

Inferring Genetic Relationship Among Haplomes in *Triticeae*: The Utility of the 5S DNA Units with Examples

B. R. BAUM

*Environmental Health program, Biodiversity Theme, Agriculture & Agri-Food Canada,
Neatby Building, Ottawa, Ontario, Canada, K1N 9M5, e-mail: baumbr@agr.gc.ca*

Abstract: In higher plants nuclear rRNA is encoded by multiple copies of rDNA genes, arranged in arrays of tandem repeats at one or more loci. Each unit comprises a coding region of ca. 120 bp and a non-transcribed spacer (NTS) that contains regulatory signals for transcription. In the *Triticeae* two separate major loci containing tandem arrays coexist, differentiated in the main by the length and nucleotide sequence of the NTS. Such groups of sequences were named unit classes. Unit classes, alone or in combination, and the differences between them are useful in representing haplomes of taxa within the *Triticeae*. Moreover, phylogenetic relationships among unit classes and the haplomes that they signify may be inferred from the nucleotide sequences of the NTS. Several examples will be presented and discussed along with some issues relating to multigene families and phylogenetic inference.

Keywords: cloning; sequence alignment; maximum likelihood; Bayesian analysis; evolutionary inference; unit classes; twat of orthology

The 5S DNA gene codes for the 5S ribosomal portion. The 5S DNA units in the *Triticeae* are organized in arrays of tandem repeats with the highly conserved genes separated by the more variable, non-transcribed spacer region (henceforth NTS). In a number of publications (e.g., BAUM & BAILEY 1997, 2000, 2001; BAUM & JOHNSON 1994, 1996, 1998, 1999, 2000, 2002, 2003; BAUM *et al.* 2001, 2003), we have described the molecular diversity of 5S DNA sequences in species within the genera *Elymus*, *Hordeum*, *Kengyilia*, and *Triticum* and based on their sequences classified the 5S DNA units into putative orthologous groups, which we called unit classes. In addition we found that we could assign the different unit classes to haplomes. For example, in *H. vulgare* L. we found sequences belonging to a unit class we labelled "short I1" to represent the I haplome identified in this taxon. Subsequent analyses (*ibid*) led to the assignment of unit classes to the other haplomes or four "basic genomes" in *Hordeum* (BOTHMER *et al.* 1986, 1987).

Studies by APPELS and BAUM (APPELS & BAUM 1992; BAUM & APPELS 1992) tentatively divided the

5S rRNA genes in the *Triticeae* into two types – the short type ranging in size from 327 to 468 base pairs (bp) and the long type ranging from 469–500 bp – and we initially adopted their terminology. As described in our previous publications (*ibid.*), duplications, insertions and/or deletions can have a profound effect on this simple division. The effect is actually so profound that the length alone does not necessarily determine the unit class or the assignment of a sequence to a particular unit class and haplome. It is the sequence divergence and the pattern of blocks of sequences that are crucial for their correct assignment (BAUM *et al.* 2001). Most taxa that have been investigated to date contain two unit classes per haplome but several exceptions have been identified. For instance in bread wheat, a hexaploid, we found five unit classes assignable to the three haplomes (BAUM & BAILEY 2001), the expected number would have been six; it may be that we failed to capture the sixth. We are able to use the sequence divergence of the 5S DNA NTS, and the assignment of unit classes to haplomes,

to infer phylogenetic relationships among the various haplomes.

MATERIALS AND METHODS

Cloning and sequencing. The materials investigated and the isolation of genomic DNA, PCR amplification of the 5S DNA genes, cloning of PCR products and sequencing of plasmid DNA have been described, e.g., in BAUM and BAILEY (1997, 2000, 2001), BAUM and JOHNSON (1994, 1996, 1998, 1999, 2000, 2002, 2003), BAUM *et al.* (2001, 2003, 2005). The PCR primers target the coding regions in tandem repeats and amplify a sequence starting from 5' from the BamH1 site within the transcribed region, through the NTS, to a site 3' of the BamH1 site within the adjacent unit in the array. Amplimers were either digested with BamH1, cloned into the BamH1 site of pUC19 (YANISCH-PERRON *et al.* 1985), and transformed into *Escherichia coli* strain DH5 α or latterly ligated directly into pGEM-T Easy (Promega Biotech) and transformed into DH5 α . For each sequence from hundreds of clones, both strands were sequenced.

Determination of putative orthology. Alignments and manual refinement were carried out as detailed in BAUM *et al.* (2001) to determine unit classes, i.e. putative orthologous groups. Based on BLAST (basic local alignment search tool, ALTSCHUL *et al.* 1990) searches, these classes were labeled to reflect known haplomes in *Triticeae*.

Test of orthology. To test for orthology of the units within unit classes Maximum Likelihood (ML) analysis was carried out using fastDNAm1[®] (OLSEN *et al.* 1994). The results have also been used to assess diversity among the units within each unit class. Prior to the ML analysis the alignments were subjected to likelihood ratio tests of 56 different evolutionary models (POSADA 2003; FELSENSTEIN 2004) to choose the best fitting model and parameters given the data in conjunction with PAUP (SWOFFORD 1998) version 4.0b10 and using MODELTEST (POSADA & CRANDALL 1998).

Phylogenetic inference among unit classes. Long H1, short I1, long H2 and long Y2 unit classes in *Hordeum*. First, selection of exemplar sequences was made from among the putative orthologous sets of sequences; ML analyses using PAUP and Bayesian analyses using MrBayes (HUELSENBECK & RONQUIST 2001) were then carried out.

Haplome relationships in *Triticeae*: Example (1) – Unit classes in *Hordeum*. Unit classes' data were

first summarized based on their presence/absence within taxa. A neighbor-joining (SAITOU & NEI 1987) analysis (NJ) was then performed and the tree was then rooted by a hypothetical ancestor, the choice of which is discussed below. The rooted tree was then subjected to a tree analysis using parsimony in order to obtain and describe the unit class changes on the NJ tree.

Haplome relationships in *Triticeae*: Example (2) – *Secale* and related haplomes in *Triticeae*. Conducted as in 4 above, i.e. by selecting exemplar sequences first and then conducting ML and Bayesian analyses.

RESULTS AND DISCUSSION

Long H1 and short I1 unit classes in *Hordeum vulgare*

Results from our initial analysis of a large numbers of clones isolated from *Hordeum vulgare* L. showed that the two classes recommended by APPELS and BAUM (1992), i.e., the long and the short types, vary so much in size that there is a substantial overlap between the two. For example the alignment in Figure 1 depicts the two unit classes and displays the variation between them. In this example some of the "short" units are actually longer than some "long" units in part due to the presence of (TAG) repeats within the NTS of the "short" units. In several publications we have extensively documented the effects of duplications, insertions and/or deletions of the length of the NTS that render the simple division of units into "short" and "long" difficult.

Furthermore, the combination of these two unit classes was found to be characteristic of the species containing the I haplome, viz. *Hordeum vulgare*, *H. spontaneum* C. Koch, and both diploid and tetraploid *H. bulbosum* L. (BAUM & JOHNSON 1996). The naming of the unit classes reflects the haplomes. Thus the short I unit class in this example is characterized by a contiguous chain of two to many TAG repeats (the top sequences in Figure 1). We found that the establishment, and thus recognition of classes, depends on the pattern of the sequences which may be revealed by careful alignment. We will return to the problem of the recognition of classes in the section "Detection of Orthology" below, but these results suggest that the sequences of the different NTSs can be used to define unit classes and to determine relationships among haplomes and species of the *Triticeae*.

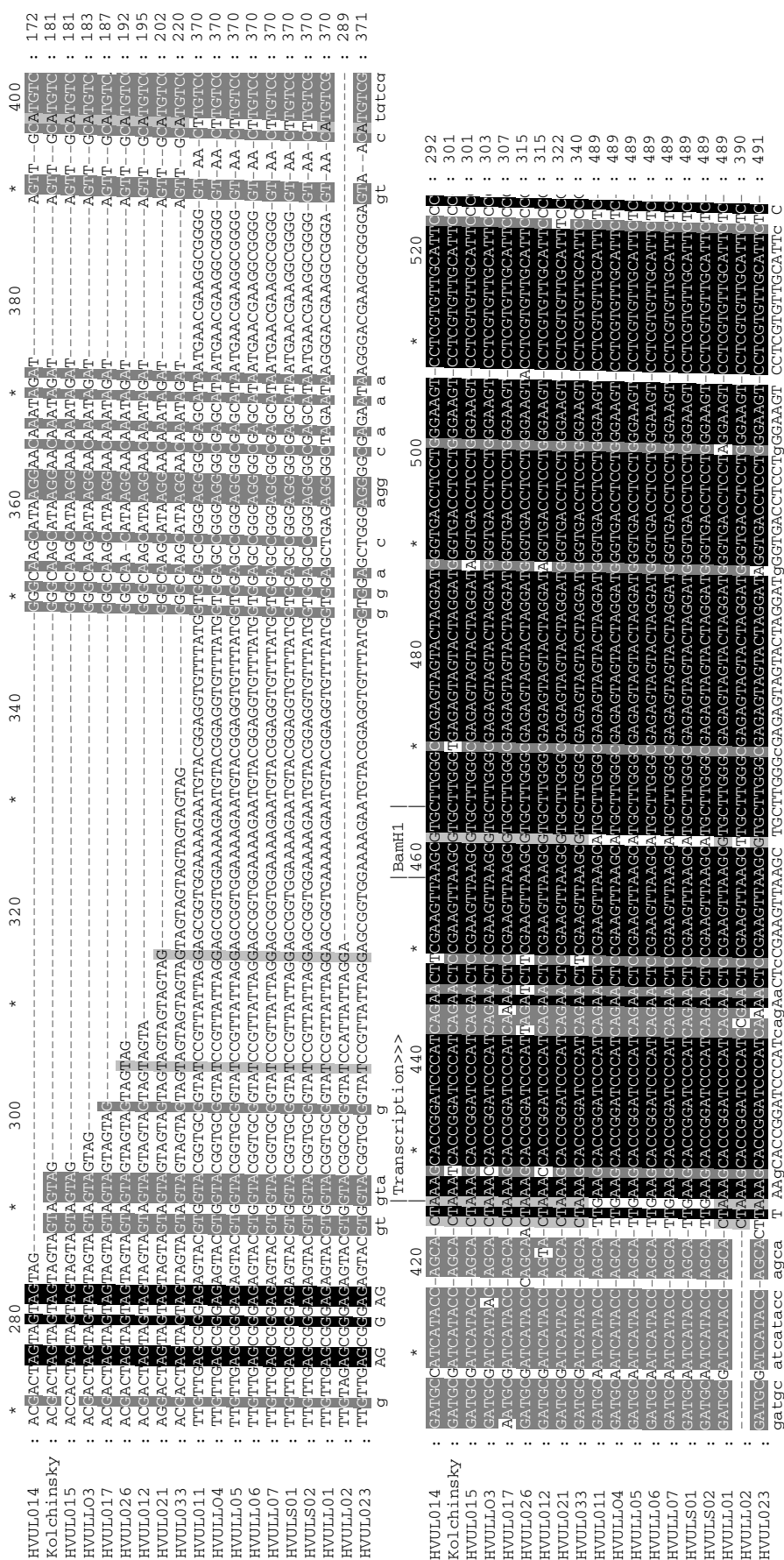


Figure 1. Alignment of selected 5S DNA sequences in *H. vulgare* depicting the differences between two putative unit classes: the short I1 (first 10 sequences from the top) and the long H1 (the remaining sequences). The likely start site for transcription and BamH1 site are noted

Long H1, short I1, long H2 and long Y2 unit classes in *Hordeum*

The South American diploid *Hordeum* species belong to the **HH** genome species (BOTHMER & JACOBSEN 1991). Based on 374 sequences of 12 taxa we found that two different unit classes characterize them, viz the long H2 and long Y2 (only a small fraction of the alignment is shown in Figure 2, where the two unit classes differ obtained after deleting the sequences from top and from bottom, for illustration only). Based upon LRT tests, the data best fit the HKY+G, i.e. the Hasegawa model (HASEGAWA *et al.* 1985) with the Gamma distribution rates of nucleotide substitutions (YANG 1994). ML analyses and various tests including the molecular clock, as well as Bayesian evolutionary inference analysis implied that the long H1 and short I unit classes found in the **II** genome diploids diverged from each other at the same rate as the long H2 and long Y2 unit classes found in the **HH** genome diploids (Figure 3). The divergence among the unit classes, estimated to be circa 7 MY, suggests that the genus *Hordeum* may be a paleopolyploid (BAUM *et al.* 2005). Figure 3 also depicts the test of orthology (see Discussion). Once more these results suggest that analysis of the NTS can be useful for investigating relationships between haplomes in *Hordeum*.

Haplome relationships in *Triticeae*: example (1) – Unit classes in *Hordeum*

The resulting NJ tree (Figure 4) is shown with the inferred unit class changes. This tree was rooted at a hypothetical ancestor containing both the long H1 and a long Y2 unit class, as no outgroup was contemplated. The key point here is that results based on the analysis of 5S DNA unit classes could be used to infer the evolutionary path among the *Hordeum* taxa and could bear directly on the relationship among the haplomes in the genus. An example of our analysis of the *Triticeae* tribe based upon the results from the time calibration analysis was recently presented (BAUM *et al.* 2005). *Hordeum* as we know it today was most likely different from the ancestral stock that may have originated at about the start of the drift of Africa from South America (at start of the Cretaceous). The discovery that *H. capense* (S. Africa) and *H. depressum* (N. America) are the only extant species found to contain both the long H1 and long Y2 unit classes

provides support for this hypothesis; subsequent analysis of the DMC1, EF-G and *rbcL* genes by PETERSEN and SEBERG (2004), also supports the idea that *H. capense* is an ancient relict. The several more *Hordeum* species currently being investigated may help solidify this interpretation.

Haplome relationships in *Triticeae*: example (2) – *Secale* and related haplomes in *Triticeae*

The *Secale* sequence analysis identified two unit classes, the long R1 and short R1. The test of orthology of these unit classes is depicted in the ML tree (Figure 5). A BLAST search for 5S DNA sequences from known unit classes most closely similar to the long R1 unit class contained sequences of the long P1 unit class from *Agropyron* (**PP** haplomes) and from *Kengyilia* (**StStYYPP** haplomes), long J1 from *Thinopyrum* (**J** haplome), whereas the search for sequences of the short R1 included the long S1 from *Pseudoroegneria* (**St** haplomes) and *Kengyilia* (**StStYYPP** haplomes), the short J1 from *Thinopyrum* (**J** haplomes) and the short V1 from *Dasypyrum* (**V** haplomes). ML and Bayesian analyses yielded a tree with the long R1 units where the long P1 and long J unit classes were closest to the R1 unit class, whereas they yielded a tree with the short R1 units where the S1 and short J1 unit classes were closest to the short R1 unit class. This result indicates a possible close relationship between the **St**, **J** and **R** haplomes (not shown) and again indicates how analysis of the NTS can be used for formulate hypotheses for future study.

Detection of orthology

Determination of unit classes. Central to this discussion is the detection of orthologous groups of sequences and their grouping into unit classes. The determination of putative orthologous groups of sequences is much more advanced for protein than for DNA sequences in part because orthology analysis is becoming an important aspect of gene function prediction. The use of phylogenetic information in genome annotation is known as phylogenomics (EISEN 1998). Conventional phylogenomics methodology employs mostly manual approaches; however, recent advances have been made in automating protein phylogenomics, based on similarity clustering, such as the COGs database (TATUSOV *et al.* 2001). Recently, attempts have been made to use explicit phylogenetic tree

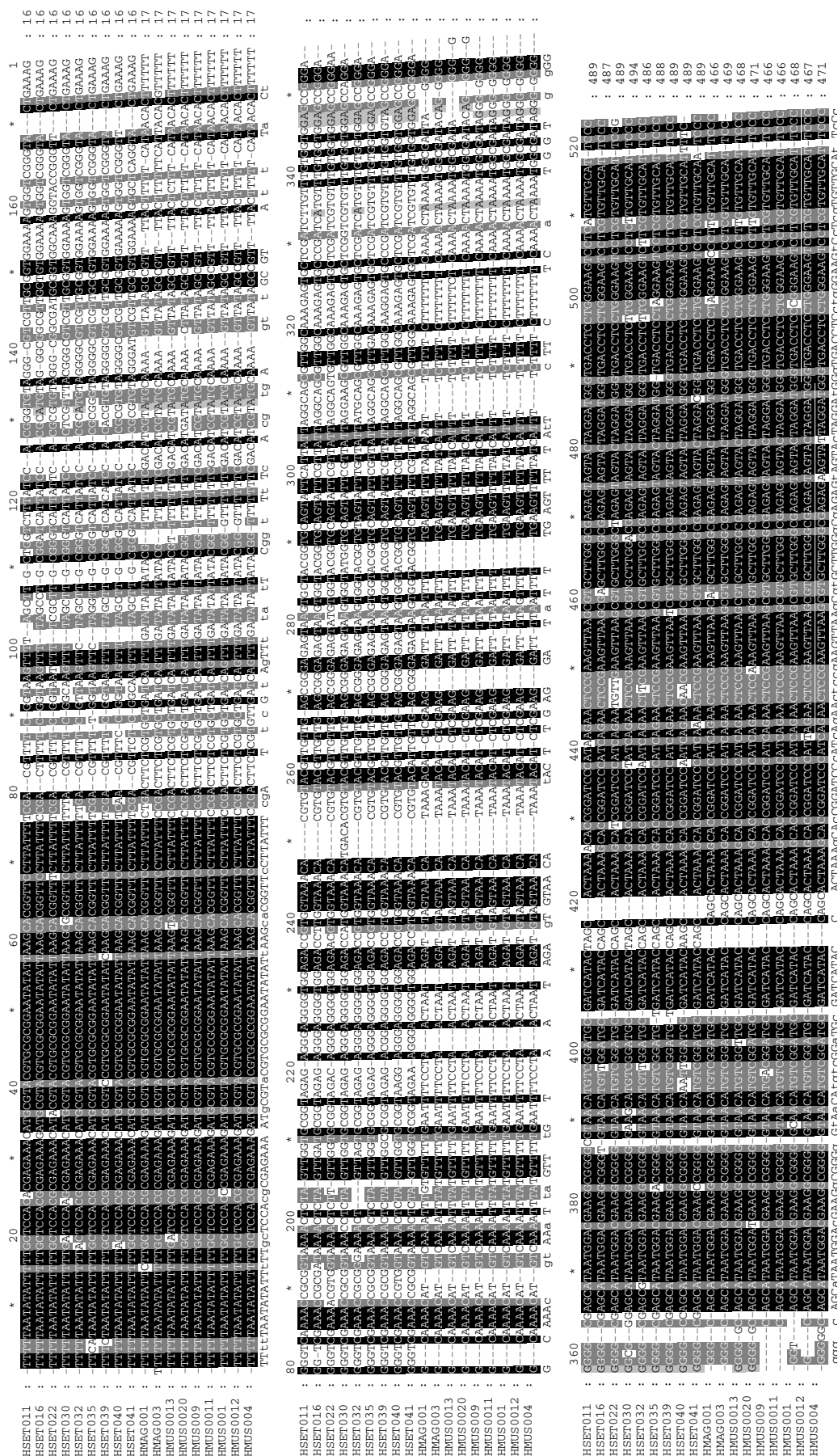


Figure 2. Window in the alignment of 374 units found in the South American native diploid *Hordeum* species, at the demarcation line between the long H2 (top 9) and the long Y2 (bottom 9) unit classes sequences. Both unit classes are present in all the species, however it is by coincidence that in this window the long H2 units are from *H. setifolium* and the long Y2 units are from *H. patagonicum* ssp. *magellanicum*

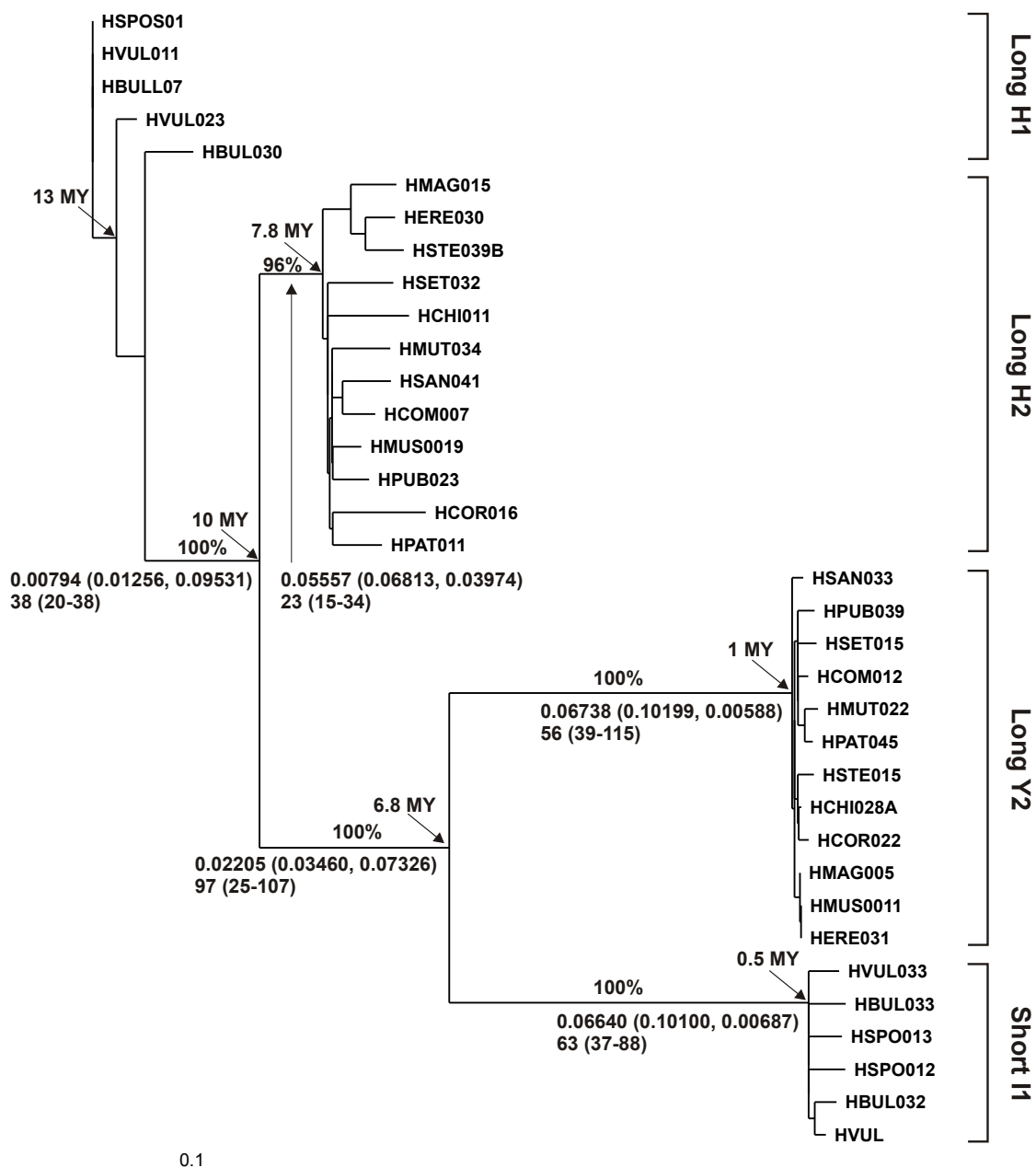


Figure 3. Best maximum likelihood tree, the molecular clock, parsimony trees superimposed on each other, of exemplars of the following unit classes in *Hordeum*, long H1, short I1, long H2 and long Y2. The first two are found in the **I** haplome species whereas the last two in the South American diploid **H** haplome species. MY – Million Years since divergence; values above major branches – bootstrap support (%); values below major branches: branches lengths (distance from root, distance from tip) and values further below – assigned branch length under the parsimony criterion (Min. possible length–Max. possible length); scale bar – distance from minimum evolution distances; on the right the four unit classes. For example, in the most basal branch: major branch length = 0.00794 (distance from root = 0.01256, distance from tip = 0.09531); and further below on the same branch: branch length = 38 base changes (minimum length changes = 20 and maximum branch changes = 38)

analysis instead to place or classify a sequence in a subfamily or group of a gene tree of known sequences. See for example the protein sequence analyses of ZMASEK and EDDY (2002) and ARVESTAD

et al. (2003). Conventionally in protein analysis one uses BLAST or similar programs. When one encounters a sequence that is not similar to a previously known group a new group is created. We



Figure 4. Relationships among all the unit classes thus far investigated in the genus *Hordeum*. The scale bar indicates the distance in the NJ tree. The unit classes were obtained from the analysis of the NJ tree by maximum parsimony of the unit class presence/absence data. See text

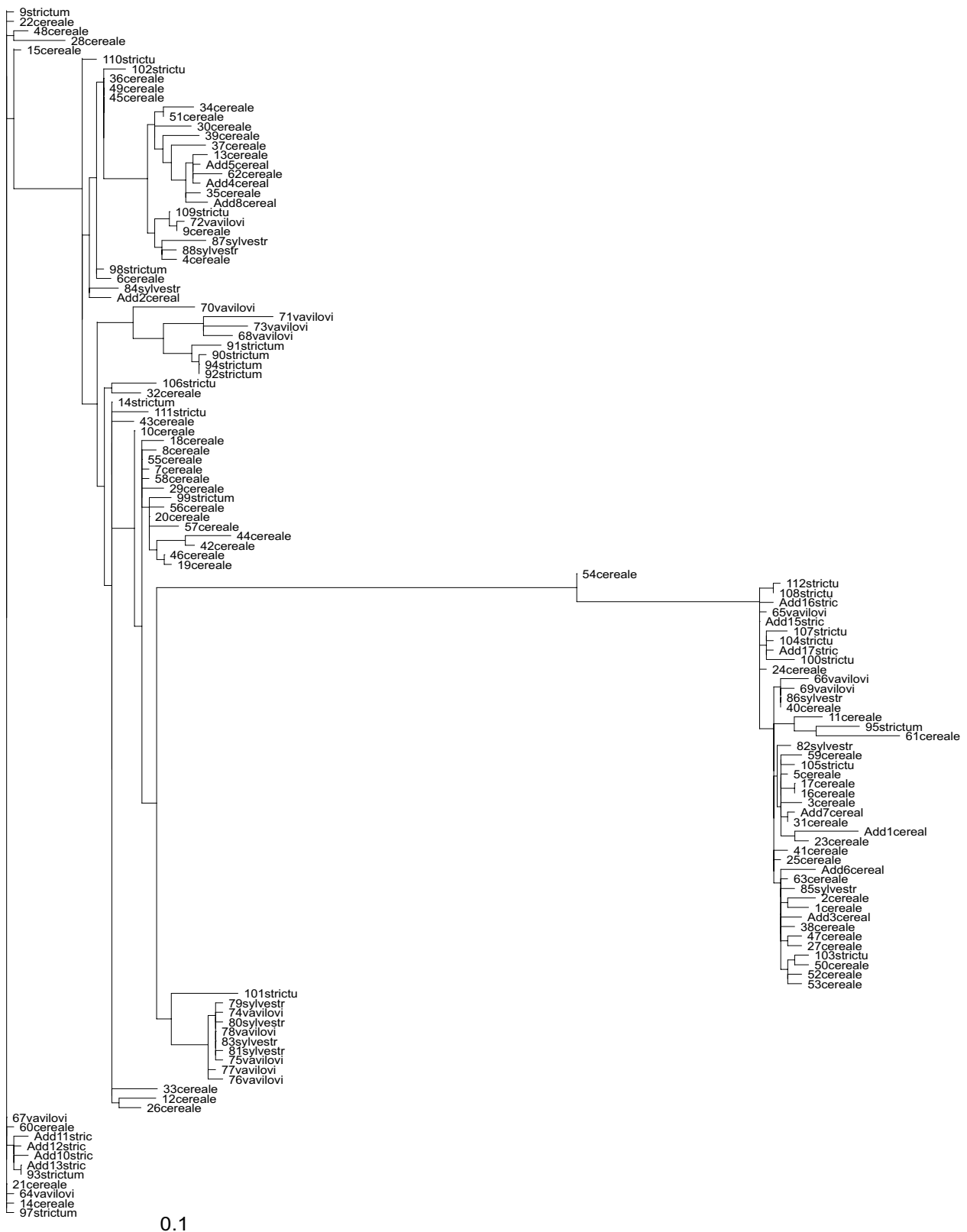


Figure 5. Test of orthology of the two unit classes in *Secale*. The sequences at the far right on the long branch were assigned to the long R1 unit class, whereas the left at the base of fastDNAMl tree were classed as the short R1 unit class

have taken the same manual and conventional approach to determine unit classes in the *Triticeae* (BAUM *et al.* 2001), except that we validate them *a posteriori* by phylogenetic analysis (as illustrated in Figures 3 and 5).

With respect to DNA sequences orthology analysis is first based on the phylogeny of the sequences, and not with respect to function, allowing the use of either coding or noncoding regions or both. This is usually done manually although attempts have been made to combine alignments with phylogenetic analysis in one step, such as the POY (Phylogeny Reconstruction via Optimization of DNA and other Data) program (WHEELER *et al.* 2003) for which the methods are based on WHEELER (1996, 1999). This method does not yet carry out orthologous analysis and, being based upon parsimony, may therefore yield erroneous results. As far as I know, no attempts were made to automate the classifying of multigenes into orthologous sequences. One needs a species tree to rigorously identify orthologous genes, but it is impossible to find a species tree unless the orthologous sequences of the species are known. With one gene we obtain a gene tree, not a species tree. In a multigene family, such as the 5S rDNA the situation is more complicated as there are different unit classes of DNA units which are paralogous with respect to each other.

Wendel and associates in an excellent review on the utility of nuclear genes for phylogeny reconstruction (SMALL *et al.* 2004) emphasized the necessity of cloning prior to sequencing, as we recommended for the *Triticeae* (BAUM *et al.* 2001) but only when polymorphism is detected at the gene amplification stage by PCR for orthology assessment. They did not take into consideration that sequence polymorphism may occur even when the PCR products appear uniform. They also advocated the use of BLAST as one of the steps in the assessment of putative orthology, as we had done. When PCR amplification reveals two or more types (bands on a gel), then clearly direct sequencing of the unpurified PCR products is not realistic. It is less well recognized that sequence polymorphism may occur even when the PCR products appear uniform in size (BAUM *et al.* 2001) and that cloning of PCR products remains necessary in this case too. SMALL *et al.* (2004) also recommended developing locus specific primers once "types" had been defined. Although we have successfully used such probes for the analysis of different unit classes via FISH (BAUM *et al.* 2004), we advocate sequencing of many clones in order to

establish putative orthology classes and to provide strong support for them.

As described above, multiple 5S rDNA unit classes are seen in the grasses. Wendel and associates found only one putative orthologous group of sequences per haplome in *Gossypium*, perhaps because of the nature of this genus, or because of gene loss due to a deletion or a failure to sequence enough samples. Sufficient sampling remains a vexatious problem (BAUM *et al.* 2001) for which we have no solution.

ML analyses

Testing orthology and haplome divergence. Whether sequence alignment it is carried by automatic means or by a combination of programs such as BLAST and alignment programs followed by manual refinement, as is conventionally carried out, it is obviously the most important step in determining putative orthology. This is as true for single copy genes as it is for multicopy genes such as the 5S DNA unit classes in the *Triticeae*. Tests of orthology rely on phylogenetic analysis of the putative orthologs, e.g., unit classes in *Triticeae*. Parsimony, although useful under certain conditions, lacks an explicit model of evolution (GOLDMAN 1990). In recent years great progress has been made especially in ML algorithms. ML methods "allow both a wide variety of phylogenetic inferences from sequence data and robust statistical assessment of all results" (WHELAN *et al.* 2001). The authors went so far as to express the opinion that "it cannot remain acceptable to use outdated data analysis techniques when superior alternatives exist" (*ibid*) as some have done.

In the tests of putative orthology the DNA sequences that belong to the same unit class were mostly, if not all, found on small branches of the tree compared to the much longer branches which subtended the "clusters" of the orthologs, i.e. the unit classes (Figures 3 and 5 for example). To carry out the relationship among the groups of orthologs we first selected exemplar sequences from each orthologous group (unit class) and then subjected the data to tests of fitting the substitution model from among the different models of evolution so far defined; and then subjected the data to ML analyses with the parameters of the models. An assessment of the robustness of the resulting trees was achieved by non-parametric bootstrapping (FELSENSTEIN 1985; SWOFFORD *et al.* 1998) and including Bayesian analysis (HUELSENBECK & RONQUIST

2001). Using this approach we can estimate the phylogenetic relationships among haplomes in the *Triticeae*, in other words haplome relationships can be estimated by the phylogenetic relationships among the unit classes and is thus also achieved with strong statistical support. An example of this procedure was described above for the **I** haplome and **H** haplome diploid *Hordeum* species.

Total evidence versus supertrees

The different unit classes in the *Triticeae*, i.e. the different groups of orthologs are paralogous groups. They need to be combined for a global analysis. Inferring phylogeny relationships by analysing combined data of different kinds, e.g. morphology and gene sequences, sequences from different genes, DNA-DNA hybridization with DNA sequences or serological data with any of these or any combination, requires comparison of like with like. This is a controversial issue, because gene phylogenies may be incongruent with organismal phylogenies. Some authors like to make separate phylogeny estimates from different data sets, and then test their congruence as in "total evidence", i.e. the matrix of evidence is analyzed as one whole without being partitioned (KLUGE 1989, 2004). The advantage of using supertrees instead is that these methods, such as Matrix Representation using Parsimony (MRP) (BAUM 1992; RAGAN 1992; BAUM & RAGAN 2004) retain the information contained in each of the different genes (or paralogs) when combining them. Analysis by supertrees enables analysis of paralogous sets of sequences from a multigene family such as the 5S DNA gene.

CONCLUSION

While the use of sequence data from multigene families to infer phylogeny is not without challenge, the methods that we have described in several publications and summarized here, provide a sound framework for such analyses. The results to date based mainly upon sequences of the 5S rDNA NTS are proving to be useful for inferring possible relationships among haplomes in the *Triticeae*, and for constructing hypotheses about their evolution. These approaches should be applicable to other multigene families

Acknowledgements. I thank Dr. D.A. JOHNSON, University of Ottawa (UO) and Mr. L.G. BAILEY, Agriculture &

Agri-Food Canada (AAFC). Without their cooperation in generating sequence data and Dr. JOHNSON in co-authoring numerous papers this work could have never been achieved. I thank collaborators HAI-YING SHANG and YU-MING WEI, of the Triticeae Research Institute, Sichuan Agricultural University, China, for the *Secale* sequences. An earlier version of the manuscript benefited from the comments by Drs. JOHNSON, UO, E. SMALL and B. MIKI both AAFC.

References

- ALTSCHUL S.F., GISH W., MILLER W., MYERS E.W., LIPMAN D.J. (1990): Basic local alignment search tool. *Journal of Molecular Biology*, **215**: 403–410.
- APPELS R., BAUM B.R. (1992): Evolution of the Nor and 5SDna loci in the Triticeae. In: SOLTIS P.S., SOLTIS D.E., DOYLE J.J. (eds.): *Molecular Systematics of Plants*. Chapman & Hall, New York, U.S.A.: 92–116.
- ARVESTAD L., BERGLUND A.-C., LAGERGREN J., SENNBLAD B. (2003): *Bioinformatics*, **10** (Suppl. 1): i7–i15.
- BAUM B.R. (1992): Combining trees as a way of combining datasets for phylogenetic inference, and the desirability of combining gene trees. *Taxon*, **41**: 3–10.
- BAUM B.R., APPELS R. (1992): Evolutionary change at the 5S DNA loci of species in the Triticeae. *Plant Systematics and Evolution*, **183**: 195–208.
- BAUM B.R., BAILEY L.G. (1997): The molecular diversity of the 5S rRNA gene in *Kengyilia alata* (Drobov) J.L. Yang, Yen & Baum (Poaceae: Triticeae): potential genomic assignment of different rDNA units. *Genome*, **40**: 215–228.
- BAUM B.R., BAILEY L.G. (2000): The 5S rDNA units in *Kengyilia* (Poaceae: Triticeae): diversity of the non-transcribed spacer and genomic relationships. *Canadian Journal of Botany*, **78**: 1571–1579.
- BAUM B.R., BAILEY L.G. (2001): The 5S rRNA gene sequence variation in wheats and some polyploid wheat progenitors (Poaceae: Triticeae). *Genetic Resources and Crop Evolution*, **48**: 35–51.
- BAUM B.R., JOHNSON D.A. (1994): The molecular diversity of the 5S rRNA gene in barley (*Hordeum vulgare*). *Genome*, **37**: 992–998.
- BAUM B.R., JOHNSON D.A. (1996): The 5S rRNA gene units in ancestral two-row barley (*Hordeum spontaneum* C. Koch) and bulbous barley (*H. bulbosum* L.): sequence analysis and phylogenetic relationships with the 5S rDNA units in cultivated barley (*H. vulgare* L.). *Genome*, **39**: 140–149.
- BAUM B.R., JOHNSON D.A. (1998): The 5S rRNA gene in sea barley (*Hordeum marinum* Hudson *sensu lato*): sequence variation among repeat units and relation-

- ship to the X haplome in barley (*Hordeum*). *Genome*, **41**: 652–661.
- BAUM B.R., JOHNSON D.A. (1999): The 5S rRNA gene in wall barley (*Hordeum murinum* L. sensu lato): sequence variation among repeat units and relationship to the Y haplome in the genus *Hordeum* (Poaceae: Triticeae). *Genome*, **42**: 854–866.
- BAUM B.R., JOHNSON D.A. (2000): The 5S rRNA gene units in the native New World annual *Hordeum* species (Triticeae: Poaceae). *Canadian Journal of Botany*, **78**: 1590–1602.
- BAUM B.R., JOHNSON D.A. (2002): A comparison of the 5S rDNA diversity in the *Hordeum brachyantherum-californicum* complex with those of the eastern Asiatic *Hordeum roshevitzii* and the South American *Hordeum cordobense* (Triticeae: Poaceae). *Canadian Journal of Botany*, **80**: 752–762.
- BAUM B.R., JOHNSON D.A. (2003): The South African *Hordeum capense* is more closely related to some American *Hordeum* species than to the European *Hordeum secalinum*: a perspective based on the 5S DNA units (Triticeae: Poaceae). *Canadian Journal of Botany*, **81**: 1–11.
- BAUM B.R., RAGAN M.A. (2004): The MRP method. In: BININDA-EMONDS O.R.P. (ed.): *Phylogenetic Super-trees. Combining information to reveal the tree of life*. Kluwer, Dordrecht, the Netherlands: 17–34.
- BAUM B.R., BAILEY L.G., JOHNSON D.A., AGAFONOV A.V. (2003): Molecular diversity of the 5S rDNA units in the *Elymus dahuricus* complex (Poaceae: Triticeae) supports the genomic constitution of St, Y, and H haplomes. *Canadian Journal of Botany*, **81**: 1091–1103.
- BAUM B.R., BAILEY L.G., BELYAYEV A., RASKINA O., NEVO E. (2004): The utility of the nontranscribed spacer of 5S rDNA units grouped into unit classes assigned to haplomes – a test on cultivated wheat and wheat progenitors. *Genome*, **47**: 590–599.
- BAUM B.R., JOHNSON D.A., BAILEY L.G. (2001): Defining groups among multicopy genes prior to inferring phylogeny, with special emphasis on the Triticeae (Poaceae). *Hereditas*, **135**: 123–138.
- BAUM B.R., JOHNSON D.A., BAILEY L.G. (2005): Ancient differentiation of the H and I haplomes in diploid *Hordeum* species based on 5S rDNA. *Genome*, **48**: 610–618.
- BOTHMER R. VON, FLINK J., LANDSTRÖM T. (1986): Meiosis in Interspecific *Hordeum* hybrids. I. Diploid combinations. *Canadian Journal of Genetics and Cytology*, **28**: 525–535.
- BOTHMER R. VON, FLINK J., LANDSTRÖM T. (1987): Meiosis in Interspecific *Hordeum* hybrids. II. Triploid hybrids. *Evolutionary Trends in Plants*, **1**: 41–50.
- BOTHMER R. VON, JACOBSEN N. (1991): Interspecific hybrids within the genus *Hordeum*. In: GUPTA P.K., TSUCHIYA T. (eds.): *Chromosome Engineering in Plants: Genetics, Breeding, Evolution, Part A*. Elsevier, Amsterdam, the Netherlands: 411–431.
- EISEN J.A. (1998): Phylogenomics: improving functional predictions for uncharacterized genes by evolutionary analysis. *Genome Research*, **8**: 163–167.
- FELSENSTEIN J. (1985): Confidence limits on phylogenies: an approach using the bootstrap. *Evolution*, **39**: 783–791.
- FELSENSTEIN J. (2004): *Inferring phylogenies*. Sinauer Associates, Sunderland.
- GOLDMAN N. (1990): Maximum likelihood inference of phylogenetic trees, with special reference to a Poisson process model of DNA substitution and to parsimony analysis. *Systematic Zoology*, **39**: 345–361.
- HASEGAWA M., KISHINO K., YANO T. (1985): Dating the human-ape splitting by a molecular clock of mitochondrial DNA. *Journal of Molecular Evolution*, **22**: 160–174.
- HUELSENBECK J.P., RONQUIST F. (2001): MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics*, **17**: 754–755.
- KLUGE A.G. (1989): A concern for evidence and a phylogenetic hypothesis of relationships among *Epicrates* (Boidae, Serpentes). *Systematic Zoology*, **38**: 7–25.
- KLUGE A.G. (2004): On total evidence: for the record. *Cladistics*, **20**: 205–207.
- OLSEN G.J., MATSUDA H., HAGSTROM R., OVERBEEK R. (1994): FastDNAm1: A tool for construction of phylogenetic trees of DNA sequences using maximum likelihood. *Computer Applications in the Biosciences*, **10**: 41–48.
- PETERSEN G., SEBERG O. (2004): On the origin of the tetraploid species *Hordeum capense* and *H. secalinum* (Poaceae). *Systematic Botany*, **29**: 862–873.
- POSADA D. (2003): Selecting models of evolution. Theory. In: SALEMI M.; VANDAMME A.-M. (eds.): *The Phylogenetic Handbook. A practical Approach to DNA and Protein Phylogeny*. Cambridge University Press, Cambridge, U.K.
- POSADA D., CRANDALL K.A. (1998): Modeltest: Testing the model of DNA substitution. *Bioinformatics*, **14**: 817–818.
- RAGAN M.A. (1992): Phylogenetic inference based on matrix representation of trees. *Molecular Phylogenetics and Evolution*, **1**: 53–58.
- SAITOU N., NEI M. (1987): The neighbour-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution*, **4**: 406–425.

- SMALL R.L., CRONN R.C., WENDEL J.F. (2004): Use of nuclear genes for phylogeny reconstruction in plants. *Australian Journal of Botany*, **17**: 145–170.
- SWOFFORD D.L. (1998): PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. Sinauer Associates, Sunderland. Massachusetts.
- TATUSOV R.L., NATALE D.A., GARKAVTSEV I.V., TATUSOVA T.A., SHANKAVARAM U.T., RAO B.S., KIRYUTIN B., GALPERIN M.Y., FEDOROVA N.D., KOONIN E.V. (2001): The COG database: new developments in phylogenetic classification of protein from complete genomes. *Nucleic Acid Research*, **29**: 22–28.
- WHELAN S., LIÒ, GOLDMAN N. (2001): Molecular phylogenetics: state-of-the-art methods for looking into the past. *Trends in Genetics*, **17**: 262–272.
- WHEELER W. (1996): Optimization alignment: the end of multiple sequence alignment in phylogenetics? *Cladistics*, **12**: 1–9.
- WHEELER W.C. (1999): Fixed character States and the optimization of molecular sequence data. *Cladistics*, **15**: 379–385.
- WHEELER W., GLADSTEIN D., DE LAET J. (2003): POY version 3.011 available at <http://research.amnh.org/scicomp/projects/poy.php>.
- YANG Z. (1994): Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. *Journal of Molecular Evolution*, **39**: 306–314.
- YANISCH-PERRON C., VIEIRA J., MESSING J. (1985): Improved M13 phage cloning vectors and host strains: nucleotide sequence of M13mp18 and pUC19 vectors. *Gene*, **33**: 103–119.
- ZMASEK C.M., EDDY S.R. (2002): RIO: Analyzing proteomes by automated phylogenomics using re-sampled inference orthologs. *BMC Bioinformatics*, **3**: 14–33.