**Aliso: A Journal of Systematic and Evolutionary Botany**

Volume 23 | Issue 1 Article 29

2007

# Phylogeny of Triticeae (Poaceae) Based on Three Organelle Genes, Two Single-Copy Nuclear Genes, and Morphology

Ole Seberg
*Botanical Institute, Copenhagen, Denmark*

Gitte Petersen
*Botanical Institute, Copenhagen, Denmark*

*Aliso* 23, pp. 362–371


# PHYLOGENY OF TRITICEAE (POACEAE) BASED ON THREE ORGANELLE GENES, TWO SINGLE-COPY NUCLEAR GENES, AND MORPHOLOGY

OLE SEBERG[1,2] AND GITTE PETERSEN[2]

*Department of Evolutionary Botany, Botanical Institute, Gothersgade 140, DK-1123, Copenhagen K, Denmark*
[1]*Corresponding author (oles@snm.ku.dk)*

ABSTRACT

Triticeae are renowned for their complicated taxonomy, but their phylogeny is equally intricate and perplexing, and remains largely unresolved. Based on morphology and nucleotide sequences from two plastid genes (*rbc*L, *rpo*A), one mitochondrial gene (*cox*II), and two single-copy nuclear genes (*DMC1, EF-G*), the most comprehensive hypothesis (both with respect to taxa and data points) of the phylogeny of diploid Triticeae to date is presented. The incongruence length difference tests clearly indicate that the four logical data partitions (morphology and the three genome compartments) are mutually incongruent, except the mitochondrial and nuclear sequences. Nonetheless, a total evidence approach results in a highly resolved, strongly supported consensus tree, though partitioned Bremer support points to a high level of conflict among the individual data sets.

Key words: *cox*II, *DMC1, EF-G*, morphology, phylogeny, Poaceae, *rbc*L, *rpo*A, Triticeae.

## INTRODUCTION

The grass tribe Triticeae includes some of the world's most important cereals—barley (*Hordeum*), rye (*Secale*), and wheat (*Triticum*)—and a considerable number of important forage grasses.

There is general agreement that the tribe is monophyletic (Watson et al. 1985: 993; Kellogg 1989; Soreng et al. 1990; Hsiao et al. 1995*a*) and not surprisingly, a substantial number of phylogenetic hypotheses have been made for the tribe using a wide variety of different data and methodologies. Within the past decade most of these phylogenies have been based on different kinds of nucleotide sequence data (Baum and Appels 1992; Kellogg 1992; Monte et al. 1993; Kellogg and Appels 1995; Hsiao et al. 1995*b*; Kellogg et al. 1996; Mason-Gamer and Kellogg 1996*b*; Petersen and Seberg 1997, 2000; Mason-Gamer et al. 1998, 2002; Baum et al. 2001). Phylogenies based upon morphology have recently been summarized by Seberg and Frederiksen (2001).

As molecular data have become more and more widely used in Triticeae for reconstructing phylogeny, it has become apparent that different genes and sequences point to broadly different relationships among the genera (Kellogg et al. 1996; Mason-Gamer and Kellogg 1996*a*). This incongruence goes beyond the use of different methodologies in phylogeny reconstruction and points to a real significant, biological problem.

The extensive conflict among the phylogenies is in severe contrast to the commonly held belief that given enough data, the individual phylogenies will eventually converge on the same phylogeny and that the differences among individual data may be attributed to noise or sampling error (Miyamoto and Cracraft 1992; Miyamoto and Fitch 1995; Grant and Kluge 2003).

Despite this we still believe that the only sensible way forward is to combine the data in a total evidence analysis. At the same time, we acknowledge that much useful information may be extracted from the individual data sets (Miyamoto and Cracraft 1992; Miyamoto and Fitch 1995; Grant and Kluge 2003).

The present analysis combines morphological data with nucleotide sequence data from five different genes from all three genome compartments—two chloroplast genes, a mitochondrial gene, and two single-copy nuclear genes. A few existing data sets (Hsiao et al. 1995*b*; Mason-Gamer et al. 2002) were not included in the present analyses mainly because of a restricted overlap in taxonomic coverage.

Kellogg et al. (1996) have attempted to explain the incongruence among different phylogenies as being caused by individual taxa. However, conflicts in data can always be reduced by removing problematic taxa. Nonetheless, in this paper we will look briefly at the contribution of different data partitions as they relate to conflicting results.

## MATERIALS AND METHODS

### *Taxa Sampled*

For reasons given by Seberg and Frederiksen (2001: 87–88), the analyses were restricted to diploid taxa. Thirty-one diploid taxa representing 30 species and 19 genera of Triticeae (including the 24 monogenomic genera originally recognized by Löve 1984), including four species with unknown genome constitution (*Eremopyrum triticeum, Hordeum marinum* subsp. *gussoneanum, H. murinum* subsp. *glaucum,* and *Peridictyon sanctum*), were analyzed. Two species (*Lophopyrum elongatum* and *Thinopyrum bessarabicum*) believed to carry modifications of the same genome (Wang et al. 1996) were also included. In some cases, monophyly of traditionally accepted genera, e.g., *Aegilops* (including several different genomes), *Australopyrum, Eremopyrum* (*E. distans* and *E. triticeum* are occasionally thought to carry different genomes; see Sakamoto 1979), *Festucopsis, Hordeum* (including two different genomes and two un-

[2] Present address: Botanical Garden and Museum, Natural History Museum of Denmark, Sølvgade 83, Opg. S, DK-1307, Copenhagen K, Denmark.

known genomes), and *Psathyrostachys* was tested by including two or more species of each genus in the analyses. The circumscriptions of taxa follow Seberg and Frederiksen (2001) except for the few taxa indicated below that were not included in the original study. Two species of *Bromus* were used as outgroups as most recent investigations indicate that the genus is the most likely sister group to Triticeae (Watson et al. 1985; Kellogg 1989; Hsiao et al. 1999; Grass Phylogeny Working Group [GPWG] 2001). The studied taxa, their accession numbers, and genome constitutions are given in Table 1.

*Morphological Data*

The morphological characters are the same as those scored by Seberg and Frederiksen (2001: 78–79). However, due to a different choice of taxa in the molecular studies, a few new taxa had to be scored: *Bromus sterilis, Festucopsis festucoides, Hordeum erectifolium, Psathyrostachys fragilis* subsp. *fragilis, P. fragilis* subsp. *villosus,* and *P. stoloniformis.* The scoring was primarily based on the literature: *F. festucoides* was scored on the basis of Seberg et al. (1990); *H. erectifolium* shares its morphology with the other *Hordeum* species of the traditional H-genome group (Seberg and Frederiksen 2001); and *Psathyrostachys* was scored on the basis of Baden (1990) and Frederiksen and Seberg (1992). *Bromus sterilis* was scored on the basis of two herbarium sheets at C (*Larsen et al. 193, Strid et al. 42821*) and checked against live plants. Due to the changed taxon sampling, two characters, rachis disarticulation (char. 14) and starch grain type (char. 32) (see Seberg and Frederiksen 2001), became parsimony uninformative and were excluded. The reduced morphological matrix will be published elsewhere.

*Molecular Data*

Of the five molecular data sets, two are from the plastids, one from the mitochondria, and two from single-copy nuclear genes. The origin of the samples used may be found in Petersen and Seberg (1997, 2000) except for one accession not included in these papers (C 618, *Bromus arvensis,* Germany, Hessen; voucher at C). Total genomic DNA was extracted either from fresh leaves following the procedure of Doyle and Doyle (1987) or from dried leaves using the DNeasy plant extraction kit (QIAGEN Ltd., Crawley, West Sussex, UK), following the manufacturer's instructions.

PCR was performed under standard conditions, and the products were purified using the QIAquick PCR purification kit (QIAGEN) according to the manufacturer's instructions. Cycle sequencing was performed using the ABI PRISM Dye Terminator Cycle Ready Reaction kit with AmpliTaq DNA Polymerase, FS (Perkin-Elmer Applied Biosystems, Wellesley, Massachusetts, USA), and the products were purified as above. DNA fragments were separated on an ABI 377 (Perkin-Elmer Applied Biosystems) automated sequencer, and sequence editing was done using Sequencher 3.1 (Gene Codes Corporation, Ann Arbor, Michigan, USA).

The two plastid regions sequenced are the α-subunit of RNA polymerase (*rpo*A) and the large subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase (*rbc*L). The amplified fragments of *rbc*L include the whole gene (1428 base pairs [bp]) except the ultimate part of the 3′-end plus sequence from the intergenic spacers (IGS) upstream of the gene. The primers used for PCR and sequencing may be found in Petersen and Seberg (2003: Table 2). Most of the *rbc*L sequences were produced for the present study. For *rpo*A, most sequences were published previously (Petersen and Seberg 1997). The fragment sequenced spans the whole of *rpo*A, the 3′-end of *pet*D encoding subunit IV of the cytochrome bf complex, the 3′-end of *rps*11 encoding ribosomal protein S11, and the two IGSs between *rpo*A and *pet*D and between *rpo*A and *rps*11.

The mitochondrial *cox*II sequences were amplified in two fragments, the first fragment using primers COX2A3F and COX2SR, and the second using primers COX2SF and COX2B2R (Table 2). The region sequenced is a small part of exon 1 (ca. 55 bp) and nearly the whole (ca. 1200 bp) intron of *cox*II. All the mitochondrial sequences were produced for the present study.

The two nuclear regions sequenced are from disrupted meiotic cDNA (*DMC1*) and elongation factor G (*EF-G*), and, at least in *Hordeum vulgare,* belong to two different linkage groups, as *DMC1* is located on the top arm of chromosome 3 (V. Klimyuk pers. comm.), whereas *EF-G* is located on the long arm of chromosome 2 (Komatsuda et al. 1999). The *DMC1* sequences included in this study have been published previously, except the sequence of *Bromus arvensis,* which was amplified using the primers TDMC1e10 and BDMC1e15R (Petersen and Seberg 2000). The amplified region spans exon 10 to exon 15, numbered according to the gene structure of *Arabidopsis thaliana* (L.) Heynh. (Klimyuk and Jones 1997). For sequencing, additional primers TDMC1e13 and TDMC1e13R were used (Petersen and Seberg 2000).

Most *EF-G* sequences were amplified using primers cMWG699-T7-2 and cMWG699-T3-2 (Komatsuda et al. 1998). For amplification of *Bromus* and *Aegilops speltoides* sequences, the latter primer was exchanged with BE-F and EFG1R, respectively (Table 2). For sequencing of some samples additional primers were used, cMWG699-T7–3 and cMWG699-T3–3 (Komatsuda et al. 1999). Amplification priming sites are located in exon 3 and downstream of exon 4 (in the IGS region). For both nuclear regions, it has been impossible to sequence *Festucopsis festucoides.* Only a few of the *EF-G* sequences have been published previously (Komatsuda et al. 1999; Petersen and Seberg 2003).

*Data Analysis*

Sequence alignment was done by eye for all data sets, and gaps coded according to the principles developed by Simmons and Ochoterena (2000), using their simple gap coding method.

Phylogenetic analyses were done on each partition and on the combined data set. All analyses were performed using PAUP* vers. 4.0b8 (Swofford 2001) and the following options: heuristic search, 100 random-addition replicates, five trees held each step, tree-bisection-reconnection (TBR) swapping, and steepest descent. PAUP* vers. 4.0b8 has been preferred to the most recent vers. 10, as the latter outputs erroneous tree lengths and an excessive number of tree islands. Uninformative characters and sites within recoded

Table 1. Taxa studied, accession numbers, and genome designations according to Wang et al. (1996). *Bromus* is the outgroup. For details of vouchers, collectors, etc., see Petersen and Seberg (1997, 2000).

| Taxon | Accession number | Genome designation |
|---|---|---|
| *Bromus* L. | | |
| *B. arvensis* L. | C 618 | — |
| *B. sterilis* L. | OSA 420 | — |
| *Aegilops* L. | | |
| *A. comosa* Sibth. & Sm. | H 6673 | **M** |
| *A. speltoides* Tausch | H 10681 | **S** |
| *A. tauschii* Coss. | H 6668 | **D** |
| *Agropyron* Gaertn. | | |
| *A. cristatum* (L.) Gaertn. | H 4349 | **P** |
| *Amblyopyrum* (Jaub. & Spach) Eig | | |
| *A. muticum* (Boiss.) Eig | H 5572 | **T** |
| *Australopyrum* (Tzvelev) Á. Löve | | |
| *A. pectinatum* (Labill.) Á. Löve | H 6771 | **W** |
| *A. retrofractum* (Vickery) Á. Löve | H 6723 | **W** |
| *A. velutinum* (Nees) B. K. Simon | H 6724 | **W** |
| *Crithopsis* Jaub. & Spach | | |
| *C. delileana* (Schult.) Roshev. | H 5558 | **K** |
| *Dasypyrum* (Coss. & Durieu) T. Durand | | |
| *D. villosum* (L.) P Candargy | H 5561 | **V** |
| *Eremopyrum* (Ledeb.) Jaub. & Spach | | |
| *E. distans* (C. Koch) Nevski | H 5552 | **F** |
| *E. triticeum* (Gartn.) Nevski | H 5553 | **Xe** |
| *Festucopsis* (C. E. Hubb.) Melderis | | |
| *F. festucoides* (Maire) Á. Löve | H 6731 | **L** |
| *F. serpentinii* (C. E. Hubb.) Melderis | H 6511 | **L** |
| *Henrardia* C. E. Hubb. | | |
| *H. persica* (Boiss.) C. E. Hubb. | H 5556 | **O** |
| *Heteranthelium* Hochst. | | |
| *H. piliferum* (Banks & Sol.) Hochst. | H 5557 | **Q** |
| *Hordeum* L. | | |
| *H. brachyantherum* Nevski subsp. *californicum* (Covas & Stebbins) Bothmer, N. Jacobsen & Seberg | H 1942 | **H** |
| *H. erectifolium* Bothmer, N. Jacobsen & R. B. Jørg. | H 1150 | **H** |
| *H. marinum* Huds. subsp. *gussoneanum* (Parl.) Thell. | H 299 | **Xa** |
| *H. murinum* L. subsp. *glaucum* (Steud.) Tzvelev | H 801 | **Xu** |
| *H. vulgare* L. subsp. *spontaneum* (K. Koch) Asch. & Graebn. | H 3139 | **I** |
| *Lophopyrum* Á. Löve | | |
| *L. elongatum* (Host) Á. Löve | H 6692 | **E**$^{e}$ |
| *Peridictyon* Seberg, Fred. & Baden | | |
| *P. sanctum* (Janka) Seberg, Fred. & Baden | H 5575 | **Xp** |
| *Psathyrostachys* Nevski | | |
| *P. fragilis* (Boiss.) Nevski subsp. *fragilis* | H 917 | **Ns** |
| *P. fragilis* subsp. *villosus* Baden | H 4372 | **Ns** |
| *P. stoloniformis* Baden | H 9182 | **Ns** |
| *Pseudoroegneria* (Nevski) Á. Löve | | |
| *P. spicata* (Pursh) Á. Löve | H 9082 | **St** |
| *Secale* L. | | |
| *S. strictum* (C. Presl) C. Presl | H 4342 | **R** |
| *Taeniatherum* Nevski | | |
| *T. caput-medusae* (L.) Nevski | H 10254 | **Ta** |
| *Thinopyrum* Á. Löve | | |
| *T. bessarabicum* (Săvul. & Rayss) Á. Löve | H 6725 | **E**$^{b}$ |
| *Triticum* L. | | |
| *T. monococcum* L. | H 4547 | **A**$^{m}$ |

gaps were excluded, and informative characters were treated as equally weighted and nonadditive.

Support measures were computed for the individual partitions and the combined data set. Bremer support (BS) (Bremer 1994) was calculated for the morphological and combined data only, using converse constraint (Eernisse and Kluge 1993). Jackknifing was not done for the morphological data set, only BS, as morphological data do not meet the basic statistic requirements for resampling. For reasons given by Farris et al. (1996) jackknifing was preferred to bootstrapping. Jackknifing was performed with PAUP* emulating ''Jac'' resampling (Farris et al. 1996) with a removal rate of $1\text{-}e^{-1}$ (ca. 0.37), running 1000 replicates. Jackknife values 50–67% are considered weak, 68–84% moderate, and 85–100% strong. To examine possible conflicts among the data sets, incongruence length difference (ILD) tests (Mickevich and Farris 1981; Farris et al. 1994) were conducted on the various character partitions, i.e., between the individual data sets, and between the plastid, mitochondrial, nuclear, and morphological data. The ILD tests were performed using PAUP* running 1000 replicates. In tests involving the nuclear data, *Festucopsis festucoides* was removed.

To evaluate the contributions by each data set to the individual clades in the strict consensus tree based on the combined analysis, partitioned Bremer support values were calculated as outlined by Baker et al. (1998). See also Gatesy et al. (1999) for an extensive discussion of different support measures.

## RESULTS

There is no length variation in *rbc*L and only a limited number of insertions/deletions (indels) in the IGS surround-

Table 2. Primers used for sequencing *EF-G* and *cox*II.

| Name | Gene | Sequence |
|---|---|---|
| BE-F | *EF-G* | CAG CTT GCA AAA TTA TGG AC |
| EFG1R | *EF-G* | ATT ATG GAC ATT ATC AGG TCC |
| COX2A3F | *cox*II | ACG AGG TAG TCG TAG ATC CAG |
| COX2SR | *cox*II | GGA TCA CTG TAT TCA GAG GTC |
| COX2SF | *cox*II | GAC CTC YGA ATA CAG TGA TCC |
| COX2B2R | *cox*II | CGT CGG ACC TAT TAT AGT CC |

Table 3. Statistics for the four major partitions and combined data matrix, excluding uninformative characters. *Festucopsis festucoides* was not included in comparisons involving nuclear DNA.

| | Morphology | cpDNA | mtDNA | nDNA | Combined |
|---|---|---|---|---|---|
| Parsimony-informative characters | 31 | 77 | 13 | 338 | 459 |
| Number of trees | ca. 4442 | 112 | >50,000 | 6 | 4 |
| Tree length | 95 | 127 | 19 | 654 | 986 |
| Consistency index | 0.38 | 0.64 | 0.68 | 0.64 | 0.56 |
| Retention index | 0.72 | 0.85 | 0.88 | 0.75 | 0.70 |

ing the *rpo*A gene. Most indels can be interpreted as short duplications and alignment was trivial. However, downstream of *rpo*A there is a short problematic T-rich region that was here aligned, but potentially may be a microsatellite. The *rpo*A sequences vary in length from 1343 to 1355 bp. The alignments of *rbc*L and *rpo*A are 1478 and 1375 bp long, respectively.

The *cox*II sequences vary in length from 1260 to 1319 bp, and were relatively straightforward to align. The alignment is 1336 bp long.

The alignment of *DMC1* shows a few long indels in intron 14. The most notable are a 44 bp length difference between the outgroup and the ingroup, a difference of 31 bp between *Secale strictum* and all other taxa, and a difference of 55 bp between *Aegilops speltoides* and *Amblyopyrum muticum* and the remaining taxa. Additionally, *DMC1* includes a transposable element, a MITE, in intron 14 in several taxa (Petersen and Seberg 2000). The MITE sequences, plus the footprints left after excision of the MITE in other taxa, were removed from the analyses. The sequences vary in length from 991 to 1035 bp (excluding the MITE and footprints) and the overall alignment is 1129 bp long. The most prominent feature in the alignment of *EF-G* is the lack of ca. 260 bp in intron 3 of *Bromus* compared to Triticeae. Triticeae sequences vary between 850 and 874 bp in length and *Bromus* sequences are 634 bp long. The alignment is 906 bp long. A full table of GenBank accession numbers will be published in a future paper.

In the combined, total evidence matrix (excluding the gap coding) 2.2% of the cells were scored as ambiguous data (? or N). However, if *Festucopsis festucoides* is excluded from the matrix this drops to 1.3%.

The statistics for the four major data partitions and combined data matrix used in the present analyses are provided in Table 3. In addition, numbers of parsimony-informative sites for the individual markers *rbc*L, *rpo*A, *DMC1,* and *EF-G* are 32, 45, 171, and 167, respectively. The partition homogeneity (ILD) tests unambiguously show that nearly all data partitions are mutually incongruent ($P = 0.01$ or 0.02) except the nuclear and mitochondrial data sets ($P = 0.11$) (Table 4). On a broader scale, the molecular and morphological data partitions are incongruent ($P = 0.01$), and the two nuclear genes ($P = 0.01$) are mutually incongruent, too. The two chloroplast genes, however, are congruent ($P = 0.10$).

When the morphological matrix is analyzed separately (length [L] = 95, consistency index [CI] = 0.38, retention index [RI] = 0.72), few weakly supported clades (BS = 1–2) remain on the strict consensus tree (based on ca. 4442 most-parsimonious trees; Fig. 1). Clades that remain in the ingroup all correspond to the clades found in the original analysis (Seberg and Frederiksen 2001). The largest clade includes *Crithopsis, Hordeum, Psathyrostachys,* and *Taen-*

Table 4. Incongruence length difference (Mickevich and Farris 1981; Farris et al. 1994) measures in comparisons between the four major partitions. *Festucopsis festucoides* was not included in the comparisons involving nuclear DNA. Numbers given are *P* values. An asterisk (*) indicates that the two partitions are significantly different (i.e., mutually incongruent).

| | cpDNA | mtDNA | nDNA | Morphology |
|---|---|---|---|---|
| cpDNA | — | 0.02* | 0.01* | 0.01* |
| mtDNA | — | — | 0.11 | 0.01* |
| nDNA | — | — | — | 0.01* |
| Morphology | — | — | — | — |

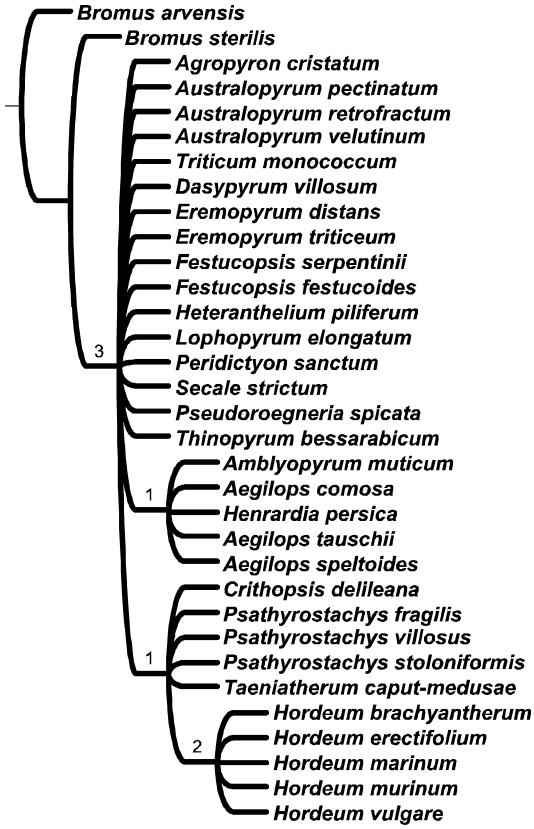


Fig. 1.—Strict consensus of 4442 most-parsimonious trees (L = 95, CI = 0.38, RI = 0.72) based on morphology alone. The numbers above the branches are Bremer support values.

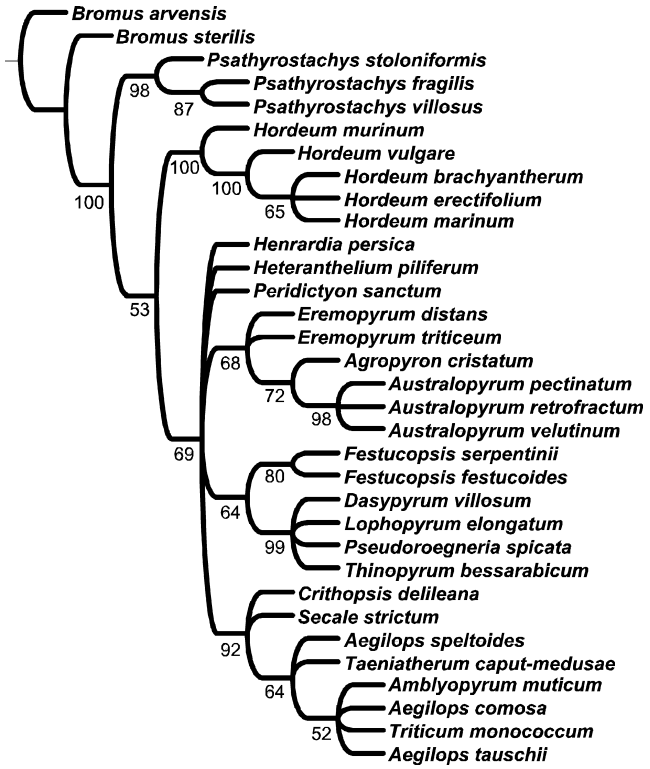


Fig. 2.—Strict consensus of 112 most-parsimonious trees (L = 127, CI = 0.64, RI = 0.85) based on the chloroplast data (*rbc*L and *rpo*A) alone. The numbers below the branches are jackknife values.

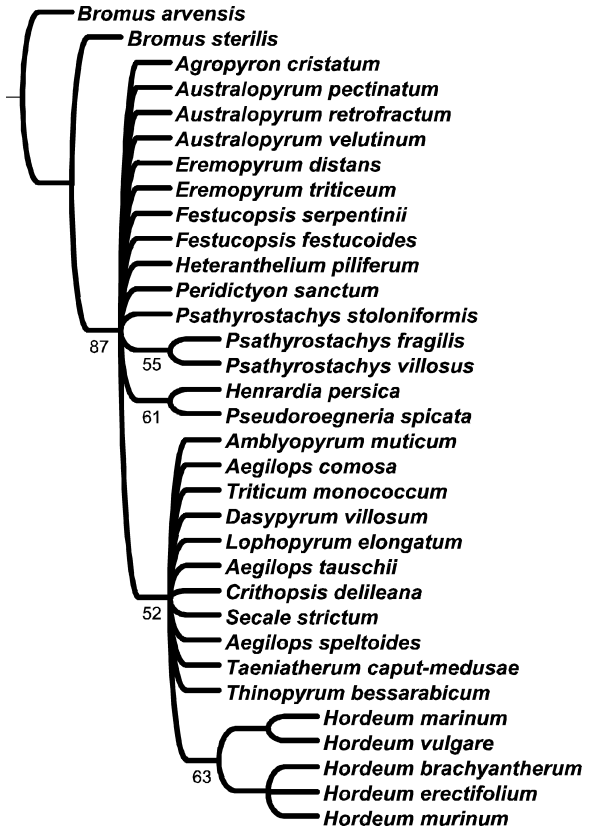


Fig. 3.—Strict consensus of >50,000 most-parsimonious trees (L = 19, CI = 0.68, RI = 0.88) based on the mitochondrial data (*cox*II) alone. The numbers below the branches are jackknife values.

*iatherum,* with *Hordeum* as a clade nested within this larger clade. A third clade, comprising *Aegilops comosa, A. speltoides, A. tauschii, Amblyopyrum,* and *Henrardia,* is also consistent with the *Aegilops* clade as defined by Seberg and Frederiksen (2001), though the number of taxa in the clade is smaller than in the original analysis. However, the only clade that is shared between the morphology tree (Fig. 1) and the strict consensus tree from the total evidence analysis (Fig. 5) is the *Hordeum* clade.

Analyzing the chloroplast data alone results in 112 most-parsimonious trees (L = 127, CI = 0.64, RI = 0.85) and a fairly well-resolved strict consensus tree (Fig. 2). However, it is largely in conflict with the total evidence tree (Fig. 5). On the chloroplast tree, *Psathyrostachys* forms a strongly supported clade (jackknife = 98%) that is, with weak support (jackknife = 53%), sister to the remaining Triticeae, with a strongly supported (jackknife = 100%), monophyletic *Hordeum* diverging next. Both the *Hordeum* and *Psathyrostachys* clades are also strongly supported (jackknife = 100%) on the total evidence tree (Fig. 5), but as sister groups embedded inside Triticeae with strong support (jackknife = 100%). Other moderately to strongly supported clades shared with the total evidence tree include *Australopyrum* (jackknife = 98%) and *Festucopsis* (jackknife = 80%).

The analysis of the mitochondrial data set was not run to completion and the strict consensus tree (Fig. 3), based on more than 50,000 most-parsimonious trees (L = 19, CI = 0.68, RI = 0.88), shows very limited and weakly supported resolution. In comparison with the total evidence tree (Fig. 5), *Hordeum* forms a clade in both, though only weakly supported by the mitochondrial data (jackknife = 63%), as do *Henrardia* plus *Pseudoroegneria* (jackknife = 61%) and *Psathyrostachys fragilis* subspp. *fragilis* and *villosus* (jackknife = 55%). A large unresolved clade found on the total evidence tree (Fig. 5), and consisting of species traditionally considered belonging to the wild wheats (*Aegilops, Amblyopyrum, Triticum*) plus *Crithopsis, Dasypyrum, Lophopyrum, Secale, Taeniatherum,* and *Thinopyrum,* is not contradicted by the mitochondrial data.

Analysis of the nuclear data results in two TBR islands, one of four trees and one of two trees (L = 654, CI = 0.64, RI = 0.75), each resulting in an almost completely resolved strict consensus tree, and the combined consensus of all six is shown in Fig. 4. The consensus tree is well resolved and nearly all relationships are strongly supported. Several similarities are shared with the total evidence tree (Fig. 5). On both trees, with 95–100% jackknife support, *Australopyrum, Hordeum,* and *Psathyrostachys* are each monophyletic, and *Agropyron* plus *Eremopyrum, Henrardia* plus *Pseudoroegneria,* and *Festucopsis* plus *Peridictyon* are resolved as clades. Except for *Hordeum,* the three clades that each comprise more than two taxa (*Festucopsis festucoides* was not included in the nuclear data set; see above) even show the same internal structure on both trees. In contrast, relationships among most clades are very different, except for the strongly supported (jackknife = 97–100%) positions of the *Henrardia* plus *Pseudoroegneria* and *Festucopsis* plus *Peridictyon* clades.

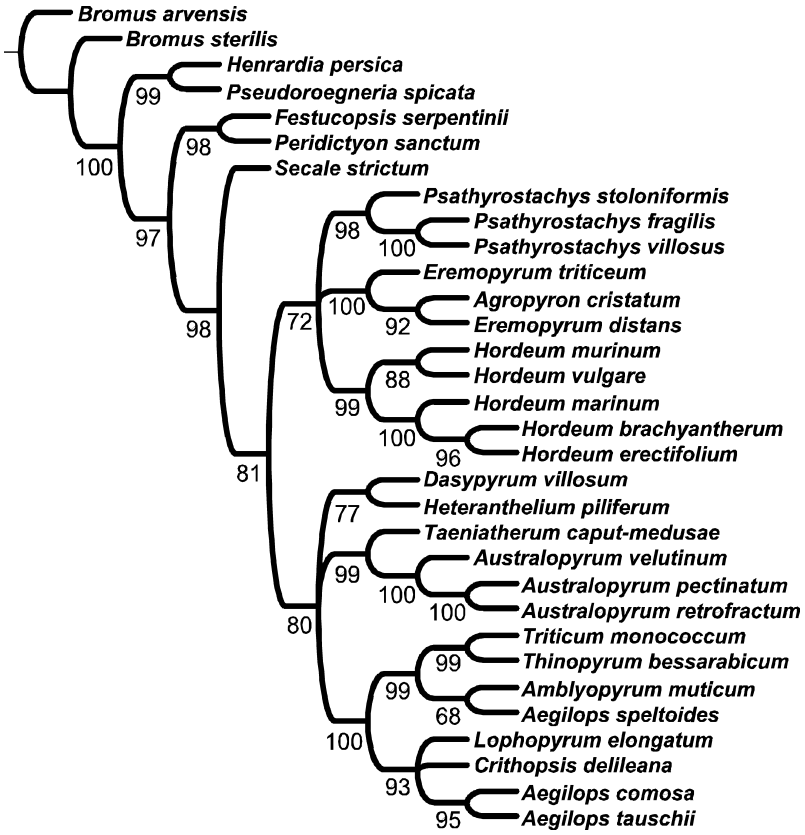


Fig. 4.—Strict consensus of six most-parsimonious trees (L = 654, CI = 0.64, RI = 0.75) is based on the nuclear data (*DMC1* and *EF-G*) alone. The numbers below the branches are jackknife values.

The total evidence analysis results in only four most-parsimonious trees (L = 986, CI = 0.56, RI = 0.71) and a highly resolved strict consensus tree (Fig. 5). The majority of the clades are moderately to strongly supported. The *Hordeum* clade, which was found in all the individual analyses (Fig. 1–4), is strongly supported (jackknife = 100%; BS = 19). Its sister relationship to *Psathyrostachys,* a genus itself strongly supported (jackknife = 100%; BS = 10), is also strongly supported (jackknife = 100%; BS = 6). The *Hordeum*/*Psathyrostachys* clade is part of a trichotomy involving a strongly supported clade (jackknife = 100%; BS = 9) composed of *Agropyron* plus *Eremopyrum* and a clade composed of the remaining Triticeae except *Festucopsis, Henrardia, Peridictyon,* and *Pseudoroegneria. Agropyron* is embedded within *Eremopyrum,* but this relationship is only moderately supported (jackknife = 72%; BS = 1). The third clade in the trichotomy is weakly supported (jackknife = 62%; BS = 1). Within this clade *Australopyrum* forms a strongly supported clade (jackknife = 100%; BS = 14) and, though weakly supported (jackknife = 57–62%; BS = 1), is sister to the remaining taxa. Many relationships among the wild wheats (*Aegilops, Amblyopyrum, Triticum*), *Crithopsis, Dasypyrum, Heteranthelium, Lophopyrum, Secale, Taeniatherum,* and *Thinopyrum* are unresolved or weakly to moderately supported. However, exceptions include the relationships between *Aegilops speltoides* and *Amblyopyrum* (jackknife = 100%; BS = 6), and between *Aegilops comosa* and *A. tauschii* (jackknife = 99%; BS = 5). Branching sequentially as strongly supported sister groups to the rest of Triticeae are first a clade composed of *Henrardia* plus *Pseudoroegneria* (jackknife = 98%; BS = 5) and a clade consisting of *Festucopsis* plus *Peridictyon* (jackknife = 95%; BS = 4). Monophyly of Triticeae is strongly supported (jackknife = 100; BS = 29). In general, the low support of many clades is caused by data conflict, not lack of variation (see Fig. 6).

## DISCUSSION

Numerous accounts have been published on the phylogeny of Triticeae based either on individual data sets (e.g., single gene sequences, cpDNA restriction sites) or various combinations of data sets (see above). A few of these differ from our study in having a less complete taxon sampling (5S RNA arrays, Kellogg and Appels 1995) or employing, in our opinion, unacceptable data analysis (Monte et al. 1993). These studies will not be dealt with here. The results from other published analyses will be discussed below in relation to our own results from different data partitions and combined analysis of all data, unless the data constitute a subset of the data used here or elsewhere; e.g., we will not discuss our own *rpo*A data (Petersen and Seberg 1997) as they are included both in an analysis by Mason-Gamer et al. (2002) involving three chloroplast regions and in the present analyses.

### *Phylogeny Based on Morphology*

When the morphology matrix is analyzed alone, few clades remain on the strict consensus tree (Fig. 1) in com-

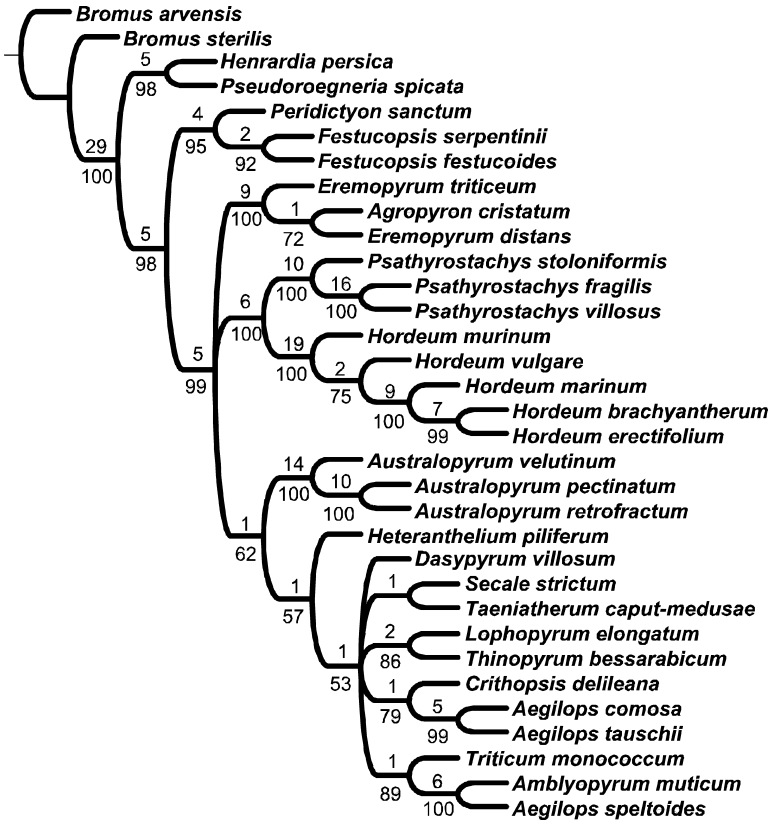


Fig. 5.—Strict consensus of four most-parsimonious trees (L = 986, CI = 0.56, RI = 0.70) based on all data—morphology, two chloroplast genes (*rbc*L and *rpo*A), one mitochondrial gene (*cox*II), and two nuclear genes (*DMC1* and *EF-G*). The numbers above and below the branches are Bremer support and jackknife values, respectively.

parison with the similar analysis done by Seberg and Frederiksen (2001). However, the taxon sampling has been changed in comparison to the previous study (see Seberg and Frederiksen 2001: Fig. 1 and pp. 81–82). Most notably, all species belonging to the *Aegilops* clade in the previous analysis are here treated separately, and not combined into a hypothetical taxonomic unit. A thorough discussion of all previous morphological analyses of Triticeae may be found in Seberg and Frederiksen (2001).

### *Chloroplast Phylogeny*

In the following discussion, published individual analyses of chloroplast data (RFLPs, Kellogg 1992; *rpo*A, Petersen and Seberg 1997) will not be dealt with separately, as they are both integrated and expanded upon by Mason-Gamer et al. (2002). Mason-Gamer et al.'s (2002) analyses were made in connection with a study of the origin of the genus *Elymus* L. and include a phylogenetic analysis of most of the monogenomic Triticeae based on three chloroplast data sets (Mason-Gamer et al. 2002: Fig. 1). As indicated above, their analyses were based on previously published data sets supplemented with added taxon sampling and a new data set, the chloroplast tRNA genes, *trn*T, *trn*L–3′, *trn*L–5′, and *trn*F, and their intervening noncoding regions. In all instances, *Psathyrostachys* was chosen as the outgroup (Mason-Gamer 2001: 993). A largely congruent tree based on the same data, but with a slightly different taxon sampling and using maximum likelihood, has subsequently been published by Mason-Gamer (2004).

The strict consensus of the 112 most-parsimonious chloroplast trees found here is rather well resolved (Fig. 2). The following discussion is restricted to comparisons with Mason-Gamer et al.'s (2002) combined analysis. The position of *Hordeum* as sister to the remaining Triticeae (excluding *Psathyrostachys*) is shared between the consensus trees (Fig. 2; Mason-Gamer et al. 2002: Fig. 1D), though with much stronger support on Mason-Gamer et al.'s tree (bootstrap = 100%) than on ours (jackknife = 53%). Additionally, a number of moderately to strongly supported relationships are shared between the trees. *Agropyron, Australopyrum,* and *Eremopyrum* are part of the same clade on both. However, on Mason-Gamer et al.'s (2002) tree, which has notably greater taxon sampling in *Agropyron, Henrardia,* and *Peridictyon* are included in this clade. This is among the possibilities on our tree (Fig. 2), although relationships among the taxa are different. Both trees also share a clade that includes *Aegilops, Secale,* and *Taeniatherum,* though this clade in our analyses also includes *Amblyopyrum* and *Triticum,* which were not sampled by Mason-Gamer et al. (2002). Furthermore, both trees share an unresolved clade consisting of *Dasypyrum, Lophopyrum* (as *Thinopyrum elongatum* (Host) D. R. Dewey in Mason-Gamer et al. 2002), *Pseudoroegneria,* and *Thinopyrum. Pseudoroegneria* is at best paraphyletic on Mason-Gamer et al.'s tree, but as our analyses include only

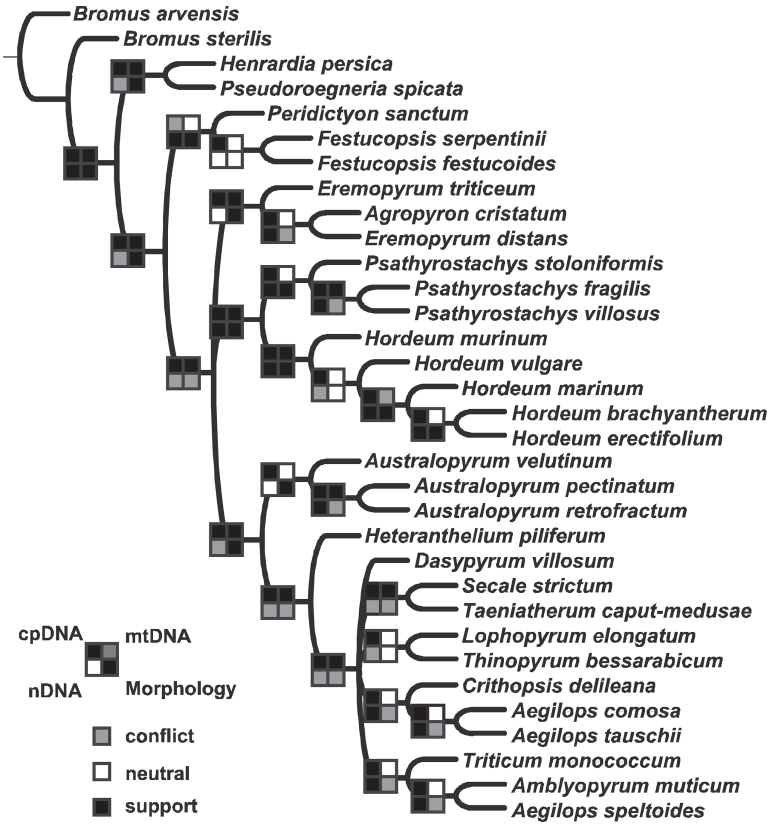


Fig. 6.—Partitioned Bremer support (Baker et al. 1998) of individual nodes grouped according to whether or not the four data partitions conflict, support, or are silent about the contribution of each data set to the resolution of a particular node.

one accession of *Pseudoroegneria,* it has not been possible to address monophyly of the genus.

### *Mitochondrial Phylogeny*

As the mitochondrial (*cox*II) data presented here are the first to be used in a phylogenetic context, no comparisons with other such studies are possible.

### *Nuclear Phylogeny*

The structure of the strict consensus tree based on the two putatively single-copy nuclear genes (*DMC1, EF-G*) analyzed in combination is almost completely resolved (Fig. 4). The only other phylogenetic analyses of Triticeae based on sequences from single or low-copy-number nuclear genes are by Mason-Gamer et al. (1998) and Mason-Gamer (2001, 2004) based on granule-bound starch synthase (*waxy*) using maximum likelihood, and by Petersen and Seberg (2000, 2002) based on parsimony analyses of *DMC1* sequences. The use of fundamentally different types of data analysis of the two genes in these studies makes direct comparison very difficult, as this is likely to further aggravate any difference in data structure, and the resulting topologies are nearly without similarity. The only significant similarity is that *Agropyron* and *Eremopyrum* constitute a clade in all analyses, but this relationship, which is strongly supported in the present analysis (Fig. 5; jackknife = 100%; BS = 9) is unsupported in the *waxy* tree (Mason-Gamer et al. 1998; Mason-Gamer 2001, 2004). However, due to the inclusion of paralogous sequences from polyploid species, the exact relationships in Mason-Gamer's (2001, 2004) later analyses are difficult to compare with the others. Whether incongruence is due to insufficient or different taxon sampling, methodology, or different gene histories will not be pursued further here. However, nearly all relevant clades in Mason-Gamer et al.'s (1998) tree have bootstrap values below 50%. The few that have higher bootstrap support are all moderately or highly supported and reflect undisputed, traditionally accepted relationships verging on the trivial: *Hordeum* plus *Critesion* Rafin. (*Critesion* is usually part of *Hordeum*; bootstrap = 69%) and *Triticum* (bootstrap = 92%), though *Aegilops* is unsupported (bootstrap $\leq$ 50%).

The data from *DMC1* (Petersen and Seberg 2000, 2002) are included in the analyses shown here and will not be discussed in isolation, but only in conjunction with the other nuclear gene (*EF-G*). For reasons given above, the data derived from the 5S RNA gene (Kellogg and Appels 1995) have not been dealt with either. Consequently, the following comparison is restricted to a discussion of Hsiao et al.'s (1995*b*) parsimony-derived phylogenies based on data from the multicopy nuclear array, ITS.

The strict consensus of the two TBR islands found here (Fig. 4), based on data from *DMC1* and *EF-G,* deviates considerably from the trees obtained by Hsiao et al. (1995*b*). While acknowledging *Bromus* as the most likely outgroup of Triticeae, Hsiao et al. (1995*b*) experimented with different outgroups, but eventually chose *B. tectorum* L. as the outgroup as it ''gave much better resolution of ingroup rela-

tionships'' than did either of the other outgroups (*Avena longiglumis* Durieu and *Brachypodium sylvaticum* (Huds.) P. Beauv.), and by using a transition : transversion : gap ratio of 3 : 1 : 1 recovered only a single tree. One of the most striking dissimilarities between the single most-parsimonious tree found by Hsiao et al. (1995*b*: Fig. 2) and our tree based on the nuclear genes (Fig. 4) is perhaps the position of *Henrardia* and *Pseudoroegneria* as sister to the rest of Triticeae in our tree. In Hsiao et al.'s (1995*b*) tree *Hordeum* is sister to the remaining Triticeae, but with a bootstrap value ≤50%, and *Psathyrostachys* is the next taxon to diverge (bootstrap = 78%), followed by *Australopyrum* (bootstrap = 63%) and *Agropyron* (bootstrap ≤ 50%). *Pseudoroegneria* is embedded within the tree in a clade with moderate support (bootstrap = 74%). In the tree based on nuclear data presented here (Fig. 4), *Australopyrum* is sister to *Taeniatherum,* and a clade consisting of *Agropyron, Eremopyrum, Hordeum,* and *Psathyrostachys* is embedded inside the tree. However, both analyses point to a close relationship between *Dasypyrum* and *Heteranthelium.* In contrast to the strict consensus tree based on cpDNA (Fig. 2), *Hordeum* and *Psathyrostachys* form a clade with *Agropyron* and *Eremopyrum* that is embedded within the tree derived from nuclear data (Fig. 4).

In equally weighted reanalyses of Hsiao et al.'s (1995*b*) data by Kellogg et al. (1996: Fig. 2c) and Mason-Gamer and Kellogg (1996*a*), a large part of the resolution disappears or is changed. However, the relationships of *Hordeum* (including *Critesion,* which is treated as a separate genus by Hsiao et al. 1995*b*) and *Psathyrostachys* remain the same, though the relationship between *Critesion* and *Hordeum* is unresolved, and the branch separating *Psathyrostachys* from the remaining Triticeae species has bootstrap support of 58%. Otherwise, most of the supported relationships are rather trivial; e.g., *Agropyron* (bootstrap = 98%), *Australopyrum* (bootstrap = 91%), and *Secale* (bootstrap = 100%) are all monophyletic. The remaining topology, though unsupported (bootstrap ≤ 50%), is similar to that found by Hsiao et al. (1995*b*), in conflict with the topology found in the present study using nuclear sequences.

A new phylogeny of the monogenomic Triticeae based on β-amylase and published in a recent paper by Mason-Gamer (2005) is in line with previous phylogenies by showing some traditional relationships and a number of unexpected new ones.

### *Phylogenetic Incongruence*

Incongruence among different data sets has been dealt with before (Kellogg et al. 1996) and is also evident from the present analysis (Table 4). Kellogg et al. (1996) and Mason-Gamer and Kellogg (1996*a*) have tried to pinpoint the reasons behind the incongruence. However, both of these analyses rely heavily on comparisons between data derived from short and long 5S RNA arrays and chloroplast DNA data. In addition to the limited taxon sampling mentioned above, there are several inherent problems related to the use of 5S RNA gene data, including alignment that is far from straightforward (especially when true outgroup taxa are included), often extensive intra- and interspecific polymorphisms, and in most instances only one of the two arrays known to occur in Triticeae (Kellogg et al. 1996) has been sequenced (either due to technical problems or because they are truly absent) (see Petersen and Seberg 2004).

In our analysis, it is evident that different data sets make different contributions to resolve or contradict a given node, at least as measured by partitioned Bremer support (Gatesy et al. 1999). Of the 26 resolved nodes found in the total evidence strict consensus tree (Fig. 6):

(1) Morphology supports 12, is in conflict with 12, and silent in 3.
(2) cpDNA data support 24, are in conflict with 1, and silent in 1.
(3) mtDNA data support 12, are in conflict with 1, and silent in 13.
(4) nDNA data support 13, are in conflict with 9, and silent in 2.

However, all data sets have considerable hidden support or conflict (Gatesy et al. 1999). A future paper will explore the contribution of individual data sets to specific nodes in greater detail, and thereby attempt to add to our understanding of data conflict.

A larger paper, including a new chloroplast data set (*ndh*F, subunit of NADH dehydrogenase) and additional ITS sequences, supplementing the existing ITS data set (Hsiao et al. 1995*b*) such that the taxon sampling agrees with the sampling used here, is in progress.

ACKNOWLEDGMENTS

We are grateful to the botanical gardens and seed banks that have added to our own extensive collection of Triticeae of wild origin. This paper has benefited from comments by Frank Blattner. We thank Charlotte Hansen and Lisbeth Knudsen for skillful technical assistance. Financial support was provided by the Danish Natural Science Research Council (grant 50-00-0271).

LITERATURE CITED

BADEN, C. 1990. A taxonomic revision of *Psathyrostachys* (Poaceae). *Nordic J. Bot.* **11**: 3–26.

BAKER, R. H., X. YU, AND R. DESALLE. 1998. Assessing the relative contribution of molecular and morphological characters in simultaneous analysis trees. *Molec. Phylogen. Evol.* **9**: 427–436.

BAUM, B. R., AND R. APPELS. 1992. Evolutionary change at the 5S DNA loci of species in the Triticeae. *Pl. Syst. Evol.* **183**: 195–208.

———, D. A. JOHNSON, AND L. G. BAILEY. 2001. Defining orthologous groups among multicopy genes prior to inferring phylogeny, with special emphasis on the Triticeae (Poaceae). *Hereditas* **135**: 123–138.

BREMER, K. 1994. Branch support and tree stability. *Cladistics* **10**: 295–304.

DOYLE, J. J., AND J. F. DOYLE. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem. Bull.* **19**: 11–15.

EERNISSE, D. J., AND A. G. KLUGE. 1993. Taxonomic congruence versus total evidence, and amniote phylogeny inferred from fossils, molecules, and morphology. *Molec. Biol. Evol.* **10**: 1170–1195.

FARRIS, J. S., V. A. ALBERT, M. KÄLLERSJÖ, D. LIPSCOMB, AND A. G.

KLUGE. 1996. Parsimony jackknifing outperforms neighbor-joining. *Cladistics* **12**: 99–124.

———, M. KÄLLERSJÖ, A. G. KLUGE, AND C. BULT. 1994. Testing significance of incongruence. *Cladistics* **10**: 315–319.

FREDERIKSEN, S., AND O. SEBERG. 1992. Phylogenetic analysis of the Triticeae (Poaceae). *Hereditas* **116**: 15–19.

GATESY, J., P. O. O'GRADY, AND R. H. BAKER. 1999. Corroboration among data sets in simultaneous analysis: hidden support for phylogenetic relationships among higher level artiodactyl taxa. *Cladistics* **15**: 271–313.

GRANT, T., AND A. G. KLUGE. 2003. Data exploration in phylogenetic inference: scientific, heuristic, or neither. *Cladistics* **19**: 379–418.

GRASS PHYLOGENY WORKING GROUP [GPWG]. 2001. Phylogeny and subfamilial classification of the grasses (Poaceae). *Ann. Miss. Bot. Gard.* **88**: 373–457.

HSIAO, C., N. J. CHATTERTON, K. H. ASAY, AND K. B. JENSEN. 1995*a*. Molecular phylogeny of the Pooideae (Poaceae) based on nuclear rDNA (ITS) sequences. *Theor. Appl. Genet.* **90**: 389–398.

———, ———, ———, AND ———. 1995*b*. Phylogenetic relationships of the monogenomic species of the wheat tribe, Triticeae (Poaceae), inferred from nuclear rDNA (internal transcribed spacer) sequences. *Genome* **38**: 211–223.

———, S. W. L. JACOBS, N. J. CHATTERTON, AND K. H. ASAY. 1999. A molecular phylogeny of the grass family (Poaceae) based on the sequences of nuclear ribosomal DNA (ITS). *Austral. Syst. Bot.* **11**: 667–688.

KELLOGG, E. A. 1989. Comments on genomic genera in the Triticeae (Poaceae). *Amer. J. Bot.* **76**: 796–805.

———. 1992. Restriction site variation in the chloroplast genomes of the monogenomic Triticeae. *Hereditas* **116**: 43–47.

———, AND R. APPELS. 1995. Intraspecific and interspecific variation in 5S RNA genes are decoupled in diploid wheat relatives. *Genetics* **140**: 325–343.

———, ———, AND R. J. MASON-GAMER. 1996. When genes tell different stories: the diploid genera of Triticeae (Gramineae). *Syst. Bot.* **21**: 321–347.

KLIMYUK, V. I., AND J. D. G. JONES. 1997. *AtDMC1,* the *Arabidopsis* homologue of the yeast *DMC1* gene: characterization, transposon-induced allelic variation and meiosis-associated expression. *Plant J.* **11**: 1–14.

KOMATSUDA, T., I. NAKAMURA, F. TAKAIWA, AND S. OKA. 1998. Development of STS markers closely linked to the *vrs1* locus in barley, *Hordeum vulgare*. *Genome* **41**: 680–685.

———, K. I. TANNO, B. SALOMON, T. BRYNGELSSON, AND R. VON BOTHMER. 1999. Phylogeny in the genus *Hordeum* based on nucleotide sequences closely linked to the *vrs1* locus (row number of spikelets). *Genome* **42**: 973–981.

LÖVE, Á. 1984. Conspectus of the Triticeae. *Feddes Repert.* **95**: 425–521.

MASON-GAMER, R. J. 2001. Origin of North American *Elymus* (Poaceae: Triticeae) allotetraploids based on granule-bound starch synthase gene sequences. *Syst. Bot.* **26**: 757–768.

———. 2004. Reticulate evolution, introgression, and intertribal gene capture in an allohexaploid grass. *Syst. Biol.* **53**: 25–37.

———. 2005. The β-amylase genes of grasses and a phylogenetic analysis of the Triticeae (Poaceae). *Amer. J. Bot.* **92**: 1045–1058.

———, AND E. A. KELLOGG. 1996*a*. Testing for phylogenetic conflict among molecular data sets in the tribe Triticeae (Poaceae). *Syst. Biol.* **45**: 524–545.

———, AND ———. 1996*b*. Chloroplast DNA analysis of the monogenomic Triticeae: phylogenetic implications and genome-specific markers, pp. 301–325. *In* P. P. Jauhar [ed.], Methods of genome analysis in plants. CRC Press, Boca Raton, Florida, USA.

———, 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.

———, C. F. WEIL, AND E. A. KELLOGG. 1998. Granule-bound starch synthase: structure, function, and phylogenetic utility. *Molec. Biol. Evol.* **15**: 1658–1673.

MICKEVICH, M., AND J. S. FARRIS. 1981. The implications of congruence in *Menidia*. *Syst. Zool.* **30**: 351–370.

MIYAMOTO, M. M., AND J. CRACRAFT. 1992. Phylogenetic inference, DNA sequence analysis, and the future of molecular systematics, pp. 3–17. *In* M. M. Miyamoto and J. Cracraft [eds.], Phylogenetic analysis of DNA Sequences. Oxford University Press, Oxford, UK. 368 p.

———, AND W. M. FITCH. 1995. Testing species phylogenies and phylogenetic methods with congruence. *Syst. Biol.* **44**: 64–76.

MONTE, J. V., C. L. MCINTYRE, AND J. P. GUSTAFSON. 1993. Analysis of phylogenetic relationships in the Triticeae tribe using RFLPs. *Theor. Appl. Genet.* **86**: 649–655.

PETERSEN, G., AND O. SEBERG. 1997. Phylogenetic analysis of the Triticeae (Poaceae) based on *rpo*A sequence data. *Molec. Phylogen. Evol.* **7**: 217–230.

———, AND ———. 2000. Phylogenetic evidence for excision of *Stowaway* miniature inverted-repeat transposable elements in Triticeae (Poaceae). *Molec. Biol. Evol.* **17**: 1589–1596.

———, AND ———. 2002. Molecular evolution and phylogenetic application of *DMC1*. *Molec. Phylogen. Evol.* **22**: 43–50.

———, AND ———. 2003. Phylogenetic analyses of the diploid species of *Hordeum* (Poaceae) and a revised classification of the genus. *Syst. Bot.* **28**: 293–306.

———, AND ———. 2004. On the origin of tetraploid species *Hordeum capense* and *H. secalinum* (Poaceae). *Syst. Bot.* **29**: 862–873.

SAKAMOTO, S. 1979. Genetic relationships among four species of the genus *Eremopyrum* in the tribe Triticeae, Gramineae. *Mem. Coll. Agric. Kyoto Univ.* **114**: 1–27.

SEBERG, O., AND S. FREDERIKSEN. 2001. A phylogenetic analysis of the monogenomic Triticeae (Poaceae) based on morphology. *Bot. J. Linn. Soc.* **136**: 75–97.

———, ———, C. BADEN, AND I. LINDE-LAURSEN. 1990. *Peridictyon,* a new genus from the Balkan Peninsula, and its relationship with *Festucopsis* (Poaceae). *Willdenowia* **21**: 87–104.

SIMMONS, M. P., AND H. OCHOTERENA. 2000. Gaps as characters in sequence-based phylogenetic analyses. *Syst. Biol.* **49**: 369–381.

SORENG, R. J., J. I. DAVIS, AND J. J. DOYLE. 1990. A phylogenetic analysis of chloroplast DNA restriction site variation in Poaceae subfam. Pooideae. *Pl. Syst. Evol.* **172**: 83–97.

SWOFFORD, D. L. 2001. PAUP*: phylogenetic analysis using parsimony (*and other methods), vers. 4.0b8. Sinauer Associates, Inc., Sunderland, Massachusetts, USA.

WANG, R.-R. C., R. V. BOTHMER, J. DVORAK, G. FEDAK, I. LINDE-LAURSEN, AND M. MURAMATSU. 1996. Genome symbols in the Triticeae (Poaceae), pp. 29–34. *In* R.-R. C. Wang, K. B. Jensen, and C. Jaussi [eds.], Proceedings of the 2nd International Triticeae Symposium 20–24 Jun 1994. Publication, Design, & Production of Utah State University, Logan, USA.

WATSON, L., H. T. CLIFFORD, AND M. J. DALLWITZ. 1985. The classification of the Poaceae: subfamilies and supertribes. *Austral. J. Bot.* **33**: 433–484.