels, 1995), GBSSI (Mason-Gamer, 2001), DMCI (Petersen and Seberg, 2002), and PEPC (Helfgott and Mason-Gamer, 2004) trees, but on most of them, as on the β-amylase tree, the *Triticum/Aegilops* clade includes additional genera (discussed in more detail later). Thus, a merger of *Triticum* and *Aegilops* based on phylogeny would require an expanded definition of the group. Furthermore, there are a few trees, including the 5S long spacer (Kellogg and Appels, 1995) and morphology-based (Seberg and Frederiksen, 2001) trees, on which *Triticum* and *Aegilops* are not closely related, suggesting that portions of their genomes are evolutionarily distinct.

One unique feature of the β-amylase tree is the polyphyletic placement of two *Aegilops* sequences. One of these, *A. comosa*, also has a sequence in the main *Triticum/Aegilops* group, while only the more unusual sequence has been recovered from *A. uniaristata*. Introgression appears to be a more likely explanation for this pattern than gene duplication, since a duplication within *Aegilops* would yield a closely related sequence. A duplication early enough in the history of the tribe to explain this pattern would be evident in other taxa as well.

With the current sampling, however, the identity of the donor of the outlying *Aegilops* sequences remains unknown.

*Placement of Taeniatherum within the Triticum/Aegilops group*—As discussed earlier, many molecular trees suggest that the *Triticum/Aegilops* group is paraphyletic. Several genera appear repeatedly within this group, including *Taeniatherum*, *Crithopsis*, and *Thinopyrum*, and for each of these, there is independent evidence both for and against their close relationship to *Triticum/Aegilops* (Table 5). On the  $\beta$ -amylase tree, four *Taeniatherum caput-medusae* individuals are nested within the *Triticum/Aegilops* group. The close relationship between *Taeniatherum* and the *Triticum/Aegilops* group was also demonstrated by the cpDNA (Mason-Gamer et al., 2002), integrated highly repetitive gene (Kellogg et al., 1996), and PEPC (Helfgott and Mason-Gamer, 2004) trees. However, other Triticeae trees do not support this relationship, including the GBSSI (Mason-Gamer, 2001) and *DMCI* (Petersen and Seberg, 2002) trees, neither of which place *Taeniatherum* in or near *Triticum* or *Aegilops*.

A possible close relationship between *Triticum/Aegilops* and *Taeniatherum*, a small genus with a single polytypic species (Frederiksen, 1986), is at odds with the morphological analysis (Seberg and Frederiksen, 2001) and with traditional taxonomic treatments, most of which place *Taeniatherum* near *Hordeum* (e.g., Clayton and Renvoize, 1986). Crossing studies involving *Taeniatherum* do not point to any likely close relatives; crosses with 30 species representing 11 genera (including 14 *Hordeum* species but none from *Triticum* or *Aegilops*) yielded very few hybrids, and in these, meiotic chromosome pairing was very low (Frederiksen and von Bothmer, 1989). In another study (Frederiksen, 1994), occasional meiotic pairing in hybrids between *Taeniatherum* and hexaploid *Triticum aestivum* was interpreted as autosyndetic pairing among homoeologous wheat genomes, although this was partly based on the morphologically reasonable a priori assumption that *Taeniatherum* was closely related to *Hordeum* and not a close relative of *Triticum/Aegilops*. In spite of their overall morphological dissimilarity, the accumulated molecular data strongly suggest that a significant portion of the *Taeniatherum* genome is closely related to that of the *Triticum/Aegilops* group.

*Placement of Crithopsis within the Triticum/Aegilops group*—A second genus closely associated with *Triticum/Aegilops* on the  $\beta$ -amylase tree is the annual, monotypic *Crithopsis*. Its placement within or very near *Aegilops* is supported by most of the molecular data sets that included *C. delileana* (Table 5), including the cpDNA restriction sites (Mason-Gamer et al., 2002) and *rpoA* gene (Petersen and Seberg, 1997), the integrated highly repetitive genes (Kellogg et al., 1996), and the *DMCI* gene (Petersen and Seberg, 2002). The morphological analysis (Seberg and Frederiksen, 2001), on the other hand, did not place *Crithopsis* with *Triticum/Aegilops* but instead with *Taeniatherum*, a result that was further supported by the application of Giemsa C-banding to several Triticeae genera (Linde-Laursen et al., 1999). These two studies are consistent with many of the molecular analyses in suggesting a close relationship between *Crithopsis* and *Taeniatherum* (Table 5), but they do not support the two species' close relationship to *Triticum/Aegilops*.

*Monophyly of Thinopyrum and its relationship to Triticum/Aegilops*—The third genus with a close affinity to *Triticum/Aegilops*

on the  $\beta$ -amylase tree is *Thinopyrum*. The relationship between these groups is economically relevant, because *Thinopyrum* is a potentially valuable source for improvement of hexaploid wheat, as a contributor of genes for resistance to several viral, fungal, and insect infestations, as well as increased tolerance to salt, low temperature, and drought (e.g., Tang et al., 2000, and references therein). The  $\beta$ -amylase gene tree does not support a sister relationship between *T. bessarabicum* and *T. elongatum*, which are the only diploid species in the genus. Instead, *T. elongatum* is in a well-supported clade with *Aegilops*, *Triticum*, *Crithopsis*, and *Taeniatherum*, while *Thinopyrum bessarabicum* is placed at the base of this clade, although the support for the latter placement is very weak (<50% bootstrap; <95% posterior probability).

The  $\beta$ -amylase tree is consistent with an isozyme analysis in which UPGMA phenograms place both *T. bessarabicum* and *Th. elongatum* next to or within a *Triticum/Aegilops* cluster, but not as sister to one another (McIntyre, 1988). Several gene trees suggest similar relationships (Table 5), including the integrated repetitive nuclear loci (Kellogg et al., 1996), which placed *Thinopyrum bessarabicum* in a weak clade with *Triticum*, *Aegilops*, *Crithopsis*, and *Taeniatherum*, with *Thinopyrum elongatum* at the base of that clade. The *DMCI* gene sequence data separate the *Triticum/Aegilops* sequences into two well-supported clades, one of which includes *Thinopyrum bessarabicum*, while the other includes *T. elongatum* and *Crithopsis* (Petersen and Seberg, 2002). Thus, although the details differ, several data sets agree that *T. bessarabicum* and *T. elongatum* do not form a monophyletic group and that they are both closely related to *Triticum/Aegilops* and *Crithopsis*. This result is also partly consistent with the morphological study (Seberg and Frederiksen, 2001) in which *Thinopyrum elongatum* was closely related (though not sister) to the *Triticum-Aegilops* clade, while *Thinopyrum bessarabicum* was placed at the base of the Triticeae. In contrast, the GBSSI gene sequence data supported a monophyletic (weakly supported) *Thinopyrum* (Mason-Gamer, 2001), but agree with the other data sets in placing them near (in this case within) *Triticum/Aegilops*. The cpDNA data are unique in that they not only strongly support the monophyly of *Thinopyrum bessarabicum* and *T. elongatum* (along with tetraploid *T. scirpeum*), but they also nest the clade within *Pseudoroegneria*, not in or near *Triticum-Aegilops* (Mason-Gamer et al., 2002).

The taxonomy of *Thinopyrum* has been debated, with particular emphasis on the relationship between the two diploids, *T. bessarabicum* and *T. elongatum*. Based on chromosome pairing data, some workers assign separate genome designations to the two species (genomes **J** and **E**, respectively) and place the **E**-genome species in a separate genus, *Lophopyrum* A.Löve (Löve, 1984; Jauhar, 1988, 1990). Others suggest that the genomes of *T. bessarabicum* and *T. elongatum* are similar enough to be assigned a single genome designation (Dvorák, 1981; Dewey, 1984; McGuire, 1984; Wang, 1985; Wang and Hsiao, 1989). In light of the equivocal results and interpretations of the cytogenetic studies, it is perhaps not surprising that different molecular phylogenetic data sets support different conclusions regarding the close relationship between these two species. Together they suggest that the genomes of *T. elongatum* and *T. bessarabicum* are mostly evolutionarily distinct and are both related to *Triticum/Aegilops*. However, portions of their genomes (most notably, the chloroplast genome) are unrelated to *Triticum/Aegilops*, probably reflecting past introgression.

**Conclusions**—The  $\beta$ -amylase gene is a potentially valuable source of phylogenetic data. It has been sequenced from several crop relatives and, in some grasses, its copy number and location(s) within the genome are relatively well understood. The results of a routine analysis of a small number of grass coding sequences, while seemingly straightforward, were complicated by underlying conflicting signal, which would have gone undetected without the examination of individual characters. The conflict, and the resulting uncertainty regarding the root of the tree, may result from the use of too-distant outgroups and thus might be eliminated with the availability of an outgroup more closely related to grasses. There are several copies of the  $\beta$ -amylase gene in the Triticeae, and the incorporation of existing genetic information allowed the successful (by all appearances) targeting of a single member of the family. Conflict among Triticeae gene trees has been demonstrated in the past, and the  $\beta$ -amylase gene tree indeed supports some new, unique relationships. Thus, on one hand, the addition of new Triticeae gene trees in the future might add new conflicting relationships, complicating the overall phylogenetic picture. On the other hand, comparisons among larger numbers of data sets potentially allow the identification of points of broad consensus. If Triticeae genomes are thought of as fluid, with potential genetic exchange among lineages, these points of consensus can be thought of as phylogenies of portions of genomes, while well-supported conflicts may be indicators of historical reticulate events.

#### LITERATURE CITED

- BARKWORTH, M. E. 1998. Grasses of the tribe Hordeae in North America. 3. Comments. *Botanical Electronic News* 199, website <http://www.uo.edu/cas/botany-micro/ben/ben199.html>.
- BUREAU, D., C. LAURIÈRE, C. MAYER, J. SADOWSKI, AND J. DAUSSANT. 1989. Post-translation modifications of  $\beta$ -amylases during germination of wheat and rye seeds. *Journal of Plant Physiology* 134: 44–50.
- BUREAU, T. E., AND S. R. WESSLER. 1994. *Stowaway*: a new family of inverted repeat elements associated with the genes of both monocotyledonous and dicotyledonous plants. *Plant Cell* 6: 907–916.
- CLARK, S. E., P. M. HAYES, AND C. A. HENSON. 2003. Effects of single nucleotide polymorphisms in  $\beta$ -amylase1 alleles from barley on functional properties of the enzymes. *Plant Physiology and Biochemistry* 41: 798–804.
- CLAYTON, W. D., AND S. A. RENVOIZE. 1986. Genera Graminum. Grasses of the world. Royal Botanical Gardens, London, UK.
- DAUSSANT, J., J. SADOWSKI, AND P. ZIEGLER. 1994. Cereal  $\beta$ -amylases: diversity of the  $\beta$ -amylase isozyme status within cereals. *Journal of Plant Physiology* 143: 585–590.
- DEWEY, D. R. 1984. The genomic system of classification as a guide to intergeneric hybridization within the perennial Triticeae. In J. P. Gustafson [ed.], *Gene manipulation in plant improvement*, Proceedings of the 6th Stadler genetics symposium, 209–279. Columbia University Press, New York, New York, USA.
- DUNN, G. 1974. A model for starch breakdown in higher plants. *Phytochemistry* 13: 1341–1346.
- DVORÁK, J. 1981. Genome relationships among *Elytrigia* (= *Agropyron*) *elongata*, *E. stipifolia*, “*E. elongata* 4 $\times$ ,” *E. caespitosa*, *E. intermedia*, and “*E. elongata* 10 $\times$ ”. *Canadian Journal of Genetics and Cytology* 23: 481–492.
- FELSENSTEIN, J. 1978. Cases in which parsimony or compatibility methods will be positively misleading. *Systematic Zoology* 27: 401–410.
- FELSENSTEIN, J. 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. *Journal of Molecular Evolution* 17: 368–376.
- FESCHOTTE, C., N. JIANG, AND S. R. WESSLER. 2002. Plant transposable elements: where genetics meets genomics. *Nature Reviews Genetics* 3: 329–341.
- FRATI, F., C. SIMON, J. SULLIVAN, AND D. L. SWOFFORD. 1997. Evolution of the mitochondrial cytochrome oxidase II gene in Collembola. *Journal of Molecular Evolution* 44: 145–158.
- FREDERIKSEN, S. 1986. Revision of *Taeniatherum* (Poaceae). *Nordic Journal of Botany* 6: 389–397.
- FREDERIKSEN, S. 1991. Taxonomic studies in *Eremopyrum* (Poaceae). *Nordic Journal of Botany* 11: 271–285.
- FREDERIKSEN, S. 1994. Hybridization between *Taeniatherum caput-medusae* and *Triticum aestivum* (Poaceae). *Nordic Journal of Botany* 14: 3–6.
- FREDERIKSEN, S., AND R. VON BOTHMER. 1989. Intergeneric hybridization between *Taeniatherum* and different genera of Triticeae, Poaceae. *Nordic Journal of Botany* 9: 229–240.
- GANA, J. A., N. E. KALENGAMALIRO, S. M. CUNNINGHAM, AND J. J. VOLENEC. 1998. Expression of  $\beta$ -amylase from alfalfa taproots. *Plant Physiology* 118: 1495–1505.
- GIBSON, T. S., V. SOLAH, M. R. GLENNIE-HOLMES, AND H. R. TAYLOR. 1995. Diastatic power in malted barley: contributions of malt parameters to its development and the potential of barley grain *beta*-amylase to predict malt diastatic power. *Journal of the Institute of Brewing* 101: 277–280.
- GRAHAM, S. W., R. G. OLMSTEAD, AND S. C. H. BARRETT. 2002. Rooting phylogenetic trees with distant outgroups: a case study from the commelinoid monocots. *Molecular Biology and Evolution* 19: 1769–1781.
- GRASS PHYLOGENY WORKING GROUP (GPWG). 2001. Phylogeny and subfamilial classification of the grasses (Poaceae). *Annals of the Missouri Botanical Garden* 88: 372–457.
- GU, X., Y.-X. FU, AND W.-H. LI. 1995. Maximum likelihood estimation of the heterogeneity of substitution rate among nucleotide sites. *Molecular Biology and Evolution* 12: 546–557.
- HASEGAWA, M., H. KISHINO, AND T. YANO. 1985. Dating the human-ape split by a molecular clock of mitochondrial DNA. *Journal of Molecular Evolution* 22: 160–170.
- HELFGOTT, D. M., AND R. J. MASON-GAMER. 2004. The evolution of North American *Elymus* (Triticeae, Poaceae) allotetraploids: evidence from phosphoenolpyruvate carboxylase gene sequences. *Systematic Botany* 29: 850–861.
- HENDY, M. D., AND D. PENNY. 1989. A framework for the quantitative study of evolutionary trees. *Systematic Zoology* 38: 297–309.
- HIGGINS, D. G., A. J. BLEASBY, AND R. FUCHS. 1992. CLUSTAL V: improved software for multiple sequence alignment. *Computer Applications in the Biosciences* 8: 189–191.
- HSIAO, C., N. J. CHATTERTON, K. H. ASAY, AND K. B. JENSEN. 1995a. Molecular phylogeny of the Pooideae (Poaceae) based on nuclear rDNA (ITS) sequences. *Theoretical and Applied Genetics* 90: 389–398.
- HSIAO, C., N. J. CHATTERTON, K. H. ASAY, AND K. B. JENSEN. 1995b. Phylogenetic relationships of the monogenomic species of the wheat tribe, Triticeae (Poaceae), inferred from nuclear rDNA (internal transcribed spacer) sequences. *Genome* 38: 211–223.
- HUELSENBECK, J. P. 1998. Systematic bias in phylogenetic analysis: is the Strepsiptera problem solved? *Systematic Biology* 47: 519–537.
- HUELSENBECK, J. P., AND K. A. CRANDALL. 1997. Phylogeny estimation and hypothesis testing using maximum likelihood. *Annual Review of Ecology and Systematics* 28: 437–466.
- HUELSENBECK, J. P., AND B. RANNALA. 1997. Phylogenetic methods come of age: testing hypotheses in an evolutionary context. *Science* 276: 227–232.
- HUELSENBECK, J. P., AND F. RONQUIST. 2001. MrBayes: Bayesian inference of phylogenetic trees. *Bioinformatics* 17: 754–755.
- JAUHAR, P. P. 1988. A reassessment of genome relationships between *Thinopyrum bessarabicum* and *T. elongatum* of the Triticeae. *Genome* 30: 903–914.
- JAUHAR, P. P. 1990. Multidisciplinary approach to genome analysis in the diploid species, *Thinopyrum bessarabicum* and *Th. elongatum* (*Lophopyrum elongatum*), of the Triticeae. *Theoretical and Applied Genetics* 80: 523–536.
- JUKES, T. H., AND C. R. CANTOR. 1969. Evolution of protein molecules. In H. N. Munro [ed.], *Mammalian protein metabolism*, 21–132. Academic Press, New York, New York, USA.
- JUNG, W., R. W. SKADSEN, AND D. M. PETERSON. 2001. Characterization of a novel barley  $\beta$ -amylase gene expressed only during early grain development. *Seed Science Research* 11: 325–334.
- KELLOGG, E. A., AND R. APPELS. 1995. Intraspecific and interspecific variation in 5S RNA genes are decoupled in diploid wheat relatives. *Genetics* 140: 325–343.
- KELLOGG, E. A., R. APPELS, AND R. J. MASON-GAMER. 1996. When gene trees tell different stories: the diploid genera of Triticeae (Gramineae). *Systematic Botany* 21: 321–347.

- KIM, J. 1996. General inconsistency conditions for maximum parsimony: effects of branch lengths and increasing numbers of taxa. *Systematic Biology* 45: 363–374.
- KIMURA, M. 1980. A simple method for estimating evolutionary rate of base substitution through comparative studies of nucleotide sequences. *Journal of Molecular Evolution* 16: 111–120.
- KISHINO, H., AND M. HASEGAWA. 1989. Evaluation of the maximum likelihood estimate from the evolutionary tree topologies from DNA sequence data, and the branching order in the Hominoidea. *Journal of Molecular Evolution* 29: 170–179.
- KREIS, M., M. WILLIAMSON, B. BUXTON, J. PYWELL, J. HEJGAARD, AND I. SVENDSEN. 1987. Primary structure and differential expression of  $\beta$ -amylase in normal and mutant barleys. *European Journal of Biochemistry* 179: 517–525.
- LI, C. D., P. LANGRIDGE, X. Q. ZHANG, P. E. ECKSTEIN, B. G. ROSSNAGEL, R. C. M. LANCE, E. B. LEFOL, M. Y. LU, B. L. HARVEY, AND G. J. SCOLES. 2002. Mapping of barley (*Hordeum vulgare* L.)  $\beta$ -amylase alleles in which an amino acid substitution determines  $\beta$ -amylase isoenzyme type and the level of free  $\beta$ -amylase. *Journal of Cereal Science* 35: 39–50.
- LINDE-LAURSEN, I., S. FREDERIKSEN, AND O. SEBERG. 1999. The giemsa C-banded karyotype of *Crithopsis delileana* (Poaceae: Triticeae). *Hereditas* 130: 51–55.
- LÖVE, Á. 1984. Conspectus of the Triticeae. *Feddes Repertorium* 95: 425–521.
- LUNDGARD, A., AND B. SVENSSON. 1987. The four major forms of barley  $\beta$ -amylase. Purification, characterization and structural relationship. *Carlsberg Research Communications* 52: 313–326.
- MACFARLANE, T. D., AND L. WATSON. 1982. The classification of Poaceae subfamily Pooideae. *Taxon* 31: 178–203.
- MADDISON, W. P., AND D. R. MADDISON. 2000. MacClade. Sinauer, Sunderland, Massachusetts, USA.
- MASON-GAMER, R. J. 2001. Origin of North American species of *Elymus* (Poaceae: Triticeae) allotetraploids based on granule-bound starch synthase gene sequences. *Systematic Botany* 26: 757–768.
- MASON-GAMER, R. J., AND E. A. KELLOGG. 1996a. Testing for phylogenetic conflict among molecular data sets in the tribe Triticeae (Gramineae). *Systematic Biology* 45: 524–545.
- MASON-GAMER, R. J., AND E. A. KELLOGG. 1996b. Chloroplast DNA analysis of the monogenomic Triticeae: phylogenetic implications and genome-specific markers. In P. P. Jauhar [ed.], *Methods of genome analysis in plants*, 301–325. CRC Press, Boca Raton, Florida, USA.
- MASON-GAMER, R. J., N. L. ORME, AND C. M. ANDERSON. 2002. Phylogenetic analysis of North American *Elymus* and the monogenomic Triticeae (Poaceae) using three chloroplast DNA data sets. *Genome* 45: 991–1002.
- MCGUIRE, P. E. 1984. Chromosome pairing in triploid and tetraploid hybrids in *Elytrigia* (Triticeae: Poaceae). *Canadian Journal of Genetics and Cytology* 26: 519–522.
- MCINTYRE, C. L. 1988. Variation at isozyme loci in Triticeae. *Plant Systematics and Evolution* 160: 123–142.
- MIYAMOTO, M. M., AND S. M. BOYLE. 1989. The potential importance of mitochondrial DNA sequence data to eutherian mammal phylogeny. In B. Fernholm, K. Bremer, and H. Joernvall [eds.], *The hierarchy of life*, 437–450. Elsevier, Amsterdam, Netherlands.
- PETERSEN, G., AND O. SEBERG. 1997. Phylogenetic analysis of the Triticeae (Poaceae) based on *rpoA* sequence data. *Molecular Phylogenetics and Evolution* 7: 217–230.
- PETERSEN, G., AND O. SEBERG. 2002. Molecular evolution and phylogenetic application of *DMC1*. *Molecular Phylogenetics and Evolution* 22: 43–50.
- QIU, Y.-L., J. LEE, B. A. WHITLOCK, F. BERNASCONI-QUADRONI, AND O. DOMBROVSKA. 2001. Was the ANITA rooting of the angiosperm phylogeny affected by long-branch attraction? *Molecular Biology and Evolution* 18: 1745–1753.
- QIU, Y.-L., AND J. D. PALMER. 1999. Phylogeny of early land plants: insights from genes and genomes. *Trends in Plant Science* 4: 26–30.
- RAJAPAKSE, S., S. D. NILMALGODA, M. MOLNAR, R. E. BALLARD, D. F. AUSTIN, AND J. R. BOHAC. 2004. Phylogenetic relationships of the sweetpotato in *Ipomoea* series *Batatas* (Convolvulaceae) based on nuclear  $\beta$ -amylase sequences. *Molecular Phylogenetics and Evolution* 30: 623–632.
- RODRÍGUEZ, F., J. L. OLIVER, A. MARÍN, AND J. R. MEDINA. 1990. The general stochastic model of nucleotide substitution. *Journal of Theoretical Biology* 142: 485–501.
- RORAT, T., J. SADOWSKI, F. GRELLER, J. DAUSSANT, AND M. DELSENY. 1991. Characterization of cDNA clones for rye endosperm  $\beta$ -amylase and analysis of  $\beta$ -amylase deficiency in rye mutant lines. *Theoretical and Applied Genetics* 83: 257–263.
- SADOWSKI, J., T. RORAT, R. COOKE, AND M. DELSENY. 1993. Nucleotide sequence of a cDNA clone encoding ubiquitous  $\beta$ -amylase in rye. *Plant Physiology* 102: 315–316.
- SANDERSON, M. J. 1998. Estimating rate and time in molecular phylogenies; beyond the molecular clock. In D. E. Soltis, P. S. Soltis, and J. J. Doyle [eds.], *Molecular systematics of plants II. DNA sequencing*, 242–264. Kluwer, Boston, Massachusetts, USA.
- SEBERG, O., AND S. FREDERIKSEN. 2001. A phylogenetic analysis of the monogenomic Triticeae (Poaceae) based on morphology. *Botanical Journal of the Linnean Society* 136: 75–97.
- SHIMODAIRA, H., AND M. HASEGAWA. 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Molecular Biology and Evolution* 16: 1114–1116.
- SMALL, R. L., R. C. CRONN, AND J. F. WENDEL. 2004. Use of nuclear genes for phylogeny reconstruction in plants. *Australian Systematic Botany* 17: 145–170.
- SORENG, R. J., AND J. I. DAVIS. 2000. Phylogenetic structure in Poaceae subfamily Pooideae as inferred from molecular and morphological characters: misclassification versus reticulation. In S. W. L. Jacobs and J. Everett [eds.], *Grasses: systematics and evolution*, 61–74. CSIRO, Melbourne, Victoria, Australia.
- SORENG, R. J., J. I. DAVIS, AND J. J. DOYLE. 1990. A phylogenetic analysis of chloroplast DNA restriction site variation in Poaceae subfam. Pooideae. *Plant Systematics and Evolution* 172: 83–97.
- SULLIVAN, J., Z. ABDO, P. JOYCE, AND D. L. SWOFFORD. In press. Evaluating the performance of a successive approximations approach to parameter optimization in maximum-likelihood parameter estimation. *Molecular Biology and Evolution*.
- SULLIVAN, J., K. E. HOLSINGER, AND C. SIMON. 1996. The effect of topology on estimates of among-site rate variation. *Journal of Molecular Evolution* 42: 308–312.
- SWOFFORD, D. L. 2002. PAUP\*: phylogenetic analysis using parsimony (\*and other methods). Sinauer, Sunderland, Massachusetts, USA.
- SWOFFORD, D. L., G. J. OLSEN, P. J. WADDELL, AND D. M. HILLIS. 1996. Phylogenetic inference. In D. M. Hillis, C. Moritz, and B. K. Mable [eds.], *Molecular systematics*, 407–514. Sinauer, Sunderland, Massachusetts, USA.
- TANG, S., Z. LI, X. JIA, AND P. J. LARKIN. 2000. Genomic in situ hybridization (GISH) analyses of *Thinopyrum intermedium*, its partial amphiploid Zhong 5, and disease-resistant derivatives in wheat. *Theoretical and Applied Genetics* 100: 344–352.
- TEMPLETON, A. R. 1983. Phylogenetic inference from restriction endonuclease cleavage site maps with particular reference to the evolution of humans and the apes. *Evolution* 37: 221–244.
- TERRELL, E. E., AND P. M. PETERSON. 1993. Caryopsis morphology and classification in the Triticeae (Pooideae: Poaceae). Smithsonian Institution Press, Washington, D.C., USA.
- VAN SLAGEREN, M. W. 1994. Wild wheats: a monograph of *Aegilops* L. and *Amblyopyrum* (Jaub. & Spach) Eig (Poaceae). Wageningen Agricultural University, Wageningen, Netherlands.
- WADDELL, P. J., AND D. PENNY. 1996. Evolutionary trees of apes and humans from DNA sequences. In A. J. Locke and C. R. Peters [eds.], *Handbook of symbolic evolution*, 53–73. Clarendon Press, Oxford, UK.
- WAGNER, G., L. ZEMANOVA, K.-P. HÄGER, AND P. ZIEGLER. 1999. The major  $\beta$ -amylase isoforms of wheat leaves correspond to one of two ubiquitously expressed  $\beta$ -amylase genes. *Plant Physiology and Biochemistry* 37: 515–530.
- WANG, R. R.-C. 1985. Genome analysis of *Thinopyrum bessarabicum* and *T. elongatum*. *Canadian Journal of Genetics* 27: 722–728.
- WANG, R. R.-C., AND C. HSIAO. 1989. Genome relationships between *Thinopyrum bessarabicum* and *T. elongatum*: revisited. *Genome* 32: 802–809.
- WANG, S. M., W. L. LUE, S. Y. WU, H. W. HUANG, AND J. C. CHEN. 1997. Characterization of a maize  $\beta$ -amylase cDNA clone and its expression during seed germination. *Plant Physiology* 113: 403–409.
- WHEELER, W. C. 1990. Nucleic acid sequence phylogeny and random outgroups. *Cladistics* 6: 363–368.

- XIANG, Q.-Y., M. L. MOODY, D. E. SOLTIS, C. FAN, AND P. S. SOLTIS. 2002. Relationships within Cornales and circumscription of Cornaceae—*matK* and *rbcL* sequence data and effects of outgroups and long branches. *Molecular Phylogenetics and Evolution* 24: 35–57.
- YAMAGUCHI, J., S. ITOH, T. SAITOH, A. IKEDA, T. TASHIRO, AND Y. NAGATO. 1999. Characterization of  $\beta$ -amylase and its deficiency in various rice cultivars. *Theoretical and Applied Genetics* 98: 32–38.
- YANG, Z. 1993. Maximum-likelihood estimation of phylogeny from DNA sequences when substitution rates differ over sites. *Molecular Biology and Evolution* 10: 1396–1401.
- YOSHIDA, N., K. HAYASHI, AND K. NAKAMURA. 1992. A nuclear gene encoding  $\beta$ -amylase of sweet-potato. *Gene* 120: 255–259.
- YOSHIGI, N., Y. OKADA, H. SAHARA, AND T. TAMAKI. 1995. A structural gene encoding beta-amylase of barley. *Bioscience, Biotechnology, and Biochemistry* 59: 1991–1993.
- ZIEGLER, P. 1999. Cereal *beta*-amylases. *Journal of Cereal Science* 29: 195–204.