Theor Appl Genet (1997) 95 : 1161–1168

J. W. Schut · X. Qi · P. Stam

Association between relationship measures based on AFLP markers, pedigree data and morphological traits in barley

Received: 24 February 1997 / Accepted: 23 June 1997

Abstract Thirty one barley lines were used to investigate the agreement between three relationship measures: genetic similarities based on 681 AFLP-markers, coefficients of co-ancestry based on pedigree data, and generalised distance based on 25 morpological characters (morphological distance). Bootstrap analysis was used to estimate the accuracy of the correlation estimates. AFLP-based genetic similarities showed a poor-to-moderate correlation with the coefficients of co-ancestry within the core set of 25 European two-row spring barleys. Morphological distance was not significantly correlated with either genetic similarity or the coefficient of co-ancestry. The precision of all correlation-coefficient estimates, however, was low. The inclusion of two European winter barleys, two North American two-row spring barleys, and two North American six-row spring barleys in the AFLP-analysis resulted in a much stronger correlation between genetic similarity and the coefficient of co-ancestry. This suggests good opportunities for the use of AFLP-markers to assess genetic diversity by distinguishing between the major ecotypes of barley. Additionally, each of the eight primer combinations used in the AFLP-analysis was able to identify all 31 lines uniquely, showing the usefulness of AFLPs for cultivar identification. Because of the inaccuracy of the investigated relationship measures, resulting in low values of the correlation-coefficient estimates, prediction of the breeding behaviour of parent combinations may be improved by the use of a combination of relationship measures, thus decreasing the effect of their individual independent errors.

Communicated by F. Salamini

J. W. Schut (✉) · X. Qi · P. Stam
C.T. de Wit Graduate School for Production Ecology,
Department of Plant Breeding, Wageningen Agriculture University,
P.O. Box 386, 6700 AJ Wageningen, The Netherlands
Fax: +31 317 48 34 57
E-mail: johan.schut@users.pv.wau.nl

Key words Bootstrap analysis · Coefficient of co-ancestry · Cultivar identification · Genetic similarity · *Hordeum vulgare*

Introduction

Knowledge about relationships between genotypes that may be used in new crosses, and about genetic diversity in available germplasm, is very useful for plant breeders. It supports their decisions on the selection of crossing combinations from large sets of parent genotypes and is helpful when they want to widen the genetic basis of a breeding program. The selection of crossing combinations is supported by prediction of the performance of offspring resulting from crossing combinations between inbred parents. Cowen and Frey (1987) distinguished three types of breeding behaviour that can be predicted: heterosis, transgressive segregation, and genetic variance among offspring. However, higher transgressive segregation fractions are a direct result of higher genetic variances among offspring, taking into account the difference between the performances of the parents. Therefore, they do not have to be considered as separate phenomena. On the other hand, heterosis and genetic variance are a direct and an indirect result of heterozygous loci in the F_1 and their effects. The degree of relationship between two parent genotypes is mainly expected to predict the number of heterozygous loci in their hybrid.

Knowledge about genotype relationships is usually based on three sources of information: (1) geographic information about the origin of genotypes, (2) pedigree information, and (3) information about plant characteristics. Geographic information is helpful in most cases. It is specifically used when other information on genotypes is either not available or else is very sparse. This is often the case for gene-bank material. Pedigrees of varieties and breeding lines are often well

documented. They trace back to landraces and wild accessions. However, pedigrees sometimes contain erroneous or incomplete information. Plant characteristics are the only source of relationship information that is, or can be made, available for any set of genotypes. Such characteristics can be divided into four arbitrary groups: agronomic characters, morphological characters (used to distinguish between varieties), biochemical characters (e.g. storage proteins, isozymes), and molecular (DNA) markers. Differences between genotypes with regard to any of these characteristics are either indirect or direct representations of differences at the DNA level and are therefore expected to provide information about genetic relationships.

A range of measures is available to quantify relationship information. For pedigree information Malécot (1948) presented the coefficient of co-ancestry (f), also known as the kinship coefficient or the coefficient of parentage. For agronomic and morphological traits measured on a continuous or ordinal scale one can use multivariate statistical techniques and construct a p-dimensional space, where p is the number of traits. The Euclidean distance between the points representing the genotypes may be used as a measure of relatedness (Goodman 1972). Generalised distance is an extension of Euclidean distance correcting for correlations between traits (Mahalanobis 1936). Plant characteristics, like isozymes and molecular markers, are scored as binary data. A commonly used similarity measure was presented by Dice (1945). Nei and Li (1979) demonstrated the usefulness of this genetic similarity for isozyme and molecular-marker data.

In winter wheat, Cox et al. (1985 b) found a poor correlation between the coefficient of co-ancestry and the genetic distance based on storage-protein data. Genetic distance based on combined isozyme and morphological data showed a moderate correlation with pedigree-based distance in soybean (Cox et al. 1985 a). However, Souza and Sorrells (1991 a, b) concluded that distance measures based on quantitative and qualitative morphological characters (the latter including isozyme characters) in oats did not correspond very well with pedigree data. The introduction of molecular markers like RFLP (restriction fragment length polymorphism; Botstein et al. 1980) and RAPD (random amplified polymorphic DNA; Williams et al. 1990) created the opportunity to assess genetic relationships directly at the DNA level. A priori, similarities at the DNA level are expected to be in better agreement with pedigree information than similarities based on morphological traits or gene products, whose expression can be influenced by the environment and/or epistatic interactions. However, results based on RFLP- and RAPD-markers are quite variable in this respect. Tinker et al. (1993) showed a moderate correlation between RAPD-based genetic distance and the coefficient of co-ancestry when considering 27 Canadian spring barley lines. Graner et al. (1994) used RFLP data to estimate genetic distances between 48 European barley varieties. They found poor-to-moderate correlations between marker-based distance and the coefficient of co-ancestry. Correlations were higher in spring barley than in winter barley. Autrique et al. (1996) obtained a moderate correlation between RFLP-based genetic distance and distance based on 16 agronomic and morphological traits in durum wheat. Correlations between these two relationship measures and the coefficient of co-ancestry were poor.

Recently AFLP, a PCR-based molecular marker technology, was introduced by Vos et al. (1995). Using PCR-amplification, genomic restriction fragments are selectively multiplied to adequate detection levels, producing reproducible DNA-fingerprint patterns. The fast and reliable production of many marker data points is an advantage of AFLP over RFLP and RAPD.

The aim of the present study is to investigate the agreement of AFLP-based genetic similarities, coefficients of co-ancestry, and generalised distance based on morphological characters. Special attention will be paid to the precision of the correlation estimates. We also discuss, in brief, the usefulness of AFLPs for variety identification as well as opportunities for the prediction of the breeding behaviour of crosses and the assessment of genetic diversity.

Materials and methods

Plant materials

Thirty one barley (*Hordeum vulgare* L.) lines were used in this study. The core set consisted of 25 European two-row spring barley varieties and breeding lines. They were chosen to represent parent populations employed in commercial spring barley breeding programs in Northwest Europe over the last 20 years. Firstly, relationship measures were compared for this core set. Secondly, a set of six cultivars consisting of two European winter barleys (one two-row; one six-row) and four North American spring barleys (two two-row; two six-row) were added as representatives of some other major barley groups. This offered an opportunity to investigate possibilities for the assessment of genetic diversity using AFLPs. The names and details of the 31 lines are presented in Table 1.

AFLP analysis

DNA extraction followed the CTAB-method described by Van der Beek et al. (1993).

The AFLP technique is described by Vos et al. (1995): DNA restriction uses the enzyme combination *Eco*RI/*Mse*I. After adapter ligation, DNA fragments are amplified using PCR. Primer annealing is targeted at the adapter and restriction-site sequence. Three-nucleotide extensions on both *Eco*RI and *Mse*I primers cause selective amplification of fragments. The AFLP-analysis followed the protocol described by Van Eck et al. (1995) with modifications by Qi and Lindhout (1997).

Primer combinations were chosen that produce a high number of unambiguous polymorphisms in a wide range of barley germplasms (Qi and Lindhout 1997). The eight primer combinations that were used are presented in Table 2.

Table 1 Genotypes used in AFLP analysis, their pedigree, country of origin, type (2 = two-row; 6 = six-row; s = spring barley; w = winter barley), the possibility to trace the pedigree to original ancestors (x-mark means: more than 75% of the pedigree can be traced to original ancestors) and the availability of morphological trait data (x-mark means: data available). The dotted line divides the core set of European two-row spring barleys from the rest

Genotype	Pedigree	Country of origin	2-row/ 6-row type	Spring/ winter type	> 75% known pedigree	Morphological-trait data available
Apex	Aramix × (CEB-6711 × (Julia(3) × (Volla × L-100))	The Netherlands	2	s		x
Aramir	Volla × Emir	The Netherlands	2	s	x	
Baronesse	5238/8-74 × 754465	Germany	2	s		x
Bonaire	E-77040-8107 × (CEB-8188 × Apex)	The Netherlands	2	s	x	x
CEB-9079	Robin × (CEB-8498 × Efron)	The Netherlands	2	s	x	x
CEB-9186	CEB-8187 × Golf	The Netherlands	2	s	x	x
Drossel	(FLO-1625/56 × Union) × Ingrid	Germany	2	s	x	x
Forester	CSBM2 × Sherpa	United Kingdom	2	s		x
GEI-119	Aramir-EI × Goldmarker (3)	The Netherlands	2	s		x
Georgie	Vada × Zephyr	United Kingdom	2	s	x	x
Gunhild	(Algerian × Lone) × MGH-63199	Denmark	2	s	x	x
IVP9211327	(GEI-119 × Gunhild) × (Prisma × Apex)	The Netherlands	2	s	x	
IVP9211510	(Prisma × Apex) × (GEI-119 × Gunhild)	The Netherlands	2	s	x	
Karat	K-1443-70 × I-2931170	Czech Republic	2	s	x	x
Kenia	Binder × Gull	Denmark	2	s	x	
Midas	(Proctor × Wong) × mildew-res.A) × mutant of Maythorpe	United Kingdom	2	s	x	x
Nudinka	Emir × Weihenstephan 1606 Nackt	Germany	2	s		
Porthos	Lignee-207 × Emir	France	2	s		x
Prisma	(Trumpf × Cambrinus) × Piccolo	The Netherlands	2	s	x	x
Proctor	Kenia × Plumage Archer	United Kingdom	2	s	x	
Riff	(VDH-240-79 × Karat) × Apex	The Netherlands	2	s	x	x
Triangel	((Mazurka × Ofir) × SVP-6045-66/25) × ((Villa × (Agio × Piroline)) × Carlsberg)	The Netherlands	2	s	x	x
Vada	Hord.laevigatum × Gull	The Netherlands	2	s	x	
Yriba	(Maris Yak × (Rika × Baladi-16-133)) × Rika	France	2	s	x	x
ZE-87-3414	Efron × (Aramir-EI × Iraq-10922)	The Netherlands	2	s	x	x
Franka	((Vogelsanger Gold × Senta) × (Dura × Dea)) × Vogelsanger Gold	Germany	6	w	x	
Harrington	Klages × ((Gazelle × Betzes) × Centennial)	Canada	2	s	x	
Igri	(Malta × Carlsberg 1427) × Ingrid	Germany	2	w		
Morex	Cree × Bonanza	USA	6	s	x	
Steptoe	Wash.Sel.3565 × Unitan	USA	6	s		
TR-306	(Abee × TR451) × WM 793-1776	Canada	2	s	x	

Table 2 Primer combinations used in AFLP analysis

Number	E + 3/M + 3 nucleotide extensions
E33M54	E + AAG/M + CCT
E33M61	E + AAG/M + CTG
E35M54	E + ACA/M + CCT
E35M61	E + ACA/M + CTG
E38M50	E + ACT/M + CAT
E38M59	E + ACT/M + CTA
E38M60	E + ACT/M + CTC
E38M62	E + ACT/M + CTT

Genetic-similarity estimation

AFLP bands were scored as present (1), absent (0), or as a missing observation (−1), for the different genotypes. Often several AFLP-markers within a primer combination show pleiotropic behaviour or very close linkage (Qi and Lindhout 1997). Likewise, in our set of genotypes, polymorphic markers with identical polymorphism patterns were found within primer combinations. We also found markers within primer combinations that seemed to be allelic. In all of these cases a second marker does not add any new independent information to a genetic-similarity estimate. Therefore, these redundant polymorphic markers within primer combinations were discarded before calculating genetic similarities.

The genetic similarities (*gs*) are calculated following Nei and Li (1979):

$$gs_{ij} = \frac{2N_{ij}}{N_i + N_j},$$

where N_{ij} is the number of bands present in both genotypes i and j, N_i is the number of bands present in genotype i, and N_j is the number of bands present in genotype j. In the case of a missing observation for a marker in genotype i and/or j, this marker was not included in the calculation of gs_{ij}. The accuracy of *gs* estimates as influenced by sampling and missing marker data was assessed by taking bootstrap samples (Efron and Tibshirani 1993) from all 681 markers, including polymorphic as well as monomorphic markers. Bootstrap standard-deviation estimates were based on 1000 samples.

Principal-coordinate analysis (Gower 1966) was used to obtain a graphic representation of the relationship structure of the 31 genotypes. Computations were performed using the MDS procedure in SAS (SAS Institute Inc 1992).

Pedigree analysis

Pedigrees of the genotypes were gathered from several sources in the literature (Arias et al. 1983; Baum et al. 1985) and from personal communication with breeders and researchers. The coefficient of co-ancestry f between two genotypes, as defined by Malécot (1948), was calculated. This is the probability that a random allele at a random locus is one genotype is identical by descent to a random allele at the same locus in the other genotype (Cox et al. 1985 b). A FORTRAN-program obtained from Van Hintum (CGN, CPRO-DLO, Wageningen) was used to calculate f. The underlying assumptions are given by Van Hintum and Haalman (1994): (1) a genotype receives half of its genes from each parent; (2) parents involved in crosses are homozygous and homogeneous; (3) ancestors for which no pedigree is available are unrelated; and (4) if selections are made from a cultivar, this cultivar is assumed to be the variable offspring of a cross between two unrelated lines. A selection from the cultivar is one of the offspring lines; if the cultivar itself is said to be used as a parent in a cross, then in fact one of the offspring lines has been used.

Genotypes often lacked some pedigree information. For 23 genotypes (Table 1), of which 19 were core-set genotypes, more than 75% of the pedigree could be traced back to original ancestors, e.g. landraces. The f value of a combination of 2 of these 23 genotypes was defined as 'well known' (f_{wk}) or complete (Graner et al. 1994). The f values of two combinations between a parent and its offspring line were also defined as 'well known', despite the fact that 75% or less of the parent pedigree could be traced back to original ancestors. Due to the direct relationship in this type of combination, the lack of pedigree information about the parent genotype does not have a strong effect on the f value.

Morphological trait analysis

Out of 34 morphological traits in barley described by the international union for the protection of new varieties of plants (UPOV 1981), we had at our disposal data on 25 traits in only 18 lines (Table 1) from the core set of 25 European two-row spring barleys. These data were obtained at our Wageningen site in 1994 in the presence of the relevant UPOV reference cultivars using one-row plots and two replicates. The data were confirmed by a similar trial in 1996. The traits are listed in Table 3.

The observed data were standardised per trait and a principal-components analysis was performed. The principal components having an eigenvalue greater than an arbitrary value $K = 1.0$ were used to calculate the generalised distances (morphological distance, md) between the lines (Goodman 1972).

Bootstrap analysis of correlation coefficients

Simple (r) and rank (r_s) correlation coefficients between genetic similarities (gs), coefficients of co-ancestry (f), and morphological distances (md) were calculated. To test whether correlations were significant we used a bootstrap procedure (Efron and Tibshirani 1993) to estimate 95%-confidence intervals for r. Bootstrap samples were produced by sampling with replacement from the set of genotypes (Schut 1997). Then the gs, f, and md matrices were constructed with rows and columns based on the genotype bootstrap sample. Due to re-sampling of the same genotype, some matrix cells contained a similarity or a distance between a genotype and itself. The contents of these cells were discarded before the calculation of the bootstrap correlation coefficient. For each correlation coefficient a 95%-confidence interval was constructed based on 2000 bootstrap samples. The BC_a method (Efron and Tibshirani 1993) was used to correct for bias and unequal variance in order to obtain a higher accuracy of the interval estimation.

Table 3 Twenty five morphological traits (UPOV 1981) used to calculate morphological distances

Number	Trait
1	Plant: growth habit
2	Lower leaves: hairiness of leaf sheaths
3	Flag leaf: attitude
4	Flag leaf: anthocyanin colouration of auricles
5	Flag leaf: intensity of anthocyanin colouration of auricles
6	Flag leaf: glaucosity of leaf sheath
7	Time of ear emergence (first spikelet visible in 50% of ears)
8	Awns: anthocyanin colouration of the tips
9	Awns: intensity of anthocyanin colouration of the tips
10	Ear: glaucosity
11	Ear: attitude
14	Ear: shape
15	Ear: density
16	Awn: length compared to ear
18	Rachis: length of first segment
19	Rachis: curvature of first segment
20	Rachis: humping of segments (in mid-third of ear)
22	Sterile spikelet: attitude
23	Sterile spikelet: length of lemma
24	Sterile spikelet: shape of tip
25	Median spikelet: length of glume and awn relative to grain
28	Grain: anthocyanin colouration of nerves of lemma
29	Grain: spiculation of inner lateral nerves of lemma
30	Grain: hairiness of ventral furrow
31	Grain: disposition of lodicules

Results

Genetic-similarity estimation

A total of 681 markers were used to estimate genetic similarities and 43.3% of them showed polymorphism in the complete set of 31 genotypes. Restricting the set to 25 European two-row spring barleys yielded a smaller percentage of polymorphic markers: 37.9%. However, each of the eight primer-combination sets of markers could discriminate all 31 barley genotypes.

Genetic similarities among all genotypes ranged from 0.857 to 0.978 with a mean of 0.919. Within the group of European two-row spring barleys the average gs was 0.932 ranging from 0.901 to 0.978.

Principal-coordinate analysis resulted in a three-dimensional graphic representation of the relationships between the genotypes (Fig. 1). The correlation coefficient between genetic similarities and Euclidean distances in the graph was -0.86.

Pedigree analysis

Coefficients of co-ancestry that were defined as 'well known' (f_{wk}), ranged from 0 to 0.623 with a mean of 0.132. Within the core set of European two-row spring

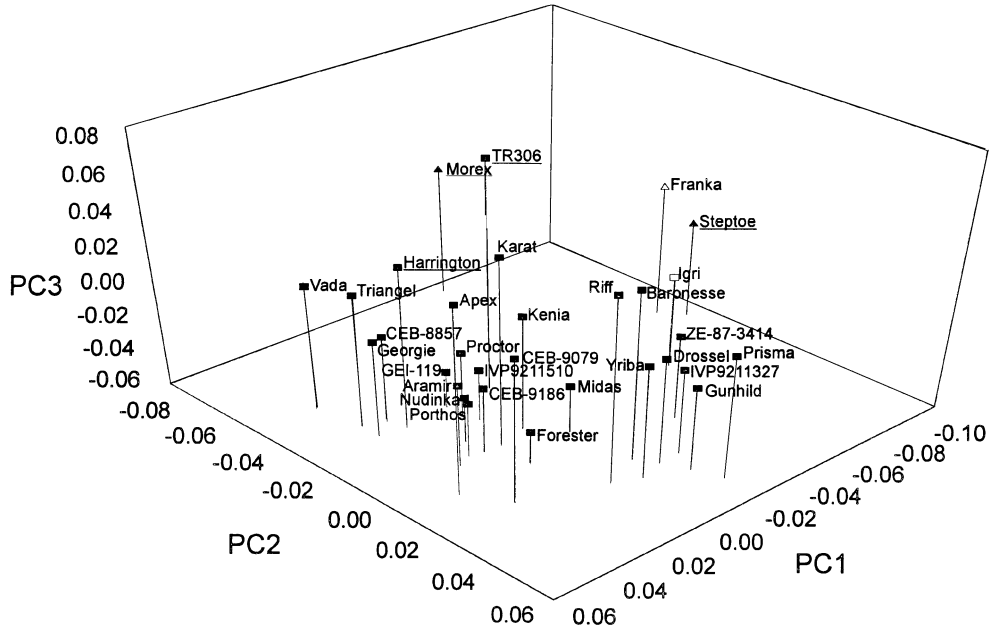

Fig. 1 Relationships between 31 barley lines visualised by principal-coordinate analysis using AFLP-based genetic similarities. ■ = two-row spring type; ▲ = six-row spring type; □ = two-row winter type; △ = six-row winter type. North American lines are *underlined*. *PC1*, *PC2 and PC3*: first, second and third principal coordinates

barleys, f_{wk} had an average of 0.176 ranging from 0.039 to 0.623.

Morphological-trait analysis

After standardisation and principal-component analysis, the first ten principal components, explaining about 87% of the variation, were used to calculate morphological distances (*md*) between the genotypes. The *md* values ranged from 1.88 to 5.86 with a mean of 4.32.

Comparison of relationship measures

Simple (r) and rank (r_s) correlation coefficients between genetic similarity (*gs*) and the 'well known' coefficient of co-ancestry (f_{wk}) were 0.404 (r) and 0.393 (r_s) within the core set of European two-row spring barleys (19 lines). The bootstrap 95%-confidence interval for r was [0.134, 0.642], indicating that r deviates significantly from 0. Including the six other barley lines from the core set, which had less-complete pedigrees, resulted in correlation coefficients of 0.389 (r), with a bootstrap 95%-confidence interval of [0.135, 0.600], and 0.334 (r_s).

The correlation between *gs* and f_{wk} for the total set of barley genotypes (23 lines), including the North American and the winter barleys, is much higher. The value of r is 0.652 with a bootstrap 95%-confidence interval [0.401, 0.803], and the value of r_s is 0.711.

The relationship between *gs* and *f* can also be assessed by comparing the principal-coordinate graph (Fig. 1), based on the *gs* estimates, with the genotype pedigrees, in a more qualitative manner. The first observation that can be made is the clear separation of four of the six North American and winter barleys. Only the Canadian two-row spring variety Harrington and the German two-row winter variety Igri are positioned relatively close to the European two-row spring barleys. Harrington's pedigree, containing several European two-row spring barleys, confirms its position in the graph. Igri is more or less positioned in between the six-row winter variety Franka and a group of European two-row spring barleys. This is consistent with Igri's origin, i.e. a cross between a six-row winter barley hybrid and a two-row spring barley named Ingrid. The latter is also a parent of the cultivar Drossel which is positioned relatively close to Igri. Offspring of Emir, a Dutch two-row spring barley which was frequently used as a parent, appears to be concentrated in the front part of the graph: Aramir, Nudinka and Porthos. The Vada-, Aramir- and Isaria/Union/Volla/Trumpf-groups, as distinguished for the European two-row spring barleys by Melchinger et al. (1994), seem to emerge here as well. Furthermore, parent-offspring combinations are not very distant in the graph: Vada-Georgie, Kenia-Proctor, Aramir-Apex. However, the combination Apex-Riff seems to be rather distant. This picture may be confirmed by the above-average morphological distance of 4.59 between Apex and Riff and by personal communication with Dutch breeders who emphasised the clear agronomic differences between the two cultivars. It seems that selection against Apex traits took place during the selection of Riff.

Simple (r) and rank (r_s) correlation coefficients between genetic similarity (*gs*) and morphological distance (*md*) were -0.124 and -0.142 within the core

set of European two-row spring barleys for which morphological trait data were available (18 lines). The bootstrap 95%-confidence interval for r was [− 0.362, 0.123], indicating no significant correlation between *gs* and *md*.

Correlation coefficients between f_{wk} and *md* could only be based on the 13 lines that had 'well known' pedigrees, as well as morphological trait data, available. The value of r was − 0.117 and the value of r_s was − 0.189. From the bootstrap 95%-confidence interval for r[− 0.363, 0.198] it was concluded that there was no significant correlation between f_{wk} and *md*.

Discussion

The degree of AFLP polymorphism does not appear to be very large in the set of barley genotypes we used. However, each primer-combination set of markers could discriminate all genotypes. This may be a result of the choice of primer combinations which yield high numbers of unambiguous polymorphisms. It is not likely that a set of AFLP markers based on a randomly chosen primer combination will always be able to discriminate barley genotypes similarly well.

Although it does not have any direct effect on correlation estimates between genetic similarity (*gs*) and other relationship measures (*f*, *md*), it was decided to include monomorphic markers in the genetic-similarity estimation. One advantage of doing so is that the addition of extra genotypes in which a band of a so-far monomorphic marker is absent, making it a polymorphic marker, does not change 'existing' *gs*-estimates. If monomorphic markers are excluded, such an addition will result in a change of 'existing' *gs* estimates. Similarly, by ignoring the simultaneous absence of a band in two genotypes, the addition of extra genotypes that have bands in 'new' positions will not change 'existing' *gs* estimates.

The values of the correlation coefficients between genetic similarity and the coefficient of co-ancestry are significant but not very high. This is an agreement with the poor-to-moderate correlations that were found between RFLP-based and RAPD-based *gs* estimates and *f* (Tinker et al. 1993; Graner et al. 1994). One of the causes for this poor relationship may be inaccuracy in *gs* and *f* estimates.

The accuracy of *gs* estimates depends on the number of markers, their distribution over the genome, and the independent information (Messmer et al. 1993) provided by the AFLP markers. For the last reason, redundant markers with identical or allelic patterns within primer combinations have been discarded. Bootstrap analysis by sampling from all 681 markers resulted in standard-deviation estimates for *gs* ranging from 0.006 to 0.012. An extra source of inaccuracy may be errors in scoring AFLP-bands. We tried to prevent part of these errors by scoring a data point as missing in case of doubt. The lack of information due to missing observations is included in the bootstrap standard-deviation estimates.

The assumptions underlying the calculation of coefficient of co-ancestry may cause quite some inaccuracy in *f* estimation (Messmer et al. 1993). The assumption that original ancestors are equally unrelated with $f = 0$ will probably not hold. It is quite likely that some pairs of ancestors, e.g. genotypes descending from the same region, are more related than others. Also the assumption that a genotype receives half of its genes from each parent is very doubtful. As a result of natural, or breeder's, selection during the inbreeding phase, alleles of one parent may have had the advantage over alleles of the other parent. As a result of this the estimated coefficients of co-ancestry may show substantial deviations from the true *f* values.

The absence of a significant relationship between morphological distance (*md*) and *gs* or *f* estimates within the European two-row spring barleys may be a result of inadequate representation of genetic relationships by the observed morphological traits. Reasons for this could be: the limited number of traits observed, the limited variation for these traits, the number of underlying genes for these traits, which may also be limited, and possible epistatic interactions among these genes. Also the distribution of the underlying genes over the genome may be quite irregular. Finally, most data were measured on the rather coarse ordinal UPOV-scales (UPOV, 1981), which may have caused some inaccuracies in the *md* estimates. The poor within-group correlation can be said to agree with the results of Souza and Sorrells (1991 b) in oats. The moderate correlation between *gs* and distance based on agronomic and morphological traits found by Autrique et al. (1996) in durum wheat is a result of the wider range of genotypes under investigation, representing more than one ecotype and resulting in much more variation among distance estimates. Also, most of the observed traits were measured on a continuous scale, probably resulting in a higher accuracy of the distance estimates.

The accuracy of the correlation-coefficient (r) estimates cannot be assessed straightforwardly, because the usual assumption of independent samples of data-pairs from a bivariate normal distribution does not hold. The data-pairs, consisting of relationship measures, are dependent and have a non-normal distribution. In our case they are based on a genotype sample from the population of European two-row spring barleys. To avoid complex analytical approaches, bootstrap sampling from the genotypes can be used to approximate the proper confidence intervals for r. Inaccuracy appears to be larger than one would expect on the basis of the usual, but false, distributional assumptions. The addition of genotypes that did not have 'well known' pedigrees slightly decreased r, showing the

effect of inaccuracy of f due to incomplete pedigree information.

Including genotypes from other barley groups, e.g. European winter barleys and North American spring barleys, resulted in a much larger estimate of r. The main reason for this bias is the simultaneous study of within- and between-group (gs, f) pairs. The higher value of r shows that AFLP-based gs estimates can be used to distinguish between major groups of barley and suggests that genetic diversity in barley may very well be assessed with AFLPs.

The prediction of the breeding behaviour of offspring from parent combinations may be improved by the simultaneous use of AFLP-based genetic similarities and coefficients of co-ancestry. A preliminary standardisation could be helpful in this respect to take account of the differing gs and f ranges. The combination of the gs and f estimates is expected to decrease the effect of their independent inaccuracies. The weights given to both relationship measures may depend on the number of markers and perhaps also on the approximate inaccuracy of f (Cox et al. 1985 a). However, the expected improvement of a combined measure can be made ineffective if gs or f estimates are biased (Souza and Sorrells 1991 b). Whether morphological distances have any predictive value on breeding behaviour remains questionable.

Conclusions

The AFLP fingerprint technique can be used for cultivar identification in barley. One primer combination may often be sufficient to identify lines uniquely.

Genetic similarities (gs), based on AFLP markers, show a poor-to-moderate correlation with pedigree-based coefficients of co-ancestry (f) within the group of European two-row spring barleys. This poor relationship may be caused by inappropriate assumptions in the calculation of f as well as marker-sampling error and biased representation of genomic differences revealed by AFLPs. Morphological distances (md) show no significant relationship with gs or f. This may be caused by biased and insufficient representation of the genome using morphological traits. The inaccuracy of the correlation coefficients between relationship measures, e.g. gs, f and md, can be assessed using bootstrap sampling of genotypes.

The clear distinction between major barley groups, based on gs estimates, suggests opportunities for the use AFLP markers in the assessment of genetic diversity. For the prediction of the breeding behaviour of parent combinations, simultaneous use of several relationship measures (gs, f) in a combined index, as proposed by Cox et al. (1985 a), may probably improve results if large biases in the gs and f estimates are absent. This improvement will be a result of the decreased effect of the individual inaccuracies.

Acknowledgements The authors thank Izak Bos for critically reviewing the manuscript. They also thank all breeders and researchers who contributed pedigree information for this study.

References

Arias G, Reiner A, Penger A, Mangstl A (1983) Directory of barley cultivars and lines. Ulmer Press, Stuttgart

Autrique E, Nachit MN, Monneveux P, Tanksley SD, Sorrells ME (1996) Genetic diversity in durum wheat based on RFLPs, morphological traits and coefficient of parentage. Crop Sci 36:735–742

Baum BR, Bailey LG, Thompson BK (1985) Barley register. Agric Canada Publ 1783B. Ottawa, Canada

Botstein D, White RL, Skolnick M, Davis RW (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am J Hum Genet 32:314–331

Cowen NM, Frey KJ (1987) Relationships between three measures of genetic distance and breeding behavior in oats (*Avena sativa* L.). Genome 29:97–106

Cox TS, Kiang YT, Gorman MB, Rodgers DM (1985 a) Relationship between the coefficient of parentage and genetic similarity indices in the soybean. Crop Sci 25:529–532

Cox TS, Lookhart GL, Walker DE, Harrell LG, Albers LD, Rodgers DM (1985 b) Genetic relationships among hard red winter wheat cultivars as evaluated by pedigree analysis and gliadin polyacrylamide gel-electrophoretic patterns. Crop Sci 25:1058–1063

Dice LR (1945) Measures of the amount of ecologic association between species. Ecology 26:297–302

Efron B, Tibshirani RJ (1993) An introduction to the bootstrap. Chapman and Hall, New York

Goodman MM (1972) Distance analysis in biology. Syst Zool 21:174–186

Gower JC (1966) Some distance properties of latent root and vector methods used in multivariate analysis. Biometrika 53:325–338

Graner A, Ludwig WF, Melchinger AE (1994) Relationships among European barley germplasm. II. Comparison of RFLP and pedigree data. Crop Sci 34:1199–1205

Mahalanobis PC (1936) On the generalized distance in statistics. Proc Natl Inst Sci India 2:49–55

Malécot G (1948) Les mathématiques de l'hérédité. Masson et Cie, Paris

Melchinger AE, Graner A, Singh M, Messmer MM (1994) Relationships among European barley germplasm. I. Genetic diversity among winter and spring cultivars revealed by RFLPs. Crop Sci 34:1191–1199

Messmer MM, Melchinger AE, Herrmann RG, Boppenmaier J (1993) Relationship among early European maize inbreds. II. Comparison of pedigree and RFLP data. Crop Sci 33:944–950

Nei M, Li WH (1979) Mathematical model for studying genetic variation in terms of restriction endonucleases. Proc Natl Acad Sci USA 76:5269–5273

Qi X, Lindhout WH (1997) Development of AFLP markers in barley. Mol Gen Genet 254:330–336

SAS Institute Inc (1992) SAS technical report P-229, SAS/STAT software: changes and enhancements, release 6.07. SAS Institute Inc., Cary, North Carolina

Schut JW (1997) Application of the bootstrap in plant genetics. In: Stein A, Plenning de Vries FWT, Schut JW (eds) Data in action. Proceedings of a seminar series 1996/1997. DLO Research Institute for Agrobiology and Soil Fertility and C.T. de Wit Graduate School for Production Ecology, Wageningen (Quantitative Approaches in Systems Analysis No. 12, pp 71–79)

Souza E, Sorrells ME (1991 a) Relationships among 70 North American oat germplasms. I. Cluster analysis using quantitative characters. Crop Sci 31:599–605

Souza E, Sorrells ME (1991 b) Relationships among 70 North American oat germplasms. II. Cluster analysis using qualitative characters. Crop Sci 31 : 605–612

Tinker NA, Fortin MG, Mather DE (1993) Random amplified polymorphic DNA and pedigree relationships in spring barley. Theor Appl Genet 85 : 976–984

UPOV (1981) Guidelines for the conduct of tests for distinctness, homogeneity and stability: barley. UPOV, Geneva

Van der Beek JG, Verkerk R, Zabel P, Lindhout WH (1993) Mapping strategy for resistance genes in tomato based on RFLPs between cultivars: *Cf9* (resistance to *Cladosporium fulvum*) on chromosome 1. Theor Appl Genet 84 : 106–112

Van Eck HJ, Van der Voort JR, Draaistra J, Van Zandvoort P, Van Enckevort E, Segers B, Peleman J, Jacobsen E, Helder J, Bakker J (1995) The inheritance and chromosomal localization of AFLP markers in a non-inbred potato offspring. Mol Breed 1 : 397–410

Van Hintum TJL, 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

Vos P, Hogers R, Bleeker M, Reijans M, Van de Lee T, Hornes M, Frijters A, Pot J, Peleman J, Kuiper M, Zabeau M (1995) AFLP: a new technique for DNA fingerprinting. Nucleic Acids Res 23 : 4407–4414

Williams JGK, Kubelik AR, Livak KJ, Rafalski JA, Tingey SV (1990) DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucleic Acids Res 18 : 6531–6535