L. Petersen · H. Østergård · H. Giese

# **Genetic diversity among wild and cultivated barley as revealed by RFLP**

Received: 15 December 1993 / Accepted: 17 May 1994

**Abstract** Genetic variability of cultivated and wild barley, *Hordeum vulgare* ssp. *vulgare* and *spontaneum*, respectively, was assessed by RFLP analysis. The material consisted of 13 European varietes, single-plant offspring lines of eight land races from Ethiopia and Nepal, and five accessions of ssp. *spontaneum* from Israel, Iran and Turkey. Seventeen out of twenty-one studied cDNA and gDNA probes distributed across all seven barley chromosomes revealed polymorphism when DNA was digested with one of four restriction enzymes. A tree based on genetic distances using frequencies of RFLP banding patterns was estimated and the barley lines clustered into five groups reflecting geographical origin. The geographical groups of land-race lines showed less intragroup variation than the geographical groups *ofspontaneum* lines. The group of European varieties, representing large variation in agronomic traits, showed an intermediate level. The proportion of gene diversity residing among geographical groups  $(F_{ST})$ varied from 0.19 to 0.94 (average 0.54) per RFLP pattern, indicating large diversification between geographical groups.

Key words *Hordeum vulgare* ssp. *spontaneum Hordeum vulgare ssp. vulgare* · Genetic distance Fixation index. Restriction fragment length polymorphism

# **Introduction**

For conservation of genetic resources and for breeding, it is important to know about the genetic variation in cultivated barley, *Hordeum vulgare* ssp. *vulgare*, as compared to that of land races of ssp. *vulgare* and of its wild ances-

Communicated by R M. A. Tigerstedt

L. Petersen  $\cdot$  H. Østergård  $\cdot$  H. Giese ( $\boxtimes$ ) Plant Genetics Section, Environmental Science and Technology Department, Rise National Laboratory, DK-4000 Roskilde, Denmark

tor, *Hordeum vulgare* ssp. *spontaneum.* An interesting question is to what extent the genes in the wild ancestors have contributed to the present gene pool in the currentlycultivated barley. Given the wide geographical range of wild barley, one would expect a highly-diverse gene pool in this subspecies. This may to some extent also be true for land races due to the way early domestication has taken place, not being a single event occurring in a confined area but rather a continuous process involving several areas (Harlan 1979; Allard 1988).

Assessments of the genetic variation in the ssp. *vulgare*  and *spontaneum* have been carried out by a number of researchers. These assessments have been based on variation in hordeins (Doll and Brown 1979), isozymes (Nevo et al. 1979; Brown and Munday 1982; Nevo et al. 1986; Allard 1988), morphological traits (Nevo et al. 1979; Brown and Munday 1982), and rDNA loci (Allard 1988; Allard et al. 1990; Saghai Maroof et al. 1990). In studies including land races, different levels of variation have been found. Thus, Jana and Pietrzak (1988) found a similar level of isozyme variation in wild barley and in land races, whereas Brown and Munday (1982) and Nevo et al. (1986) reported a level of diversity in land races between that of the wild and of the modern cultivated barley. The latter authors also found that cultivated barley was characterized by a higher level of diversity with respect to morphological traits than was the case in wild barley.

The use of restriction fragment length polymorphism (RFLP) provides the opportunity of assessing variation at the molecular level. This has been exploited by Graner et al. (1990) in a study of some barley varieties and by Zhang et al. (1993) for populations of wild barley. In addition, the high degree of polymorphism generally displayed by RFLP markers ensures a higher level of differentiation of the plant material analysed than is possible using classical markers such as isozymes. The two methods may, however, uncover different classes of variation (Zhang et al. 1993).

The present paper reports on the genomic variation measured by RFLP markers in *H. vulgare* ssp *vulgare* and *spontaneum.* The purpose is to obtain an indication of whether the level of genetic variation within modern, partly-unrelated varieties is lower than that within land races or that within the wild subspecies. Further, the problems encountered when interpreting RFLP data for studies of genetic variation are discussed.

# **Materials and methods**

# Plant material

The plant material used in this study include accessions/lines/varieties of *H. vulgare* ssp. *spontaneum* and ssp. *vulgare* (Table 1). For each accession/line/variety a random sample of approximately 50 kernels is considered. The *spontaneum* ssp. is represented by five accessions of diverse origin acquired from the USDA barley world collection. Twenty-one samples of *vulgare* ssp. are included. They consist of eight lines derived from accessions of land races from Ethiopia and Nepal acquired from various sources (Weibullsholm Plant Breeding Institute, Landskrona, Sweden; Institute for Agronomy and Plant Breeding, Braunschweig, Germany; and Plant Breeding Institute, Cambridge, UK) as well as 13 European spring and winter varieties. Homogeneity of the lines derived from the land races was ensured by three generations of single-plant selection. (This material has kindly been provided by J. Helms Jørgensen, Risø). The European varieties are chosen to represent a broad range of variation. In

Table 1 Description of accessions/varieties included

Taxa	No. 1n text	Geographical origin	Year of collection/ entry
H. vulgare ssp. spontaneum			
Accessions			
PI 245739	1	Ceylanpinal, Turkey	1958
PI 296897	$\overline{c}$	Eshtaol, Israel	1964
PI 296926	3	Kinneret, Israel	1964
PI 227019	$\overline{\mathbf{4}}$	Shustar, Iran	1955
PI 268242	5	Kerman, Iran	1960
H. vulgare ssp. vulgare			
Accessions			
K385	10	Shoa, Ethiopia	1973
K446	11	Genu Gofa, Ethiopia	1973
FAL015657	12	Ubarmer Baco, Ethiopia	Unknown
FAL004779	13	Cherem, Nepal	Unknown
FAL004782	14	Pangu, Nepal	Unknown
FAL004789	15	Gonba, Nepal	Unknown
PBI5654	16	Nachipundo, Nepal	1971
PBI4770	17	Salung, Nepal	1971
Spring variety			
Hanna	20	Austria	1884
Prentice Tystofte	21	Denmark	1900
Pallas	22	Sweden	1961
Jupiter	23	United Kingdom	1979
Corgi	24	United Kingdom	1985
Regatta	25	United Kingdom	1985
Digger	26	United Kingdom	1987
Derkado	27	East Germany	1989
'Segu' Sejet	28	Denmark	1989
Winter variety			
Vogelsanger Gold	30	West Germany	1965
Igri	31	West Germany	1980
Sonate	32	Denmark	1983
Marinka	33	The Netherlands	1985

the following the term line will be used to refer to the pooled sample of 50 kernels of an accession, line or variety.

### DNA probes

Variation in the plant material was studied by means of ten cDNA and 11 gDNA probes (see Table 2). The cDNA probes derived from wheat, *Triticum aestivum,* were kindly provided by M. D. Gale, Cambridge, UK. (Sharp et al. 1989). The gDNA probes were constructed as described in Giese et al. (1993), and two probes, MSU12 and MSU21, were kindly provided by T. Blake, Montana State University, USA (Shin et al. 1990).

#### DNA preparations

Fifty seeds of each accession were grown in a greenhouse for 10 days. Barley seedlings were harvested, frozen in liquid nitrogen, and stored at -80°C. DNA was extracted from the *pooled* sample by the method described by Sharp *et al.* (1988).

#### Restriction enzyme digestion, electrophoresis, and Southern blotting

The purified high-molecular-weight DNA was digested to completion with the following four restriction enzymes: *BamHI, EcoRI, EcoRV,* and *HindIII,* according to manufacturer's instructions (Amersham). Horizontal gel electrophoresis and Southern blotting were performed according to Maniatis et al. (1982). Nylon membranes (Amersham) were used for the transfer of DNA.

#### Hybridization methods

Before hybridization, the blotted membranes were prehybridized at 42°C for  $1-24$  h in a 45% formamide,  $4 \times$ SSPE ( $1 \times$ SSPE=180 mM NaCl, 10 mM sodium phosphate, 1 mM EDTA, pH 7.5), 6× Denhard's solution (0.025% BSA, PVP360 and Ficoll400, w/v), 200 gg/ml salmon sperm, and 0.1% SDS. The hybridization was performed overnight at  $42^{\circ}$ C in an identical buffer except that  $10\%$  dextran sulphate was added (Christiansen and Giese 1990). The denatured insert DNA was labelled with <sup>32</sup>P-deoxynucleotides by random priming for 4 h at room temperature or 1 h at  $37^{\circ}$ C (Feinberg and Vogelstein 1983). The hybridized membranes were washed once for 15 min and twice for 30 min with  $2 \times$  SSC and 0.1% SDS (sodium dodecyl sulphate) at  $42^{\circ}$ C. Occasionally, it was necessary to carry out an additional wash, either by using  $0.2 \times SSC$  or increasing the temperature to 65°C. The membranes were exposed to X-ray films (Hyperfilm-MP, Amersham) at  $-70^{\circ}$ C for  $7-14$  days, typically.

#### Data analysis

Standard genetic parameters were used for describing diversity (Hartl and Clark 1989). The following assumptions were made: one probe corresponds to one locus, only one allele per locus is present in each line, and alleles are defined by the entire banding pattern obtained using a specific probe.

Genetic distances  $(D_{XY})$  between lines were calculated using the formula:

$$
D_{XY} = -\ln\left(J_{XY} / \sqrt{J_X J_Y}\right)
$$

where  $J_X$ ,  $J_Y$ , and  $J_{XY}$  are the means of  $\Sigma p_i^2$ ,  $\Sigma q_i^2$ , and  $\Sigma p_i q_i$  over all loci,  $p_i$  and  $q_i$  being the frequencies of the *i*th allele in the populations  $\hat{X}$  and  $\hat{Y}$ , respectively. This can be reduced to  $-\ln[J_{XY}]$  according to the above assumptions. Based on these genetic distances, an unrooted tree was estimated using the programmes FITCH and DRAWTREE (Felsenstein 1993). These programmes use the Fitch-Margoliash procedure, which successively includes all lines and finds the tree for which the sum of the length of all branches between two lines is closest to the observed genetic distance (Fitch and Margoliash 1967). Several runs were carried out with the lines being included in different order.

Based on the branching of the observed tree, the lines were grouped and the variation within and between groups was considered. The variation within a group, using the different samples as "individuals" and the group as the "population", was described using four parameters: the percentage of polymorphic loci (P), the average number of alleles per locus (A), the average gene diversity  $(H_s; 1-\Sigma p_i^2$  for each locus), and the maximal possible gene diversity per locus when all lines in the group have different alleles  $(maxH<sub>s</sub>)$ .

For each locus a measure of gene diversity between groups, the fixation index  $F_{ST}$ , was calculated:

 $F_{ST} = (H_T - \overline{H}_S)H_T$ 

where  $\tilde{H}_{S}$  is the average gene diversity within group, and  $H_{T}$  is the total gene diversity calculated from the pool of all lines,

The estimates of genetic parameters were not corrected for different numbers of "individuals" in the different groups as these numbers were of similar order, and the estimates were used solely to obtain comparisons within this study and not for making general statements.

# **Results**

Twenty-one probes (Table 2) are used to study genetic variation in 26 lines of/-/, *vulgare* ssp *vulgare* and *spontaneum*  (Table 1). In most cases, the cDNA probes reveal a lower degree of polymorphism than the gDNA probes. The cDNA probes give rise to approximately three different restriction patterns on the average (Table 2), while the average number of restriction patterns obtained with the gDNA probes is seven. The probe size has no clear-cut influence on the number of restriction patterns, as highly-polymorphic patterns are found using both large and small probes. Furthermore, an increase in probe size is not related to an increased number of fragments per pattern.

For each line, the distribution of restriction patterns for each probe (Table 3) is used to estimate the genetic distance between any two lines and, based on these distances, a tree can be drawn (Fig. 1). The tree illustrates the genetic relationship between the lines and shows that the lines clearly cluster into separate groups which correspond to the geographical origin of the lines. The *spontaneum* accessions split into two groups: the two Israeli accessions (nos. 2 and 3) and the three accessions collected in Iran and Turkey (nos. l, 4, and 5). The lines of land races also split into two groups: Nepal (nos. 13-17) and Ethiopia (nos. 10-12). Finally, the European varieties (nos. 20-28 and 30-33) may all be assigned to a single group. However, a slight tendency towards a subgroup consisting of three of the four included winter varieties (nos. 30, 31 and 33) is present. The most similar lines are found among the Nepalese assessions and among the spring varieties. Two winter varieties differ as much as different accessions from Israel. When the genetic variation is considered in detail (Table 3), it is seen that different alleles are often found in lines from different geographical groups and this results in a high value of the fixation index  $(F_{ST})$  for each of the polymorphic loci as well as for the average over polymorphic loci ( $F_{ST}$ =0.54).



\* Sharp et al. i989; \*\* Giese et al. (1994); \*\*\* Shin et al. 1990



Table 3 The RFLP pattern of all loci for the 26 accessions/lines Table 3 The RFLP pattern of all loci for the 26 accessions/lines





Fig. 1 Tree based on genetic distances demonstrating the clustering of lines. The numbers refer to Table 1

Table 4 Genetic variation for each geographical group described as the observed percentage of polymorphic loci (P), the observed average number of alleles per locus  $(A)$ , the average gene diversity  $(H<sub>S</sub>)$ , the theoretical maximal gene diversity (max  $H<sub>S</sub>$ ) and the ratio of the latter two

Geographical group <sup>a</sup>		А	$H_{\rm c}$	$maxH_s$	$H_s/maxH_s$
Israel $(2)$	43	1.4	0.21	0.50	0.42
Iran and Turkey (3)	57	1.8	0.31	0.66	0.47
Ethiopia (3)	38	1.4	0.19	0.66	0.29
Nepal $(5)$	38	1.4	0.17	0.80	0.21
Europe $(13)$	71	2.6	0.33	0.92	0.36

Number in parenthesis gives the number of lines of that origin

Within each geographical group a different degree of variation is found (Table 4). The Nepalese lines display the least variation. Only one allele is found in more than half the examined loci and only in a single case does the number of alleles within a locus exceed two (Table 3). The two most similar lines in this group, nos 15 and 17, can only be distinguished by two probes, while the most distant, e. g., nos. 15 and 16, only differentiate at five loci. The results are similar for the Ethiopian lines, which show nearly the same low variation taking into account the number of lines. Regarding the European varieties, only about a third of the loci are common for the 13 varieties. However, in many of the variable loci the same allele is found in nearly all lines implying that the relative gene diversity is between that of the land race and the *spontaneum* lines. The most different spring varieties (nos. 21 and 26) differ at ten loci, in contrast to the most similar (nos. 23 and 25)

which differ at three loci. The Israeli and the Turkish/ Iranian lines, respectively, differ in 9-10 loci, and are the most variable groups taking the number of lines into account (Table 3).

# **Discussion**

In this study, restriction patterns are used as a basis for calculating genetic distances, thus excluding any attempt to graduate the degree of genetic similarity by means of single-band differences. No matter whether the unit chosen to reflect genetic relationships is the restriction patterns or single restriction fragments, false conclusions may emerge, as illustrated in Fig. 2. Three different patterns are depicted, which could be interpreted as the result of single nucleotide differences between A and B, and between A and C, respectively. Based on a calculation of shared fragments, however, the A and C genotypes would be concluded to be more closely related than A and B, whereas, based on the comparison of patterns, all three genotypes would be concluded to be equally related. This example illustrates that a conclusive measure of genetic diversity can only be obtained if it is possible to establish the specific mutations that have generated the actual patterns.

One of the assumptions of this study, namely that one probe corresponds to one locus, is not true in all cases. The probe Xris25 is known to recognize three different loci in the varieties 'Vogelsanger Gold' and 'Alf' (Giese et al. 1994). This may also be true for other gDNA probes. When this is the case individual allele patterns may represent the combined allele patterns from more than one locus. The unambiguity of RFLP patterns can further be questioned as it cannot be excluded that the fragments are incompletely resolved, thus concealing small differences in fragment length. Also, two bands of similar length may not actually be identical (Gepts and Clegg 1989).

The major result emerging from this study of genetic variation is that the lines cluster into distinct groups which correspond to the geographical origin of the lines. This could reflect either that each group is to some extent derived from different ancestors, or that different alleles have been retained from an original gene pool due to processes such as selection and drift. The latter is the more plausible as the nodes splitting the branches are rather close.

As in previous studies of genetic variation, the wild barley lines of ssp. *spontaneum* are more variable (H<sub>s</sub>/maxH<sub>s</sub>) equals 0.42 and 0.47, respectively) than the cultivated lines of ssp. *vulgare*  $(H<sub>S</sub>/maxH<sub>S</sub>$  equals 0.21, 0.29 and 0.36, respectively) (Table 4). This is illustrated by the considerable genetic distance between the lines of different origin (Fig. 1). Contrary to previous findings, the lines derived from land races display a level of variation similar to, or even less than, that of the European varieties  $(H_s/maxH_s)$ equals 0.21, 0.29 compared to 0.36). This is also in disagreement with the general belief that Ethiopia is an especially important centre of genetic variation in barley. However, the variation between the European varieties is based



Fig. 2 RFLP pattern for three hypothetical genotypes *A, B, and C*  as generated by different point mutations in B and C

on varieties selected to reflect different genetic material whereas the lines derived from the land races have been selected randomly from the respective geographical areas. These considerations address the important question of how to sample plant material that allows a comparable estimate of genetic variation in wild lines and lines that have been under selection. Selection for different characters during a breeding programme quickly leads to verydivergent lines, exemplified here by the degree of variation between two related winter barleys 'Marinka' (no. 33) and 'Sonate' (no. 32). Regardless of the possible bias in the sampling of the analysed barley lines, it is clear, that there is considerable genetic variation in cultivated barley. Thus, there is no indication that the gene pool in cultivated barley is a limiting factor in generating new barley varieties.

Variation within lines was ignored in this study as pooled DNA from 50 kernels of each line was analysed. Cultivated barley is predominantly self-pollinating and lines were maintained so as to retain homogeneity, e.g., the land-race lines were subjected to three cycles of singleplant selection. However, it cannot be excluded that variation is present due to homozygous variants or rare heterozygotes in the lines. However, this, and the inclusion in the analysis of ,,alleles" which represent more than one locus, is not considered to have seriously biased the results. Thus, the RFLP technique can be concluded to represent a powerful tool for the estimation of genetical diversity in *H. vulgare.* The analysis of only a few lines resulted in clear geographical groups suggesting a potential for the use of RFLP markers in assessing genetic variation for conservation as well as for breeding purposes.

Acknowledgements The excellent technical assistance of Bente Andersen is gratefully acknowledged, and Hans Siegismund is thanked for his help in the estimation of the genetic tree and for carefully reviewing an earlier version of the manuscript.

## **References**

Allard RW (1988) Genetic changes associated with the evolution of adaptedness in cultivated plants and their wild progenitors. J Hered 79:225-238

- Allard RW, Saghai Maroof MA, Zhang Q, Jørgensen RA (1990) Genetic and molecular organization of ribosomal DNA (rDNA) variants in wild and cultivated barley. Genetics 126:743- 751
- Brown AHD, Munday J (1982) Population-genetic structure and optimal sampling of land races of barley from Iran. Genetica 58: 85-96
- Christiansen SK, Giese H (1990) Genetic analysis of the obligate parasitic barley powdery mildew fungus based on RFLP and virulence loci. Theor Appl Genet 79:705-712
- Doll H, Brown AHD (1979) Hordein variation in wild *(Hordeum spontaneum*) and cultivated *(H. vulgare)* barley. Can J Genet Cytol 21:391-404
- Feinberg A, Vogelstein B (1983) A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity. Anal Biochem 132:6-13
- Felsenstein J (1993) PHYLIP (Phylogeny Inference Package) version 3.5c. Distributed by the author. Department of Genetics, University of Washington
- Fitch WM, Margoliash E (1967) Construction of phylogenetic trees. Science 155:279-284
- Gepts R Clegg MT (1989) Genetic diversity in pearl millet *[Pennisetum glaucum* (L.) R. Br.]at the DNA sequence level. Heredity 80:203-208
- Giese H, Holm-Jensen AG, Jensen HP, Jensen J (1993) Localization of the Laevigatum powdery mildew resistance gene to barley chromosome 2 by the use of RFLP markers. Theor Appl Genet 85:897-900
- Giese H, Holm-Jensen AG, Mathiassen H, Kjær B, Rasmussen SK, Bay H, Jensen J (1994) Distribution of RAPD markers on a linkage map of barley. Hereditas 120:267-273
- Graner A, Siedler H, Jahoor A, Herrmann RB, Wenzel G (1990) Assessment of the degree and the type of restriction fragment length polymorphism in barley *(Hordeum vuIgare).* Theor Appl Genet 80:826-832
- Hartl DL, Clark AG (1989) Principles of population genetics, 2nd edn. Sinauer Associates, Sunderland, Massachusetts, USA
- Harlan JR (1979) On the origin of barley. In: Barley: origin, botany, culture, utilization, pests. Agriculture Handbook 338, US Department of Agriculture, pp 10-36
- Jana S, Pietrzak LN (1988) Comparative assessment of genetic diversity in wild and primitive cultivated barley in a center of diversity. Genetics 119:981-990
- Maniatis T, Fritsch EF, Sambrook I (1982) Molecular cloning: a laboratory manual. Cold Spring Harbour Laboratory, Cold Spring Harbour, New York
- Nero E, Zohary D, Brown AHD, Haber M (1979) Genetic diversity and environmental associations of wild barley, *Hordeum spontaneum,* in Israel. Evolution 33:815-833
- Nevo E, Beiles A, Zohary D (1986) Genetic resources of wild barley in the Near East: structure, evolution and application in breeding. Biol J Linn Soc 27:355-380
- Saghai Maroof MA, Allard RW, Zhang Q (1990) Genetic diversity and ecogeographical differentiation among ribosomal DNA alleles in wild and cultivated barley. Proc Natl Acad Sci USA 87: 8486-8490
- Sharp PJ, Kreis M, Shewry PR, Gale MD (1988) Location of  $\alpha$ -amylase sequences in wheat and its relatives. Theor Appl Genet 75: 286-290
- Sharp PJ, Chao S, Desai S, Gale MD (1989) The isolation, characterization and application in the *Triticeae* of a set of wheat RFLP probes identifying each homoeologous chromosome arm. Theor Appl Genet 78:342-348
- Shin JS, Chao S, Corpuz L, Blake T (1990) A partial map of the barley genome incorporating restriction fragment length polymorphism, polymerase chain reaction, isozyme, and morphological marker loci. Genome 33:803-810
- Zhang Q, Saghai Maroof MA, Kleinhofs A (1993) Comparative diversity analysis of RFLPs and isozymes within and among populations of *Hordeum vulgare* ssp. *spontaneum*. Genetics 134: 909-916