Theor Appl Genet (1996) 93: 1103–1111

J. Lu · M. R. Knox · M. J. Ambrose · J. K. M. Brown
T. H. N. Ellis

Comparative analysis of genetic diversity in pea assessed by RFLP- and PCR-based methods

Received: 22 January 1996 / Accepted: 14 June 1996

Abstract DNA-based molecular-marker techniques have been proven powerful in genetic diversity estimations. Among them, RFLP was the first and is still the most commonly used in the estimation of genetic diversity of eukaryotic species. The recently developed PCR-based multiple-loci marker techniques, which include RAPD, AFLP, Microsatellite-AFLP and inter-SSR PCR, are playing increasingly important roles in this type of research. Despite the wide application of these techniques, no direct comparison of these methods in the estimation of genetic diversity has been carried out. Here we report a direct comparison of DNA-based RFLP with various PCR-based techniques regarding their informativeness and applicability for genetic diversity analysis. Among ten pea genotypes studied, all the PCR-based methods were much more informative than cDNA-RFLP. Genetic diversity trees were derived from each marker technique, and compared using Mantel's test. By this criterion, all trees derived from the various molecular marker techniques, except for the tree derived from inter-SSR PCR, were significantly correlated, suggesting that these PCR-based techniques could replace RFLP in the estimation of genetic diversity. On the basis of this result, AFLP analysis was applied to assess the genetic diversity of a sample of accessions representing the various species and subspecies within the genus *Pisum*.

Key words Diversity · Molecular-markers · Pea (*Pisum*) · Relatedness-trees · Mantel's test

Introduction

Characterization and quantification of genetic diversity has long been a major goal in evolutionary biology and plant breeding. DNA-based polymorphisms have allowed direct comparisons of the variation in nucleotide sequences, and have proved to be a powerful tool in various types of genetic analysis (Brown 1992; Monckton and Jeffreys 1993).

The first, and the most common, DNA polymorphic marker method used for genetic mapping and the estimation of genetic diversity in eukaryotic species is restriction fragment length polymorphism (RFLP) analysis (Botstein et al. 1980; Beckman and Soller 1986; Tanksley et al. 1989). The RFLP approach relies on the cleavage of genomic DNA by restriction enzymes; length polymorphism between a given pair of sites is then detected by hybridization to a labelled DNA probe. Species relationships and taxonomic studies on the basis of single-copy RFLP markers have been reported for several crop species (Havey and Muehlbauer 1989; Miller and Tanksley 990; Kesseli et al. 1991; Moser and Lee 1994). Although these single-copy RFLP marker analyses are valuable for assessing the genetic variation, the major disadvantage is that they are single-locus-specific and detect polymorphism directly related to the DNA sequence used as probe; in addition, RFLP analysis is time consuming and expensive.

The discovery of minisatellites (Jeffreys et al. 1985) created multilocus RFLPs by using the minisatellite core sequences as probes. Because of their powerful utility in genetic identification, this technique was designated 'genomic fingerprinting', and was soon extended to some other repetitive DNAs. In general, these markers are classified into three groups: (1) M13 repeat probes (Rogstad et al. 1988), (2) simple repetitive sequences (Tzuri et al. 1991), and (3) minisatellite DNAs (Dallas 1988). Due to their repetitive nature, all of these kinds of markers generate complex banding patterns after Southern hybridization and all of them have been successfully employed for revealing genetic variation in plant nuclear genomes (Rogstad et al. 1988; Nybom 1990; Amos and Pemberton 1992; Beyermann et al. 1992; Zhou and Gustafson 1995) and for phylogenetic studies (Gebhardt et al. 1989; Jung et al. 1993).

Various PCR-based polymorphic marker techniques have been used successfully in the genomic fingerprinting

Communicated by G. E. Hart

J. Lu · M. R. Knox · M. J. Ambrose · J. K. M. Brown · T. H. N. Ellis (✉)
John Innes Centre, Colney Lane, Norwich NR4 7UH, UK

of plant genomes (Welsh and McClelland 1990; Kesseli et al. 1994; Rus-Kortekaas et al. 1994) and in genetic diversity studies (Tinker et al. 1993; N'goran et al. 1994). Among them, RAPD analysis is fast and these markers are easily generated by PCR (Welsh and McClelland 1990; Williams et al. 1990). This method is based on the amplification of genomic DNA fragments with a single short primer of arbitrary nucleotide sequence. The amplified fragments are separated by electrophoresis and used to reveal polymorphisms. The inter-SSR PCR (Zietkiewicz et al. 1994) is another newly developed method which relies on one primer for PCR. The primer used in this method is a simple sequence repeat (SSR). This kind of repeat consists of tandemly repeated short-sequence motifs which are widely dispersed among all eukaryotic genomes. AFLP is another recently developed multilocus polymorphic marker technique. The procedure to generate AFLPs is called Selected Restriction Fragment Amplification (SRFA) and is applied to a portion of the genomic DNA digest using PCR (Zabeau and Vos 1993; Vos et al. 1995). A method which combines the AFLP technique with simple sequence repeats is also available (microsatellite-AFLP); here one of the two AFLP amplification primers is replaced by a compound simple sequence repeat in the PCR, so the polymorphisms detected are presumed to be contributed by the simple sequence repeats (Vogel 1995). The AFLP multilocus polymorphic marker technique has been used in a phylogeny study of *Lens* spp. (Sharma et al. 1996); microsatellite-AFLP would also be a good candidate for genetic diversity studies.

The advantages of these PCR-based methods over RFLP and the related genomic fingerprinting techniques are: they can be undertaken from minute amounts of template DNA, coupled with the speed with which results can be obtained and the large number of loci detected. These methods are much more sensitive than RFLP-based analysis, which makes it possible to detect low-frequency polymorphism (Brown 1992). A major advantage of PCR-based methods is that they can easily be applied to a large number of samples, and can be automated.

Although there are several reports of the use of PCR-based multilocus markers in assessing genetic variation and some comparisons between the RFLP and RAPD techniques (Dos Santos et al 1994; N'goran et al. 1994; Rus-Kortekaas et al. 1994), there is no direct comparison of all of these methods with typical RFLP-based markers, or a comparison of the PCR-based marker techniques with each other. In this article we report the comparison of the RFLP and PCR-based methods in assessing genetic diversity in pea, both with respect to their power to detect polymorphisms and the frequencies of polymorphism they identify. We have derived genetic diversity trees from the data of each of these methods and applied Mantel's (1987) test to compare the similarities between the genotypes found using different marker systems. The difference in the frequency of polymorphisms obtained and their nature, is discussed. A general conclusion concerning the applicability of these methods in assessing genetic diversity is presented.

Materials and methods

Plant material

All pea lines are designated by their accession number and were obtained from the John Innes Germ-plasm collection. The lines used for comparative purposes have previously been used for the generation of linkage maps in this laboratory (Ellis et al. 1992). Of these lines, four are cultivars or selections from cultivars (JI399 cv Cennia all sweet type carrying *rb*, JI430 cv Greenshaft and JI540 cv Dark Skinned Perfection, both vining peas, JI813 carries *yp* from cv Vinco); four are genetic stocks (JI15 is a multiply marked stock, JI61 is a cytogenetic marker stock, and JI1194 and JI1201 are a near-isogenic pair for the markers *af, tl* and *st*, developed by G.A. Marx in Cornell, in a background thought to be similar to JI540), JI126 is derived from a cross between two cultivated types (by H. Lamprecht) and JI281 is a primitive cultivated form. Plants were grown in a glasshouse, in pots, containing John Innes potting compost No. 1 supplemented with 30% grit.

Nucleic acid manipulation

Conditions for DNA preparation, restriction-enzyme digestion, gel electrophoresis for RFLP, Southern transfer, hybridization and autoradiography have been presented by Ellis et al. (1992) and Ellis (1994).

RAPD

Ten-base oligonucleotide primers (A–F) were obtained from Operon Technologies, Alameda, California. DNA amplification reactions contained 10 mM Tris-HCl pH 9.0, 50 mM KCl, 1.5 mM $MgCl_2$, 0.1% (w/v) Gelatin, 0.1% Triton (v/v) X-100, 20 mM of each dNTP, 20 ng (0.2–0.3 mM) primer, 10 ng genomic DNA, and 0.6 U *Taq* polymerase (Promega). DNA amplifications were performed in a Perkin-Elmer Cetus DNA thermal cycler 480 using the following conditions: 95°C for 3 min, followed by 40 cycles at 94°C for 45 s, 37°C for 1 min, 72°C for 2 min, utilizing the fastest-available temperature transition. These cycles were followed by an additional 10 min at 72°C. Amplification products were separated by electrophoresis in 2% agarose gels in 1×TAE (0.04 M Tris-acetate, 0.01 M EDTA pH 7.9) which were stained with ethidium bromide and viewed under UV light and photographed using a polaroid camera.

Inter-SSR PCR

The simple sequence primers used in this experiment were $(GACA)_5$, $(GATA)_5$ and $(GAAA)_5$. The genomic DNAs were digested with *Alu*I (BMG). The digested genomic DNAs were used as PCR templates employing the same machine and components as above, except that 3.5 mM $MgCl_2$ was used. PCRs were performed with the following programme: 95°C for 3 min, 30 cycles of 94°C for 45 s, 32°C for 2 min, 72°C for 2 min; with a final cycle at 72°C for 5 min. The amplification products were separated and viewed as above.

AFLP

Restriction of genomic DNA and ligation of adapters

Genomic DNA (2.5 µg) was digested with 25 U *Pst*I (Pharmacia) and 12.5 U of *Mse*I (NEB) within a Restriction Ligation (RL) buffer (10 mM Tris acetate pH 7.5, 10 mM Mg acetate, 50 mM K acetate, 5 mM DTT) in a total volume of 50 µl at 37°C for 2 h. Then 5 pMoles of 5′-biotinylated *Pst*I adapter, 50 pM *Mse*I adapter (sequences shown in Table 1), 1.2 mM ATP, and 5 U T4 DNA Ligase (Gibco BRL), all in a total volume of 10 µl, was added to the digestion mixture and incubated at 37°C for 3 h.

Table 1 Adapter and primer sequences used in the AFLP and microsatellite-AFLP experiments

*Pst*I adapter		5′ – CTCGTAGACTGCGTACATGCA – 3′ 3′ – CATCTGACGCATGT – 5′
Pst primers	A	5′ – GACTGCGTACATGCAGCC – 3′
	B	5′ – GACTGCGTACATGCAGGC – 3′
	C	5′ – GACTGCGTACATGCAGGG – 3′
*Mse*I adapter		5′ – GACGATGAGTCCTGAG – 3′ 3′ – TACTCAGGACTCAT – 5′
Mse primer	2	5′ – GATGAGTCCTGAGTAACAC – 3′
	3	5′ – GATGAGTCCTGAGTAATGC – 3′
	5	5′ – GATGAGTCCTGAGTAATAT – 3′
	9	5′ – GATGAGTCCTGAGTAAGCT – 3′
Micro-satellite primers	1	5′ – GTGTGTGTGTGTCTGTCT – 3′
	2	5′ – GAGAGAGAGAGTGTGTGTGT – 3′
	3	5′ – GAAAGAAAGAAAGACAGACA – 3′
	4	5′ – TCTCTGTGTGTGTGTGTG – 3′

Selection of biotinylated DNA-fragments

Fragments ligated with the biotinylated *Pst*I adaptors were separated from non-biotinylated fragments by binding to streptavidin-coated magnetic beads (Dynal). Then 10 µl of beads were washed once in 100 µl STEX (100 mM NaCl, 10 mM Tris-HCl, 1 mM EDTA, 0.1% Triton X-100, pH 8.0) and re-suspended in 140 µl STEX. The beads were subsequently added to the DNA mixture to give a final volume of 200 µl, then incubated for 30 min at room temperature with gentle agitation to ensure proper binding of the biotinylated DNA. The beads were collected with a magnet, and the supernatant was removed, washed with 200 µl STEX once and transferred to a fresh tube. The beads were washed a further three times with 200 µl STEX and finally re-suspended in 200 µl T0.1E (10 mM Tris, 0.1 mM EDTA, pH 8.0), and transferred to a fresh tube. The resulting template DNA was stored at 4°C.

SRFA-reaction

The primers for PCR in the SRFA reaction are complementary to the adapter sequences but extend into the ligated fragments by two additional selective bases at the *Pst*I site and by three at the *Mse*I site. The three pairs of primer combinations used in this experiment are PstA/Mse3, PstA/Mse5 and PstA/Mse9, with the sequences shown in Table 1.

The PCR mixture was composed of 15 ng of *Pst* primer (γ^{33}P labelled in a kinase reaction), 15 ng *Mse* primer, 0.2 mM of each dNTP, 1 unit of Ampli*taq* (Perkin-Elmer) or equivalent, 1 µl of template DNA, to a total volume of 10 µl in PCR buffer (50 mM KCl, 10 mM Tris-HCl pH 8.4, 1.5 mM $MgCl_2$, 0.1 mg/ml Gelatin). The PCR was performed in a PTC-100 Programmable Thermal Controller (MJ Research, Inc.) with temperature cycling as follows: denature at 94°C for 30 s, anneal at 65°C for 30 s, extend at 72°C for 60 s, one cycle; the annealing temperature was then owered 1°C over the next nine cycles to 56°C with no change of other conditions; the annealing temperature remained at 56°C for a further 24 cycles.

Gel analysis

Gels were as for standard DNA sequencing, but with the exception that 4.5% acrylamide was used.

AFLP for genetic diversity analysis in the genus Pisum

The AFLPs for the genetic diversity analysis were set up as described above except that the SRFA reactions were fluorescently labelled using 1 µM of Fluorescein-11-dUTP (FluoroGreen, Amersham Life Science) and 100 µM of each dNTP. The PCR-amplified products were run out on 5% Longranger (Flowgen) acrylamide using an Automated Laser Fluorescent ALF electrophoresis unit (Pharmacia Biotech). Data was analysed using Fragment Manager Software (Pharmacia Biotech).

Microsatellite AFLP

This method was based on AFLP, and introduced by Vogel (1995). The procedures of restriction of genomic DNA and ligation of adaptors, selection of biotinylated DNA-fragments and the gel analysis, were as for AFLP.

In the SRFA-reaction a pre-amplification was applied, with partial selection, to the template DNA using primer *Pst*A and the *Mse*I adapter, upper-strand sequence, (see Table 1) without selective bases; the selective restriction fragment amplifications were performed with 1 µl of a ten-fold dilution of the pre-amplification product, a microsatellite primer replacing the *Mse* primer and the *Pst*A primer in the final 10 µl reaction mixture. The PCRs were performed with same conditions as for SRFA. The four microsatellite primers are listed in Table 1.

Data analysis

The presence or absence of each specific band was scored for each genotype and for every method used in this study. The cDNA-RFLP data were also scored according to the hybridisation pattern, whereby each pattern was assigned a unique code number [i.e. the pattern was designated 1 for those lines where the hybridisation pattern was indistinguishable from that of the lowest line number, or 2 for those indistinguishable from that of the line with the second largest line number (when this was not scored as 1), and so on]. The neighbour-joining method (Saitou and Nei 1987) was used to carry out a cluster analysis of *Pisum* accessions. The relationships between the results of the different marker methods were analysed using Mantel's test for correlation between matrices (Mantel 1967; Manly 1985). Pairwise distances between accessions were calculated as the proportion of markers which differed. A correlation coefficient between the elements of the matrices was calculated for each pair of matrices. The significance of this correlation was then tested according to Mantel (1967). This test is a randomization procedure that compares the correlation between two matrices with the correlation between one of these and randomizations of the other; 100 randomizations were carried out for each pairwise comparison. The randomized correlation coefficients were assumed to be normally distributed and confidence intervals were calculated to assess the significance of the correlation between the matrices derived by the different DNA marker techniques. Calculations were carried out using the Genstat 5 package (Numerical Algorithms Group, Oxford).

Results

Levels of polymorphisms

The level of polymorphism detected with different markers was assessed (Table 2). The variation detected by cDNA probes was increased by pre-selecting the probes that would produce polymorphism. Approximately 2/3 rds of the randomly-chosen cDNA probes detected at least one difference between these pea lines after digestion of the genomic DNA with *Eco*RI. Based on the observation that most of the cDNA probes produce simple fingerprint patterns, and in order to obtain minimal bias, five cDNA mark-

Table 2 The polymorphic level of PCR-based and cDNA-RFLP markers. The table shows the number and proportions of informative markers from the various procedures. The AFLP primer combinations designated A3, A5 and A9 refer to the combination of the *Pst* primer A with *Mse* primers 3, 5 and 9 respectively (see Table 1). The microsatellite primers MS1 to MS4 are as designated in Table 1, and these were used in conjunction with the *Pst* primer A. The RAPD primers A to F were from Operon, and the inter-SSR PCR primers designated A, T or C corresponded to the sequences $(GAAA)_5$, $(GATA)_5$ and $(GACA)_5$ respectively. The cDNA probes were selected as informative probes, approximately 2/3rds of the cDNA probes detected at least one difference in the set of ten lines used in this analysis after screening for differences with a single restriction enzyme (*Eco*RI)

Markers:	AFLP			Microsatellite-AFLP				RAPD					Inter-SSR PCR			cDNA RFLP
Primers or primer combinations	A3	A5	A9	MS1	MS2	MS3	MS4	A	C	D	E	F	A	T	C	49 probes
Total band number	39	45	22	17	24	27	16	12	12	13	15	9	24	16	14	167
Polymorphic band number	19	22	5	9	14	20	13	8	8	8	11	7	14	9	10	
		46				56				42				33		167
Polymorphic percentage		47.8				68.7				68.9				6.1		/
Polymorphic bands per primer or probe		15.3				14				8.4				11		3.34

ers which showed complicated band patterns were not selected for further analysis in this study. Even with selection, the cDNA probes still showed the lowest level of polymorphism among the methods compared. All of the PCR-based methods showed high levels of polymorphism. All RAPD primers detected variation among these pea genotypes, except for primer B which detected several monomorphic bands. There were some bands which were difficult to score, because they were either faint or close to other bands, and these were ignored. Some of the bands generated by inter-SSR PCR also could not be scored. The three microsatellite primers chosen for this study were based on the observation that they are highly conserved in most plant genomes. In the AFLP analysis, the three AFLP primer pairs were chosen randomly. Due to the expected product size these fragments were separated in polyacrylamide gels, the bands were easier to score than RAPDs and inter-SSR PCRs, except for the very faint ones. In the Microsatellite-AFLP analysis, pre-amplification of the template DNA with a *Pst*A and *Mse*I adapter upper strand (see Table 1) was applied; without such pre-amplification the SRFA products were not reproducible.

The genetic diversities assessed

The genetic diversity trees produced from each method are shown in Fig. 1. Except for AFLP, all methods could distinguish all ten pea genotypes; AFLP markers could only distinguish eight of the ten genotypes. The two genotypes (JI 1194 and JI 1201) which AFLP markers could not distinguish are near-isogenic lines developed by Marx (1987). The variation observed between these two lines with the other methods was higher than expected, especially for RAPD analysis. The frequency distribution of band occurrence per accession for each method is given in Fig. 2.

The cDNA data were scored by two different methods which agree well. The tree derived from inter-SSR PCR is

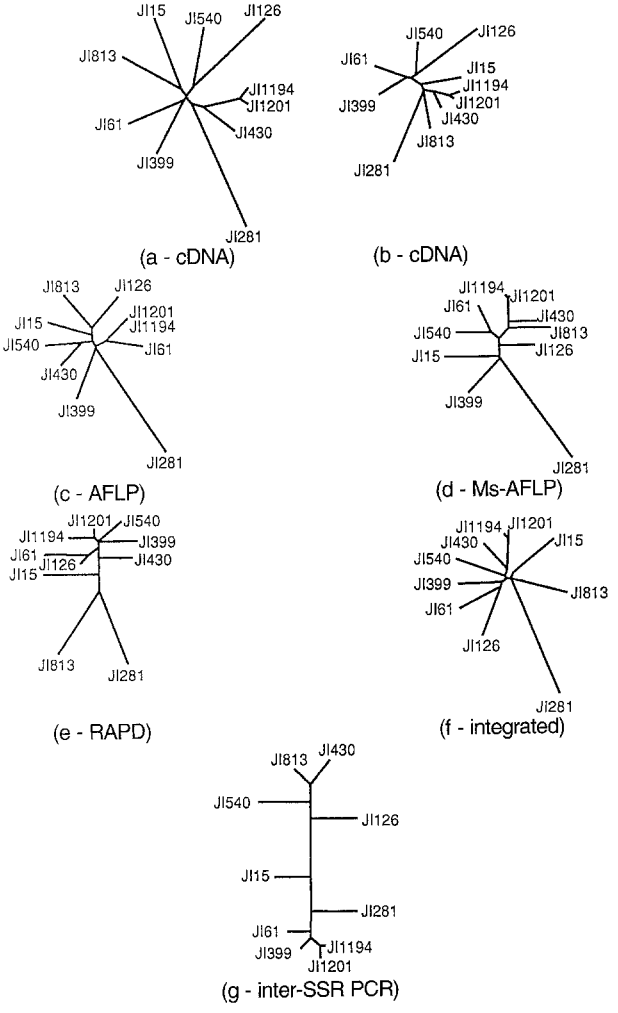

Fig. 1a–f Genetic diversity trees generated using the neighbour-joining method. (**a**) cDNA RFLP (the data scored as band patterns), (**b**) cDNA RFLP (the data scored as band presence or absence), (**c**) AFLP, (**d**) microsatellite-AFLP, (**e**) RAPD, (**f**) the integration cDNA, AFLP, microsatellite-AFLP and RAPD data, (**g**) inter-SSR PCR

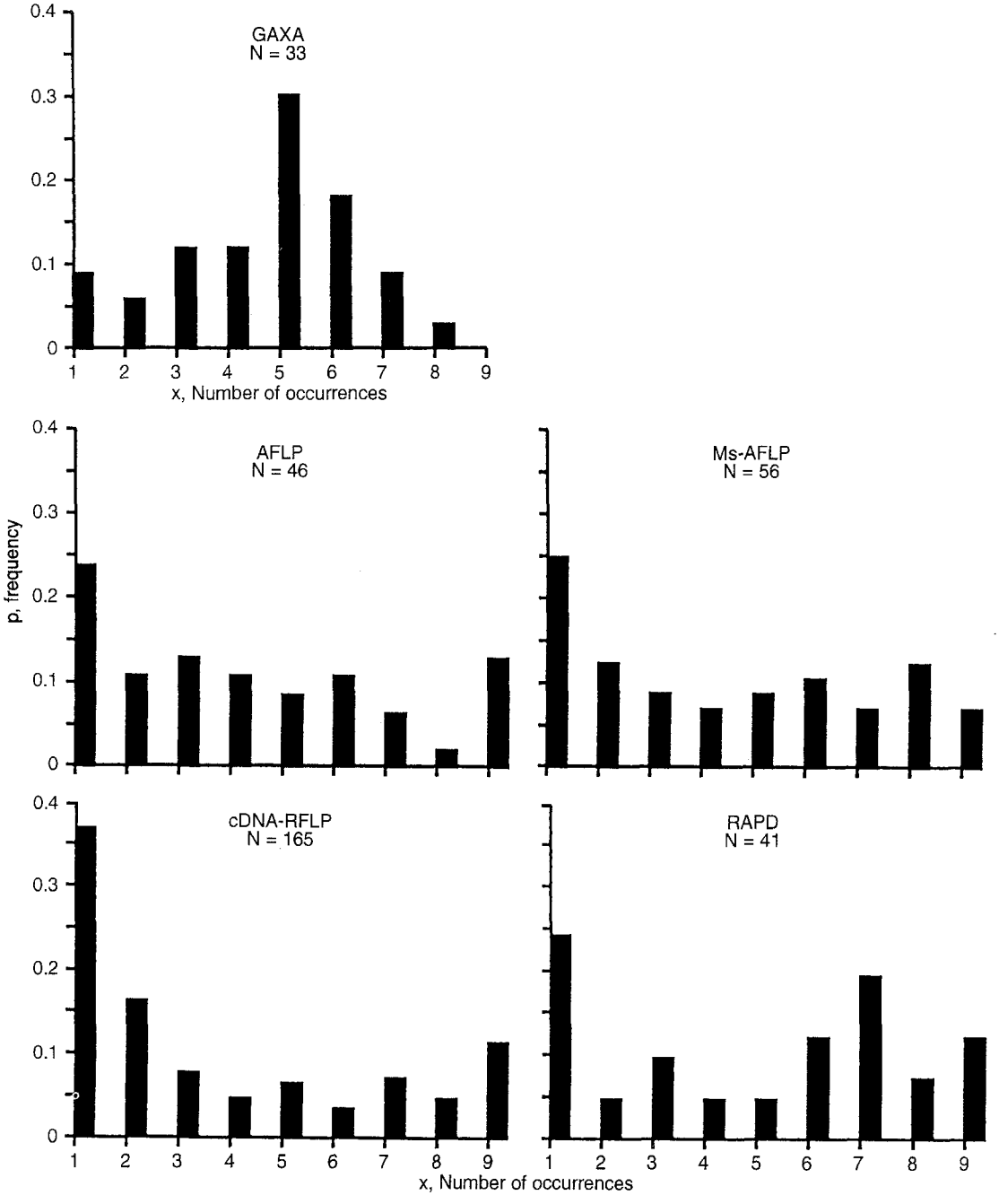

Fig. 2 Frequency distributions of band occurrence. These values are given as proportions (p). N is the total number of bands for each method, x is the number of accessions in which the band occurred (as monomorphic bands are excluded p varies from 1 to 9). *GAXA*: the inter-SSR PCR method; *AFLP*: AFLP analysis; *Ms-AFLP*: the microsatellite-AFLP method; *cDNA RFLP*: cDNA RFLP analysis with data scored as band presence or absence; *RAPD*: RAPD analysis

different from the others as follows: (1) in all methods except inter-SSR PCR the genotypes appear in a "star phylogeny", (2) JI 281 is clearly the most divergent as assessed by all methods except inter-SSR PCR. Mantel's-test results for the correlations between the distance matrices obtained by different methods are shown in Fig. 3. Correlation coefficients which fell outside the 95% confidence intervals showed positive correlation and those within the 95% confidence region showed no significant correlation. Additionally, no significant negative correlation was detected in this study. Mantel's test shows that the genetic diversity tree detected by $(GAXA)_5$ inter-SSR PCRs is not significantly correlated with the ones detected by other methods, whereas all of the others are significantly positively correlated. In other words, the differences between accessions determined from the inter-SSR PCR technique were not in agreement with those from the other marker techniques. All the data, except for the inter-SSR PCR analysis, were combined to construct an integrated genetic diversity tree which is shown in Fig. 1f.

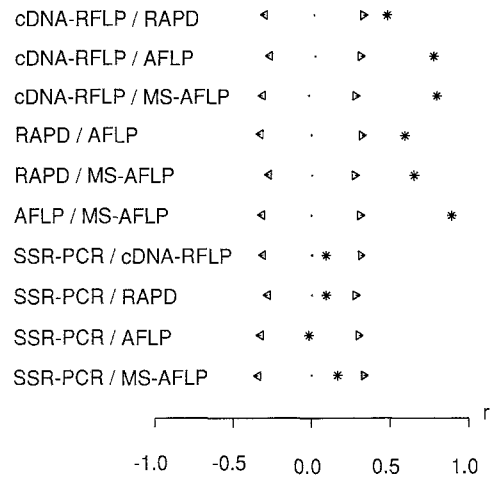

Fig. 3 Mantel's test results for the significance of the correlation coefficients for pairs of molecular marker methods. *r* correlation coefficient, * observed value of *r* for the pair of methods, • mean of *r* for randomized data, ◁ ▷ upper and lower 95% confidence interval for randomized correlation coefficients

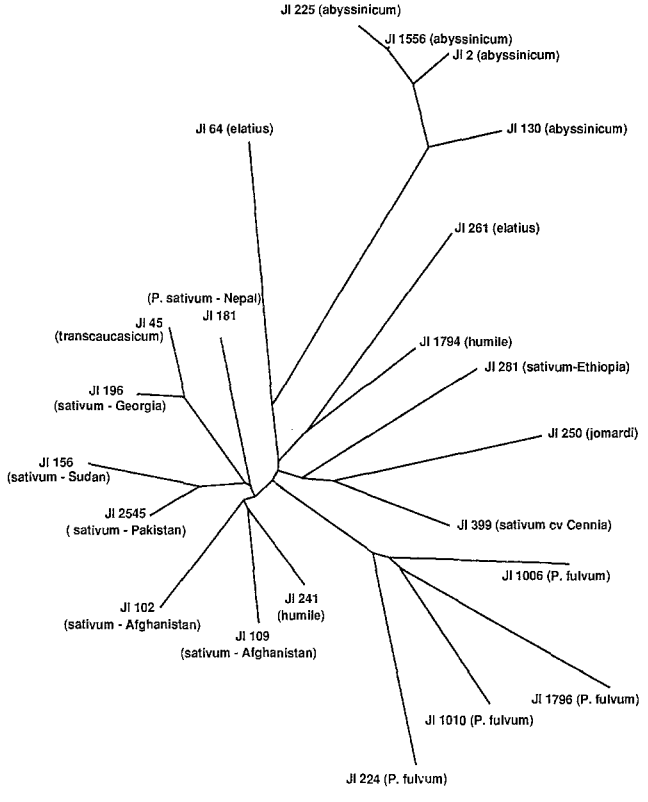

Fig. 4 Genetic diversity tree of accessions from throughout the genus *Pisum*

Genetic diversity in the genus *Pisum* assessed by AFLP

A genetic diversity tree from a selection of various species and subspecies within the genus *Pisum* was obtained from AFLP. The three primer combinations employed in this study were an arbitrary choice (*Pst*A/*Mse*3, *Pst*B/*Mse*5 and *Pst*C/*Mse*2). The primer sequences shown in Table 1 revealed 132 informative bands; the pattern of similarity and difference obtained among these lines generated the neighbour joining tree shown in Fig. 4.

Discussion

The levels of polymorphism

In this study, we compared the cDNA-RFLP method and PCR-based methods as molecular markers in genetic diversity assessment in pea. One of the goals of the study was to investigate the differences in the levels of polymorphism detected by these methods and to evaluate the potential of the PCR-based methods to replace RFLP in genetic assessment. The polymorphism levels from different methods indicated that the cDNA-RFLP is the least informative of the methods, even though the value was enlarged by pre-selection for informative clones. Most of the cDNA probes detected single- or low-copy loci markers. This is one of the major reasons to compare cDNA-RFLP with the PCR-based methods. The PCR-based methods gave similar high levels of polymorphism, except for the RAPD primer B which was monomorphic. The RAPD analysis of pea gave an exceptionally high level of polymorphism compared to other species; in some species, pre-selection of the RAPD primers is necessary to maximize informativeness (Kesseli et al. 1994; Rus-Kortekaas et al. 1994). It was also observed that the primers D, E or F could distinguish the ten pea genotypes. This suggests that there is generally a high level of diversity in pea. Williams et al. (1990) showed that RAPD markers cover the entire genome, revealing coding or noncoding regions, repeated or single-copy sequences. It has been suggested that repeated sequences may be preferentially amplified by the RAPD technique (Arnau et al. 1994; N'goran et al. 1994). The number of repeated sequences is known to be correlated with the size of the genome (Vedel and Delseny 1987; Lapitan 1992). As the pea genome size is 4–5 pg and contains abundant repeated sequences (Murray and Thompson 1982), one could expect a relatively high content of repeated sequence to be amplified. With the AFLP analysis, the extra selective bases in the AFLP primers were chosen randomly, so the fragments amplified by this method were assumed to be a nearly random subset of the restriction fragments. However, long tandem arrays of repeat sequences will probably be under-represented in these restriction fragments.

With the microsatellite-AFLP technique, because the microsatellite sequences were included in the amplified fragments, the polymorphism level was expected to be higher than for AFLP. The results shown in this study agree with this hypothesis. In the microsatellite AFLP method, because of replication slippage during the PCR, the microsatellite bands are expected to be accompanied by faint bands (Perry et al. 1994). Some of the bands from this study

do not behave in this way, suggesting that some of the fragments amplified from the simple sequence primers were not microsatellites, and consequently the polymorphism observed in these cases was not due to microsatellite, but rather to conventional, AFLP. The design of the compound microsatellite primer could be a major factor affecting the nature of the sequences amplified from the PCR, for example as a consequence of the length or sequence of the 5′ or 3′ portion and the relationship of the two components. The primer design is intended to control the annealing site, allowing detection of microsatellite length polymorphism (Vogel 1995).

With the inter-SSR PCR, the polymorphism level was similar to the other PCR-based methods. Because the primers used in the study were similar, the polymorphism detected reflects the variation of one simple sequence family and the associated sequences. The polymorphisms detected by inter-SSR PCR are from the sequences between the two simple sequence primer sites and are not necessarily simple sequence length polymorphisms. Like RAPDs, the source of polymorphisms may include deletion of a priming site, insertion or deletion that changes the DNA fragment size, insertion that renders priming sites too distant to support amplification, or an alteration to the restriction site used in the pre-digestion. The nature of the amplified fragments therefore depends on both the simple sequence primers used and on genome structure.

Comparison of various markers in genetic diversity

Molecular markers represent a sample of a plant genome, yet they are used to infer relationships of the entire genome among a set of lines or populations. The distribution of the loci detected by individual methods will affect the precision of the resulting estimates of genetic distance (Nei 1987; Leberg 1992). For a general comparison of genetic diversity, it is ideal that the loci detected are randomly dispersed and sample the whole genome. The comparison of distributions of the different individual markers needs to be based on the genetic map, but this is very time consuming. An alternative approach may be to estimate the relationships of the genetic diversity derived from several marker techniques. Furthermore, in order to investigate the potential of using PCR-based methods to replace the RFLP method, a comparison of the relationships of these diversity data is also necessary. The genetic diversity trees assessed by the different methods were statistically compared using Mantel's test. The results showed that, except for the inter-SSR PCR, the trees from other molecular marker techniques are significantly correlated, indicating that the trees from these PCR-based methods agree with that from cDNA-RFLP. Furthermore, this result suggests that these PCR-based methods may be used as alternatives to replace RFLP in genetic diversity assessment.

Presumably the different behaviour of the inter-SSR PCR reflects some feature of this class of sequences in the pea genome. The simple sequence primers used in this study belong to the same family, so the fragments amplified from these primers were associated with one family of repetitive DNA sequences. Consequently, the relationships inferred from this technique mainly reflect the diversity of this family and its associated DNA sequences in the pea genotypes assessed in this study. The organization of this family of simple sequences in the pea genome has been studied in detail (manuscript in preparation); the relevant point being that these kinds of repetitive sequences are clustered together and associated with other specific repetitive sequences. Similar observations have been reported for other species (Epplen 1988; Condit and Hubbell 1991; Schmidt et al. 1993; Poulsen et al. 1994; Arens et al. 1995; Hamann et al. 1995). According to Delsney et al. (1988) repetitive sequences do not undergo the same evolutionary processes as the rest of the genome. Repetitive sequences could evolve more rapidly than single-copy sequences by amplification and transposition. The different genetic diversity tree derived from this method was not a surprise, as we knew that simple sequences and other associated complex repetitive sequences were involved in this amplification, and the organization of this family of sequences in pea genome is complex. Although the evolutionary processes affecting this simple sequence family remain to be explored, this method provides a potential way for analysing the genetic diversity of a specific repeated sequence family in relation to that of the genome as a whole. Further analyses are needed, exploring other families of simple sequence repeats, to determine whether the relationships detected by this class of sequences is representative of the evolution of the genome as a whole. If other dispersed simple sequence repeats, or a large number of primers, had been used, they may have generated results in agreement with the other methods.

The number of markers needed

An additional factor affecting the genetic diversity assessed by different marker techniques is the number of markers or probes used in an analysis (Messmer et al. 1991; Smith et al. 1992). Generally, precision improves as more probes or marker loci are detected in the analysis (Moser and Lee 1994; Tivang et al. 1994). For the marker techniques employed in this study, different numbers of markers were involved. The significant correlations among the techniques obtained from Mantel's test provided indirect evidence to confirm the significance of the trees. Mantel's test is not a confirmation of the individual pairwise distances between accessions but between the organisation of the trees; re-testing sub-populations of our data sets confirmed that the numbers of bands we compared were sufficient to test for correlations between the methods used. In order to obtain a more precise estimate of genetic relationship among the ten pea genotypes, the significantly correlated data were pooled to obtain an integrated genetic diversity tree (Fig. 1f) which was considered to represent the most accurate estimate of genetic relatedness obtained by this study.

The scoring methods

The scoring methods employed for different data may also affect the comparison of results. The cDNA data were first scored by the individual band patterns detected by each probe, which considered every probe as a single marker. In fact, this scoring method partially under-estimates the effects of the probes which detected multiple loci. For this reason, the cDNA markers were also scored from the same data according to the presence or absence of an individual band. From the data scored by this method, another genetic diversity tree was derived. The correlation of these two was analyzed using Mantel's test; the two methods are significantly correlated within an r value of about 0.95, $p \sim 7 \times 10^{-9}$. This indicates that the scoring methods did not much affect the results, perhaps because most of the cDNA probes detected a single locus with two alleles; and in this specific case the two methods are indistinguishable.

The genetic diversity in genus *Pisum*

It is clear from this work that AFLP analysis is consistent with other molecular methods for the assessment of genetic relatedness. The validity of this method in relation to accepted taxonomy (Blixt 1972; Ben-Ze'ev and Zohary 1973) has been investigated by the characterization of a sample of accessions representing the various species and subspecies within the genus *Pisum.*

This neighbour joining tree (Fig. 4) shows that the accessions from throughout the genus appear to radiate from a common centre. *Pisum fulvum* formed a single, although diverse, group but it is not markedly exceptional. *P. sativum* ssp. *abyssinicum* appears to form a more discrete group clustering with subspecies *elatius* and at least one accession of ssp. *humile*. The remainder of the accessions form a loose assemblage and the pattern of branching at the centre of the tree is probably not very reliable as the distances between these nodes are short. The pattern of relatedness shown in Fig. 4 suggests that genetic diversity is relatively evenly distributed within the genus and that there is no exceptionally marked differentiation between these species and subspecies, an observation which is consistent with the classical taxonomy of the genus (Blixt 1972; Ben-Ze'ev and Zohary 1973).

General conclusions

The results of this study within the genus *Pisum*, and the comparison between the methods employed, demonstrated that, with the exception of inter-SSR PCR, the various PCR-based DNA molecular markers are suitable to address the questions of relevance to pea systematics and breeding. This finding is expected to be generally applicable to other plant species, although inter-SSR PCR may be valuable when the level of diversity is much lower than in *Pisum*. We have shown that PCR-based methods can be good candidates to replace RFLP markers in genetic diversity estimations and that they have comparable accuracy. They are generally much simpler to apply and more sensitive than the traditional morphological and biochemical methods or the RFLP-based fingerprinting techniques because they are more polymorphic; yet are generally correlated with RFLP analysis. The inter-SSR PCR method was not well correlated with the other techniques and this may mean that the method is more appropriate for microevolutionary studies. A major advantage of these PCR-based methods is that they can be automated, and so have great potential in large-scale population genetics and plant breeding.

Acknowledgements The authors are grateful to Dr. Rod Casey for reading and commenting on this manuscript, and to Dr. C. M. Thomas for introducing us to the AFLP technique. J. Lu was supported by John Innes Foundation. The John Innes Centre, which has provided the resources to undertake this work, is supported by a grant in aid from BBSRC.

References

Amos B, Pemberton J (1992) DNA fingerprinting in a non-human population. Curr Opin Genet Dev 2:857–860

Arens P, Odinot P, Heusden AW van, Lindhout P, Vosman B (1995) GATA- and GACA-repeats are not evenly distributed throughout the tomato genome. Genome 38:84–90

Arnau J, Housego AP, Oliver RP (1994) The use of RAPD markers in the genetic analysis of the plant pathogenic fungus *Cladosporium fulvum.* Curr Genet 25:438–444

Beckman JS, Soller M (1986) Restriction fragment length polymorphisms and genetic improvement of agricultural species. Euphytica 35:111–124

Ben-Ze'ev N, Zohary D (1973) Species relationships in the genus *Pisum* L. Israel J Bot 22:73–91

Beyermann B, Nurnberg P, Weihe A, Meixner M, Epplen JT, Borner T (1992) Fingerprinting plant genomes with oligonucleotide probes specific for simple repetitive DNA sequences. Theor Appl Genet 83:691–694

Blixt S (1972) Mutation genetics in *Pisum.* Agric Hort Genet 30:1–293

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

Brown PTH (1992) DNA fingerprinting in plant breeding. Agro-Food-Industry Hi-Tech November/December

Condit R, Hubbell SP (1991) Abundance and DNA sequence of two-base repeat regions in a tropical tree genome. Genome 34:66–71

Dallas JF (1988) Detection of DNA "fingerprints" of cultivated rice by hybridization with a human minisatellite DNA probe. Proc Natl Acad Sci USA 85:6831–6835

Delsney M, Grellet F, Tremousaygue D, Raynal M, Panabieres F (1988) Structure, evolution et expression de l'ADN nucleaire. Bull Soc Bot Fr 135:23–38

Dos Santos JB, Nienhuis J, Skroch P, Tivang J, and Slocum MK (1994) Comparison of RAPD and RFLP genetic markers in determining genetic similarity among *Brassica oleracea* L. genotypes. Theor Appl Genet 87:909–915

Ellis THN (1994) Approaches to the genetic mapping of pea. Mod Meth Plant Anal 16:117–160

Ellis THN, Turner L, Hellens RP, Lee D, Harker CL, Enard C, Domoney C, Davies DR (1992) Linkage maps in pea. Genetics 130:649–663

Epplen JT (1988) On simple GAT/CA sequences in animal genomes: a critical reappraisal. J Hered 79:409–417

Gebhardt C, Ritter E, Debener T, Schachtschabel U, Walkemeier B (1989) RFLP analysis and linkage mapping in *Solanum tuberosum.* Theor Appl Genet 78:65–75

Hamann A, Zink D, Nagl W (1995) Microsatellite fingerprinting in the genus *Phaseolus*. Genome 38:507–515

Havey MJ, Muehlbauer FJ (1989) Variability for restriction fragment lengths and phylogenies in lentil. Theor Appl Genet 77:839–843

Jeffreys AJ, Wilson V, Thein SL (1985) Individual-specific "fingerprints" of human DNA. Nature 316:76–79

Jung C, Pillen K, Frese L, Fahr S, Melchinger AE (1993) Phylogenetic relationships between cultivated and wild species of the genus *Beta* revealed by DNA "fingerprinting". Theor Appl Genet 86:449–457

Kesseli R, Ochoa O, Michelmore R (1991) Variation at RFLP loci in *Lactuca* spp. and origin of cultivated lettuce (*L. sativa*). Genome 34:430–436

Kesseli RV, Paran I, Michelmore RW (1994) Analysis of a detailed genetic linkage map of *Lactuca sativa* (Lettuce) constructed from RFLP and RAPD markers. Genetics 136:1435–1446

Lapitan NLV (1992) Organization and evolution of higher plant nuclei genomes. Genome 35:171–181

Leberg PL (1992) Effects of population bottlenecks on genetic diversity as measured by allozyme electrophoresis. Evolution 46:477–494

Manly BFJ (1985) The statistics of natural selection. Chapman and Hall, London

Mantel N (1967) The detection of disease clustering and a generalized regression approach. Cancer Res 27:209–220

Marx GA (1987) A suite of mutants that modify pattern formation in pea leaves. Plant Mol Biol Rep 5:311–335

Messmer MM, Melchinger AE, Woodman WL, Lee EA, Lamkey KR (1991) Genetic diversity among progenitors and elite lines from the Iowa Stiff Stalk Synthetic (BSSS) maize population: comparison of allozyme and RFLP data. Theor Appl Genet 83:97–107

Miller JC, Tanksley SD (1990) RFLP analysis of phylogenetic relationships and genetic variation in the genus *Lycopersicon*. Theor Appl Genet 80:437–448

Monckton DG, Jeffreys AJ (1993) DNA profiling. Curr Opin Biotechnol 4:660–664

Moser H, Lee M (1994) RFLP variation and genealogical distance, multivariate distance, heterosis, and genetic variation in oats. Theor Appl Genet 87:947–956

Murray MG, Thompson WF (1982) Repeat sequence interspersion in coding DNA of peas does not reflect that in total pea DNA. Plant Mol Biol 1:143–153

Nei M (1987) Molecular evolutionary genetics. Columbia University Press, New York

N'goran JAK, Laurent V, Risterucci AM, Lanaud C (1994) Comparative genetic diversity studies of *Theobroma cacao* L. using RFLP and RAPD markers. Heredity 73:589–597

Nybom H (1990) Genetic variation in ornamental apple trees and their seedlings (*Malus* Rosaceae) revealed by DNA "fingerprinting" with M13 repeat probe. Hereditas 133:17–28

Perry BC, Mahinur SA, Arvind AB, Lavi U, Jiang R (1994) Length polymorphisms of simple sequence repeat (SSR) DNA as molecular markers in plants. In: Gresshoff PM (ed) Plant genome analysis. CRC Press, Boca Raton, London, pp 47–56

Poulsen GB, Kahl G, Weising K (1994) Differential abundance of simple repetitive sequences in species of *Brassica* and related *Brassicaceae*. Plant Syst Evol 190:21–30

Rogstad SH, Patton JC, Schaal BA (1988) M13 repeat probe detects DNA minisatellite-like sequences in gymnosperms and angiosperms. Proc Natl Acad Sci USA 85:9176–9178

Rus-Kortekaas W, Smulders MJM, Arens P, Vosman B (1994) Direct comparison of levels of genetic variation in tomato detected by a GACA-containing microsatellite probe and by random amplified polymorphic DNA. Genome 37:375–381

Saitou N, Nei M (1987) The neighbour-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425

Schmidt T, Boblenz K, Metzlaff M, Kaemmer D, Weising K, Kahl G (1993) DNA fingerprinting in sugar beet (*Beta vulgaris*) – identification of double haploid breeding lines. Theor Appl Genet 85:653–657

Sharma (1996). Theor Appl Genet (in press)

Smith OS, Smith JSC, Bowen SL, Tenborg RA (1992) Numbers of RFLP probes necessary to show associations between lines. Maize Genet Coop Newslett 66:66

Tanksley SD, Young ND, Paterson AH, Bonierbale MW (1989) RFLP mapping in plant breeding – new tools for an old science. Bio Technol 7:257–264

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

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

Tzuri G, Hillel J, Lavi U, Harberfeld A, Vainstein A (1991) DNA fingerprint analysis of ornamental plants. Plant Sci 76:91–97

Vedel F, Delseny M (1987) Repetitive and variability of higher-plant genomes. Pl Physiol Biochem 25:191–210

Vogel JM (1995) A microsatellite-based multiplexed genome assay (abtracts). In: Plant Genome. Scharago International, New York, p 84

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

Welsh J, McClelland M (1990) Fingerprinting genomes using PCR with arbitrary primers. Nucleic Acids Res 18:7213–7218

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

Zabeau M, Vos P (1993) Selective restriction fragment amplification: a general method for DNA fingerprinting. European Patent Application 924026 29.7 (publication number: 0534 858 A1)

Zhou Z, Gustafson JP (1995) Genetic variation detected by DNA fingerprinting with a rice minisatellite probe in *Oryza sativa*. Theor Appl Genet 91:481–488

Zietkiewicz E, Rafalski A, Labuda D (1994) Genome fingerprinting by simple sequence repeat (SSR) – anchored polymorphism chain reaction amplification. Genomics 20:176–183