

Marker-assisted acquisition and core collection formation: a case study in barley using AFLPs and pedigree data

R. van Treuren*, I. Tchoudinova, L.J.M. van Soest and Th.J.L. van Hintum
*Centre for Genetic Resources, The Netherlands, Wageningen University and Research Centre, P.O. Box 16, 6700 AA Wageningen, The Netherlands; *Author for correspondence (e-mail: Robbert.vanTreuren@WUR.NL; phone: +31-317-477078; fax: +31-317- 418094)*

Received 5 December 2003; accepted in revised form 9 April 2004

Key words: Acquisition, AFLP, Conservation, Core collection, *Hordeum*, PGR management

Abstract

A problem that often occurs in deciding which germplasm should be acquired to expand the diversity of a plant genetic resources collection, and which accessions should be included in a core collection, is the lack of proper data. The usefulness of an AFLP-based protocol to assist in acquisition decisions and in core collection formation was examined by using 52 barley cultivars. For validation purposes, pedigree data of the cultivars were used to calculate the 'effective number of origin lines' (n_{oi}), a parameter introduced in earlier research that was defined as the number of alleles per locus, not identical by descent, in a set of lines. Two AFLP primer combinations were able to distinguish all 52 cultivars from each other, and to discriminate between spring and winter crop types. Using the year of origin of the cultivars, the historical development of n_{oi} showed a stepwise pattern, indicating the periodical release of genetically similar cultivars, alternated by the incorporation of new material. Comparison of AFLP data between cultivars and both their parents was possible in five cases. These comparisons revealed a high likelihood that the correct parents were involved but a rather skewed contribution of parents to offspring, suggesting that backcrossing had been applied. Treating the 25 cultivars that were released before 1980 and played an important role in barley cultivation as a basic collection, and the 27 more recent cultivars as potential candidates for acquisition, n_{oi} values generated by a marker-based approach largely followed those using a random approach. Given this poor performance, a marker-based protocol to assist in acquisition decisions was not considered useful for the analysed material. If the 52 cultivars were considered to be the collection from which a core collection had to be selected, the marker-based selection showed much better results compared to a random selection. About half of the total number of origin lines could be captured with a quarter of the collection, indicating the potential utility of AFLPs in core collection formation.

Introduction

According to the Convention on Biodiversity as formulated in Rio de Janeiro in 1992, countries are committed to conserve the genetic resources of their important cultivated crops. Rapid expansion of the number and size of collections has resulted in more than 1300 genebank collections, collec-

tively conserving more than six million accessions (FAO 1996). Therefore, gaining knowledge about these collections and improving the efficiency of plant genetic resources conservation are crucial for effective management. To this end, molecular marker technologies are increasingly playing a role in supporting traditional managerial approaches (Bretting and Widrechner 1995).

Acquisition of genebank accessions is an important aspect of plant genetic resources management. In many cases, genebank collections are expanded by including obsolete varieties, landraces and wild crop relatives obtained from expeditions to natural distribution areas. In the latter case, molecular studies directed to the distribution of genetic variation over the natural areas may help in developing guidelines to optimise the diversity captured during collection missions (McGregor et al. 2002). Because of ongoing plant breeding activities, new varieties are released regularly. After a period of utilisation, varieties become obsolete, and are removed from variety lists. Without conservation, obsolete varieties may be lost for future utilisation in plant breeding. Varieties are often developed from existing varieties, but also by the use of 'exotic' material. In the latter case, valuable genetic variation may be involved that is not yet represented in genebank collections, and therefore could improve the genetic diversity of the collection. However, relevant information about varieties, such as pedigree data, is often lacking or insufficient. Moreover, passport data are often difficult to translate in genetical terms. Given the large numbers of varieties that are developed, genebanks are faced with the question how to select varieties that can improve the genetic composition of existing collections.

Given the large sizes of collections, the core collection concept was introduced to facilitate the utilisation of plant genetic resources (Frankel and Brown 1984). A core collection can be defined as a subset of a collection that contains the major part of the diversity present within the collection (Hintum et al. 2000). In the assembly of core collections, passport and evaluation data are generally used to select accessions that are expected to represent a cross section of the diversity within the collection. However, missing, incomplete or unreliable passport data, and evaluation data with insufficient discriminating power often hamper the development of core collections.

If information about genetic resources is insufficient, molecular analysis may assist genebank curators in making decisions about germplasm acquisition and the formation of core collections. The added value of accessions to a collection may be estimated by comparing the fingerprints of candidate accessions with those of the current

genebank accessions. Candidates that add the most in terms of diversity can then be included in the collection. In the assembly of core collections, the fingerprints of available accessions may be used to select those subsets with the highest diversity. Since both for acquisition as for core collection formation the objective is to maximise the allelic diversity of useful traits, the question is to what extent this diversity is reflected by molecular marker data. Using pedigree information, a measure for allelic diversity can be obtained by quantifying the expected number of alleles not identical by descent.

To calculate the probability of non-identical descent of alleles using pedigree data, Hintum and Haalman (1994) introduced the 'effective number of origin lines' (n_{ol}), which was defined as the average number of alleles per locus, not identical by descent, in a set of lines. Correlations between the diversity estimated from DNA fingerprinting and the effective number of origin lines calculated from pedigree data could therefore be used to validate the results obtained from molecular data. This validation approach was followed in the present study, which employed an AFLP analysis of 52 barley cultivars, well documented with pedigree data. Our main goal was to evaluate the utility of a marker-based protocol to assist genebank curators in acquisition decisions and core collection assembly.

Materials and methods

Study material

The barley collection of the Centre for Genetic Resources, the Netherlands presently consists of 3436 accessions, of which 757 are advanced cultivars. Fifty two of these cultivars were examined in the present study. The analysed material can be divided into two groups. The first group consisted of 27 modern cultivars (year of release ≥ 1980) that have been important for cultivation in The Netherlands, and a second group of 25 older cultivars (year of release < 1980) consisting of well-known varieties that have been historically important both for cultivation and plant breeding in The Netherlands (Table 1). Information about the cultivars can be accessed via the Internet (<http://www.cgn.wageningen-ur.nl/pgr/>). Seedlings from

Table 1. Investigated varieties with their country of origin, and ordered according to the year of origin. The varieties are subdivided into spring and winter barleys.

Variety name	Country of origin	Year of origin
<i>Spring barley</i>		
Kenia	Denmark	1932
Balder	Sweden	1942
Agio	The Netherlands	1950
Volla	Germany	1957
Cambrinus	The Netherlands	1962
Emir	The Netherlands	1962
Delisa	The Netherlands	1965
Diamant	Czech Republic	1965
Zephyr	The Netherlands	1965
Sultan	The Netherlands	1966
Julia	The Netherlands	1968
Mazurka	The Netherlands	1969
Ofir	The Netherlands	1970
Aramir	The Netherlands	1972
Maris Mink	United Kingdom	1973
Trumpf	Germany	1973
Piccolo	The Netherlands	1977
Havila	The Netherlands	1979
Melody	The Netherlands	1979
Claret	United Kingdom	1980
Iban	The Netherlands	1981
Apex	The Netherlands	1982
Bellona	The Netherlands	1982
Efron	The Netherlands	1985
Prisma	The Netherlands	1986
Robin	The Netherlands	1986
Grosso	The Netherlands	1988
Magda	The Netherlands	1990
Triangel	The Netherlands	1990
Riff	The Netherlands	1993
Reggae	The Netherlands	1994
Ardila	The Netherlands	1996
Extract	United Kingdom	1997
Video	The Netherlands	1997
Luzon	The Netherlands	1998
<i>Winter barley</i>		
Herfordia	Germany	1950
Pella	The Netherlands	1964
Malta	Germany	1968
Alpha	France	1972
Banteng	The Netherlands	1973
Birgit	Germany	1978
Tapir	The Netherlands	1980
Flamenco	France	1982
Masto	The Netherlands	1983
Marinka	The Netherlands	1985
Oribi	The Netherlands	1988
Alpaca	The Netherlands	1989
Intro	The Netherlands	1990
Vectra	The Netherlands	1990
Noveta	The Netherlands	1991
Anoa	The Netherlands	1995
Cumbia	The Netherlands	1995

each of the selected cultivars were grown under greenhouse conditions, and leaf material was collected from about 2-week-old plants. Because barley cultivars can be expected to be largely homogeneous (e.g. Treuren and Hintum 2001), leaf material (ca. 100 mg) from 10 different individuals was bulked together to represent a cultivar sample. From two cultivars ('Extract' and 'Noveta'), a second bulk of 10 was made in order to test reproducibility. In addition, five individual plants from each of three cultivars ('Julia', 'Mazurka' and 'Sultan') were sampled in order to check for cultivar homogeneity. All samples were collected in 2-ml Eppendorf tubes and stored at -80°C . DNA isolation basically followed the protocol of Fulton et al. (1995). Further details can be found in Treuren and Hintum (2001).

AFLP fingerprinting

Basic AFLP procedures were carried out according to the methods described by Vos et al. (1995), and in further detail by Treuren and Hintum (2001). All samples were analysed for the primer *EcoRI* + AC (E12) in combination with each of the primers *MseI* + CCA (M51) and *MseI* + CCT (M54). These two pairs were found to be appropriate primer combinations for AFLP analysis in barley in earlier research (Treuren 2001). PCR products radiolabelled with P^{33} were separated by polyacrylamide gel-electrophoresis.

Data analysis

AFLP data

Autoradiograms were manually scored for the presence or absence of AFLP fragments in the range of 50–500 bp. Differently sized AFLP fragments were assumed to represent different loci, each potentially having two alleles, i.e. band presence and absence. Variation in band presence at a locus was recorded as a genetic polymorphism. Genetic relationships between cultivars were expressed by Jaccard's similarity values, and visualised by a Principal Coordinate Analysis (PCO) using the Genstat 6 software package (release 6.1.0.200). For groups of cultivars, two measures of genetic diversity were estimated; first, the total number of genetic polymorphisms (P);

and second, the average gene diversity per locus (H), calculated as $2*(p*(1-p))$, in which p represents the frequency of a band at a locus. The parameter average gene diversity is analogous to the mean expected heterozygosity under Hardy-Weinberg Equilibrium (e.g. Nei 1987). P and H were used to relate the molecular data to pedigree information.

Pedigree data

Most of the pedigree data of the 52 cultivars and their parents were taken from Baumer and Cais (2000). Some additional data were extracted from CGN's database (www.cgn.wageningen-ur.nl/pgr/) and the database of the USA National Plant Germplasm System (www.ars-grin.gov/npgs). This resulted in 135 pedigrees involving 218 cultivars. These data were used to calculate the 'effective number of origin lines' (n_{ol}) in sets of cultivars via tailor-made computer programmes written in MS Visual Basic in an MS Excel environment. Details about the model and calculation methods are given by Hintum and Haalman (1994).

Relationships between the data types

To investigate the reliability of the pedigree information and the AFLP fingerprints, five varieties were examined for which both the parents had also been fingerprinted in the present study. These were the two winter varieties, 'Alpaca' (= 'Banteng' * 'Tapir') and 'Anoa' (= 'Masto' * 'Birgit'), and the three spring varieties, 'Aramir' (= 'Volla' * 'Emir'), 'Ofir' (= 'Volla' * 'Emir') and 'Robin' (= 'Trumpf' * 'Maris Mink'). In all cases, both parents had the same seasonality as the offspring. The number of cases where both parents showed a band, while the offspring did not, was determined, along with the opposite case. These cases were considered 'errors'. To estimate the probability that these observations could be explained by chance alone, it was calculated how often an observed number of errors, or less, would result from considering all other varieties of the same seasonality as offspring, while assuming the same parents. Additionally, for each case where parents were polymorphic, the parent that contributed to the offspring was determined. The probability was then estimated that these or more skewed contributions would occur by chance assuming an equal contribution of both parents to the offspring and a binomial distribution.

The usefulness of a marker-based protocol to assist in acquisition decisions and the formation of core collections was investigated by relating two molecular diversity measures, P and H, with the effective number of origin lines (n_{ol}) in sets of cultivars. First, n_{ol} values were calculated for cultivar groups of different size, starting with two cultivars, and followed by the addition of one cultivar after each simulation. Simulations were performed in different ways. In the 'random' approach, cultivars were chosen at random, and the average n_{ol} of the group calculated over 10,000 replicated simulations. This approach showed the development of n_{ol} with increasing group size if no external information was applied to the decision. In the 'maximum' approach, varieties were not chosen at random, but groups were each time extended with the variety that increased n_{ol} the most. In the 'minimum' approach, the opposite procedure was followed, i.e. each time a variety was added that increased n_{ol} the least. The 'maximum' and 'minimum' approach represented the best and worst case scenarios and set the boundaries within a manager could operate. In the 'historical' approach, cultivars were added based on their year of origin, showing a time trend in the development of n_{ol} when cultivars were added in order of year of appearance.

Second, P, H and n_{ol} were calculated for the group of cultivars released before 1980. This set of varieties was regarded as a small collection, to which obsolete varieties from 1980 or later were to be added. N_{ol} values were again calculated following the 'random' (10,000 simulations), 'maximum', 'minimum' and 'historical' approach, but also following an optimisation procedure by using P and H, respectively. In the latter case, varieties were added sequentially to the collection, their order being determined by the largest increase in the genetic diversity of the extended collection. The added value of molecular data in acquisition decisions was then measured by comparing the marker-based approach with the random approach following the addition of cultivars to the pre-1980 collection.

Third, by using the entire set of 52 cultivars, n_{ol} values were calculated for groups of cultivars optimised with P and H, respectively, and compared to those based on the random approach. This comparison was conducted to reveal the added value of molecular data in core collection formation.

Results

AFLP data

A total of 144 bands was scored for the two primer combinations, of which 75 (52%) were polymorphic within the total sample analysed. All 52 cultivars could be distinguished based on the two primer pairs used. The two pairs of bulks of 10 plants showed identical results for all 75 polymorphic bands for each of the cultivars 'Extract' and 'Noveta'. For both 'Julia' and 'Sultan' the bulk of 10 plants displayed identical AFLP profiles to all five individually sampled plants. These results indicated the robustness of the AFLP analyses and a high level of homogeneity within these tested cultivars. Only in case of the cultivar 'Mazurka' was heterogeneity detected; two AFLP polymorphisms were observed between the bulk of 10 plants and the five individually sampled plants which did not show any variation *inter se*.

A PCO cluster analysis separated the 52 cultivars in two main groups on the first principal axis, explaining 27.8% of the observed variation (Figure 1). The first principal axis distinguished all but one of the winter types from the spring type cultivars, indicating the strength of AFLPs to discriminate between the two crop types. Only the cultivar Alpha, registered as a winter barley,

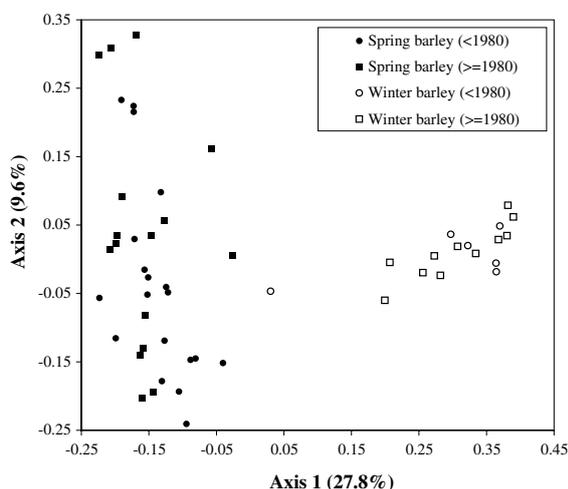


Figure 1. PCO cluster analysis of the 52 investigated barley varieties based on AFLP data. Crop type (spring or winter barley) and year of origin (<1980 or ≥1980) are depicted by different symbols. The percentage of variation explained by each axis is given between parentheses in the axis legends.

displayed a somewhat intermediate position between the spring and winter barleys. For the investigated cultivars a discriminating band was observed using the primer combination E12/M51, as that fragment was present in all spring barleys but absent in all winter barleys. A lower level of genetic heterogeneity was observed among the winter types, suggesting a narrower gene pool within this group of barley cultivars. No clear differences were found between the material originating from before 1980 and the more recent cultivars, suggesting a relative lack of divergence between the older and more recent cultivars. Inspection of the AFLP data showed that 61 of the 75 polymorphisms were represented in the 25 varieties from before 1980, and hence that the 27 more recent cultivars contributed 14 additional polymorphisms to this set. A randomly chosen set of 25 varieties resulted in an average number of polymorphisms of 64.73. The fact that this value was only slightly higher than the number of polymorphisms within the set of 25 varieties from before 1980 again indicated limited genetic differences between the cultivars released before and after 1980.

Pedigree data

By definition, the effective number of origin lines (n_{ol}) equals 1 for a single homogeneous, homozygous cultivar (Hintum and Haalman 1994). Each time the group size is increased by one, the maximum increase in n_{ol} is 1, namely when the new cultivar is completely unrelated to the others. Thus, if the 52 analysed cultivars would be completely unrelated, the n_{ol} of the entire set would be 52. The n_{ol} of the total set of 52 cultivars analysed was 23.9, indicating overlap in genetic background. The increase in n_{ol} with increasing group membership following the maximum, minimum, random and historical approaches is presented in Figure 2. Because of the increase of group size with randomly chosen varieties and the large number of replications performed, the random line followed a smooth, steady increase of n_{ol} from 1 to 23.9. Evidently, all approaches converged to this same value. Instead of a steady increase of n_{ol} , the historical line rather followed a stepwise pattern, indicating the periodic release of cultivars with similar genetic backgrounds, alternating with the

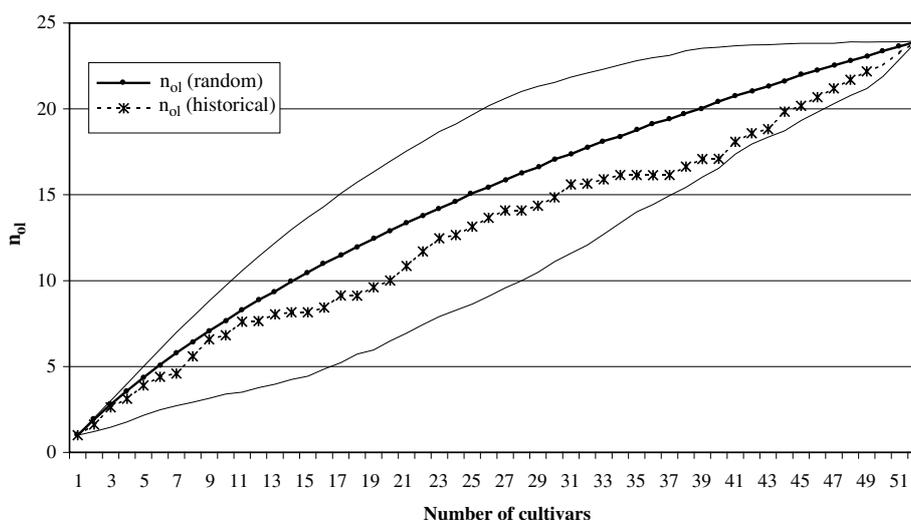


Figure 2. Scatter diagram of the effective number of origin lines (n_{oi}) against the number of cultivars. The top and bottom lines represent the maximum and minimum n_{oi} if cultivars are added that result in the largest and smallest increase in n_{oi} , respectively. For the random line, cultivars are added at random, and, for the historical line, cultivars are added in order of their year of origin.

incorporation of novel germplasm. The periodic incorporation of new material was also shown by the fact that the historical line did not resemble the development of n_{oi} following to the maximum approach. For the group of 25 varieties from before 1980 the n_{oi} was 13.2. Thus, the n_{oi} of this set was increased by 10.7 by adding the 27 more recent cultivars.

Reliability of pedigree and AFLP data

In five cases, a comparison of AFLP data was possible between varieties and both of their parents (Table 2). In three cases ('Alpaca', 'Anoa' and 'Robin'), polymorphisms were observed in the offspring that could not be explained based on the parental AFLP profiles. These results may

be attributed to the occurrence of 'noise' due to intra-varietal variation, mutation, technical inconsistencies or scoring errors of the fingerprints. In all five cases, the low probabilities of achieving identical results based on random sampling indicated a high likelihood that the correct parents were involved. Except in the case of 'Robin', the contribution of the parents to the offspring was found to be rather skewed, suggesting the disproportionate selection of genes from the parents or the involvement of backcrossing during the ensuing breeding process.

Acquisition

To investigate the utility of AFLP data in support of acquisition, the 25 varieties from before 1980

Table 2. AFLP-based validation of pedigree data for varieties that had both parents (P_1 and P_2) included in the study.

Variety	P_1	P_2	# errors	p_{errors}	# Polymorphisms P_1	# Polymorphisms P_2	$p_{\text{distribution}}$
Alpaca	Banteng	Tapir	2	0.07	22	6	0.00
Anoa	Masto	Birgit	2	0.07	15	3	0.00
Aramir	Volla	Emir	0	0.03	5	12	0.07
Ofir	Volla	Emir	0	0.03	16	1	0.00
Robin	Trumpf	Maris Mink	1	0.03	10	8	0.41

The number of AFLP Polymorphisms in a variety that could not be explained based on the AFLP profile of the parents is given by '# errors'. ' p_{errors} ' denotes the probability that a randomly chosen variety of the same seasonality, while assuming the same parents, would result in an equal or lower '# errors'. The numbers of polymorphisms in the offspring resulting from P_1 and P_2 are indicated by # Polymorphisms P_1 and P_2 . $p_{\text{distribution}}$ denotes the probability of this distribution, or a more skewed one, assuming an equal contribution from both parents.

were regarded as a basic collection and the 27 more recent cultivars as potential candidates for acquisition. Since in most cases, potential candidates will not be completely identical to already accessed material, new accessions could be expected to increase the diversity of a collection, even if they were chosen at random. Therefore, the question was not so much whether additions to a collection result in an increase of genetic diversity, but rather whether the slope of the random line, representing the development of genetic diversity until 1980, was significantly increased thereafter. A slight increase of this slope after 1980 was observed for both the parameters P and H (results not shown), reflecting the potential value of recent cultivars in diversifying the basic collection to some extent. Notably, the 14 polymorphisms that are new to those contained in the basic collection could be achieved by including only nine new accessions.

For validation purposes, the marker-based approach was related to the development of n_{oi} for increased collection sizes (Figure 3). These analyses showed that, in general, the marker-based approach did not perform substantially better than did the random approach. Only during the first two extensions of the basic collection of 25 accessions, n_{oi} values based on a P-optimised approach

followed the maximum approach. The P-optimised approach could not be continued after the addition of nine accessions because at that point the total number of 75 polymorphisms was reached. At this point, the P-optimised strategy performed equally well as the random strategy, increasing the n_{oi} value of the extended collection to only 16.93. The H-optimised approach basically followed the random line throughout. A problem that was encountered in using parameter H was the fact that cultivar additions may lead to a reduction in genetic diversity. For a single AFLP locus, H has its maximum value if band absence and presence occur in equal frequency. Therefore, the addition of cultivars that increase frequency differences will result in a reduction of H. Given the poor performance of the P- and H-optimised approaches, a marker-based protocol to assist in acquisition decisions was not considered particularly useful for the analysed material.

Core collection formation

In the formation of core collections, the aim is to assemble a subset of accessions that contains the major part of the diversity present within a collection. To investigate the utility of AFLPs in core

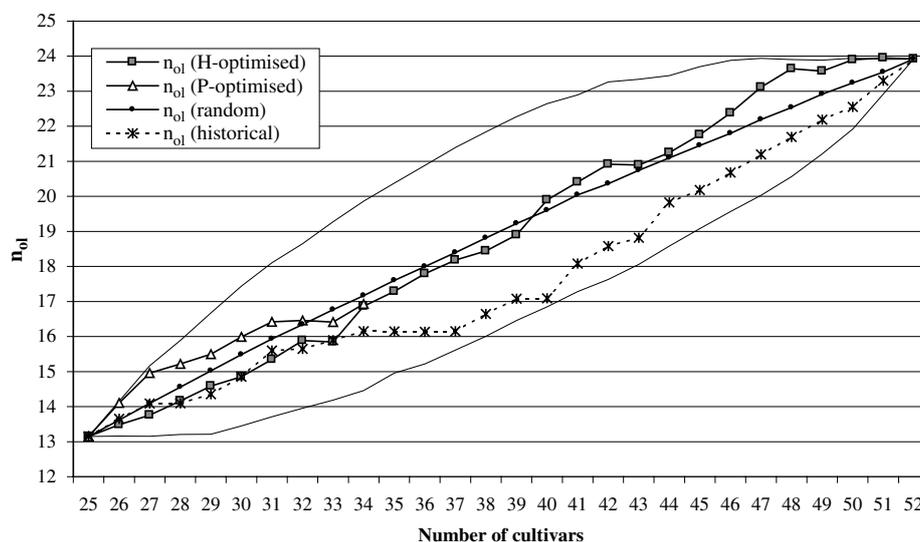


Figure 3. Scatter diagram of the effective number of origin lines (n_{oi}) against the number of cultivars from 1980 or later, added to a basic collection consisting of 25 cultivars released before 1980. The top and bottom lines follow the maximum and minimum increase in n_{oi} . The increase of n_{oi} was also optimised based on the diversity measures, P and H, as calculated from the AFLP data. Also, the random and historical lines are depicted.

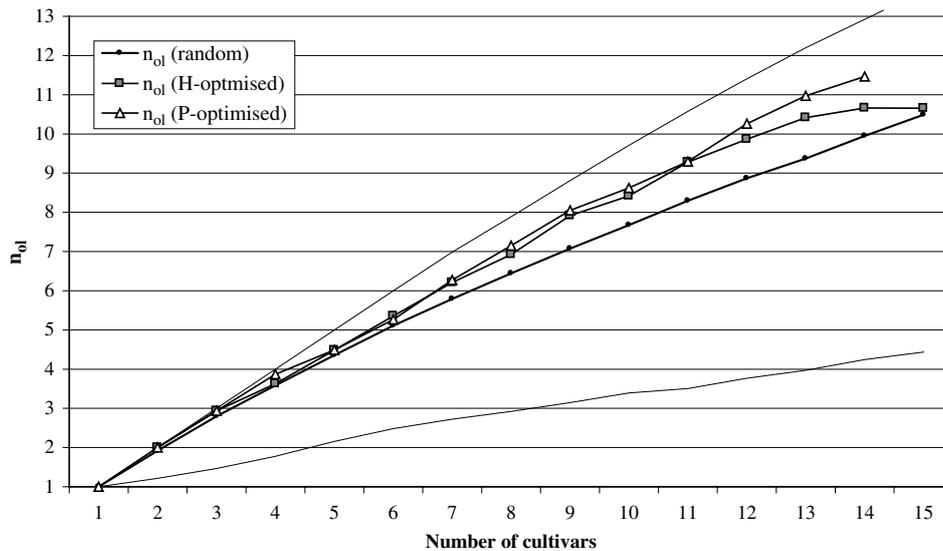


Figure 4. Scatter diagram of the effective number of origin lines (n_{ol}) against the number of cultivars. The top and bottom lines follow the maximum and minimum increase in n_{ol} . The increase in n_{ol} is also given for the 'random' approach and for the strategies using an optimisation procedure with P and H, respectively, as calculated from the AFLP data. The lines truncate at the point where optimisation was no longer possible.

collection formation, the 52 barley cultivars were regarded as a collection. It was then examined whether AFLPs could assist in optimising the effective number of origin lines within subsets of cultivars (Figure 4). Particularly for parameter P, the marker-based approach performed substantially better than did the random approach. The P-optimised line showed n_{ol} values that were halfway in between the random and maximum approach. All 75 AFLP polymorphisms could be captured by the selection of 14 cultivars, resulting in a n_{ol} value of 11.46, being 88.4% of the maximum n_{ol} (12.97) at this point. Using the P-optimised approach, 48% of the total number of origin lines (23.92) could be captured with 27% of the collection, indicating the value of AFLPs in core collection formation from the investigated material.

Discussion

The utility of AFLP-based protocols to assist in acquisition decisions and core collection formation was assessed by measuring the effectiveness of various protocols in relation to the effective number of origin lines, as calculated from pedigree data. The outcome of such analyses depends on the resolving power of the AFLPs, the appropri-

ateness of n_{ol} to reflect true co-ancestry and the correlation between the AFLP and pedigree data.

The resolving power of AFLPs was demonstrated in previous research, as 12 different genotypes could be distinguished among 36 individuals from a barley landrace by using just one primer combination (Treuren and Hintum 2001). In another study using eight primer pairs on 31 barley cultivars, one primer combination was generally found to be sufficient to distinguish all analysed cultivars, and AFLPs could easily discriminate between spring and winter crop types (Schut et al. 1997). Similar results were observed in the present study by using 75 markers from two AFLP primer combinations. A strikingly similar separation of winter and spring barley has also been reported based on RFLPs (Melchinger et al. 1994; Casas et al. 1998). For an overview of the use of various marker technologies for the description of diversity in barley see Graner et al. (2003).

Concerning crop-type discrimination, only the cultivar 'Alpha' displayed a somewhat intermediate position between the spring and winter crop types. It appeared from CGNs database that this cultivar is represented twice in the collection. One accession is documented as a winter type originating from France, whereas the other is listed as a spring type originating from the US. Moreover,

analysis of the pedigree data showed that the winter type 'Alpha' had spring type barley in its pedigree. Therefore, the analysed cultivar 'Alpha' may indeed represent an intermediate type. The AFLP results also indicated a narrower gene pool for the winter type than for the spring type barley. These results were in accordance with those of a study screening 48 barley cultivars with RFLPs (Melchinger et al. 1994).

In the calculation of the effective number of origin lines, violations of the assumptions underlying the calculation might have compromised the appropriateness of n_{oi} to reflect true co-ancestry. These assumptions were: (i) cultivars receive half of their genes from each parent, (ii) parents in crosses are homozygous and homogeneous, (iii) ancestors are considered unrelated if pedigree data are lacking, and (iv) the coefficient of parentage between a cultivar and a selection from that cultivar equals 0.75. As pointed out by Hintum and Haalman (1994) all four assumptions can be disputed to varying extents. For example, because of selection, cultivars may not always receive half of the genes from each parent. Also, original ancestors may come from similar collection areas and hence share some genetic background, although they are assumed to be completely unrelated (see also Schut et al. 1997). In addition to violation of the underlying assumptions, inaccurate pedigree data may result in incorrect estimates of n_{oi} . As was shown from the results of Table 2, back-crossing in plant breeding is often not mentioned. Nevertheless, calculation of the effective number of origin lines showed an overlap of only 2.9 between the spring and winter types in a study using 85 modern barley cultivars, indicating the appropriateness of n_{oi} to reflect genetic differentiation at the level of crop type (Hintum and Haalman 1994).

For the analysed material, the AFLP results correlated well with pedigree data in core collection formation, but poorly in extending the genetic diversity of collections. However, poor performance for the latter is partly due to the way the two groups of varieties were constituted. Because the group of older varieties consisted of the more common varieties that have been important for plant breeding in The Netherlands, the two groups could be expected to be genetically related, rather than representing two independent samples. Our results indicated that AFLP analysis in barley is a

quite appropriate tool to select genetically distinct material, but suggested less applicability when dealing with more closely related germplasm. These findings were in line with the significant, but low, correlation coefficient that was found between genetic similarity based on 681 AFLP markers and the coefficient of co-ancestry based on pedigree data in a study using 25 European two-row spring barley cultivars (Schut et al. 1997). The correlation in that study appeared to improve considerably when European winter barleys and North American spring barleys were included. Poor-moderate correlations with pedigree data have also been found for RFLP data in barley (Ellis et al. 1997; Casas et al. 1998). The poor correlation between the AFLP results and the effective number of origin lines may be partly due to the fact that violation of the assumptions in the calculation of n_{oi} tends to underestimate relatedness (higher n_{oi}), but the extent of this effect will differ depending on the composition of the group studied.

The 14 accessions of the core collection assembled by the P-optimised approach included the cultivars 'Mazurka', 'Noveta', 'Ardila', 'Marinka', 'Balder', 'Apex', 'Triangel', 'Luzon', 'Masto', 'Alpha', 'Banteng', 'Tapir', 'Claret' and 'Riff'. Most curators would consider this selection to be over-represented by the modern varieties because they usually tend to emphasise the presence of well-known and older cultivars in a core. However, a questionnaire sent by CGN to Dutch breeders in 2001 revealed that the older varieties are not regarded particularly useful for breeding purposes (Soest 2001). Breeders often prefer to use modern cultivars in their crosses because they are the most elite and may contain novel genetic diversity from crop-related wild species. In that sense, the use of AFLPs in the investigated group of barley cultivars contributed to the formation of a core preferred by end-users.

Acknowledgements

This work was supported by a fellowship (project 18844, CA/29505) granted to I. Tchoudinova, by the International Agricultural Centre on behalf of The Netherlands Ministry of Agriculture, Nature Management and Fisheries. The authors are grateful to Peter Goossens and Noor Bas for their technical involvement during the project. We

thank two anonymous referees for their suggestions to improve an earlier version of the manuscript.

References

- Baumer M. and Cais R. 2000. Abstammung der Gerstensorten. Bayerische Anstalt für Bodenkultur und Pflanzenbau. Freising, Germany.
- Bretting P.K. and Widrechner M.P. 1995. Genetic markers and plant genetic resource management. *Plant Breed. Rev.* 13: 11–86.
- Casas A.M., Igartua E., Valles M.P. and Molina-Cano J.L. 1998. Genetic diversity of barley cultivars grown in Spain, estimated by RFLP, similarity and coancestry coefficients. *Plant Breed.* 117: 429–435.
- Ellis R.P., McNicol W., Baird E., Booth A. and Lawrence P. 1997. The use of AFLPs to examine genetic relatedness in barley. *Mol. Breed.* 3: 359–369.
- FAO. 1996. *FAO State of the World's Plant Genetic Resources for Food and Agriculture*. Food and Agriculture Organisation of the United Nations, Rome.
- Frankel O.H. and Brown A.H.D. 1984. Plant genetic resources today: a critical appraisal. In: Holden J.H.W. and Williams J.T. (eds), *Crop Genetic Resources: Conservation and Evaluation*, Allen and Unwin. Winchester, Massachusetts, USA.
- Fulton T.M., Chunwongse J. and Tanksley S.D. 1995. Micro-prep protocol for extraction of DNA from tomato and other herbaceous plants. *Plant Mol. Biol. Rep.* 13: 207–209.
- Graner A., Bjørnstad Å., Konishi T. and Ordon F. 2003. Molecular diversity of the barley genome. In: von Bothmer R., Sato K., Knüpfner H. and van Hintum Th.J.L. (eds), *Diversity in Barley (Hordeum vulgare)*. Elsevier, Amsterdam, pp. 121–142.
- van Hintum Th.J.L., Brown A.H.D., Spillane C. and Hodgkin T. 2000. Core collections of plant genetic resources. IPGRI Technical Bulletin No. 3. International Plant Genetic Resources Institute, Rome, Italy.
- van Hintum Th.J.L. and Haalman D. 1994. Pedigree analysis for composing a core collection of modern cultivars, with examples from barley (*Hordeum vulgare* s. lat.). *Theor. Appl. Genet.* 88: 70–74.
- McGregor C.E., van Treuren R., Hoekstra R. and van Hintum Th.J.L. 2002. Analysis of the wild potato germplasm of the series *Acaulia* with AFLPs: implications for *ex situ* conservation. *Theor. Appl. Genet.* 104: 146–156.
- Melchinger A.E., Graner A., Sing M. and Messmer M.M. 1994. Relationships among European barley germplasm: I. Genetic diversity among spring and winter cultivars revealed by RFLPs. *Crop Sci.* 34: 1191–1199.
- Nei M. 1987. *Molecular Evolutionary Genetics*. Columbia University Press, New York.
- Schut J.W., Qi X. and Stam P. 1997. Association between relationship measures based on AFLP markers, pedigree data and morphological traits in barley. *Theor. Appl. Genet.* 95: 1161–1168.
- van Soest L.J.M. 2001. Promotion of the utilisation of plant genetic resources in the Netherlands. In: Świecicki W., Naganowska B. and Wolka B. (eds), *Broad Variation and Precise Characterisation – Limitation for the Future*. Proceedings of the XVIth EUCARPIA Section Genetic Resources Workshop, May 16–20, 2001, Poznan, Poland, pp. 358–362.
- van Treuren R. 2001. Efficiency of reduced primer selectivity and bulked DNA analysis for the rapid detection of AFLP polymorphisms in a range of crop species. *Euphytica* 117: 27–37.
- van Treuren R. and van Hintum Th.J.L. 2001. Identification of intra-accession genetic diversity in selfing crops using AFLP markers: implications for collection management. *Genet. Res. Crop Evol.* 48: 287–295.
- Vos P., Hogers R., Bleeker M., Reijans M., van de Lee T., Hornes M., Frijters A., Pot J., Peleman J., Kuiper M. and Zabeau M. 1995. AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Res.* 23: 4407–4414.