Mahmoud Zeid · Chris-Carolin Schön ·
Wolfgang Link

Genetic diversity in recent elite faba bean lines using AFLP markers

Abstract Amplified fragment length polymorphism (AFLP) markers were used to study the genetic diversity among a large set (n = 79) of inbred lines of recent elite faba bean (*Vicia faba* L.) cultivars with Asian, European (Northern and Southern) and North African origin. The inbred lines were analyzed using eight selected AFLP primer combinations that produced 477 polymorphic fragments. Errors when scoring repeated lanes of one pre-amplification reaction on one gel were negligible, whereas errors when scoring lanes of two individuals of the same inbred line run on different gels were markedly higher. Scoring across gels should be backed by replicates and several appropriate check entries. Based on clustering with Jaccard's similarity coefficient and Principal Coordinate Analysis, only the Asian lines were distinct as a group, the other lines showed no marked further grouping. Nevertheless, several known pedigree relationships were verified. A priori grouping of inbred lines (geographic origin and seed size) and AFLP data corroborate available information on the history of spread and cultivation of faba bean in the studied regions. Based on the diversity observed, studies especially concerning the relationship between genetic similarity based on AFLP markers and hybrid performance within the European elite germplasm have been launched.

Keywords Genetic diversity · *Vicia faba* · Bootstrap · Similarity coefficients · AFLP

M. Zeid · W. Link (✉)
Institute of Agronomy and Plant Breeding,
Georg-August-University, 37075 Göttingen, Germany,
e-mail: wlink@gwdg.de

C.-C. Schön
State Plant Breeding Institute, University of Hohenheim,
70593 Stuttgart, Germany,

Introduction

Faba bean (*Vicia faba* L.) is a genetically isolated species with no successful crossing with any other *Vicia* species. Muratova, as early as 1931, recognized distinct groups within faba bean based mainly on seed size, ranging from small-seeded *minor* beans (0.2–0.8 g per seed) to medium-seeded *equina* beans and the large-seeded *major* beans (single-seed weight of up to 2.6 g) that have become known later as botanical groups. Other taxonomic methods to study the evolution and genetic diversity in the species followed and were mainly based on morphology (Muratova 1931), geographic origin (Cubero 1974) and karyotype (Yamamoto 1973). The genetic variation currently available is not fully exploited and sterility barriers do not exist between subspecies or between botanical groups (Lawes et al. 1983).

Controlled hybridization between selected faba bean genotypes is the basis of improvement programs practiced by faba bean breeders. Crosses combining winter hardiness of *equina* winter types and absence of tannin originating from vegetable *major* spring beans (Bond 1976) is one of many examples of the intense recombinations between botanical groups that have been practiced in faba bean. Most breeding programs are based mainly on elite germplasm, including locally adapted and unadapted cultivars. Reliable information on the genetic relationships between such germplasm and the amount of genetic diversity the germplasm comprises is crucial in planning breeding-programs. Such information is difficult to achieve if based only on morphological and agronomical characters that are affected by environment and the growth stage of the plant.

Molecular markers have been repeatedly applied to study the genetic diversity in faba bean. Käser and Steiner (1983) employing protein and isozyme markers reported no clear grouping of 22 cultivars and 49 landraces from a world-wide collection except for the 17 German cultivars studied. Link et al. (1995), applied the random amplified polymorphic DNA (RAPD) technique on 28 inbred lines to study the genetic diversity within and between

European *minor* and *major* germplasm and Mediterranean germplasm. They grouped the European large-seeded lines as an intermediate group between the Mediterranean and the European small-seeded groups. The amplified fragment length polymorphism (AFLP) analysis was used to analyze in detail the genetic diversity among 22 faba bean inbred lines derived from elite cultivars (Zeid et al. 2001). Several approaches were used to analyze the marker-based genetic diversity of the breeding material. If a priori defined groups exist, each entry may be characterized by its similarity to its own group compared to the other group(s) (e.g. Lübberstedt et al. 2000). Very often, clustering and distance-based ordination methods like Principal Coordinate Analysis (PCoA) were applied (e.g. Becker et al. 1995; Lübberstedt et al. 2000). A direct quantification of the degree of differentiation of a gene pool into sub-pools was proposed by Finkeldey (1994) and Excoffier (1994), focusing on codominant markers and allele frequencies.

Despite irrefutable advantages, molecular data are not free of errors. The reproducibility of the AFLP profile in sugar beet was tested by Jones et al. (1997) across a network of six European laboratories. Using two *Mse*I/*Pst*I primer combinations that yielded 16 polymorphic fragments, only a single difference was observed (absence of a band) in the produced AFLP profile. Apart from the reliability of the marker profile, its consistency, i.e. whether an accession and a primer combination give the exact banding profile when the reaction is repeated under constant laboratory conditions, is another important aspect. Running duplicates of the tested genotypes has been reported in some studies as an indication for reproducibility (Winfield et al. 1998; Hansen et al. 1999). However, it is rarely mentioned whether duplicates were re-runs of the same PCR products, re-amplifications or a replicate of a second individual within the same genotype. In most cases it was only a replicated lane in the same gel. Furthermore, the minimum number of AFLP markers required to achieve a given level of precision, estimated by re-sampling of subsets of markers through the Bootstrap (Efron and Tibshirani 1986) and/or the Jacknife approach (Sokal and Rohlf 1981), is of great interest.

The objectives of this work were: (1) to study the structure of the genetic diversity within a large sample of 79 selected elite cultivars representing the actual faba bean material used for breeding in Asia, Europe and North Africa based on AFLP, and (2) to analyze the precision of the results obtained.

Materials and methods

Plant material

This investigation was based on a large sample of (n = 79) *V. faba* L. inbred lines derived from elite cultivars that were released in various world markets (eight Asian cultivars, 40 European spring *minor* cultivars, eight European spring *major* cultivars, eight South European cultivars, seven European winter bean cultivars and eight North African cultivars; Table 1). This collection comprises actual cultivars that are available on the market today, e.g. Giza 402 (Egypt), Scirocco (Germany) and Styria (Austria), and ancestral cultivars, e.g. Cixi Dabeican (China), Herz Freya (Germany) and Minica (Netherlands), that have dominated breeding and production for a considerable period of time and are now represented in many of the pedigrees of recent germplasm (Stelling et al. 1994). Inbred lines were developed in Germany at the Institute of Agronomy and Plant Breeding, Göttingen, and at the State Plant Breeding Institute, Stuttgart-Hohenheim, for one to more than 12 generations of selfing (Table 1), employing the single-seed descent method.

DNA isolation and AFLP analysis

One individual was used to represent one inbred line (exceptions are explained below). Approximately 0.20 g of young leaves harvested from 15-day old seedlings was directly ground in liquid nitrogen. DNA was isolated using CTAB extraction buffer according to Doyle and Doyle (1990). AFLP reactions were performed according to the procedure described by Vos et al. (1995), using a commercially available kit (AFLP analysis System I, GIBCO BRL, Life Technologies, Inc., Rockville, Md.) and following the manufacturer's instructions.

Eight out of 14 tested *Eco*RI/*Mse*I primer combinations were employed (Table 2). The *Eco*RI primers were radiolabeled with [γ-^{33}P]-dATP provided from Amersham Pharmacia Biotech. After selective amplification, the polymerase chain reaction (PCR) products were mixed with an equal volume of loading buffer [98% (v/v) formamide, 10 mM EDTA, 0.05% (v/v) xylene cyanol, and 0.05% (v/v) bromophenol blue] and denatured at 95 °C for 5 min and directly placed on ice. Seven microliters of the mixture were loaded on an 8% (v/v) polyacrylamide gel (30 × 40 cm) containing 8.3 M of urea that was pre-warmed for 20 min. Gels were run with a 0.5× TBE electrophoresis buffer [50 mM Tris, 50 mM boric acid, 1 mM EDTA, pH 8.3] at 58-W constant power. After electrophoresis, gels were fixed in 10% acetic acid and dried. Dried gels were then exposed to X-ray films X-OMAT AR (Kodak) for 3–5 days depending on the intensity of the radiation signal.

Detection and scoring of AFLP fragments

AFLP amplification products, separated by gel electrophoresis along the inbred lines' lane, are located at positions on the gel that strictly correspond to the products' length. Any position on the gel which, when scored across genotypes, contains at least once a DNA amplification product will be called herein the "fragment". Whenever a given fragment contains a scoreable amplification product, this will be termed the "band". Hence, the total number of fragments is a constant number for the experiment, whereas the number of bands per genotype varies across genotypes. Monomorphic fragments contain an amplification product, a band, in each inbred line, while polymorphic fragments do not. Fragments were scored as "1" when a band was present and were scored as "0" if not. In doubtful cases, mostly where a band's intensity was markedly less-strong than that of checks but not convincingly absent, a score of "9" was given and it was considered as a missing data-point later in the analysis. Fragment scoring was performed visually from X-ray films. Throughout, two neighboring lanes on a gel were assigned to the amplified products of a single inbred line; a maximum of 24 inbred lines could be scored per gel. Thus, to test the 79 inbred lines with a single primer combination, at least four gels were run and scored. To facilitate scoring across gels, an AFLP DNA ladder (30–300 bp; GIBCO BRL, Life Technologies, Inc., Rockville, Md.) was used. Furthermore, for each of the tested primer combinations, amplification products of two inbred lines (K25 and Pel, Table 1) were included as additional checks in separate lanes on each gel.

Table 1 Ten of these lines (‡) were used with two individuals (replicate assays), one (†) was used as a replicate and as a duplicate assay (one PCR product analyzed in two neighboring lanes); two were additionally run on every gel as internal checks (§)

Faba bean cultivar	Country of origin	Year of release	Inbreeding generation
European minor			
Alfred	Netherlands	1983	F_3
Alpine‡	UK	1997	F_3
Blaze	UK	1976	F_{10}
Blitz‡	Canada	1994	F_2
Carola	Austria	1986	F_8
Caspar	Netherlands	1992	F_7
Columbo	Denmark	1996	F_3
Condor	Germany	1990	F_8
Corvette	Denmark	1996	F_3
Erfano	Germany	1988	F_5
Fatima	Canada	1993	F_2
Francks Ackerperle	Germany	–	F_{10}
Franz	Austria	1990	F_4
Geo	Germany	1989	F_6
Gloria	Austria	1994	F_8
Gobo	Germany	1987	F_8
Hedin	Germany	1986	F_9
Herz Freya	Germany	1935	F_{12}
Jantarnij	Russia	–	F_2
Karna	Austria	1983	F_6
Kornberger Klein	Austria	1956	F_{10}
Kristall (K25)†, §	Germany	1973	F_{12}
Limbo	Germany	1998	F_2
Luna	UK	1991	F_2
Maris Bead	UK	1964	F_3
Mars	Denmark	1993	F_3
Maya	France	1995	F_3
Merkur	Czech Republic	1997	F_3
Music	France	1995	F_3
Nadwislanski	Poland	1955	F_7
Nixe	Germany	–	F_{12}
Orletzkij	Russia	–	F_2
Pistache	Netherlands	1990	F_7
Scirocco	Germany	1992	F_7
Styria	Austria	1996	F_{12}
Topas	Germany	1986	F_8
Troy	Germany	1985	F_{11}
Ukko‡	Finnland	1984	F_7
Victor	Netherlands	1988	F_{10}
Wieselb. Klein	Austria	–	F_8
European major			
Canner Express	Netherlands	–	F_{12}
Con Amore	Netherlands	1955	F_{12}
Dreifachweiss	UK	1905	F_7
Hylon	UK	1980	F_7
Imperial Green W.‡	Netherlands	1930	F_7
Minica	Netherlands	1973	F_{13}
L_1*	Germany	–	F_7
L_2*	Germany	–	F_7
Winter bean			
Clipper	UK	1996	F_2
Côte d'Or‡	France	–	F_8
Hiverna	Germany	1986	F_8
Punch	UK	1988	F_2
Striker	UK	1994	F_2
Target	UK	1996	F_2
Webo	Germany	1997	F_7
South Europe			
Aquadulce	Spain	–	F_6
Enantia‡	Italy	–	F_3
Gemini	Italy	–	F_3
Peleponnes (Pel) §	Greece	–	F_{12}
Pietranera	Italy	–	F_{10}
Polycarpe	Greece	–	F_3

Table 1 (continued)

Faba bean cultivar	Country of origin	Year of release	Inbreeding generation
Sikania	Italy	–	F_3
Sikelia	Italy	–	F_3
North Africa			
F402/4	Egypt	–	F_4
Giza 3‡	Egypt	1979	F_{11}
Giza 402‡	Egypt	1979	F_4
Giza 461‡	Egypt	1990	F_4
Hudeiba	Sudan	1969	F_4
Morocco (34M)[#]	Morocco	–	F_6
Rebaya	Egypt	–	F_8
Zeidab Local‡	Sudan	–	F_4
Asia			
Cixi Dabeican	China	–	F_2
Kawachi-green	Japan	–	F_2
Mairudo-green	Japan	–	F_2
Maya-Asia	China	–	F_2
Otafuku	Japan	–	F_2
Pinghu Zaojiazhong	China	–	F_2
Pingyang Zhaodozi	China	–	F_2
Shangya Tian	China	–	F_2

[#] A line obtained from ICARDA (BPL228)

* L1 and L2 are two inbred lines originating from the same cross Minica × Canner Express Pedigree information: Troy is parent of Maya. Alfred is parent of Mars. Minica is a parent of Alfred and of Victor

Table 2 Level of polymorphism revealed by the eight *Eco*RI/*Mse*I primer combinations with 79 faba-bean inbred lines

Primer combination	Total number of fragments	Polymorphic fragments [%]
E-AAC/M-CAC	80	91.25
E-ACA/M-CTA	46	91.30
E-ACC/M-CAC	56	96.43
E-ACC/M-CAG	47	87.23
E-ACG/M-CTT	52	94.23
E-AGC/M-CTA	71	91.55
E-AGG/M-CAT	92	83.70
E-AGG/M-CTC	83	91.57

Consistency of the AFLP profiles

The consistency of AFLP profiles was assessed by comparing the marker phenotypes from one duplicate assay (inbred line K25) and from replicate assays of 11 inbred lines (cf. Table 1) using seven primer combinations. The duplicate AFLP profiles were the outcome of two aliquots of one pre-amplification reaction of the entry K25 (Table 1), each of which was run in a separate gel. Replicate AFLP profiles were the product of replicate assays. A replicate assay comprised two individuals, termed the replicate individual and the reference individual. These two were prepared on separate occasions from seed, harvested from the same selfed plant of a given inbred line. The replicate and reference individual of inbred line K25 were run together in the same gel throughout. A further ten replicate assays, as specified in Table 1, were run with their two individuals in separate gels, respectively.

Data analysis

The genetic similarity values (GS values) between the two individuals of each of the ten replicate assays were estimated employing three GS coefficients: GS_J (Jaccard 1908), GS_{SM} (Simple Matching; Sneath and Sokal 1973) and GS_{NL} (Nei and Li 1979). Analysis of data was performed for the primer combinations one at a time. It was analyzed whether the mean GS_J values across the ten replicate assays were significantly different as a result of the actual primer combination. Differences between the GS_J value of a replicate individual to any of the other 78 inbred lines and the GS_J value of its reference individual to the same 78 inbred lines were analyzed, and used to estimate an error variance σ^2_e and a least-significant difference (LSD) for these GS_J values as well as for the GS_J values of the main data set (see below).

For the 79 inbred lines, the 1/0/9 matrix that resulted from the scoring of AFLP products of eight primer combinations (main data set) was used to calculate all possible pairwise similarity coefficients between inbred lines. After excluding monomorphic fragments, GS_J were used throughout, applying the software NTSYS-pc version 2.1q (Rohlf 2000). Jaccard's coefficient and the exclusion of monomorphic fragments were implemented following the recommendation of Link et al. (1995). Bootstrapping (Felsenstein 1985) on the similarity matrix was performed. Marker samples of different sample sizes, starting from five fragments per sample and increasing in steps of 35 fragments per sample, were randomly drawn from the total number of fragments (n = 477) with replacement. Two hundred marker samples for each sample size were established for two pairs of inbred lines; the most-distant pair (Blaze and Columbo) and the most-similar pair (L$_1$ and L$_2$). The mean coefficient of variation (CV%) for Jaccard's coefficient of similarity for each sample size was calculated using the software

Gen Dist. 1.8 provided by A. Valentini, Universita' della Tuscia, upon request.

Based on the country of origin of the cultivars and on their seed size, it was obvious to divide the whole collection into six groups of germplasm, namely: Asian, European (EU *major* and EU *minor*), South European, winter beans and North African. For the large (n = 40) EU *minor* group and the other five smaller groups, the average intra-group GS_J-value was calculated $[(n^2 - n)/2$ values per group]. Moreover, for each of the lines belonging to one of the five small germplasm groups, its average GS_J-value to the n = 40 EU *minor* lines was calculated. The difference between the mean inter-group similarity of e.g. the EU *major* lines to the EU *minor* lines, and the intra-group EU *minor* similarity was tested by the *t*-test for the significance of difference; the test was conducted for all five small groups. The GS_J mean values describing the similarity between groups of the germplasm, taking the group as the smallest unit of analysis, were used to construct a dendrogram based on the unweighted pair-group method with arithmetic means (UPGMA) using NTSYS-pc version 2.1q (Rohlf 2000).

The similarity matrix of the 79 individual inbred lines (main data set) was used to construct a further dendrogram using UPGMA. Associations among the 79 inbred lines using PCoA according to Gower (1966) were revealed based on the GS_J estimates employing NTSYS-pc version 2.1q (Rohlf 2000). Bootstrap analysis was performed by drawing 400 random samples of fragments, to determine the confidence limits of the UPGMA-based dendrogram using the software package "WinBoot" developed at IRRI (Yap and Nelson 1996).

Results

Analysis of 79 faba bean lines with eight *Eco*RI/*Mse*I primer combinations (Table 2) resulted in the amplification of 527 scoreable fragments. The number of scoreable fragments per primer combination ranged from 46 to 92 for the primer pairs E-ACA/M-CTA and E-AGG/M-CAT respectively (Table 2), with an average of 66 fragments. From those, 477 were polymorphic, thus monomorphic fragments amounted to 9.5% of the total number.

Consistency of the AFLP profiles

The inbred line K25 (highly inbred; F_{12}), was used for a duplicate assay (two aliquots of one pre-amplification reaction run in different gels). Nine mismatches from a total of 477 polymorphic fragments ($GS_J = 0.97$) were found. A replicate assay of K25 (DNA from two individuals belonging to this line), run together in one gel, was scored with only three mismatches ($GS_J = 0.99$).

Examining the GS_J values between the ten replicate individuals and the 79 entries (including the corresponding ten reference individuals; Table 1) showed that each replicate individual expressed its maximum similarity with its corresponding reference individual. Results on

Table 3 Range and mean of the three genetic similarity estimators for the ten replicate assays based on results of seven primer combinations

Estimator	Min.	Max.	Mean
Simple Matching	0.75	0.83	0.80
Nei and Li	0.83	0.88	0.86
Jaccard	0.71	0.79	0.76[#]

[#] Compare Fig. 1

the GS values for the replicate assays when estimated based on the three tested coefficients of similarity are summarized in Table 3. The mean genetic similarity based on Nei and Li between a replicated individual and its reference line amounted to $GS_{NL} = 0.86$, followed by a mean $GS_{SM} = 0.80$ and a mean $GS_J = 0.76$.

Consider two GS_J values comprising one common line, this common line being one of the ten replicate assays [e.g., GS_J (Alpine × Alfred) = 0.69 and GS_J (Alpine × Blaze) = 0.62, cf. Table 1]. Hence, for the common line, data across two individuals are available. Comparing these two GS_J values, an error variance of $\sigma^2_e = 4.95 \times 10^{-4}$ was estimated, i.e. the standard error deviation amounts to 2.22×10^{-2} (cf. Table 4). The corresponding LSD $_{(0.05)}$ value amounts to 4.37×10^{-2}. Extending this to the main data set, where GS_J values are calculated from non-replicated data, the appropriate $LSD_{(0.05)}$ to compare GS_J values amounts to 6.17×10^{-2}. The error variance (4.95×10^{-4}) amounts to about 29% of the variance between all GS_J values in the main data set (17.32×10^{-4}).

Genetic similarities and PCoA

Genetic similarity values of pairs of inbred lines ranged from $GS_J = 0.88$ between the inbred lines L_1 and L_2, and $GS_J = 0.53$ between the inbred lines Blaze and Columbo. Clustering of inbred lines based on their GS_J values using the UPGMA, showed no major "ball clusters" (*sensu* Rohlf 1993). Six of the eight Asian inbred lines clustered together. Within this cluster (Fig. 1), the two Japanese inbred lines, Otafuku and Kawachi-green, clustered together and were separated from the four Chinese inbred lines (Cixi Dabeican, Shangya Tian, Pinghu Zaojiazhong and Maya-Asia). None of the other groups formed a recognizable cluster. Highly similar lines were Mars and Alfred ($GS_J = 0.79$; Alfred is a parent of Mars), Condor and K25 ($GS_J = 0.76$; Kristall is a parent of Condor), Maya and Troy ($GS_J = 0.76$; Troy is a parent of Maya). Jantarnij and Orletzkij ($GS_J = 0.81$) as well as Mars and Merkur ($GS_J = 0.79$) are seemingly related but do not

Table 4 Analysis of variance for GS_J values calculated between ten replicate assays (I) with two individuals each (R), and n = 79 inbred lines (L)

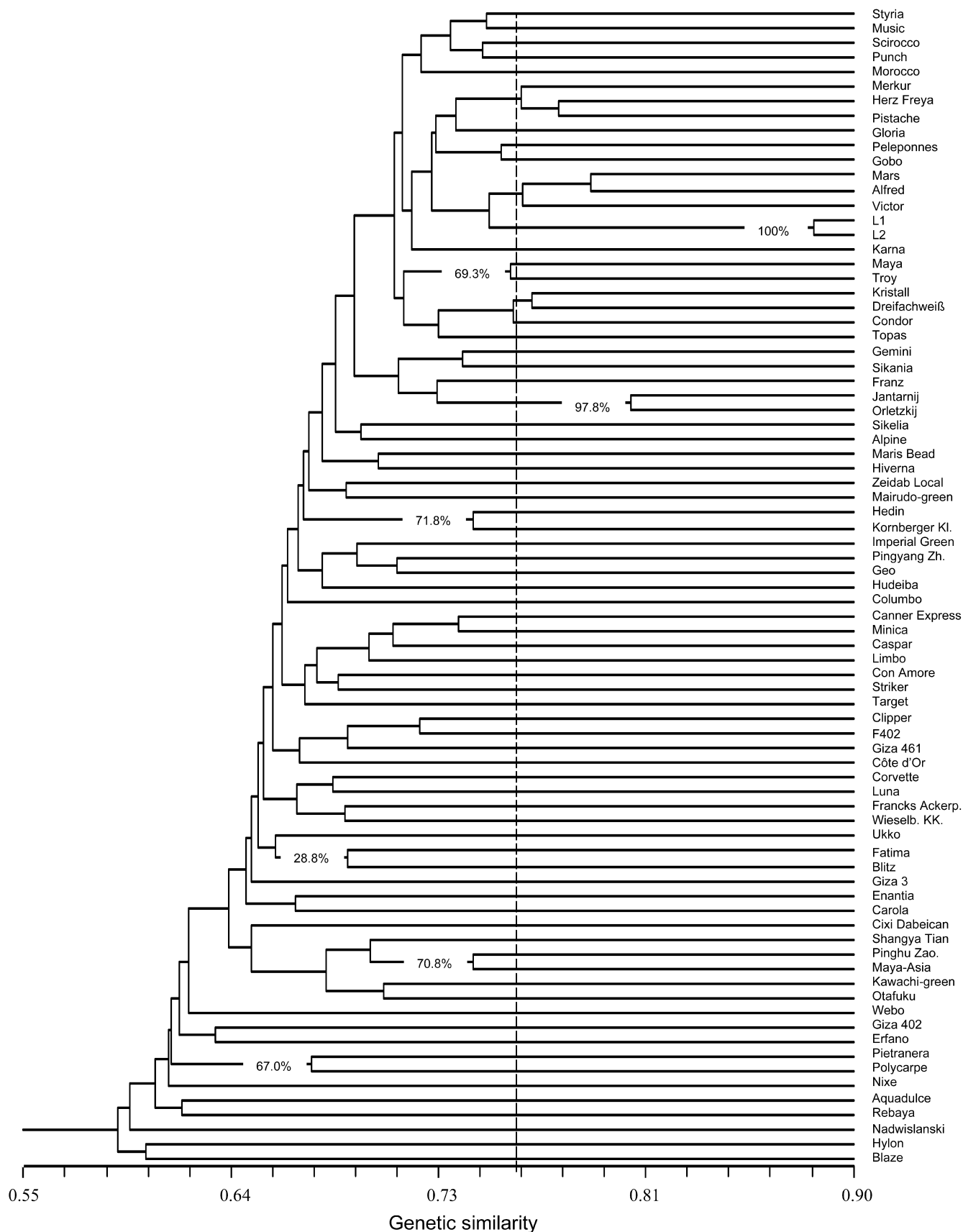| Source	df	Mean squares (×10⁻⁴)	Variance components (×10⁻⁴)	LSD (0.05) (×10⁻²)
I	9	238.78	1.48	0.49
L	78	100.33	4.77	1.38
LI	692	4.34	−0.31	4.37
LRI	780	4.95	4.95	–

Fig. 1 Grouping of 79 faba-bean lines based on 477 polymorphic fragments using Jaccard's coefficient of similarity and UPGMA clustering. The *dotted line* indicates the threshold value of 0.76 (compare Table 3), *nodes* to the right of this line present tightly related inbred lines. *Numbers* shown at different nodes represent percentage confidence limits of the bootstrap analysis
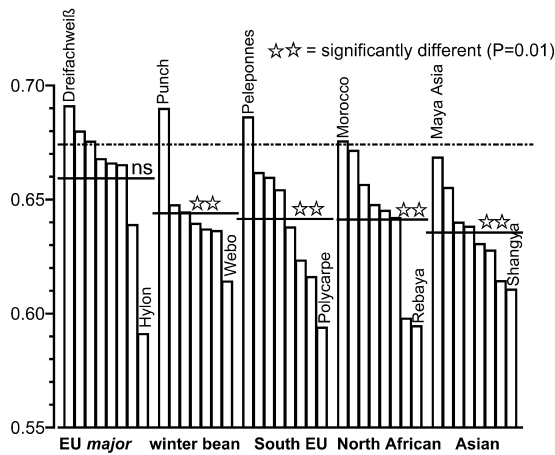
Fig. 2 Mean GS_J values between the individual non-EU *minor* inbred lines and the 40 EU *minor* lines. The *solid lines* show group mean similarities, the *broken line* shows the average intra-EU *minor* similarity for comparison. Intra-EU *minor* similarity and the mean similarities of the groups were tested by the *t*-test for significance of difference

Fig. 3 Dendrogram generated by UPGMA for mean GS values of germplasm groups

Fig. 4 Associations among 79 faba bean inbred lines revealed by Principal Coordinate Analysis performed on 477 polymorphic fragments

share common ancestors based on available pedigree data.

Bootstrap values indicated on the dendrogram showed that the two inbred lines L_1 and L_2 were grouped together in 100% of the cases. Similarly, a bootstrap value of $P = 97.8\%$ was shown for the inbred lines Jantarnij and Orletzkij (Fig. 1). Other inbred lines grouping together with a relatively high confidence level, although not reaching the 95% bootstrap P value (Felsenstein 1985), were e.g. Maya and Troy ($P = 69.3$), and Pinghu Zaoajiazhong and Maya-Asia ($P = 70.8$).

The mean intra-group similarity within the EU *minor* lines amounted to $GS_J = 0.67$; Mars, Herz Freya, Kristall, Condor and Victor showed the highest values ($GS_J \approx 0.70$), whereas Blaze, Nadwislanski, Nixe, Erfano and Fatima showed the smallest ($0.61 < GS_J < 0.67$). The same mean value ($GS_J = 0.67$) was found for the intra-group similarity within the EU *major* lines and within the Asian bean lines. The mean intra-group similarity within the winter beans amounted to $GS_J = 0.64$; the same mean value was found for the intra-group similarity within the South European and within the North African inbred lines. The mean genetic similarity between the EU *major* and EU *minor* lines amounted to $GS_J = 0.66$. This was not significantly different ($P = 0.05$) from the intra-group EU *minor* value ($GS_J = 0.67$, cf. Fig. 2). The similarities of the other small germplasm groups to the EU *minor* group lines were significantly different (P = 0.01) from the intra-group EU *minor* value (Fig. 2), the Asian beans showing the smallest value. Upon clustering of the six groups using the UPGMA based on their mean GS_J values (Fig. 3), the Asian group was convincingly separated from all other groups. These could be regarded as two further groups: EU *minor*, EU *major* and winter bean vs South European and North African. Yet, this further grouping is less convincing.
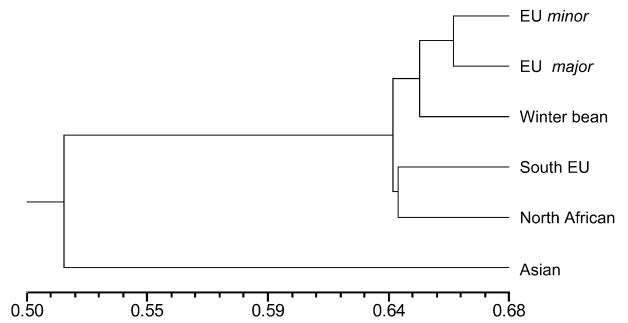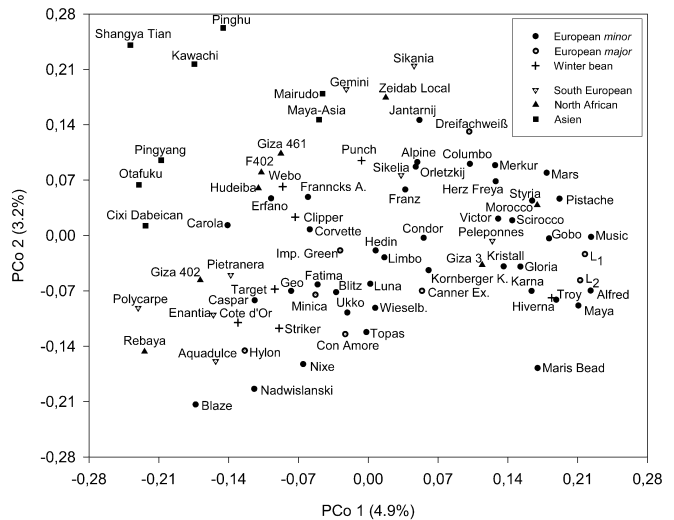
The first two principal coordinates (Fig. 4) explained 8.16% of the total variability, with the Asian group partly separated from the others. Both, the EU *major* and *minor* groups were spread across the principal coordinates and no grouping was seen. The pattern shown was not much clearer when the other three groups were regarded (Fig. 4).

Bootstrapping of AFLP fragments revealed that the coefficient of variation (CV%) of Jaccard's similarity coefficient between the inbred lines Blaze and Columbo (the most distant inbred lines) decreased from 50% for samples of size-five fragments to 5.1% for samples of n = 477 fragments (Fig. 5). A coefficient of variation of 2.1% for samples of n = 477 fragments, representing the whole data set, was achieved for the closest inbred lines: L_1 and L_2.
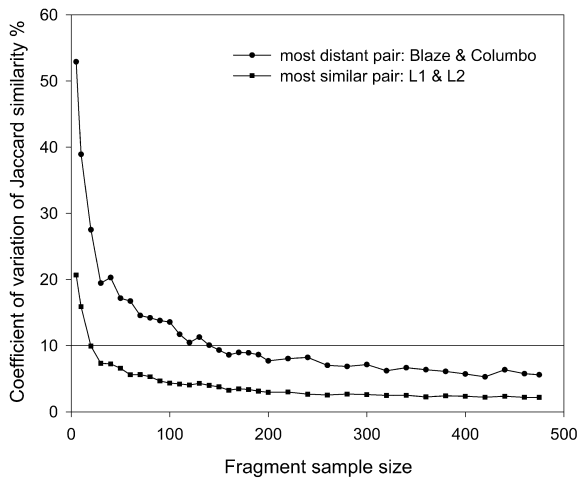
Fig. 5 Coefficient of variation of Jaccard's genetic similarity in relation to marker sample size (n = 200 bootstraps). Fragment samples of different size were sampled with replacement from two pairs of individuals: the most similar and most distant

Discussion

Error estimation

Not only the amount of diversity (fixed when sampling the inbred lines), but the amount of error in the scored data as well, are decisive for the outcome and usefulness of a diversity analysis. DNA quality differences, affecting the AFLP banding pattern, may result in errors (Vos et al. 1995). Hansen et al. (1999) evaluated the usefulness of the AFLP technique for genetic analysis of sugar beet. Gel resolution was among the sources of error reported in their study. In the present study replicate and duplicate assays were employed to allow for inspection and quantification of error in the results. The replicate (GS_J = 0.99) and duplicate assay (GS_J = 0.97) with K25 demonstrated the high consistency of the AFLP profiles obtained here. This is in agreement with Winfield et al. (1998) where estimated GS values for duplicates of *Populus nigra* ranged between 0.96 < GS < 1.00; this seems to reflect the amount of technique-born errors for AFLPs. As to the amount of error estimated, replicate assays of ten inbred lines run in different gels showed an average of 82 mismatches with an average genetic similarity GS_J = 0.76. Theoretically, this value should be near to unity as shown for the replicate assay of K25 (GS_J = 0.99). However, K25 itself was one of the two internal checks, which resulted in better scoring, since all of its bands were traced precisely across gels. Scoring of bands of the ten replicate assays across gels on the other hand, was based on the bands of the internal checks (K25 and Pel) as a reference. This obviously led to high error rates during scoring. These results are supported by the observations of Schwarz et al. (2000), showing greater inter-gel differences in the band size of identical samples than intra-gel differences for fluorescence-based semi-automated AFLP analysis in barley and wheat.

Sneath and Sokal (1973) noted the usefulness of testing the significance of the difference between similarity coefficients. The least significant difference in the main data set was $LSD_{0.05}$ = 0.062. This value focuses on differences between individuals within the same inbred line (replicate assays) caused by residual segregation within the lines and by errors after DNA extraction, especially at scoring across gels. Applying this LSD to GS_J values of both L1 and L2 to any other inbred line, showed no significant differences between these values; the same was true for the inbred line pair Jantarnij and Orletzkij. Opposed to this, 96.4% of the GS_J values of any EU *major* line to any other EU *major* lines was surpassing this difference when compared with the corresponding GS_J values to any Asian line (cf. Fig. 2), reflecting the larger AFLP similarities among EU *major* lines compared to those between EU *major* and Asian lines. These findings substantiate the usefulness of the LSD-value.

Primer combinations to employ

Ellis et al. (1997) observed that 80% of the provided genetic information could already be achieved by selecting six "good" primer combinations. Selection was based on the number of polymorphic fragments revealed and the number of genotypes uniquely identified by each primer combination. We compared the GS_J values of each of ten replicate assays for one primer-combination at a time. All primer-combinations were equally useful (Table 5), and no significant differences were observed between them. Consequently all primer-combinations were employed.

The appropriate genetic-similarity coefficient

The appropriateness of similarity coefficients has been often debated (Piepho and Laidig 1997; Robinson and Harris 1999). In the present case (Table 3), where two replicate sister individuals taken from the same selfed parental plant were studied, we suggest employing GS_{SM} that takes (0-0) negative matches into account. Negative matches here uncover the genetic similarity, since the

Table 5 Analysis of variance for GS_J values calculated for ten replicate assays (I), with two individuals each, when tested using seven different primer combinations (P)

Source	df	Mean squares ($\times 10^{-4}$)	Variance components ($\times 10^{-4}$)	F
I	9	47.1	−1.51	0.82
P	6	73.3	1.56	1.27
IP	53	58.0	58.0	–

cause for the absence of a band in these two individuals should be identical and inherited from the same ancestor (identity by descent, Smith and Smith 1992). This however is not the case when comparing non-related inbred lines, where the absence of a given band is not necessarily based on the same genetic cause: it may or may not be caused by the same DNA variant. Thus, in such cases a similarity coefficient that ignores negative matches is more appropriate. Both Jaccard and Nei and Li (1979) disregard negative matches. Piepho and Laidig (1997) demonstrated the monotonic, non-linear relationship between both namely: $GS_{Nlxy} = 2\ GS_{Jxy}/(1 + G_{Jxy})$; hence, the rank correlation between GS_{NL} values and GS_J is $r = 1.00$, which could be corroborated by our findings.

Lamboy (1994) studied the bias in an estimated GS value, defined as the difference between the estimated and true GS value, caused by false positive bands or false negative bands. He regarded dominant marker data, and used both Jaccard's and Nei and Li's coefficients of similarity. Simulations demonstrated smaller bias by Nei and Li than by Jaccard. A disadvantage of the Nei and Li coefficient in the case of dominant markers is: it is not a linear function of the coancestry coefficient of the lines under study, whereas the Jaccard coefficient indeed is a linear function; provided the entries are homozygous lines (Link et al. 1995). Decision on which of these two coefficients should be used will depend on the type of molecular marker employed as explained by Engqvist and Becker (1994) and Link et al. (1995), i.e. the Jaccard for dominant markers, and Nei and Li (1979) for codominant markers.

Here, not all lines were highly homozygous (Table 1), which could blur the results. Less inbred individuals tend to have more bands; they do share more common bands with others; but for a heterozygous AFLP locus the recessive AFLP allele ("no band") is ignored. On average, n = 311.6 bands were amplified per highly inbred line (>F9) in comparison to n = 317.1 bands for lines that were selfed for only three generations or less. The average number of bands expected from crosses (F1) among the highly inbred lines, following a dominant inheritance of AFLPs, amounted to n = 384.2 bands. Compared with this result, obviously a high-inbreeding level (311.6 ≈ 317.1 ≪ 384.2) was achieved even with our less-inbred entries. This is expected; the inbred lines were obtained by selfing individuals from elite cultivars that are already rather narrow populations or even lines.

Re-sampling

A stable classification of genotypes is not obtained based on a small number of polymorphic fragments (Sneath and Sokal 1973). Bootstrap analysis was applied to the genetic similarity matrix of the 79 inbred lines to estimate the sampling variance of the genetic similarities. Bootstrapping of 477 polymorphic fragments for all GS_J values resulted in a mean CV of less than 4% (details not shown), indicating that a sufficient number of markers

were applied here. To achieve a CV < 10%, only 110 AFLP fragments were required for the most-distant pair of inbred lines; closer pairs of inbred lines required even fewer fragments. A further increase in the number of fragments did not result in a proportional decrease in the CV value. These results are in accordance with those obtained on other marker types and species (Tivang et al. 1994; Ajmone-Marsan et al. 2001), indicating only marginal gains in precision upon increasing the number of fragments beyond a certain limit. A further point, not regarded here (because mapping data for the fragments scored here is not available), is the localization and distribution of the marker loci across the faba bean genome (Satovic et al. 1996); an even distribution should be favorable (Castiglioni et al. 1999).

Pedigree and historical aspects

Considering the long-history of the cultivation of faba bean in Central and Western Europe, the European *minor* group is represented by a dominating, large-sample of selected cultivars in this study (n = 40). Yet narrowing down this sample to eight lines (eight is the number in other groups) would not alter the final conclusion deduced from this experiment (details not shown). Central and South America were not represented here, though faba bean was introduced there in the sixteenth century by the Spaniard and the Portuguese, and surely underwent an independent development there (Lawes et al. 1983).

For the 79 inbred lines, the cluster analysis indicated that no clear grouping was detected, neither based on seed size (EU *minor*, EU *major*) nor on geographic origin. In one case only, for the Asian group, six of the respective eight inbred lines clustered together (Fig. 1). Yet, cases of known pedigree relationships like Alfred and Mars were convincingly corroborated (cf. Table 1); "non related pedigrees" were uncovered as doubtful because of temptingly high genetic similarity (e.g. Jantarnij and Orletzkij). Nadwislanski, Blaze and Hylon were uncovered as very loosely related to any other line; hence, it is very tempting to consider these for crosses with recent elite lines not being isolated, like Mars, Condor and Limbo. The PCoA did not show a more marked pattern in the AFLP-based genetic diversity either. The first two principal coordinates of the PCoA (Fig. 5) explained 8.16% of the variation. This, in spite of the first impression, is not a small quantity; a high number of entries (n = 79) drastically restrict this quantity. In maize, Messmer et al. (1992) employing RFLP markers and Lübberstedt et al. (2000) employing AFLP markers found quantities of 14.6–17.5% with smaller numbers (n = 57 and n = 51), even though lines of two clearly distinct maize gene pools were analyzed in their studies.

A rather simple approach was, to a priori group the inbred lines according to their geographic origin and their botanical group (Figs. 2, 3). Hanelt (1972) reported that in Europe, faba bean *minor* was the first group to be cultivated and surprisingly late, after about the year 500,

the *major* group did emerge. Only after the seventeenth century the first true winter hardy stocks were developed; those were most probably selected from the European *minor* group (Hanelt 1972). Our results showed a higher genetic similarity among the EU *minor*, EU *major* and the winter bean germplasm in comparison to genetic similarity to the other groups. The Near or Middle east was suggested to be the center of diversity. The spread of the crop took place towards Central and Northwest Europe through North Africa: hence, some similarity is expected between European and North African germplasm. In the Far East, faba bean was probably introduced already in about 100 BC (Tao 1981). Regarding the very far geographic distance of the European and Mediterranean area to China and Japan, the results of the molecular data (Fig. 3) are in accord with the expectations.

Results from the marker data presented in this study revealed the broad genetic base of the faba bean germplasm investigated. Although crossing between faba bean germplasm groups has occurred in the past and is still an ongoing process, a great potential for further cultivar improvement and hybrid production is still available. Wide crosses between materials from the European germplasm and that from Asia are of special importance, e.g. if traits like winter-hardiness of the Asian material are of interest. The European germplasm, although showing an elevated intra-group similarity, comprises quite diverse lines that provide a good base for utilizing readily adapted elite material for future hybrid-breeding programs. Studies concerning the relationship between genetic similarity based on AFLP markers and hybrid performance within the European elite germplasm were launched, and results are currently under evaluation.

References

Ajmone Marsan P, Negrini R, Crepalde P, Milanesi E, Gorni C, Valentini A, Cicogna M (2001) Assessing genetic diversity in Italian goat populations using AFLP markers. Int Soc Animal Genet 32:281–288

Becker HC, Engquist GM, Karlsson B (1995) Comparison of rapeseed cultivars and re-synthesized lines based on allozyme and RFLP markers. Theor Appl Genet 91:62–67

Bond DA (1976) Field bean, *Vicia faba*. In: Simmonds NW (ed) Evolution of crop plants. Longman, London, pp 179–182

Castiglioni P, Ajmone Marsan P, van Wijk R (1999) AFLP markers in a molecular linkage map of maize: codominant scoring and linkage group distribution. Theor Appl Genet 99:425–431

Cubero JI (1974) On the evolution of *Vicia faba* L. Theor Appl Genet 45:47–51

Doyle JJ, Doyle JL (1990) Isolation of plant DNA from fresh tissue. Focus 12:13–15

Efron B, Tibshirani R (1986) Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. Stat Sci 1:54–77

Ellis RP, Mc Nicol JW, Baird E, Booth A, Lawrence P, Thomas B, Powell W (1997) The use of AFLPs to examine genetic relatedness in barley. Mol Breed 3:359–369

Engqvist GM, Becker HC (1994) Genetic diversity for allozymes, RFLPs and RAPDs in re-synthesied rape. In: van Ooijen JW, Jansen J (eds) Biometrics in plant breeding: application of molecular markers. Proc 9th meeting of the EUCARPIA Section Biometrics in Plant Breeding, 6–8 July 1994, Wageningen, pp 85–90

Excoffier L (1994) The statistical analysis of molecular data for inferring population genetic structure: the AMOVA framework. In: Perrin N, Keller L, Goudet J (eds) Evolution of structured populations. Bull Soc Vaud Sci Natl 83:159–160

Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39:783–791

Finkeldey R (1994) A simple derivation of the partitioning of genetic differentiation within subdivided populations. Theor Appl Genet 89:198–200

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

Hanelt P (1972) Die infraspezifische Variabilität von *Vicia faba* L. und ihre Gliederung. Die Kulturpflanze 20:75–128

Hansen M, Kraft T, Christiansson M, Nilsson NO (1999) Evaluation of AFLP in *Beta*. Theor Appl Genet 98:845–852

Jaccard P (1908) Nouvelles researches sur la distribution florale. Bull Soc Vaud Sci Natl 44:223–270

Jones CJ, Edwards KJ, Castaglione S, Winfield MO, Sala F, van de Wiel C, Bredemeijer G, Vosman B, Matthes M, Daly A, Brettschneider R, Bettini P, Buiatti M, Maestri E, Malcevschi A, Marmiroli N, Aert R, Volckaert G, Rueda J, Linacero R, Vazquez A, Karp A (1997) Reproducibility testing of RAPD, AFLP and SSR markers in plants by a network of European laboratories. Mol Breed 3:381–390

Käser HR, Steiner AM (1983) Subspecific classification of *Vicia faba* L. by protein and isozyme patterns. Fabis Newslett 7:19–20

Lamboy W (1994) Computing genetic similarity coefficients from RAPD data: the effects of PCR artefacts. PCR Methods Appl 4:31–37

Lawes DA, Bond DA, Poulsen MH (1983) Classification, origin, breeding methods and objectives. In: Hebblethwaite PD (ed) The faba bean (*Vicia faba* L.). A basis for improvement. Butterworths, London, England, pp 23–67

Link W, Dixkens C, Singh M, Schwall M, Melchinger AE (1995) Genetic diversity in European and Mediterranean faba bean germplasm revealed by RAPD markers. Theor Appl Genet 90:27–32

Lübberstedt T, Melchinger AE, Dußle C, Vuylsteke M, Kuiper M (2000) Relationships among early European maize inbreds. IV. Genetic diversity revealed with AFLP markers and comparison with RFLP, RAPD and pedigree data. Crop Sci 40:783–791

Messmer MM, Melchinger AE, Boppenmaier J, Brunklaus-Jung E, Herrmann RG (1992) Relationships among early European maize (*Zea mays* L.) inbreds. I. Genetic diversity among flint and dent lines revealed by RFLPs. Crop Sci 32:1301–1309

Muratova V (1931) Common beans, *Vicia faba* L. Suppl Bull Appl Bot Genet Plant Breed, Leningrad

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

Piepho HP, Laidig F (1997) A review of methods for cluster analysis of marker data. UPOV document BMT/47/7. UPOV, Genf

Robinson J, Harris SA (1999) Which DNA marker for which purpose? In: Gillet EM (ed) Final compendium of the research project development, optimisation and validation of molecular tools for assessment of biodiversity in forest trees in the

European Union. DGXII Biotechnology FW IV-Research Programme of Molecular Tools for Biodiversity

Rohlf FJ (1993) NTSYS-PC. Numerical taxonomy and multivariate analysis system. Exeter Software, New York

Rohlf FJ (2000) NTSYS-pc. Numerical taxonomy and multivariate analysis system Exeter Software, New York

Satovic Z, Torres AM, Cubero JI (1996) Genetic mapping of new morphological, isozyme and RAPD markers in *Vicia faba* L. using trisomics. Theor Appl Genet 93:1130–1138

Schwarz G, Herz M, Huang XQ, Michalek W, Jahoor A, Wenzel G, Mohler V (2000) Application of fluorescence-based semi-automated AFLP analysis in barley and wheat. Theor Appl Genet 100:545–551

Smith JSC, Smith OS (1992) Fingerprinting crop varieties. Adv Agron 47:85–140

Sneath PH, Sokal RR (1973) Numerical taxonomy. WH Freeman and Company, San Francisco, USA

Sokal RR, Rohlf J (1981) Biometry. WH Freeman and Company, New York, USA

Stelling D, von Kittlitz E, Ebmeyer E, Sass O, Jaiser H, Link W (1994) Erfolge und Perspektiven der züchterischen Verbesserung von Körnererbsen und Ackerbohnen in der EU. Bericht über die 45. Arbeitstagung der Arbeitsgemeinschaft der Saatzuchtleiter im Rahmen der "Vereinigung österreischer Pflanzenzüchter", BAL Gumpenstein, Nov 1994, pp 22–24

Tao ZX (1981) Faba bean production and research in China. FABIS 3:24–25

Tivang JG, Nienhuis J, Smith OS (1994) Estimation of sampling variance of molecular-marker data using the bootstrap procedure. Theor Appl Genet 89:259–264

Vos P, Hogers R, Bleeker 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

Winfield MO, Arnold GM, Cooper F, Le Ray M, White J, Karp A, Edwards KJ (1998) A study of genetic diversity in *Populus nigra* subsp. *betulifolia* in the Upper Severn area of the UK using AFLP markers. Mol Ecol 7:3–10

Yamamoto K (1973) Karyotaxonomical studies on *Vicia*. 1. On the karyotype and character of some annual species of *Vicia*. Jpn J Genet 48:315–327

Yap IV, Nelson RJ (1996) WINBOOT: a program for performing bootstrap analysis of binary data to determine the confidence limits of UPGMA-based dendrograms. IRRI Disc Paper Ser 14

Zeid M, Schön CC, Link W (2001) Genetic diversity in a group of recent elite faba bean lines. Czech J Genet Plant Breed 37:34–40