Theor Appl Genet (1994) 89:991–997

T. J. L. van Hintum

Comparison of marker systems and construction of a core collection in a pedigree of European spring barley

Received: 5 May 1994 / Accepted: 21 June 1994

Abstract Based on data in four publications describing European barley cultivars, similarities between pairs of cultivars were calculated using individual markers and combinations of markers. These markers included 19 isoenzyme patterns, Giemsa C-banding variants of each of the seven chromosomes, hordein polypeptide patterns, DDT susceptibility type, and three morphological descriptors. The rank correlation between the coefficients of parentage and marker-based similarities is low; the highest single marker correlation is with *Est-1*, 0.41, and the highest correlation with a combination of markers is 0.58. Giemsa C-banding patterns score rather high, as opposed to the three morphological characters that score very low. Selection of core collections using the effective-number-of-origin-lines theory is successful. The average number of types found in a core collection of cultivars of given size is always considerably larger than a random set of the same size. The core collection approaches the maximum possible number of types.

Key words Genetic diversity · Marker systems
Core collections · Genetic resources · *Hordeum vulgare*

Introduction

There are several reasons why information about the genetic similarity of individuals is important. For the breeder, the larger the genetic distance the more heterosis he can expect in the F_1, and the wider is the genetic base he can select from in later generations. For the curator of a germplasm collection information about the genetic similarity of his accessions can help him decide what material should be added to his collection, or what material should be included in a core collection (Brown 1989).

There are several ways of estimating the genetic similarity of cultivars. Basically they can be divided into pedigree-based and marker-based methods.

The value of pedigree data for the estimation of the genetic similarity of cultivars can be, and is frequently, questioned. For pedigree analysis of cultivars, the following assumptions have to be made (cf. Martin et al. 1991): (1) A cultivar receives half of its genes from each parent. (2) Parents in crosses are homozygous and homogeneous. (3) Ancestors for which no pedigree information is available are unrelated. If selections made from cultivars are included in the analysis a fourth assumption has to be added indicating the relationship between a selection and the 'parent cultivar', for example a cultivar selected from a landrace. This assumption might be: (4) The coefficient of parentage between a cultivar and a selection from that cultivar is 0.75 (Martin et al. 1991). An alternative for this fourth assumption, as argued by Hintum and Haalman (1994), is: (4′) If selections are made from a cultivar, this cultivar is assumed to be the variable offspring of the cross of two unrelated lines. A selection from the cultivar is one of the offspring lines; if the cultivar itself is used as a parent in a cross, one of the offspring lines is considered to be used.

All these assumptions can be questioned. A cultivar will not receive half its genes from each parent since selection in the offspring of a cross will favour the alleles of one of the parents, which will thus contribute more. The second assumption, assuming homozygosity and homogeneity, is usually not true, especially in the case of old cultivars or landraces. But even in the case of modern cultivars heterogeneity is quite common (e.g., Linde-Laursen et al. 1982, 1987). The third assumption, assuming no relatedness if no information is available, is obviously not true. Any fourth assumption, giving the relationship between a selection and the population it was selected from, is fundamentally arbitrary, since this relation will differ in each case that is being studied.

The value of some marker systems, such as isoenzymes, for the description of genetic diversity in crops is also

Communicated by P. M. A. Tigerstedt

T. J. L. van Hintum
Centre for Plant Breeding and Reproduction Research (CPRO-DLO), Centre for Genetic Resources The Netherlands, P.O.Box 16, 6700 AA Wageningen, The Netherlands

sometimes questioned. Lack of congruence between different data sets describing barley genebank accessions was found by Cross et al. (1992) who also discussed other such cases. Generally, the potential of modern techniques involving molecular marker systems such as RFLPs is believed to be high. It can be expected that a large number of RFLPs or comparable molecular markers, uniformly distributed throughout the genome, would yield high correlations with pedigree-based estimates (Melchinger et al. 1990). There would be a limit well below 1.0 due to the limitations in both the methodologies of pedigree analysis and marker-based estimation (cf. Messner et al. 1993). Smith et al. (1990) obtained a value of 0.90 when they correlated coefficients of parentage with genetic similarities based on RFLP data in a group of related elite maize inbreds. Ajmone-Marsan et al. (1992) compared coefficients of genetic similarity based on RFLP data with coefficients of parentage in pair-wise comparisons in two maize heterotic groups and found a high correlation in one group (r=0.70) and a very low correlation in the other (r=0.07), while Messner et al. (1993) found high correlations in both groups (r=0.71, r=0.86).

In the present study an attempt is made to compare the relative value for similarity estimates of several different groups of 'classical' markers including isoenzyme patterns, Giemsa C-banding variants, hordein polypeptide patterns, DDT susceptibility type, and morphological descriptors. A methodology for developing a core collection on the basis of pedigree data is also demonstrated and evaluated using these data.

Materials and methods

Material

Data from four publications on the description of European barley cultivars (Andersen 1982; Linde-Laursen et al. 1982, 1987; Nielsen and Bay Johansen 1986) were computerized.

Both papers of Linde-Laursen et al. (1982, 1987) concerned the same 59 related spring barley cultivars, 57 two-rowed and two six-rowed. The material spanned the last 150 years of breeding efforts. The data of these papers were combined in one data set. The first paper gave data of Giemsa C-banding variants, hordein polypeptide patterns, and isozyme patterns governed by alleles at four loci. The more recent paper presented data of 17 isoenzyme systems that showed variation.

Andersen (1982) provided data on 47 cultivars from the late seventies for which he used breeders or pre-basic seed. The data concerned three morphological descriptors and some biochemical markers, namely, rachilla hair length, spicules on inner dorso-lateral nerves, hairs on margins of the ventral furrow, the alpha and beta amylase type, and four esterases which showed variation.

Nielsen and Bay Johansen (1986) described 66 cultivars which were on the 'Danish List of Varieties of Agricultural Crops 83/84'. Data on life form (seven winter types were included), 16 variable isoenzymes, and the alleles of hordein polypeptide patterns, were available.

The pedigrees of all cultivars were traced as far back as possible using several sources including Arias et al. (1983), Baum et al. (1985), EBC (1991) and Linde-Laursen et al. (1982). Only cultivars for which suitable pedigree information was available were included in the analysis. If a cultivar showed variation at one marker, it was

Table 1 Pedigree data used in the analysis

Cultivar	Pedigree
Agio[a]	Kenia×Georgine
Alf[a,d]	Bomi×Bomi
Alfa[a,b]	$Kenia_{L1}$×$Kenia_{L2}$
Algerian[a]	Unknown
Alva[a]	Sv.2148×Vada
Ark Royal[a]	Proctor×HP 5466
Balder[a]	(Gull×Scania)×Maja
Binder[a,b]	$Hanna_{L1}$×$Hanna_{L2}$
Bine[a]	Balder×Minerva
Bomi[a]	Bonus×Minerva
Bonus[a]	Maja×Sv.34/22
Carlsberg II[a,b]	$Carlsberg_{L1}$×$Carlsberg_{L2}$
$Carlsberg_{L1}$	Prentice×Maja
$Carlsberg_{L2}$	Prentice×Maja
Carlsberg[a]	Prentice×Maja
Criewener 403[b]	$Hanna_{L1}$×$Hanna_{L2}$
Delta[a,c]	Kenia×(*H. laevigatum*$_{\#1}$×Gull)
Drost[a]	Maja×Kenia
Drost A[a,b]	$Drost_{L1}$×$Drost_{L2}$
$Drost_{L1}$	Maja×Kenia
$Drost_{L2}$	Maja×Kenia
Elbo[a]	Rika×Drost
Emir[a]	Delta ×[Agio×(Kenia^3×Arabische)]
Flavina[a]	Gasell×Emir
Foma[a]	Sv.01513×Morgenrot
Freja[a]	Seger×Opal
Gasell	Herta ×[Balder ×(Opal×Binder)]
Georgie[a]	Vada×Zephyr
Georgine[a]	Isaria×Moosburger Rhätia
Hadm.4509[a]	Isaria×Russische 22
Haisa[a,b]	($Hanna_{L1}$×$Hanna_{L2}$)×Isaria
Haisa II[a]	Haisa×WMR I
Hanna[a,b]	$Hanna_{L1}$×$Hanna_{L2}$
Hannchen[a,b]	$Hanna_{L1}$×$Hanna_{L2}$
Hassan[a]	Delta×[Agio×(Kenia^3×Arabische)]
Heine 0553[b]	($Kneifel_{L1}$×$Kneifel_{L2}$)×Hadm.4509
Heine 2149	Heine 501×Haisa
Heine 2369	[(Lyallpur×Maja)×Haisa]×Pirol
Heine 501	Morgenrot×WMR I
Herta[a]	Kenia×Isaria
HP 5466	Heine 2369×Heine 05532
Isaria[a]	Bavaria×Danubia
Julia[a]	Delta×Wisa
Kenia[a]	Binder×Gull
$Kenia_{L1}$	Gull×Binder
$Kenia_{L2}$	Gull×Binder
$Kneifel_{L1}$[b]	$Hanna_{L1}$×$Hanna_{L2}$
$Kneifel_{L2}$[b]	$Hanna_{L1}$×$Hanna_{L2}$
Maja[a]	Binder×Gull
Mari[a,d]	Bonus×Bonus
Mentor[a]	Balder×WMR II
Minerva[a,c]	*H. laevigatum*$_{\#2}$×Gull
Mona[a]	Mari^2×Monte Cristo
Monte Cristo[a]	Unknown
Morgenrot[a]	Heils Franken×Australischer Frühe
Opal[a]	Gull×Binder
Pallas[a,d]	Bonus×Bonus
Pirol	WMR I×Morgenrot
Plumage Archer[a]	($Plumage_{L1}$×$Plumage_{L2}$) ×($Archer_{L1}$×$Archer_{L2}$)
$Plumage_{L1}$[b]	Plumage $Korn_{L1}$×Plumage $Korn_{L2}$
$Plumage_{L2}$[b]	Plumage $Korn_{L1}$×Plumage $Korn_{L2}$
Prentice[a,b]	$Archer_{L1}$×$Archer_{L2}$
Proctor[a]	Kenia×Plumage Archer
Rika[a]	Kenia×Isaria
Seger[a]	Gull×Hannchen
Sultan[a]	([(Arabische×Kenia^2)× Agio]× Kenia) × Balder

Table 1 (continued)

Cultivar	Pedigree
Sundance[a]	Vada×Zephyr
Sv.01513	Sv.34/22×Maja
Sv.2148	Balder ×(Seger×Binder)
Sv.34/22	Seger×Opal
Tyra[a]	(Algerian×Herta[8])×(Rika×Drost)
Vada[a,c]	*H. laevigatum*$_{\#3}$×Gull
WMR I[a]	Criewener 403×Pflugs Intensiv
WMR II[a]	WMR I×Isaria
Wisa[a]	WMR I×Isaria
Zephyr[a]	Heine 2149×Carlsberg

[a] The cultivar is part of the set that was analyzed by Linde-Laursen et al. (1982, 1987)
[b] If a selection was made from a cultivar this cultivar was treated as an offspring family of two lines designated by the cultivar name with L1 and L2 as a suffix
[c] Crosses involving *H. laevigatum* were considered to be made with unrelated lines
[d] Mutants were considered equal to the parent

treated as two entries; if there were more varying markers the cultivar was excluded from the analysis. This resulted in 62, 52 and 73 entries corresponding to respectively 55, 47 and 63 distinct cultivars. The pedigrees of the cultivars included in the analysis from the first data set (Linde-Laursen et al. 1982, 1987) and those of their ancestors are presented in Table 1.

Analysis

The coefficients of parentage (r) as defined by Kempthorne (1969) were calculated for each pair of cultivars in a data set. The r of two individuals is the probability that a random allele at a random locus in one individual is identical by descent to a random allele at the same locus in the other individual. This quantity was first described by Malecot (1948) who called it the co-ancestry coefficient f, but is now commonly known as r.

The rs of all pairs of individuals were correlated with a number of marker-based dissimilarities, i.e., dissimilarities based on single markers and dissimilarities based on combinations of markers. If a marker showed the same expression in two cultivars the dissimilarity between those two was 0.0, if it showed a different expression the dissimilarity was 1.0. If more than one marker was considered in the comparison of two cultivars, the dissimilarities based on the individual markers were summed.

Marker-based dissimilarities were compared with the coefficients of parentage by calculating the Spearmans rank correlation (r_s) between the two values of all pair-wise comparisons in a set of cultivars. So if 62 cultivars were considered, all $n*(n-1)/2=1891$ possible pair-wise comparisons were made, and the rank correlation between the 1891 marker-based dissimilarities and the coefficients of parentage was calculated.

The data set of Linde-Laursen et al. (1982, 1987) was further analyzed. The average number of types in a set of any given size was calculated by taking a random set of cultivars and counting the number of different types. Isoenzyme bands, Giemsa C-banding variants, hordein polypeptide patterns, and DDT susceptibility type, were all considered; a single individual contains 27 types, one for each character. For example, if a group of two cultivars contains 30 types this would imply that there are three polymorphic characteristics; if a group of more than two cultivars contains 30 types there could also be characters with more than two types of expression, for example 25 monomorphic, one dimorphic and one character with three states. The sampling was repeated 100 000 times per set size, and the results were averaged. The minimal and maximal number of types that a set of given size could contain were determined by running an optimization procedure. This procedure took a random set of the given size and tried, in the case where the minimum was searched, to reduce the number of types in the set by systematically exchanging each selected cultivar with each non-selected cultivar. This was repeated until no further reduction could be achieved. The optimization procedure was repeated 1000 times per set size and the extreme result was recorded.

A core collection of a given size was selected by using an optimization procedure using the linear estimation method of the effective number of origin lines (n_{OL}, Hintum and Haalman 1994), i.e., a group was selected which was expected to contain the highest average number of alleles, not identical by descent, per locus. This was repeated 10 000 times resulting in an average, minimum and maximum of counted types, all with the same n_{OL}.

All calculations were performed on a minicomputer with Fortran programmes.

Results

The Spearman rank correlation coefficients between all pair-wise comparisons of marker-based similarities and the coefficients of parentage of all three data sets are presented in Table 2; the results of the comparisons of combinations of markers with the coefficients of parentage are presented in Table 3. All coefficients are low; the highest single marker correlation, 0.41, was with *Est-1* and the highest correlation with a combination of markers, the complete set of Linde-Laursen et al. (1982, 1987), was 0.58. Giemsa C-banding scored rather high, as opposed to the three morphological characters which scored very low.

The number of types in random sets and core collections of all possible sizes are presented in Fig. 1. It can been seen that the average number of types in the core collection is closer to the maximum possible number than to the average number of a random set. There is only a small range of set sizes, i.e., 14 to 18 and 21 cultivars, where there is only one combination of cultivars with the maximal n_{OL}, and thus only one number of types per set size (see Fig. 1).

The distribution of the number of types in sets of 5, 10 and 15 cultivars is presented in Fig. 2. As can be seen from the variation in the number of types in Fig. 2 there were many different core collections of size 5 and 10 which had the highest possible n_{OL}, i.e., 5.00 and 9.25. There was only one of size 15 which had a n_{OL} of 11.05. It can also be seen in Fig. 2 that the maximal and minimal possible number of types is extremely rarely realized, which should be taken into account when studying Fig. 1.

The n_{OL} of the entire set of 55 cultivars was 11.32, indicating a very high degree of relatedness among the cultivars in the set. It was possible to select a core collection of only 21 cultivars with this maximal n_{OL}, and thus theoretically containing all the variation in the set. A core collection of 15 cultivars contained only 0.27 less. The set of 55 cultivars originated from 22 origin lines (see Table 4), in which the landraces 'Archer', 'Hanna' and 'Plumage Korn' appear twice since they are considered to be composed of the offspring of two unrelated lines (Hintum and Haalman 1994). The degree to which the origin lines contributed to the complete set and to the core collections of

Table 2 Number of types[a] per marker, mean dissimilarity and Spearman rank correlation coefficients between individual marker-based similarities and coefficients of parentage

Data source	Linde-Laursen et al. (1982, 1987)			Nielsen and Bay Johansen (1986)			Andersen (1982)		
Number of objects	61			52			73		
Number of comparisons	1891			1326			2628		
Average coefficient of parentage (r)	0.21			0.14			0.15		
	No. of types	Mean dissi.	r_s	No. of types	Mean dissi.	r_s	No. of types	Mean dissi.	r_s
Life form (winter/spring)							2	0.20	0.22
Aat-3	2	0.03	0.07						
Aco-1	3	0.33	0.15				3	0.38	–0.04
Acp-2	2	0.40	0.39				3	0.47	0.12
Acp-3	4	0.52	0.23				5	0.30	0.19
Acp-4	2	0.03	0.03				2	0.11	0.07
Amy-1	2	0.50	0.20	2	0.49	0.06	2	0.49	0.13
Amy-2				2	0.51	0.03			
Bmy-1	2	0.30	–0.01				2	0.50	0.03
Est-1	3	0.49	0.41	3	0.24	0.11	3	0.28	0.13
Est-2	2	0.15	–0.13						
Est-4	3	0.52	0.18	2	0.29	0.26	2	0.33	0.20
Est-5	4	0.26	0.14	3	0.21	–0.01	2	0.24	0.07
Est-9	2	0.46	–0.04	2	0.50	0.02	2	0.50	–0.01
Gpi-1							2	0.13	0.12
Ndh-1	2	0.03	0.07				2	0.03	0.00
Ndh-3	4	0.72	0.03				4	0.72	–0.00
Ndh-4	2	0.12	–0.08				2	0.03	0.00
Pgd-2	2	0.44	0.07				2	0.51	0.01
Prx-4	2	0.28	0.13				3	0.43	0.09
Giemsa C-banding pattern chr 1	3	0.13	0.22						
Giemsa C-banding pattern chr 2	2	0.03	0.07						
Giemsa C-banding pattern chr 3	3	0.50	0.36						
Giemsa C-banding pattern chr 4	3	0.09	0.16						
Giemsa C-banding pattern chr 5	5	0.56	0.04						
Giemsa C-banding pattern chr 6	5	0.71	0.12						
Giemsa C-banding pattern chr 7	7	0.60	0.15						
Hor-1	6	0.51	0.21				10	0.84	0.02
Hor-2	11	0.86	0.13				15	0.88	0.06
DDT resistance	2	0.50	0.12						
Rachilla hair length				2	0.14	0.06			
Spicules on inner dorso-lateral nerves				2	0.21	0.01			
Hairs on margins of the ventral furrow				2	0.27	0.03			

[a] A type usually corresponds to an allele, but can also correspond to a Giemsa C-banding pattern or any other expression of a marker

Table 3 Number of markers, maximal and mean dissimilarity and Spearman rank correlation coefficients between marker-based similarities and coefficients of parentage

Data source	Linde-Laursen et al. (1982, 1987)				Nielsen and Bay Johansen (1986)				Andersen (1982)			
	No. of markers	Max. dissi.	Mean dissi.	r_s	No. of markers	Max. dissi.	Mean dissi.	r_s	No. of markers	Max. dissi.	Mean dissi.	r_s
All markers	27	24	10.1	0.58	9	7	2.9	0.20	19	17	7.4	0.29
Isoenzymes	17	14	5.6	0.46					16	14	5.4	0.27
Esterase 1, 4, 5 and 9	4	4	1.7	0.35	4	4	1.3	0.20	4	4	1.4	0.20
Amylase 1 and 2					2	2	1.0	0.06				
Giemsa C-banding	7	7	2.6	0.44								
Hordeins	2	2	1.4	0.26					2	2	1.7	0.06
Morphology					3	3	0.6	0.03				

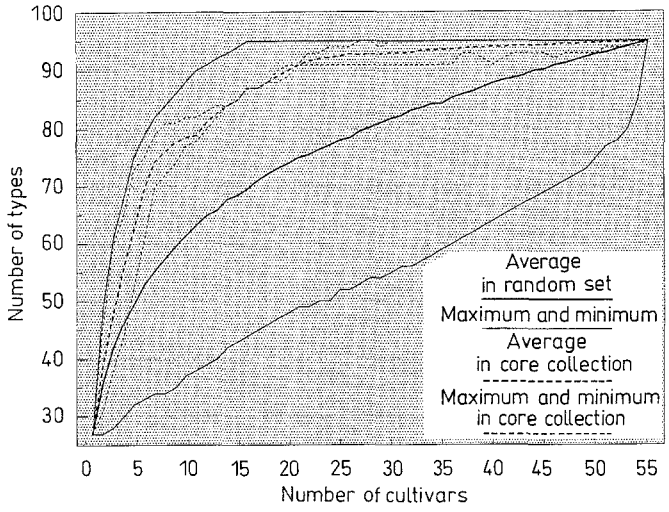

Fig. 1 Number of types in sets of given size

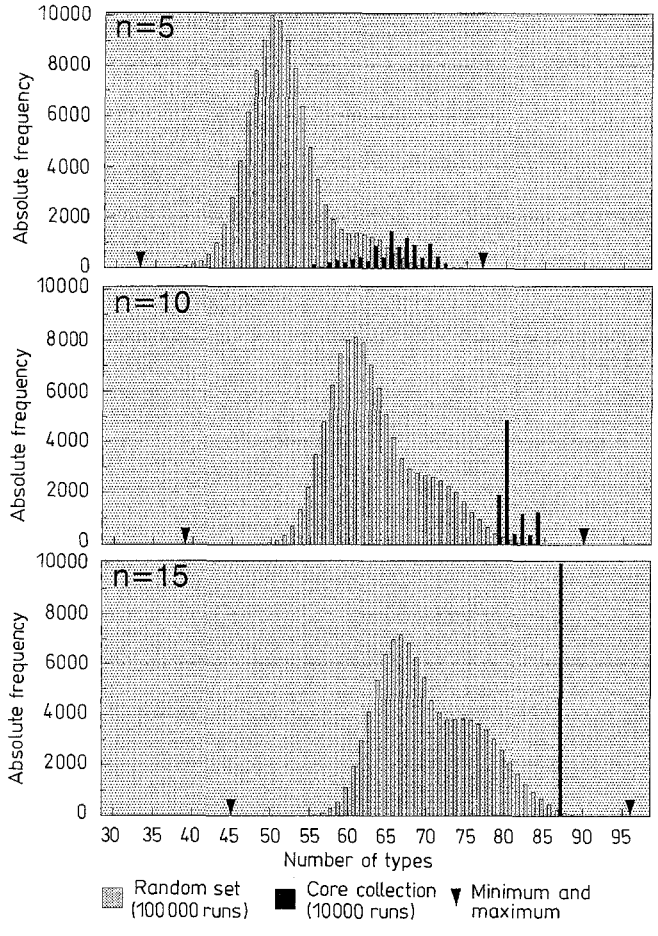

Fig. 2 Distribution of number of types in sets of 5, 10 and 15 cultivars

Table 4 The contribution[a] made by the origin lines to the complete set of cultivars and to core collections of three different sizes, and the effective number of origin lines of these sets

	Complete set	Core collections		
Cultivars in set	55[b]	15[c]	10[d]	5[e]
Types[f] in set	97	87	82	65
Algerian	1.00	1.00	1.00	1.00
Arabische	0.09	0.00	0.00	0.00
$Archer_{L1}$	0.63	0.63	0.63	0.00
$Archer_{L2}$	0.63	0.63	0.63	0.00
Australischer Frühe	0.50	0.50	0.50	0.00
Bavaria	0.50	0.50	0.25	0.50
Danubia	0.50	0.50	0.25	0.50
Gull	1.00	0.94	0.75	0.50
$Hanna_{L1}$	0.94	0.87	0.63	0.25
$Hanna_{L2}$	0.94	0.87	0.63	0.25
Heils Franken	0.50	0.50	0.50	0.00
Hordeum laevigatum$_{\#1}$	0.25	0.25	0.00	0.00
Hordeum laevigatum$_{\#2}$	0.50	0.50	0.50	0.50
Hordeum laevigatum$_{\#3}$	0.50	0.50	0.50	0.00
Lyallpur	0.02	0.02	0.00	0.00
Monte Cristo	1.00	1.00	1.00	1.00
Moosburger Rhätia	0.50	0.50	0.50	0.00
Pflugs Intensiv	0.50	0.50	0.50	0.50
Plumage $Korn_{L1}$	0.25	0.25	0.25	0.00
Plumage $Korn_{L2}$	0.25	0.25	0.25	0.00
Russische 22	0.09	0.09	0.00	0.00
Scania	0.25	0.25	0.00	0.00
n_{OL}	11.32	11.05	9.25	5.00

[a] Corresponds to the effective overlap of origin lines (r_{OL}) between the origin line and the target set
[b] The complete set (see Table 1)
[c] The core collection consisting of Algerian, Ark Royal, Balder, Delta, Georgine, Hanna, Hannchen, Isaria, Minerva, Monte Cristo, Morgenrot, Plumage Archer, Prentice, Vada and WMR I, the only one with n_{OL}=11.05
[d] The core collection consisting of Algerian, Georgine, Hannchen, Minerva, Monte Cristo, Morgenrot, Plumage Archer, Prentice, Vada and WMR I, one of the many with n_{OL}=9.25
[e] The core collection consisting of Algerian, Isaria, Minerva, Monte Cristo and WMR I, one of the many with n_{OL}=5.00
[f] A type usually corresponds to an allele, but can also correspond to a Giemsa C-banding pattern or any other expression of a marker

size 5, 10 and 15, are also presented in Table 4. These contributions correspond to the effective overlap of origin lines (r_{OL}) of the origin line and the target sets. Hintum and Haalman (1994) define r_{OL} as the average number of alleles, not identical by descent, per locus present in common in two sets of individuals. As a result, the average contribution of, for example, 'Archer' to the complete set is calculated to be 1.26 alleles per locus, which are not identical by descent to any of the other alleles in the set.

Discussion and conclusions

The coefficients of correlation between the coefficients of parentage and marker-based similarities are low. In the case of single marker-based similarities, this value can not be

expected to be extremely high since the dissimilarity of two individuals can have only two values, i.e., 0.0 (same type) or 1.0 (different type). The maximal r_s of such a binary variable with a uniformly distributed variable is 0.75. But values for single-marker-based correlations below 0.41 and for combinations of marker-based correlations below 0.58 are low.

Amongst individual markers within a group, such as the isoenzymes, the differences are large. If the mean dissimilarity for a marker is low this indicates that there are only a few deviating individuals; this marker will be able to explain only little of the variation in relatedness. But if the mean dissimilarity is close to the maximum, 0.50 in the case of two and 0.67 in the case of three types, one could expect similar values for r_s. This is clearly not the case; *Est-9* was scored in all three datasets, had similar mean dissimilarities and was always very poorly correlated with the relatedness, whereas *Est-1*, also scored in all three datasets, had moderate mean dissimilarities and high-to-moderate correlations. Relatively low correlations can be explained either by linkage to genes which are under selective pressure or by low reliability of the observations of the marker system.

The differences between the datasets can to a large extent be explained by the differences in the material studied, for example, with respect to age or distribution, but may also relate to different levels of authenticity of the material and reliability of the pedigrees.

Though the correlations between the coefficients of parentage and marker-based similarities were low, the selection of core collections using the effective-number-of-origin-lines theory was very successful. The average number of types in core collections of given size was always much higher than that of random sets. The core collection came relatively close to the maximum possible number of types, especially if the distribution of types in random sets was considered. Only very rarely did a random set manage to come close to the core collection. It can be expected that the theory of the effective-number-of-origin-lines is a much more powerful tool for analysing the degree of common ancestry of more than two cultivars as compared to approaches using, for example, the average coefficient of parentage (Cox et al. 1986; Murphy et al. 1986; Knauft and Gorbet 1989; Souza and Sorrells 1989; Martin et al. 1991). If cultivars are studied only pair-wise, the two approaches are equal; the theory of the effective-number-of-origin-lines is a generalization of the classical theory (Hintum and Haalman 1994).

If the number of cultivars in the set is low, in this data set below 13, it is possible to select several different core collections with an equal n_{OL} since the selected cultivars are not, or are hardly, related and thus exchangeable, i.e., with n_{OL} equal or close to the number of cultivars in the set. The same holds true if the maximum n_{OL} is reached; in this example if the number of cultivars is 21. From that point on any cultivar can be added to the previous core collection without changing the n_{OL}. But the number of types in these core collections with an equal n_{OL} can be different, depending on the cultivars included in them.

If a set of cultivars is analyzed which is less related than the one in this study it will be possible to select a higher number of unrelated, or hardly related, cultivars, so the zone where the number of types in a core of given size is not fixed will be longer. But, since all different core collections only contain unrelated, or hardly related, cultivars, they can all be expected to contain significantly more types than random sets that also include related cultivars.

Methodology could be further improved by making it possible to use estimates of the degree of relatedness (or similarity) of origin lines in the analysis. This degree could be estimated either based on historical knowledge about the background of the material, or based on the results of screenings such as the ones presented in the data sets used for this analysis, or else those of molecular screenings which are becoming increasingly available.

Acknowledgements The author thanks Dr. I. Linde-Laursen and Prof. R. von Bothmer for their helpful comments on the manuscript.

References

Ajmone-Marsan P, Livini C, Messmer MM, Melchiner AE, Motto M (1992) Cluster analysis of RFLP data from related maize inbred lines from BSSS and LSC heterotic groups and comparison with pedigree data. Euphytica 60:139–148

Andersen HJ (1982) Isoenzyme characters of 47 barley cultivars and their application in cultivar identification. Seed Sci Technol 10: 405–413

Arias G, Reiner L, Penger A, Mangstl A (1983) Directory of barley cultivars and lines. Technical University of Munich, Freising-Weihenstephan

Baum BR, Bailey LG, Thompson BK (1985) Barley register. Publ no 1783/B. Agriculture Canada, Ottawa

Brown AHD (1989) Core collections: a practical approach to genetic resources management. Genome 31:818–824

Cox TS, Murphy JP, Rodgers DM (1986) Changes in genetic diversity in the red winter wheat regions of the United States. Proc Natl Acad Sci USA 83:5583–5586

Cross RJ, Fautrier AG, McNeil DL (1992) IBPGR morphological descriptors – their relevance in determining patterns within a diverse spring barley germplasm collection. Theor Appl Genet 85:489–495

EBC (1991) Advances in malting barley (harvest 1990). Secretariat European Brewery Convention, Zoeterwoude, The Netherlands

Hintum ThJLvan, 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

Kempthorne O (1969) An introduction to genetic statistics. Iowa State University Press, Ames

Knauft DA, Gorbet DW (1989) Genetic diversity among peanut cultivars. Crop Sci 29:1417–1422

Linde-Laursen I, Doll H, Nielsen G (1982) Giemsa C-Banding patterns and some biochemical markers in a pedigree of European barley. Z Pflanzenzuecht 88:191–219

Linde-Laursen I, Nielsen G, Bay Johansen H (1987) Distribution of isoenzyme markers at 37 loci in a pedigree of European spring barley. Hereditas 106:241–251

Malecot G (1948) Les Mathématiques de l'hérédité. Masson et Cies, Paris

Martin JM, Blake TK, Hockett EA (1991) Diversity among North American spring barley cultivars based on coefficients of parentage. Crop Sci 31:1131–1137

Melchinger AE, Lee M, Lamkey KR, Woodman WL (1990) Genetic diversity for restriction fragment length polymorphisms: rela-

tion to estimated genetic effects in maize inbreds. Crop Sci 30:1033–1040

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

Murphy JP, Cox TS, Rodgers DM (1986) Cluster analysis of red winter wheat cultivars based upon coefficients of parentage. Crop Sci 26:672–676

Nielsen G, Bay Johansen H (1986) Proposal for the identification of barley varieties based on the genotypes for 2 hordein and 39 isoenzyme loci of 47 reference varieties. Euphytica 35:717–728

Smith OS, Smith JSC, Bowen SL, Tenborg RA, Wall SJ (1990) Similarities among a group of elite maize inbreds as measured by pedigree, F_1 grain yield, grain yield heterosis, and RFLPs. Theor Appl Genet 80:833–840

Souza E, Sorrells ME (1989) Pedigree analysis of North American oat cultivars released from 1951 to 1985. Crop Sci 29:595–601