Differences Between South American H Haplome Diploids and I Haplome Diploids from the Perspective of the *5S rDNA* Gene in the Genus *Hordeum*

Bernard R. BAUM[1] and Douglas A. JOHNSON[2]

[1]Agriculture & Agri-Food Canada, Eastern Cereal and Oilseed Centre, Ottawa, Ontario, Canada;
[2]University of Ottawa, Department of Biology, Ottawa, Ontario, Canada

Abstract: Twelve South American diploid *Hordeum* species belonging to the H genome and three diploid species belonging to the I genome (including cultivated barley) were investigated for their 5S rDNA sequence diversity. The 374 sequenced clones were assigned to classes called "unit classes" which were further assigned to haplomes. Two unit classes were found to be present in each haplome. These were labelled to reflect the haplomes, viz. the long H1 and short I1 unit classes for the I haplome diploids, and the long H2 and long Y2 unit classes for the South American H genome diploids. The aligned sequences were subjected to a series of Maximum Likelihood analyses and various tests, including molecular clock, which are presented and discussed. The divergences among the unit classes suggest that the genus *Hordeum* might be of paleopolyploid origin.

Keywords: *5S DNA* gene; molecular clock; paleopolyploid

5S rRNA genes in the *Triticeae* are organized into tandem repeats with the highly conserved genes separated by the more variable, non-transcribed spacer region (henceforth NTS). In several publications (e.g., Baum & Bailey 1997, 2000, 2001; Baum & Johnson 1994, 1996, 1998, 1999, 2000, 2002; Baum *et al*. 2001, 2003), we have described the molecular diversity of 5S rDNA sequences in species within *Elymus*, *Hordeum*, *Kengyilia* and *Triticum*. We were able to classify them into putative orthologous groups, which we call unit classes, and moreover to assign the different unit classes to haplomes. Our objectives were: (1) to examine the diversity of the *5S rRNA* gene sequences among the South American diploid species belonging to the H-genome group (Bothmer & Jacobsen 1991); (2) to classify them into unit classes; and (3) to compare the unit classes of these South American diploids with the unit classes previously found in the I-genome group of which cultivated barley (*H. vulgare* L.) is a member. The present study describes the results of these analyses with respect to the evolution and time of divergence of these genes.

MATERIALS AND METHODS

Plant material and 5S rDNA cloning

Material from the following 11 taxa, represented by 38 accessions of seeds collected in their natural habitats in Argentina and Chile, was investigated: *H. chilense* Roem. & Schult., *H. comosum* Presl, *H. cordobense* Bothmer, N. Jacobsen & Nicora, *H. erectifolium* Bothmer, N. Jacobsen & Jørgensen, *H. muticum* Presl, *H. patagonicum* (Hauman) Covas ssp. *magellanicum* (Parodi & Nicora) Bothmer, Giles & N. Jacobsen, *H. patagonicum* ssp. *mustersii* (Nicora) Bothmer, Giles & N. Jacobsen, *H. patagonicum* ssp. *patagonicum*, *H. patagonicum* ssp. *santacrusense* (Parodi & Nicora) Bothmer, Giles & N. Jacobsen, *H. patagonicum* ssp. *setifolium* (Parodi & Nicora) Bothmer, Giles & N. Jacobsen, *H. pubiflorum* Hook. f., and *H. stenostachys* Godr. The five subspecies according to the taxonomy of Bothmer *et al*. (1991) are also recognized as species especially by South American authors. Several sequences from species containing the I haplome published earlier (Baum & Johnson 1994, 1996) were retrieved from

GenBank® and used for comparisons. The isolation of genomic DNA, PCR amplification of the *5S rRNA* genes, cloning of PCR products and sequencing of plasmid DNA have been described, e.g. Baum and Johnson (1994, 1996, 1998, 1999, 2000, 2002). The PCR primers were used to 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 DH5α or latterly ligated direct into pGEM-T (Promega Biotech) and transformed into DH5α. A total of 374 clones were isolated and for each clone, both strands were sequenced.

Sequence analysis

Sequences were routinely checked to ensure removal of vector sequences using the VecScreen program at NCBI (National Center for Biotechnology Information, USA). For each sequence a search for direct and inverted repeats was carried out using DNAMAN (Lynnon Biosoft®). The sequences were then submitted to NCBI and an accession number for each was obtained.

The 374 sequences were subsequently aligned using CLUSTALW (Thompson *et al*. 1994). The alignment was further improved by visual examination and editing using GeneDoc© Version 2.6.002 (Nicholas & Nicholas 1997). GeneDoc© was used to assign similar sequences to sequence groups, i.e. unit classes, based on the refined alignments for each putative orthologous group. At this stage, the alignment would reveal any sequences that would appear to have been assigned to the wrong unit class. Several sequences representative of each unit class were then subjected to similarity searches of the GenBank® and EMBL (European Molecular Biology Laboratory) databases using the NCBI Web-based BLAST service (Altschul *et al*. 1990) to identify the unit class with an already established, i.e. published unit class. The sequence identified as being the most similar, i.e. having the highest scoring segment pairs and the lowest P(N) value (as defined in Altschul *et al*. 1990), was subsequently aligned *in toto* with the representative unit class, and also with other sequences of interest among previously defined unit class sequences mentioned above. All the 374 sequences were then re-assembled for a final step of alignment and manual refinement. This process and the method of unit class determination and recognition were discussed in more detail in Baum *et al*. (2001).

The data were subjected to the program WinModeltest© 4.b (Posada & Crandall 1998) to test the fit of various maximum likelihood (ML) models and to choose the model that best fits the data using the hierarchical likelihood ratio test. To conduct the following ML analyses, a reduced set of sequences, representative of each unit class within each taxon, was used after first subjecting them to WinModeltest in order to verify that the model with the best fit for the reduced data set was identical to the model derived for the total data set. Phylogenetic analyses were then conducted with the following ML methods and bootstrapping with neighbour-joining (NJ) search: ML clock not enforced, ML clock enforced, using both PAUP*© (Swofford 1998) and PHYLIP© (Felsenstein 1993), tree calibration by nonparametric rate smoothing (NPRS) and NJ using TREEFINDER© (Jobb 2003). The time calibration in NPRS was based on the estimated time of divergence between barley and wheat of 13 million years (MY) (Gaut 2002). All the trees were imported to PAUP* and subjected to the KH (Kishino & Hasegawa 1989) and the SH (Shimodaira & Hasegawa 1999) tests for the best tree. The latter was then subjected to tree analysis under the maximum parsimony and minimum evolution criteria using PAUP*.

RESULTS AND DISCUSSION

All of the South American diploid taxa in this study were found to possess the same two 5S rDNA unit classes, i.e. we found 159 sequences that were assigned to the long H2 and 215 sequences to the long Y2. This is consistent with our previous findings in *H. cordobense* (Baum & Johnson 2002) and *H. muticum* (Baum & Johnson 2003).

An analysis of these sequences was carried out with the addition of exemplars of two unit classes from I genome species, i.e. long H1 and short I1. Modeltest selected the following settings as best, namely the HKY+G model with transition/transversion ratio = 1.0799, the assumed nucleotide frequencies of A = 0.269, C = 0.1743, G = 0.2492 and T = 0.3075, the shape parameter = 2.9192 and the rates = gamma.

These analyses yielded eight trees – one ML tree, one molecular clock tree (PAUP and PHYLIP), one

NPRS tree and four NJ trees (TREEFINDER), and one NJ tree from PAUP.

The KH and SH tests selected the molecular clock tree as best. However, all the eight trees have identical topologies in the sense that the major branches are identical and the four unit classes (Figure 1) are clearly identified and similar. The differences relate to branch size and specific details that depend upon the optimality criteria used in the tree analyses. Some specifics apply to a particular optimality criterion, e.g. with the parsimony criterion branch lengths are expressed in the number of nucleotide step changes that clearly do not apply to the other criteria. The phylogram

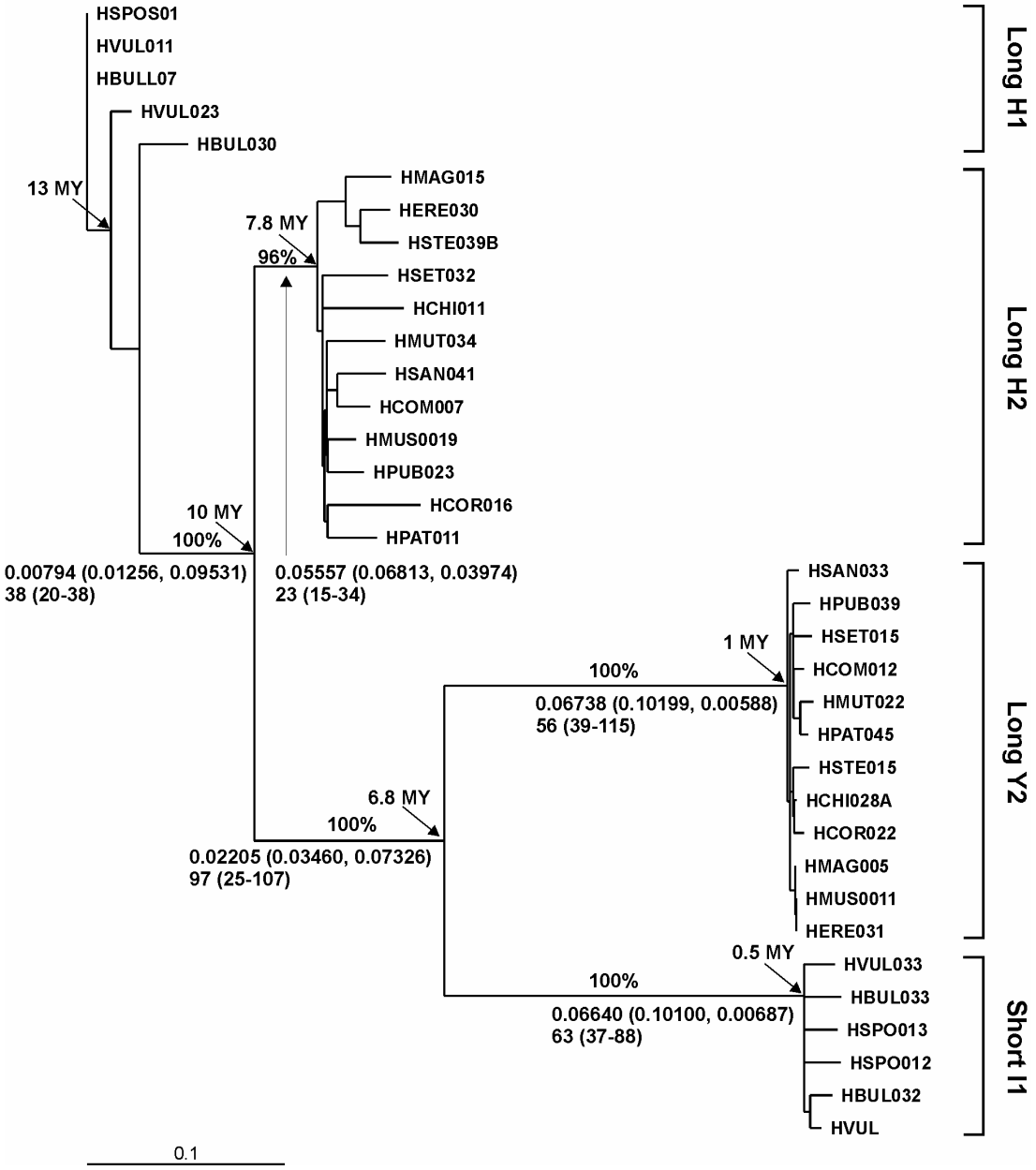

Figure 1. Phylogram of 5S rDNA sequences representative of the four unit classes, based on minimum evolution tree. The phylogram combines information of the best maximum likelihood tree, the molecular clock, parsimony and NPRS trees. MY: million years since divergence; above major branches: bootstrap support (%); below major branches: branch lengths (distance from root, distance from tip) and 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

(Figure 1) combines the results obtained from all the three criteria (parsimony, likelihood and distance) superimposed on the (distance) minimum evolution tree.

The estimated time of divergence of 13 MY between *Hordeum* and *Triticum* (Gaut 2002), likely based on comparisons of cultivated species (*H. vulgare* and *Triticum aestivum* L.), is close to the previous estimate of Wolfe *et al*. (1989) based on chloroplast DNA variation. But the genus *Hordeum*, with 31 species Bothmer *et al*. (1991), or more depending on the taxonomy, is probably much older judged from its distribution with a centre of diversity in South America, one species in South Africa, a centre in the Near East and a diffuse distribution across Eurasia. Although phylogenetic estimates of organisms form the basis for genomic and evolutionary studies, in this paper we have examined genome evolution through the 5S rDNA. Clearly different unit classes might have arisen at different times (Figure 1). What is puzzling is that the long H1 and short I1 unit class, both present in the I-genome species and characteristic to them, are the most divergent ones. Furthermore, the South American diploid species belonging to the H-genome species group (Bothmer *et al*. 1991) contain both the long H2 and long Y2 unit classes which are nearly as divergent as the former two unit classes. Does this indicate that *Hordeum* as we know it today is a diploidized paleopolyploid? (Paleopolyploids have a disomic inheritance and their progenitors cannot be detected by cytology or DNA markers.) Sequence analysis of full genomes in the classical diploids *Arabidopsis thaliana* and *Oryza sativa* has revealed that they are apparently paleopolyploids. Alternatively, could two different unit classes have evolved so rapidly in each group during roughly six MY? Both case scenarios point to an earlier origin of the genus *Hordeum*. In this study the 5S rDNA has shown itself to be a valuable tool for the detection of paleopolyploidy status, at least in the *Triticeae*. The non-transcribed spacer varies considerably more than the coding region of the 5S DNA which is consistent with the greater than 10-fold synonymous substitution rate variation among homeologous loci in *Arabidopsis* (Zhang *et al*. 2002). In addition it appears that paralogs do not evolve at different rates and that the unit classes appear to have evolved by the molecular clock fashion. Based upon these conclusions, it can be postulated that the 5S rDNA units featuring two different unit classes were present in two unidentified diploid progenitors. In the past, they contributed to the formation of a paleopolyploid through horizontal transfer of chromosomal segments or genes (Levy & Feldman 2002) giving rise to the present day diploid species.

Acknowledgements. The authors thank Drs. B. Miki and S. J. Molnar for helpful comments.

References

Altschul S.F., Gish W., Miller W., Myers E.W., Lipman D.J. (1990): Basic local alignment search tool. J. Mol. Biol., **215**: 403–410.

Baum B.R., Bailey L.G. (1997): The molecular diversity of the *5S rRNA* gene in *Kengyilia alatavica* (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. Can. J. Bot., **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). Genet. Res. Crop Evol., **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 relationship 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). Can. J. Bot., **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). Can. J. Bot., **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). Can. J. Bot., **81**: 1–11.

Baum B.R., Johnson D.A., Bailey L.G. (2001): Defining orthologous groups among multicopy genes prior to inferring phylogeny, with special emphasis on the Triticeae (Poaceae). Hereditas, **135**: 123–138.

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. Can. J. Bot., **81**: 1091–1103.

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.

Bothmer R. von, Jacobsen N., Baden C., Jørgensen R.B., Linde-Laursen I. (1991): An ecogeographical study of the genus *Hordeum*. Studies on Crop Genepools 7. International Board for Plant Genetic Resources, Rome.

Felsenstein J. (1993): PHYLIP (Phylogeny Inference Package) version 3.5c. Distributed by the author. Department of Genetics, University of Washington, Seattle.

Gaut B.S. (2002): Evolutionary dynamics of grass genomes. New Phytol., **154***:* 15–28.

Jobb G. (2003): TREEFINDER version December 2003. Munich, Germany. Distributed by the author at www.treefinder.de

Kishino H., Hasegawa M. (1989): Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in Hominoidea. J. Mol. Evol., **29**: 170–179.

Levy A.A, Feldman M. (2002): The impact of polyploidy on grass genome evolution. Plant Physiol., **130**: 1587–1593.

Nicholas K.B., Nicholas H.B. Jr. (1997): GeneDoc©: a tool for editing and annotating multiple sequence alignments. Distributed by the authors.

Posada D., Crandall K.A. (1998): Modeltest: Testing the model of DNA substitution. Bioinformatics, **14**: 817–818.

Shimodaira H., Hasegawa M. (1999): Multiple comparisons of log-likelihoods with applications to phylogenetic inference. Mol. Biol. Evol., **16**: 1114–1116.

Swofford D.L. (1998): PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. Sinauer Associates, Sunderland. Massachusetts.

Thompson J.D., Higgins D.G., Gibson T.J. (1994): CLUSTALW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res., **22**: 4673–4680.

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.

Wolfe K.H., Gouy M., Yang Y.W., Sharp P.M., Li W.H. (1989): Date of the monocot-dicot divergence estimated from chloroplast DNA sequence data. Proc. Natl. Acad. Sci. U.S.A. **86**: 6201–6205.

Zhang L., Vision T.J., Gaut V.S. (2002): Patterns of nucleotide substitution among simultaneously duplicated gene pairs in *Arabidopsis thaliana*. Mol. Biol. Evol., **19**: 1464–1473.

Received for publication March 2, 2004
Accepted March 18, 2004

Souhrn

Baum B. R., Johnson D. A. (2004): **Rozdíly mezi jihoamerickými diploidy haplomů H a diploidy haplomů I z perspektivy genu *5S rDNA* u rodu *Hordeum***. Czech J. Genet. Plant Breed., **40**: 45–50.

Zkoumali jsme variabilitu sekvencí genu *5S rDNA* z dvanácti jihoamerických diploidních druhů rodu *Hordeum*, které patří ke genomu H, a ze tří diploidních druhů, které náležejí ke genomu I (včetně kulturního ječmene). 374 sekvencovaných klonů jsme zařadili do tříd nazvaných „jednotkové třídy", které byly dále přiřazeny k jednotlivým haplomům. Zjistili jsme, že každý haplom obsahuje dvě jednotkové třídy. Ty byly označeny tak, aby charakterizovaly jednotlivé haplomy, tedy jednotkové třídy dlouhého H1 a krátkého I1 pro diploidy haplomu I,

a jednotkové třídy dlouhého H2 a dlouhého Y2 pro jihoamerické diploidy genomu H. Sekvence jsme srovnali a podrobili řadě analýz metodou maximální pravděpodobnosti a různým testům, včetně testu molekulárních hodin, které předkládáme a hodnotíme. Divergence mezi jednotkovými třídami naznačují, že by rod *Hordeum* mohl mít paleopolyploidní původ.

Klíčová slova: gen *5S DNA*; molekulární hodiny; paleopolyploid

Corresponding author:

Bernard R. Baum, Ph.D., Agriculture & Agri-Food Canada, Eastern Cereal and Oilseed Centre, 960, Carling Ave., Ottawa (Ontario), K1A0C6, Canada
tel.: + 1 613 759 1392, fax: + 1 613 759 1924, e-mail: baumbr@em.agr.ca