

Phylogenetic Analysis of Allopolyploid Species

G. PETERSEN and O. SEBERG

Botanical Garden and Museum, The Natural History Museum of Denmark, University of Copenhagen, DK-1307 Copenhagen K, Denmark, e-mail: gittep@snm.ku.dk

Abstract: In principle all methods of phylogenetic analysis presuppose that evolution is divergent, hence phylogenetic analysis of allopolyploids poses the problem that allopolyploids by definition have evolved by reticulation. The use of sequence data for phylogenetic analysis may solve part of the problem but other problems may arise. Sequences from the nuclear genome of the allopolyploids may be extracted from each of the genomes derived from the progenitor species and, provided that these genomes have not interacted, it is possible to track the evolutionary history of each genome and hence of the allopolyploid. Data from the organellar genomes, usually the plastids, may provide additional information about the direction of the initial hybridization event creating the allopolyploid, e.g., point to the maternal (in most seed plants) or paternal (in conifers) parent. A method that efficiently and selectively amplifies genome specific sequences from allopolyploids is briefly described. Examples from the *Triticeae* (i.e., the genera *Hordeum*, *Triticum*, and *Hordelymus*) are used to discuss the potential strength and possible problems of the method related to phylogenetic analysis of allopolyploids.

Keywords: allopolyploidy; reticulate evolution; phylogeny; *Triticum*; *Hordeum*; *Hordelymus*

The generally accepted methods of phylogenetic analysis produce branching diagrams (cladograms), i.e. they assume that evolution is divergent. Reticulations can only be shown superimposed on the phylogenetic trees and are as such not hypotheses of relationships created through the phylogenetic analysis, but hypotheses made on the basis of external evidence. Hypotheses about allopolyploidization reside in evolutionary theory not in phylogenetic analysis. Allopolyploid species have by definition evolved through reticulation and hence, pose a special problem for phylogenetics. In the *Triticeae*, where the frequency of allopolyploids is generally assumed to be very high, this has seriously hindered the formation of phylogenetically based hypotheses about the origin of polyploid taxa, their potential progenitors instead being mostly hypothesized on the basis of genome analysis. However, genome analysis cannot be used for revealing phylogenetic relationships and has many other inherited weaknesses (SEBERG & PETERSEN 1998).

Evolution and phylogenetic analysis of polyploids

Polyploid taxa may arise either by the merger of two fully divergent nuclear genomes (allopolyploidy), by the doubling of a single nuclear genome (autopolyploidy), or by something in between (segmental allopolyploidy), thereby doubling all or nearly all nuclear genomic sequences (WENDEL 2000). In allopolyploids – the only type of polyploids that will be discussed here – the nuclear genomes are contributed by different donor species at the time of polyploidization. The differentiated nuclear genomes are united in the same nucleus, whereas the cytoplasmic genomes are generally inherited from only one of the two parents; maternally in most seed plants with the conifers as a notable exception (MOGENSEN 1996). It is to be expected that the descendant nuclear genomes in an allopolyploid are sisters (viz. are more closely related) to their counter-parts in the

individual donor species, rather than to each other (Figure 1A), whereas the cytoplasmic genome is expected to be sister to the cytoplasmic genome of the donor of the cytoplasm (Figure 1B). Ideally, speciation occurring after allopolyploidization is expected to result in the formation of topologically similar clades formed by each of the sequence copies composing the polyploid species.

However, the compelling simplicity of the above is dependent on truly independent sequence evolution of the individual sequences co-existing in the allopolyploid (WENDEL 2000). Any interaction such as homogenization or recombination may make extraction of a phylogenetic signal intractable. Hence, sequences which are especially prone to concerted evolution, e.g. ribosomal DNA sequences, are particularly unsuitable for reconstruction of polyploid phylogeny. Decay or deletion of one or more of the sequence copies in allopolyploids may also obscure the phylogenetic signal as may further duplication of one or more of the sequences.

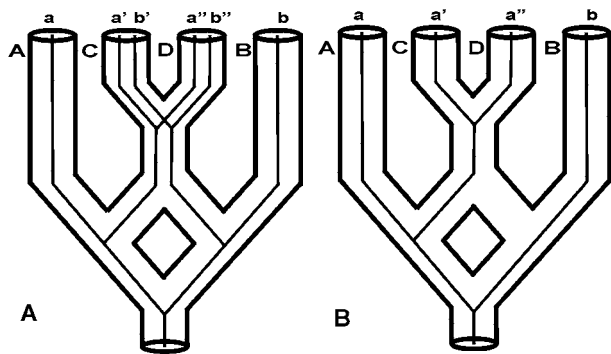


Figure 1. Theoretically expected topology of gene trees in polyploids that diverge following polyploidization. A, B, C, and D represent four extant species; A and B are diploids and C and D are polyploids that have diverged following an earlier polyploidization event between the progenitors of A and B. Nuclear genes: The two genes a' and a'' are more closely related to a than either is to b, indicating that a' and a'' diverged from a following the divergence of a and b. Similarly b' and b'' are more closely related to b than either is to a. a' and a'' and b' and b'' are paralogous, but the duplication event is a homology shared between species C and D. Plastid genes: The two genes a' and a'' are more closely related to a than either is to b, indicating that a' and a'' diverged from a following the divergence of a and b. As the plastids are – usually – maternally inherited species C and D never received a copy from the progenitor of B.

Ideally, phylogenetic analysis of polyploid species should include all diploid species belonging to the same monophyletic group of taxa as the polyploid(s) (e.g., as in the *Hordeum* case below). In practice, the number of diploids may have to be reduced, but at least all major supposedly monophyletic groups of taxa should be represented. An example would be inclusion of representatives of all monogenomic *Triticeae* in a phylogenetic analysis of a heterogenomic taxon (e.g., as in the *Triticum* and *Hordelymus* cases below). In phylogenetic analyses where the taxon sampling is reduced to include the “expected” progenitors of the allopolyploid(s), the results are immunized against the unexpected.

Phylogenetic analysis of sequence data from a particular locus results in a gene tree. However, a gene trees may for several reasons differ from the species tree (DOYLE & DAVIS 1998; WENDEL & DOYLE 1998) and it is generally assumed that the most precise approximation of the “true” phylogeny must be derived through analysis of all available evidence (total evidence) (KLUGE & WOLF 1993; NIXON & CARPENTER 1996; GRANT & KLUGE 2003). Any total evidence analysis must, as a minimum, include two sets of data, but it is not straightforward to perform a total evidence analysis including polyploid species. If the analysis includes two sets of sequences from two nuclear genes/loci from an allotetraploid the problem is how to match one sequence from one gene with one sequence from the other gene; in short which sequences stem from the same progenitor. Assuming that the two different loci originated from the same two parental species, separate phylogenetic analyses of each locus will ideally result in similar sister group relationships of each the two sequence copies from each locus. Hence, sequence copies showing similar sister group relationships can be combined accordingly (PETERSEN & SEBERG 2004). Adding a plastid data set (where the allotetraploid of course only contributes one sequence) to the analysis presents yet another problem. Provided that the plastid sequence originated from the same species as one of the nuclear sequences, separate analyses of the nuclear data and the plastid data ideally show the same sister group relationships of these sequences. Consequently, these sequences can be combined, whereas the unmatched nuclear sequences by necessity must be combined with a sequence scored entirely as missing data (PETERSEN & SEBERG 2004).

As this approach for matching sequences from one locus with sequences from another locus relies on individual gene trees, it is evident that any deviations of these from the true – but unknown – species tree are potential sources of error.

Phylogeny of polyploid *Triticeae*

In the *Triticeae* phylogenetic hypotheses have usually been constructed either only for the diploid or monogenomic taxa, or polyploid species have been included in analyses based on molecular data from the plastid genome, hence, at best revealing only one of the progenitors of the polyploids. KELLOGG (1989), FREDERIKSEN and SEBERG (1992), and SEBERG and FREDERIKSEN (2001) made phylogenetic analyses based on morphological characters for the monogenomic genera and depicted the potential origin of the heterogenomic (polyploid) genera as reticulations mapped onto the cladogram of the diploids. This approach assumes that the inferred history as polyploids is correct. Actual inclusion of polyploid taxa in phylogenetic analyses have been done by numerous authors, e.g., SEBERG and LINDE-LAURSEN (1996), PETERSEN and SEBERG (1998), and MASON-GAMER *et al.* (2002), but the data in these analyses were only from the plastid genome limiting the derived hypotheses to statements about the maternal progenitor (CORRIVEAU & COLEMAN 1988).

Only recently have phylogenetic analyses included data from each of the genomes presumed to compose the total genome in selected heterogenomic taxa. MASON-GAMER (2001, 2004) and HELFGOTT and MASON-GAMER (2004) studied the origin of North American allopolyploid species of *Elymus*, BLATTNER (2004) studied all polyploid species of *Hordeum*, whereas PETERSEN and SEBERG (2004) explored the origin of the two allotetraploids *H. capense* and *H. secalinum*, and PETERSEN *et al.* (unpubl.) the origin of allotetraploid and allohexaploid wheats, *Triticum turgidum* and *T. aestivum*, respectively. Whereas Blattner used ITS sequences, which not only are subjected to an unknown degree of concerted evolution but also present problems related to occurrence of paralogs even in the diploids species, both Mason-Gamer and Helfgott and Petersen *et al.* used single-copy nuclear sequences. Nevertheless, a fundamental methodological difference with far reaching implications for the subsequently obtained results distinguishes the works of Mason-Gamer and Helfgott from the works of Petersen *et al.* This difference is explored in detail below.

PCR recombination

Sequences brought together in allopolyploids are usually not widely different as the parental species most often will be closely related. Hence, due to sequence similarity of the primer sites, PCR amplification of nuclear sequences will usually result in simultaneous amplification of all sequence copies. Cloning of the amplification products is needed to obtain pure sequences (MASON-GAMER 2001, 2004; HELFGOTT & MASON-GAMER 2004). However, when the amplified sequences mutually have a high degree of similarity the risk of PCR recombination is imminent. PCR recombination occurs when incompletely extended PCR products anneal to a highly similar but non-identical DNA template and become extended to completion in subsequent rounds of PCR (SAIKI *et al.* 1988; CRONN *et al.* 2002) (Figure 2). Recombinant molecules or chimeras can be propagated in subsequent rounds of PCR, and they may even be subjected to additional recombination.

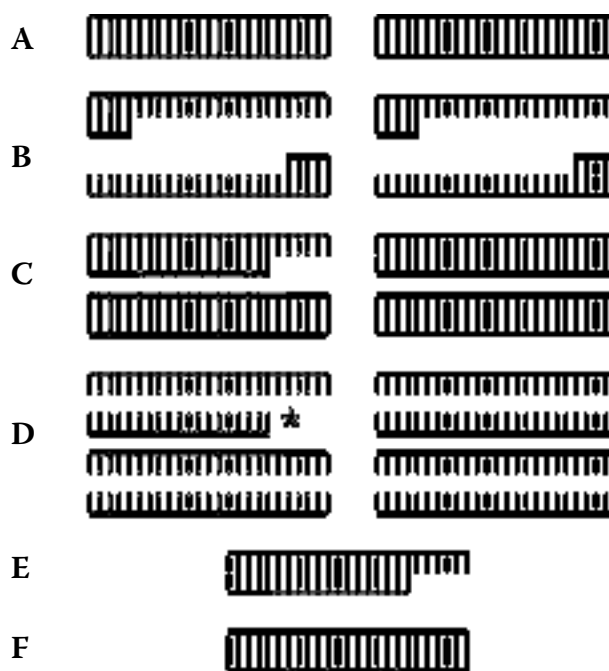


Figure 2. Formation of a chimeric DNA fragment through PCR recombination. A. The DNA sample contains a mix of two very similar but not identical templates (red and black). B. The blue primers anneal to both templates. C. During extension one strand marked with an * is not completed. D–F. In the next PCR cycle the incomplete strand (black) anneals to a similar but not identical template (red). During extension a chimeric strand of DNA is produced, and this may be propagated in subsequent PCR cycles

The magnitude of the problem depends on several factors including the degree of similarity between the initial templates, the length of the amplified product, the extension time, and the use of PCR additives. Whereas the first factor is beyond the control of the researcher the latter factors can be modified to reduce the problem. Long PCR products are more prone to recombination than small, but even with products only 242 bp long, YANG *et al.* (1996) observed that more than 5% of the products were PCR recombinants. In a small experiment we mixed DNA samples from two species of *Hordeum* (*H. vulgare* and *H. marinum*), and after PCR amplification of a 633 bp long fragment of *rbcL* (< 1% difference [5 bp] between *H. vulgare* and *H. marinum*) we observed that three out of four cloned products were chimeric. In allotetraploid *Gossypium*, CRONN *et al.* (2002) amplified four different nuclear genes varying in length from 860 bp to 4.050 bp and observed an average of 31% recombinant products. Individual amplification reactions produced from 0% to 89% chimeric products. JUDO *et al.* (1998) demonstrated that extension times up to 6 min per kb reduced the recombination frequency and SHAMMAS *et al.* (2001) showed that PCR additives such as dimethylsulphoxide (DMSO) and betaine could have a similar effect. None of the methods eliminates PCR recombination, but only reduces its frequency.

Hence, the problem of PCR recombination cannot be neglected when trying to amplify nuclear sequences from polyploid taxa. In experimental designs with mixed DNA samples the individual sequences can be precisely determined beforehand and the degree of recombination can be directly assessed. Thus, any measure taken to reduce the recombination frequency can also be assessed. However, in allopolyploids the sequences assumed to be present in the genome are *a priori* unknown. Obviously, the likelihood that they have diverged from the sequences in their ancestors increases with the age of the polyploid. Hence, *a priori* attempts to identify and discard recombinant products (HELFGOTT & MASON-GAMER 2004) reside in preconceived ideas about sequence similarity and the possible origin of sequences.

The MAMA technique

To avoid PCR recombination another technique based on sequence specific primers can be used to obtain clean sequences from polyploid species:

As previously indicated, sequences of a polyploid cannot be known *a priori*; consequently, PCR amplification is initially carried out under standard conditions and the products are directly sequenced. Contrary to normal practice it is desirable to obtain a "dirty" or mixed sequence – a chromatogram with clear indications of presence of more than one sequence. This sequence may include polymorphic sites and regions where reading may be obscured or difficult due to length mutations. Provided that the obtained sequence is still of a sufficiently high quality to allow reading, the polymorphic sites and possibly even smaller length mutations can be directly identified and subsequently used in the design of sequence specific primers. If the reading of the obtained mixed sequences is not possible, cloning of the products may be necessary, but the cloned sequences are only used for identifying possible variable sites, which can be used for the subsequent design of sequence specific primers. As an example of sequence specific primers imagine a "dirty" sequence from an allotetraploid showing both a C and a G at the same site. This site may be used for constructing two potentially sequence specific forward primers: one ending with a C, another ending with a G corresponding to the polymorphism at this site.

It is often assumed that the ultimate base (the 3'-end) of a primer needs to match perfectly for PCR amplification to occur. However, amplification frequently occurs even if this requirement is not met. Hence, mismatch of the ultimate base of a primer is not sufficient to ensure primer specificity. Instead a very efficient way of constructing sequence specific primers, the MAMA (Mismatch Amplification Mutation Assay) technique, has been developed by CHA *et al.* (1992). This technique deliberately includes an extra mismatching base at the ultimate position of the primer. Hence, imagine that both of the above sequences include a T after the C/G position. Following the MAMA technique one sequence specific primer should end with CA, CC, or CG, the other with GA, GC, or GG. Each primer will have one mismatch (the ultimate base) compared to one of the two sequences, but two mismatches (the ultimate and the penultimate bases) compared to the other sequence. Each primer is only able to amplify the sequence with which it shares only one mismatch.

In our experience one sequence specific primer in combination with a non-sequence specific primer is usually sufficient to ensure PCR amplification

of only one sequence at a time. However, do not expect a 100% success – sometimes amplification fails entirely, sometimes both sequences are still amplified, but varying the PCR conditions may produce the desired result. Allotetraploid species with only two sets of sequences are obviously easier to deal with than taxa with higher levels of ploidy. The initially obtained mixed sequence from a higher level polyploid may be too “dirty” for direct reading and identification of variable sites, and most sites may only distinguish one sequence from the remaining two or more sequences. Hence, several rounds of PCR and design of potential sequence specific primers are often needed. Below we describe three cases in which we have successfully applied the MAMA technique to amplify genome specific sequences from allotetraploid and allohexaploid *Triticeae* species.

Hordeum capense and *Hordeum secalinum*

PETERSEN and SEBERG (2003) made a phylogenetic analysis of the diploid species of *Hordeum* based on two nuclear genes, one plastid gene, and plastid RFLP data. Subsequently, the nuclear genes, DMC1 (disrupted meiotic cDNA) and EF-G (translation elongation factor G), were sequenced in the two supposedly closely related allotetraploid species *H. capense* and *H. secalinum* in order to explore their origin (PETERSEN & SEBERG 2004). Initially, cloning of the mixed PCR products for the two tetraploids obtained using the same primers as used for amplifying DMC1 sequences in the diploid species was attempted. However, the cloned products were virtually all different and we suspected that recombination between the almost 99% similar sequences had occurred. The cloned sequences were instead used to design sequence specific primers, which first were constructed by letting only the ultimate base match a variable site. Using these new primers for PCR and subsequent direct sequencing yielded two clean sequences from each of the tetraploid species of *Hordeum*. None of the sequences showed any signs of heterozygosity.

Due to our increased awareness of the problems related to PCR recombination and with knowledge about the MAMA technique it became possible to avoid the tedious step of cloning entirely in order to obtain the EF-G sequences from the two tetraploids. The mixed sequences obtained in the initial PCR amplification were directly sequenced, and the “dirty” sequence used to design sequence specific primers. These primers were designed according

to the MAMA technique and they proved more efficient in amplifying the individual sequences than the “traditional” primers previously used for amplifying the DMC1 sequences.

Phylogenetic analyses combining the newly obtained sequences from the polyploid species with the sequences from the diploid taxa were performed for each of the genes. The gene trees derived from these analyses were not entirely congruent. One sequence from both *H. capense* and *H. secalinum* was placed in a well supported group together with *H. marinum* ssp. *gussoneanum* on both gene trees, but the other set of sequences were placed either as a well supported sister group to *H. brevisubulatum* (on the DMC1 gene tree) (PETERSEN & SEBERG 2004: Figure 1) or in a large unresolved clade, which however, did not include *H. brevisubulatum* (on the EF-G gene tree) (PETERSEN & SEBERG 2004: Figure 2). The latter clade was only very poorly supported and inclusion of *H. brevisubulatum* would require only one extra step. Hence, to perform a combined analysis we had to assume that the “non-marinum”-like sequences belonged to the same parental genome. A combined analysis confirmed *H. marinum* ssp. *gussoneanum* as one of the parental species and *H. brevisubulatum* as the other (Figure 3). Plastid *rbcL* sequence data showed that *H. brevisubulatum* most likely was the female parent (PETERSEN & SEBERG 2004: Figure 3).

The origin of wheat

Hexaploid wheat, *Triticum aestivum*, with the genomic formula A^uA^uBBDD has been a classical text book example of evolution of a crop species through allopolyploidization. Whereas little doubt has existed about the origin of the A^u genome and the D genome the origin of the B genome has been more enigmatic. The traditionally envisioned evolutionary scenario starts with the formation of the allotetraploid *Triticum turgidum*, A^uA^uBB , combining the genomes of diploid *T. urartu*, A^uA^u , and an unknown diploid potentially closely related to *Aegilops speltoides*, SS. Subsequent hybridization between *Triticum turgidum* and diploid *Aegilops tauschii*, DD, resulted in the allohexaploid, common wheat.

In order to elucidate the origin of hexaploid wheat in a phylogenetic context and finally resolve the relationships of the B genome we used an approach similar to the above (PETERSEN *et al.* unpubl.). Accessions of hexaploid *Triticum aes-*

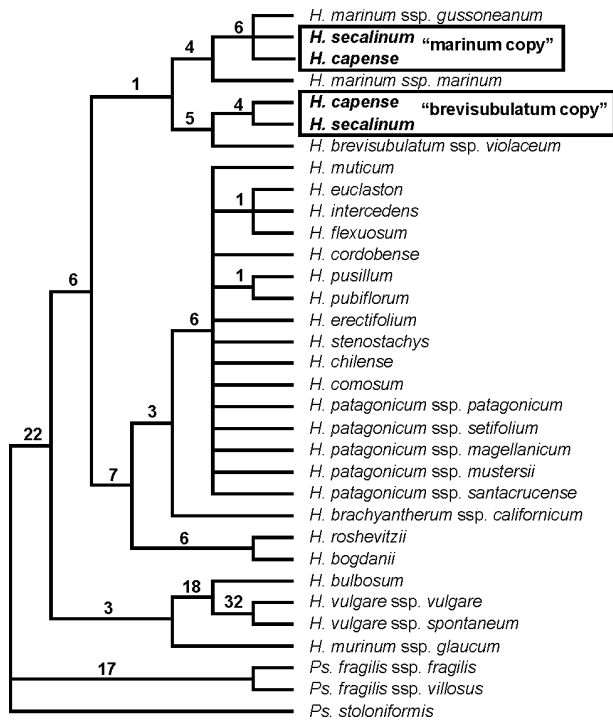


Figure 3. Strict consensus tree derived from combined analysis of sequences from two nuclear genes, DMC1 and EF-G, and one plastid gene, *rbcL*, for the diploid species of *Hordeum* plus the tetraploids *H. capense* and *H. secalinum*. Numbers above branches are Bremer support values. H. = *Hordeum*, Ps. = *Psathyrostachys*. Modified from PETERSEN & SEBERG (2004 – Figure 4)

tivum, tetraploid *T. turgidum*, all diploid species of *Triticum* and *Aegilops*, and representatives of all other monogenomic *Triticeae* were subjected to phylogenetic analyses based on two different nuclear genes and one plastid gene. The two nuclear genes were the same as used in the above mentioned *Hordeum* study, but in this case we were able to avoid cloning of both genes and all sequence specific primers were constructed according to the MAMA technique. As we predicted that tetraploid wheat, which only combines two genomes would be easier to work with than hexaploid wheat, we sequenced the tetraploid accessions first. Having successfully amplified two clean sequences from the tetraploids, the specific primers constructed for the A^u and B genomes were subsequently used for PCR amplification of hexaploid wheat. Given the assumed recent origin of *Triticum aestivum* (Cox 1998) we predicted that the A^u and B genome sequences of this hexaploid would be almost identical to the sequences of the tetraploids. Hence, according to our expectation each of the genome specific primers either amplified a clean sequence

from *T. aestivum* or amplified two sequences simultaneously when a primer also matched the D genome. In the latter case new sequence specific primers had to be constructed. In this manner we managed to successfully extract three different sequences from the hexaploid wheat.

As in *Hordeum*, the EF-G and DMC1 based gene trees were not congruent, but the sister group relationships of the three wheat genomes were consistent (PETERSEN *et al.* unpubl.). One set of sequences present in both tetra- and hexaploids was sister to *Triticum urartu* (the A^u genome), another set present in both tetra- and hexaploids was sister to *Aegilops speltoides* (the B genome), and a sequence only present in the hexaploid was sister to *Ae. tauschii* (the D genome). A combined analysis of the nuclear genes confirmed the relationships of the three genomes (Figure 4) and phylogenetic analysis of plastid *ndhF* sequences revealed *Ae. speltoides* as the plastid donor (PETERSEN *et al.* unpubl.). Hence, the hitherto accepted evolutionary history was confirmed and the identity of the B genome donor could finally be settled. Interestingly, the gene trees also suggested that *Ae. speltoides* is not closely related to the other S genome species of *Aegilops* (neither is it to any other species of *Aegilops*), and the genome of *Ae. speltoides* could more appropriately be assigned as B.

Hordelymus

LÖVE (1984) considered the monotypic genus *Hordelymus* an allopolyploid consisting of the H and T genomes derived from *Hordeum* and *Taeniatherum*, respectively. However, much of the recent literature considers the relationships between the genomes of *Hordelymus* and other *Triticeae* genomes as unknown (WANG *et al.* 1996) though BOTHMER *et al.* (1994) suggested a relationship to *Taeniatherum* and *Psathyrostachys*. Surprisingly, *Hordelymus* has never been included in a phylogenetic analysis and GenBank does not include a single sequence from the genus.

To elucidate the origin of *Hordelymus* we have sequenced two plastid genes, *rbcL* and *ndhF*, and presently we are working on sequences from two nuclear genes, DMC1 and EF-G, also used in the two above mentioned studies. Methodologically, the sequences from both genes are being extracted in exactly the same way as the sequences of the tetraploid wheats. We have recovered the predicted two copies of each gene and not observed any further polymorphism that could be due to heterozygosity.

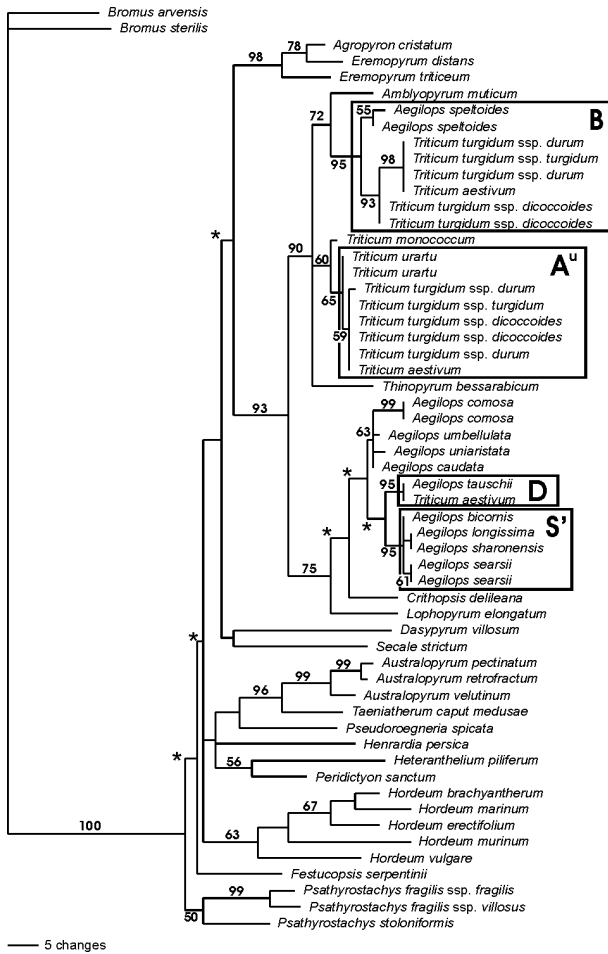


Figure 4. One of six equally parsimonious trees derived from combined analysis of nuclear DMC1 and EF-G sequences for the diploid *Triticeae* plus tetraploid *Triticum turgidum* and hexaploid *T. aestivum*. Branches that collapse in the strict consensus tree are marked with an *. Numbers above or below branches are jackknife proportions

However, preliminary phylogenetic analyses of the data from *Hordelymus* together with data from all diploid *Triticeae*, do not reveal as clear a picture as in the previous cases. The plastid data very strongly supports a sister group relationship to *Psathyrostachys* with the *Psathyrostachys* + *Hordelymus* clade being the sister group to the remaining *Triticeae* (Figure 5). On the DMC1 gene tree one sequence copy is placed as a sister group to *Psathyrostachys*, whereas the other copy is placed in an unresolved trichotomy which in addition includes the *Psathyrostachys* + *Hordelymus* clade and a clade including *Peridictyon*, *Heterantherium*, *Henrardia*, *Pseudoroegneria*, *Taeniatherum*, and *Australopyrum* (Figure 6). However, sequences from the EF-G

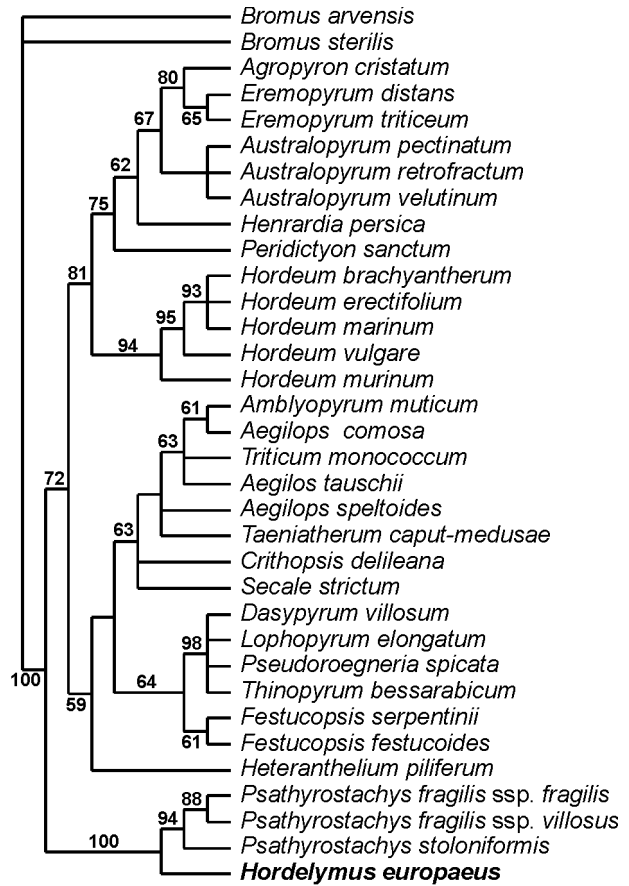


Figure 5. Strict consensus tree derived from combined analysis of plastid *rbcL* and *ndhF* sequences data for the diploid *Triticeae* plus *Hordelymus*. Numbers above or below branches are jackknife proportions

gene strongly disagree with those relationships: one EF-G sequence copy is the sister group to all other *Triticeae*, the other copy is the sister group to a clade including *Agropyron* and *Eremopyron* (Figure 7).

The observed level of incongruence is not at all unique in the *Triticeae* (e.g., KELLOGG *et al.* 1996; SEBERG & PETERSEN, in press), but it creates a serious problem for combining the data in a total evidence analysis. How can the different nuclear sequences from *Hordelymus* be matched when all sequence copies apparently have a unique origin? Based on the assumption that the DMC1 and EF-G sequences do have a common origin one approach could be to combine sequences such that topological incongruence is minimized. But how certain can we be about a common origin? The two DMC1 sequence copies are far more similar to each other and topologically much closer placed than the EF-G sequence copies. Could homogenization of

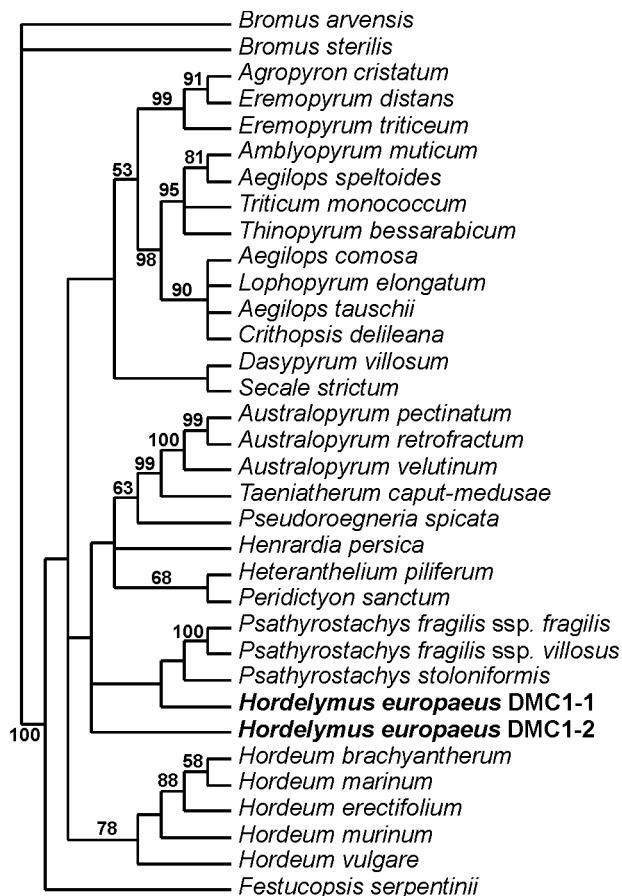


Figure 6. Strict consensus tree derived from analysis of nuclear DMC1 sequence data for the diploid *Triticeae* plus *Hordelymus*. Numbers above or below branches are jackknife proportions

the DMC1 copies have occurred obscuring their origin? Or could it be that even though the copies do not appear as sister groups on the DMC1 gene tree that they are paralogs created by duplication after the formation of *Hordelymus*? But if so, what has happened to the other "original" sequence copy? Was it deleted, has it decayed, or did we for some reason not pick it up? This of course is pure speculation, but it may illustrate some of the problems we face in particular when working with presumed ancient allopolyploids.

CONCLUSION

The above mentioned three cases illustrate various levels of success in tracing the origin of allopolyploid *Triticeae* species. In all cases the experimental design is based on the use of sequence specific primers to produce the expected sets of sequences; two in the tetraploid species

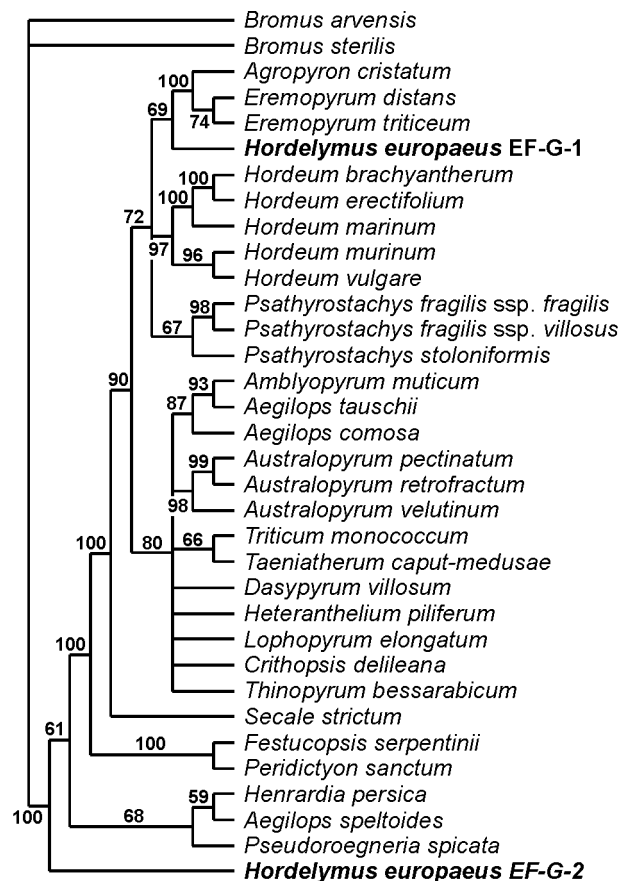


Figure 7. Strict consensus tree derived from analysis of nuclear EF-G sequence data for the diploid *Triticeae* plus *Hordelymus*. Numbers above or below branches are jackknife proportions

and three in the hexaploid species. However, it must be stressed that recovering the expected number of sequences is no guarantee that each of them originates directly from the progenitor species. The evolutionary interpretation of the cladograms was not straight forward in all instances. As evolutionary events potentially obscuring the phylogenetic signal may accumulate with time, it may become increasingly difficult to trace the origin of ancient allopolyploids. Some events like gene duplication may be hypothesized based on a gene tree, and further corroborated if more than the expected number of sequences are extracted from an allopolyploid. However, intergenomic recombination, homogenization (mimicking duplication), or any other interaction between the originally independent sequences cannot be directly traced on gene trees, and invoking any such event can never be more than an *ad hoc* explanation.

References

- BLATTNER F.R. (2004): Phylogenetic analysis of *Hordeum* (Poaceae) as inferred by nuclear rDNA ITS sequences. *Molecular Phylogenetics and Evolution*, **33**: 289–299.
- BOTHMER R. VON, LU B.-R., LINDE-LAURSEN I. (1994): Intergeneric hybridization and C-banding patterns in *Hordelymus* (Triticeae, Poaceae). *Plant Systematics and Evolution*, **189**: 259–266.
- CHA R.S., ZARBL H., KEHAVONG P., THILLY W.G. (1992): Mismatch amplification mutation assay (MAMA): application to the c-H-*ras* gene. *PCR Methods and Application*, **2**: 14–20.
- CORRIVEAU J.L., COLEMAN A.W. (1988): Rapid screening method to detect potential biparental inheritance of plastid DNA and results for over 200 angiosperm species. *American Journal of Botany*, **75**: 1443–1458.
- COX T.S. (1998): Deepening the wheat gene pool. *Journal of Crop Production*, **1**: 1–25.
- CRONN R., CEDRONI M., HASELKORN T., GROVER C., WENDEL J.F. (2002): PCR-mediated recombination in amplification products derived from polyploid cotton. *Theoretical and Applied Genetics*, **104**: 482–489.
- DOYLE J.J., DAVIS J.I. (1998): Homology in molecular phylogenetics: a parsimony perspective. In: SOLTIS D.E., SOLTIS P.S., DOYLE J.J. (eds): *Molecular Systematics of Plants II*. Kluwer Academic Publishers, Boston, Dordrecht, London: 101–131.
- FREDERIKSEN S., SEBERG O. (1992): Phylogenetic analysis of the Triticeae (Poaceae). *Hereditas*, **116**: 15–19.
- GRANT T., KLUGE A.G. (2003): Data exploration in phylogenetic inference: scientific, heuristic, or neither. *Cladistics*, **19**: 379–418.
- HELFGOTT D.M., MASON-GAMER R.J. (2004): The evolution of North American *Elymus* (Triticeae, Poaceae) allotetraploids: evidence from phosphoenolpyruvate carboxylase gene sequences. *Systematic Botany*, **29**: 850–861.
- JUDO M.S.B., WEDEL A.B., WILSON C. (1998): Stimulation and suppression of PCR-mediated recombination. *Nucleic Acids Research*, **26**: 1819–1825.
- KELLOGG E.A. (1989): Comments on genomic genera in the Triticeae (Poaceae). *American Journal of Botany*, **76**: 796–805.
- KELLOGG E.A., APPELS R., MASON-GAMER R.J. (1996): When genes tell different stories: the diploid genera of Triticeae. *Systematic Botany*, **21**: 321–347.
- KLUGE A.G., WOLF A.J. (1993): Cladistics: what's in a word? *Cladistics*, **9**: 183–199.
- LÖVE Á. (1984): Conspectus of the Triticeae. *Feddes Repertorium*, **95**: 425–521.
- MASON-GAMER R.J. (2001): Origin of North American *Elymus* (Poaceae: Triticeae) allotetraploids based on granule-bound starch synthase gene sequences. *Systematic Botany*, **26**: 757–768.
- MASON-GAMER R.J. (2004): Reticulate evolution, introgression, and intertribal gene capture in an allohexaploid grass. *Systematic Biology*, **53**: 25–37.
- MASON-GAMER R.J., ORME N.L., ANDERSON C.M. (2002): Phylogenetic analysis of North American *Elymus* and the monogenomic Triticeae (Poaceae) using three chloroplast DNA data sets. *Genome*, **45**: 991–1002.
- MOGENSEN H.L. (1996): The hows and whys of cytoplasmic inheritance in seed plants. *American Journal of Botany*, **83**: 383–404.
- NIXON K.C., CARPENTER J.M. (1996): On simultaneous analysis. *Cladistics*, **12**: 221–241.
- PETERSEN G., SEBERG O. (1998): Phylogeny and genomes in the genus *Hordeum*. In: JARADAT J.J. (ed.): *Triticeae III*. Science Publishers, Enfield: 109–118.
- PETERSEN G., SEBERG O. (2003): Phylogenetic analyses of the diploid species of *Hordeum* (Poaceae) and a revised classification of the genus. *Systematic Botany*, **28**: 293–306.
- PETERSEN G., SEBERG O. (2004): On the origin of the tetraploid species *Hordeum capense* and *H. secalinum* (Poaceae). *Systematic Botany*, **29**: 862–873.
- SAIKI R.K., GELFAND D.H., STOFFEL S., SCHARF S.J., HIGUSHI R., HORN G.T., MULLIS K.B., ERLICH H.A. (1988): Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science*, **239**: 487–491.
- SEBERG O., FREDERIKSEN S. (2001): A phylogenetic analysis of the monogenomic Triticeae (Poaceae) based on morphology. *Botanical Journal of the Linnean Society*, **136**: 75–97.
- SEBERG O., LINDE-LAURSEN I. (1996): *Eremium*, a new genus of the Triticeae (Poaceae) from Argentina. *Systematic Botany*, **21**: 3–15.
- SEBERG O., PETERSEN G. (1998): A critical review of concepts and methods used in classical genome analysis. *Botanical Review*, **64**: 373–417.
- SEBERG O., PETERSEN G. (in press): Phylogeny of the Triticeae (Poaceae) – based on three organelle genes, two single-copy nuclear genes, and morphology. In: COLUMBUS J.T., FRIAR E.A., HAMILTON C.W., PORTER J.M., PRINCE L.M., SIMPSON M.G. (eds): *Monocots: Comparative Biology and Evolution*, Vol. X. Rancho Santa Ana Botanic Garden, Claremont, California, USA.
- SHAMMAS F.V., HEIKKILA R., OSLAND A. (2001): Fluorescence-based method for measuring and determining

- the mechanisms of recombination in quantitative PCR. *Clinica Chimica Acta*, **304**: 19–28.
- WANG R.R.-C., BOTHMER R. VON, DVORAK J., FEDAK G., LINDE-LAURSEN I., MURAMATSU M. (1996): Genome symbols in the Triticeae (Poaceae). In: Wang R.R.-C., Jensen K.B., Jaussi C. (eds): Proc. 2nd Int. Triticeae Symp., Utah State University Publication, Design and Production, Logan: 29-34.
- WENDEL J.F. (2000): Genome evolution in polyploids. *Plant Molecular Biology*, **42**: 225-249.
- WENDEL J.F., DOYLE J.J. (1998): Phylogenetic incongruence: window into genome history and molecular evolution. In: SOLTIS D.E., SOLTIS P.S., DOYLE J.J. (eds): *Molecular Systematics of Plants II*. Kluwer Academic Publishers, Boston, Dordrecht, London: 265–296.
- YANG Y.L., WANG G., DORMAN K., KAPLAN A.H. (1996): Long polymerase chain reaction amplification of heterogeneous HIV type 1 templates produces recombination at a relatively high rate. *AIDS Research and Human Retroviruses*, **12**: 303–306.