Biochemical and Molecular Markers

Laurent Grivet and Jean-Louis Noyer

Biochemical and molecular markers have several applications in plant genetics. They allow us to observe closely the polymorphism of DNA sequences at a certain number of sites or loci spread across the genome. More precisely, biochemical markers reveal the polymorphism of sequences of certain proteins and thus, indirectly, the polymorphism of DNA sequences from which they are translated. Molecular markers directly reveal the polymorphism of DNA, the targeted sequences corresponding or not corresponding to the coding sequences.

Because of their properties, biochemical and molecular markers are a powerful tool to study the structure of genetic variability within a species and trace its evolutionary history. They are relatively unaffected by the environment or genetic basis. We can thus use them to compare individuals that were studied in different experiments or that are present in different collections. It is generally acknowledged that biochemical and molecular markers reveal a neutral polymorphism, i.e., one that is not subject to selection. They are relatively insensitive to homoplasy: there is little chance of observing two identical alleles that result from different mutational histories.

In this chapter we specify the characteristics of different biochemical and molecular markers to study the diversity of collections of plant genetic material. There are now more than ten techniques of genetic marking (see, for example, Weising et al., 1995; Karp et al., 1996, 1997, 1998; Santoni, 1996; de Vienne and Santoni, 1998). We describe here five widely used and promising techniques: isozymes, RFLP (restriction fragment length polymorphism), RAPD (random amplified polymorphic DNA, also called rapid), AFLP (amplified fragment length polymorphism), and microsatellites. The principle of each technique is briefly described in the appendix to this chapter. After summarizing the structure of the plant genome, we compare the five techniques in methodological terms (target sequences, nature and

level of polymorphism detected, genetic similarity) and practical terms (cost, quickness of result, infrastructure needed).

ORGANIZATION AND VARIABILITY OF PLANT GENOMES

The Organization of the Genome

The genome of a plant is distributed in three cell compartments: the mitochondria, the chloroplasts, and the nucleus.

The mitochondrial genome is composed of a circular molecule of master DNA of 200 to 2500 kilobases (1 kilobase = 1 kb = 10^3 base pairs) depending on the species, carrying 100 to 120 genes. The chloroplast genome is also circular and its size is about 150 kb. It has about 100 genes. The chloroplast and mitochondrial genomes are, most often, inherited through the maternal side.

The nuclear genome is composed of a definite number of chromosomes comprising DNA linear molecules. The genes are spread on a non-coding DNA matrix, essentially constituted of repeat DNA. The size of the nuclear genome varies considerably from one species to another: it is about 400 megabases (1 megabase = 1 Mb = 10^6 base pairs) in rice and about 16,000 Mb in wheat. The largest known nuclear genome among the angiosperms, that of *Fritillaria assyriaca*, contains 600 times more DNA than the smallest, that of *Arabidopsis thaliana* (Bennett and Smith, 1991). These variations are due to differences in ploidy and especially to differences in the quantity of non-coding dispersed repeat DNA. It is accepted that the number of genes carried by a nuclear genome on the plant is between 20,000 and 50,000. Therefore, the information contained in the nucleus, even if diluted, thus remains clearly more important than that contained in the cytoplasmic organelles. The nuclear genome has a biparental heredity via meiosis and fertilization.

DNA Polymorphism

The polymorphism of DNA results from the accumulation of mutations, that is, of modifications of sequences under the action of endogenous or exogenous factors. The mutations may appear in the form of visibly large rearrangements on the cytogenetic scale (deletion, translocation, inversion) or in the form of occasional modifications of sequences.

Biochemical and molecular markers are essentially used to detect occasional variations in sequences. Two types of variations can generally be distinguished: mutations corresponding to tmsubstitution of one base for another and mutations by insertion or deletion of a short fragment of DNA.

DNA polymorphism generated by occasional mutations is very commonly observed. It affects the entire genome, at every level, depending on the type of sequence and genomic compartment. For example, it is high in the repeat nuclear sequences of the micro- or minisatellite type, but low in the chloroplast DNA. The sequencing, over 1933 base pairs, of the nuclear gene *Opaque-2* of maize in 21 lines representing the diversity of temperate germplasm showed that the 21 sequences are entirely different. In total, 14% of the bases present a polymorphism resulting from occasional mutations and 26 phenomena of insertion-deletion have been detected (Henry and Damerval, 1997).

A particular type of DNA polymorphism is exploited by microsatellite markers. A microsatellite is generally composed of a tandem repeat of a motif formed of a small number of base pairs, generally two or three. Microsatellites are numerous in the eukaryote genomes and are spread throughout the chromosomes. Each is flanked by sequences specific to it. For each microsatellite site, mutations produced at a high frequency cause a variation in the number of repetitions of the base motif, which results in high allelic diversity. These mutations can be due to a gliding of the polymerase arising from unequal replication or crossing-over (Jarne and Lagoda, 1996, for review).

Sometimes there are significant differences in the total quantity of nuclear DNA between related interfertile species or even between genetic groups within the same species. These differences are probably due to variations in the quantity of repeat DNA. They can be revealed by flow cytometry (Dolezel, 1991; Bennett and Leitch, 1995). They may sometimes carry elements useful in differentiating groups within a species or within a species complex.

INFORMATION REVEALED BY DIFFERENT GENETIC MARKERS

In this section, the methodological aspects that differentiate the five types of genetic markers considered are covered, then the differences are examined to see how they affect the estimation of genetic diversity and the similarities between individuals.

The Nature and Genetic Interpretation of Polymorphism

Each type of marker is associated with a methodology that determines the nature of the information obtained. Two major categories of markers can be distinguished: (1) markers that can be used to reveal a series of several alleles for each locus studied (multiple allele markers) and (2) markers that allow detection of the presence or absence of a single allele for each locus, simultaneously for a large number of loci (genetic fingerprinting).

Multiple allele markers

Multiple allele markers comprise isozymes, RFLP, and microsatellites. The alleles detected are most often codominant: the two homologous alleles can be observed in heterozygous individuals.

Isozymes and RFLP

For the isozymes, a locus is defined by a catalytic function towards a specific substrate. For the RFLP, a locus corresponds to the region of the genome that hybridizes with a probe, the size of which is a few hundred to a few thousand base pairs. In both cases, the alleles are distinguished by their electrophoretic mobility. In the simplest case, an allele is materialized by a single band of specific molecular weight. Nevertheless, several factors may complicate this scenario.

- For the RFLP, an allele may correspond to a combination of several bands if there is a restriction site or sites for the restriction enzyme in the DNA sequence homologous to that of the probe.
- For the isozymes, the polymer enzymes may generate new bands in the heterozygotes.
- A substrate (isozymes) or a probe (RFLP) may simultaneously reveal the alleles of several paralogous loci, that is, of loci presenting the same DNA sequence or very similar sequences.
- For the RFLP, if several restriction enzymes are used in combination with a single probe, each allele corresponds in principle to a specific combination of bands detected with different enzymes. In practice, alleles are rarely reconstructed from analysis of multienzyme polymorphism because they are too complex. In this case, it is enough to note the presence or absence of each band or of each profile.

With isozymes and RFLP, therefore, it is often thus necessary to study the heredity of bands in controlled populations to interpret the profiles observed in terms of loci and alleles. Genetic interpretation is generally simpler in autogamous plants, for which each individual is most often homozygous for all its loci.

For the isozymes, the target sequences are by definition genes coding for the enzymes. The polymorphism is essentially due to occasional mutations that induce the replacement of one amino acid by another, modifying thus the overall electric charge of the protein or its molecular mass and, in consequence, its electrophoretic behaviour. However, several mutations are silent because of the degeneracy of the genetic code and a mutation inducing a change of amino acid does not systematically cause a modification of the electrophoretic mobility of the protein. This has been proved, for example, for the a-amylases of *Drosophila* (Inomata et al., 1995).

For the RFLP, the nature of the target sequences depends on the type of probe selected. It is possible to use probes that reveal the polymorphism of repeat sequences or single copy coding or non-coding sequences, of nuclear

or cytoplasmic sequences. We are essentially interested in single copy nuclear sequences. The polymorphism observed is due to mutations affecting the homologous sequence of the probe and above all that of its adjacent regions over about 10 kb. It may be a question of occasional mutations in the restriction sites of the enzyme or even insertions or deletions between these sites. The relative importance of the two phenomena may vary from one species to another. The occasional mutations that do not affect the restriction site of the enzyme or the small insertions-deletions may go unnoticed. Whether it is for the isozymes or for RFLP, an allele corresponds potentially to a group of several sequences, to the extent that these two techniques do not have the power of resolution to detect all the mutations that affect the sequence of the target locus.

Microsatellites

For the microsatellites, a locus is defined by a microsatellite site accompanied by its flanking sequences. Defined primers in these sequences, from either side of the microsatellite, allow us to amplify the locus to the exclusion of any other in the genome. This specificity is guaranteed by the length of the primers (20 to 25 nucleotides). It renders infinitely low the probability of amplifying another sequence in the genome purely at random. The source of the polymorphism being looked for is variation in the number of repetitions of the base motif of the microsatellite site. In most cases, each allele will be materialized by a band of specific molecular weight, which may be translated into a number of repetitions of the base motif of the microsatellite if the gel is sufficiently resolutive. In some cases, however, part of the alleles arise from variations of sequence between one of the primers and the microsatellite site (Orti et al., 1997). The rates of mutation associated with the microsatellite sites are 100 to 1000 times higher than for the isozymes (Jarne and Lagoda, 1996). Whatever the nature of the polymorphism, the profiles observed can be directly interpreted in terms of locus and alleles.

Genetic fingerprinting

Two techniques of genetic fingerprinting are presented: RAPD and AFLP. The term 'genetic fingerprinting' is commonly used because these techniques allow us to reveal simultaneously the polymorphism of a large number of loci so well that each individual has ample chance of having a multilocus profile of its own, as with human fingerprints.

With RAPD and AFLP techniques, the similarity of sequence between bands revealed by a single pair of primers does not allow us to establish relations of homology. It covers 20 base pairs at most, corresponding to the sum total of the sequence of the two primer sites, for RAPD. The similarity is less for AFLP, to the extent that the only common point between all the bands revealed by a pair of primers is the sequence of two flanking restriction sites and that of the selective bases, which represents at most 16 base pairs.

Typically, a RAPD amplification allows the detection of 5 to 20 bands and an AFLP amplification a few tens to 100 bands. For the genetic interpretation of profiles observed, each band is considered as an allele of a particular locus. The identity of the locus and of the allele depends entirely on the electrophoretic mobility of the band. For each locus, the only other allele possible is the absence of the band. The band is thus a dominant allele since the homozygous individuals that present the band cannot be distinguished from heterozygotes. For each locus, the absence of the band probably corresponds to a heterogeneous set of sequences.

Even if the bands amplified by the same pair of RAPD or AFLP primers have little chance of corresponding to homologous sequences, the two different bands may, in a small number of cases, materialize two alleles from a single locus. It generally cannot be perceived in studies of diversity, but this phenomenon can be observed when these techniques are used to make a genetic map.

In principle, AFLP and RAPD allow the targeting of all compartments of the genome and all types of sequences. The polymorphism revealed by RAPD corresponds somewhat to the modification of a primer site following the substitution of one base for another or to an insertion-deletion phenomenon. It may also be a matter of one insertion between two primer sites that are sufficiently far apart to make the amplification impossible (Williams et al., 1990). The polymorphism observed with AFLP corresponds probably to the modification of a restriction site of one of the two restriction enzymes following an occasional mutation (Vos et al., 1995).

Several other techniques of genetic fingerprinting give profiles that can be interpreted in genetic terms in the same way that RAPD and AFLP are. Among these are DAF (DNA amplification fingerprinting; Caetano-Anolles and Gresshoff, 1991), which is a variant of RAPD, ISSR (inter-simple sequence repeat analysis; Zietkiewicz et al., 1994), which consists of amplifying the sequences located between microsatellite sites, or even the use of microsatellite probes in RFLP (Rus-Koretekaas et al., 1994).

Comparison of the Five Techniques

In this section the polymorphism and genetic similarities measured with the different markers are compared on the basis of experimental data.

Polymorphism of multiple allele techniques

It is possible to compare the polymorphism obtained with different multiple allele techniques by using the usual parameters of population genetics such as the percentage of polymorphic loci, the mean number of alleles per locus, and the Nei diversity index, applied to the same sample of individuals.

Table 1. Comparison of rates of polymorphism, Nei diversity and the average number of alleles per locus between isozymes (Iso) and RFLP and between microsatellites (SSR) and RFLP, in various studies. For RFLP, a single restriction enzyme was used, unless otherwise mentioned

Species	Technique	No. of individuals studied	No. of loci studied*	Rate of polymorphism (%)	Nei diversity	Av. no. of alleles per locus	Reference
Brassica campestris	Iso	285	5	60	0.41	2.7	Mitchell and Quiros (1992)
	RFLP	277	4	100***	0.60	8.5	
Hordeum vulgare	Iso	268	7	100	0.44	4.9	Zhang et al. (1993)
	RFLP	240	13	100***	0.47	4.2	
Populus tremuloides	Iso	118	14	77	0.25	2.8	Liu and Furnier (1993)
	RFLP	91	41	71	0.25	2.7	
Populus grandidenta	Iso	96	14	29	0.08	1.4	Liu and Furnier (1993)
	RFLP	75	37	65	0.13	1.8	
Zea mays	Iso	21	22	68	-	2.1	Messmer et al. (1991)
	RFLP	21	144(79)*	94	-	3.34**	
Zea mays	Iso	31	27	-	-	2.2	Gerdes and Tracy (1994)
	RFLP	43	71	-	-	4.1	
Zea mays	Iso	445	20	75	0.23	2.4	Dubreuil and Charcosset (1998)
	RFLP	285	35	100	0.61	6.3	
Hordeum vulgare	RFLP	18	114(42)*	-	0.32	2.62**	Russell et al. (1997)
	SSR	18	13	100	0.57	5.38	
Glycine max	RFLP	19	114	100***	0.38	2.15	Morgante et al. (1994)
	SSR	19	4	100	0.52	4.25	
Glycine sp.	RFLP	12	110	25	0.41	—	Powell et al. (1996)
	SSR	12	36	100	0.60	4.28	
Oryza sativa	RFLP	14	18	33	0.10	2.3	Wu and Tanksley (1993)
	SSR	14	8	100	0.70	5.2	
Oryza sativa	RFLP	20	12	83	0.32	2.5	Olufowote et al. (1997)
	SSR	20	10	100	0.62	7.4	
Zea mays	RFLP	12	96	-	0.58	—	Taramino and Tingey (1996)
	SSR	12	34	100	0.76	6.56	

*Number of enzyme-probe combinations, value in parentheses is number of probes.
**Average number of bands per enzyme-probe combination.
***RFLP probes preselected for their polymorphism.

Examples of comparisons between isozymes and RFLP and between RFLP and microsatellites are given in Table 1. These results must be interpreted with caution to the extent that samples of individuals are generally small and to the extent that the numbers of loci observed are often not equal among the techniques. In most of the studies mentioned, the RFLP polymorphism is observed with a single enzyme. Even with this restriction, RFLP appears at least as polymorphic as the isozymes and sometimes clearly more so. This result agrees with what is known of the nature of polymorphism detected. Isozymes allow us to detect the polymorphism of gene sequences only across the fine filter of translation and degeneracy of the genetic code. The RFLP allows detection of polymorphism of coding and non-coding regions. Moreover, when the probes corresponding to the genes are used, it is essentially the polymorphism of the regions flanking the target genes that is revealed.

In Table 1, microsatellites appear systematically much more polymorphic than RFLP, which is consistent with their particularly high rates of mutation. These observations are confirmed by several other studies. Saghai-Maroof et al. (1994), for example, have found up to 28 different alleles for a microsatellite locus in barley.

Polymorphism revealed by all the techniques

The number of alleles per locus and the Nei diversity index lose part of their genetic significance when fingerprinting techniques are applied. The number of alleles per locus is by nature limited to two (presence and absence of the band). In consequence, the Nei diversity index reaches a maximum at 0.5. Moreover, calculation of the Nei index is based on the knowledge of allele frequencies. Because markers are dominant, these frequencies are directly accessible in only two particular cases: if one makes the hypothesis that there high probability of pairing individuals (the Hardy-Weinberg equilibrium is respected) or even if there is autogamy and one makes the hypothesis that all the individuals are homozygous for all the loci. If these hypotheses are not acceptable, the calculation of allelic frequencies is not possible at least to make an independent estimation of the fixation index F_{IS} (Lynch and Milligan, 1994; Kremer, 1998). If one of these hypotheses is overused, the estimation of the Nei index may be biased. This index is nevertheless sometimes used indicatively on markers of the genetic fingerprinting type. In a study by Liu and Furnier (1993) on two species of poplar, the Nei indexes calculated for the RAPD markers are about 0.3 for each species. These values are higher than those obtained for isozymes and RFLP on the same individuals (Table 1). Nevertheless, the hypothesis made on the Hardy-Weinberg equilibrium prevents a clear conclusion as to the better aptitude of RAPD in revealing a polymorphism.

In the particular case of autogamous species, the allele frequencies may be directly estimated from the frequency of genotypes and the Nei index

may be calculated. On soybean, for example, Powell et al. (1996) found Nei indexes of 0.31 and 0.32 for RAPD and AFLP, respectively. These values are lower than those obtained for RFLP and microsatellites on the same sample of individuals (Table 1).

Genetic fingerprinting techniques also reveal simultaneously the polymorphism of several loci per experimental unit. An experimental unit corresponds to a strip on a gel. To take into account this characteristic, which has great practical importance for the experimenter, it may be useful to compare the polymorphism not at the level of the locus but at the level of the experimental unit. Several measurements of diversity may thus be considered, for example, the average number of polymorphic bands per experimental unit, the PIC (polymorphism information content; Weir, 1990), and, for autogamous species, the marker index (Powell et al., 1996). The PIC is written as the Nei index, replacing the allele frequencies by frequencies of multilocus profiles. The marker index is defined as the product of the Nei diversity index by the number of polymorph bands per experimental unit. It is interpreted as the average number of bands per experimental unit differentiating two randomly selected individuals. On the basis of these criteria, comparisons of techniques used on barley (Russell et al., 1997), soybean (Powell et al., 1996), and peas (Lu et al., 1996) show that AFLP is distinguished by an average number of polymorphic bands per experimental unit, a PIC, and a marker index clearly higher than those for other techniques. AFLP is useful not because it reveals more diversity in each locus, but because it allows us to observe simultaneously a large number of polymorphic loci.

Genetic similarities

When a group of individuals is observed for several loci with a given technique, it is possible to calculate a genetic similarity for all the pairs of individuals (Perrier et al., 1999). The set of similarities can be synthesized by a multivariant analysis or a branched representation. It is legitimate to ask whether the various markers give the same estimate of the genetic proximity of individuals.

There are so far few experimental results that allow us to address these questions. Some are grouped in Table 2 in the form of correlations between indexes of similarity calculated for different pairs of techniques. Examples of relations between distributions are given in Fig. 1. The absence of homogeneity between the tests, the levels of significance used, the ranges of grouping between individuals, and the power of parameters (number of individuals studied) complicate the synthesis of observations. Nevertheless, as a general rule, the correlations between the similarities calculated from several markers are significant but of variable intensity.

The low correlation values may be explained for the most part by the locus sampling bias. This is particularly true for isozymes in that the number

Table 2. Correlation between similarities calculated from different techniques of molecular markers

Reference	Species	No. of individuals used to calculate correlation	Test*	Comparison**	Correlation (*r*)	Threshold (*P*)
Lu et al. (1996)	peas	10	Tm	RFLP-RAPD	0.5	<0.05
				RFLP-AFLP	0.7	<0.05
				RAPD-AFLP	0.6	<0.05
Powell et al. (1996)	soybean	10	Tm	RFLP-RAPD	0.24	ns
				RFLP-AFLP	0.42	<0.01
				RFLP-SSR	0.18	ns
				RAPD-AFLP	0.45	<0.01
				RAPD-SSR	0.15	ns
				AFLP-SSR	0.14	ns
Beer et al. (1993)	*Avena sterilis*	177	Tm	RFLP-Iso	0.26	<0.005
Heun et al. (1994)	*Avena sterilis*	24	Tm	RFLP-Iso	0.36	<0.01
Peakall et al. (1995)	*Buchloe dactyloides*	48	Tm	RAPD-Iso	0.63	<0.001
Messmer et al. (1991)	maize	21	Cp	RFLP-Iso	0.23	<0.01
Gerdes and Tracy (1994)	maize	31	Cp	RFLP-Iso	0.26	<0.01
Prabhu et al. (1997)	soybean	10	Cp	RFLP-DAF	0.70	<0.01
Russell et al. (1997)	barley	18	Cr	RFLP-RAPD	0.20	<0.01
				RFLP-AFLP	0.71	<0.01
				RFLP-SSR	0.51	<0.01
				RAPD-AFLP	0.11	ns
				RAPD-SSR	0.24	<0.01
				AFLP-SSR	0.52	<0.01
Dos Santos et al. (1994)	Cabbage	45	Cr	RFLP-RAPD	0.75	<0.01
Engquist and Beeker (1994)	colza	17	Cr	RFLP-RAPD	0.76	<0.01
				RFLP-Iso	0.53	<0.01
				RAPD-Iso	0.67	<0.01
Thormann et al. (1994)	*Brassica* spp.	18	Tm	RFLP-RAPD	0.93	—

*Tm, Mantel test; Cp, Pearson coefficient of correlation; Cr, Spearman coefficient of rank correlation.
**Iso, isozymes; SSR, microsatellites; DAF, DNA amplification fingerprinting.

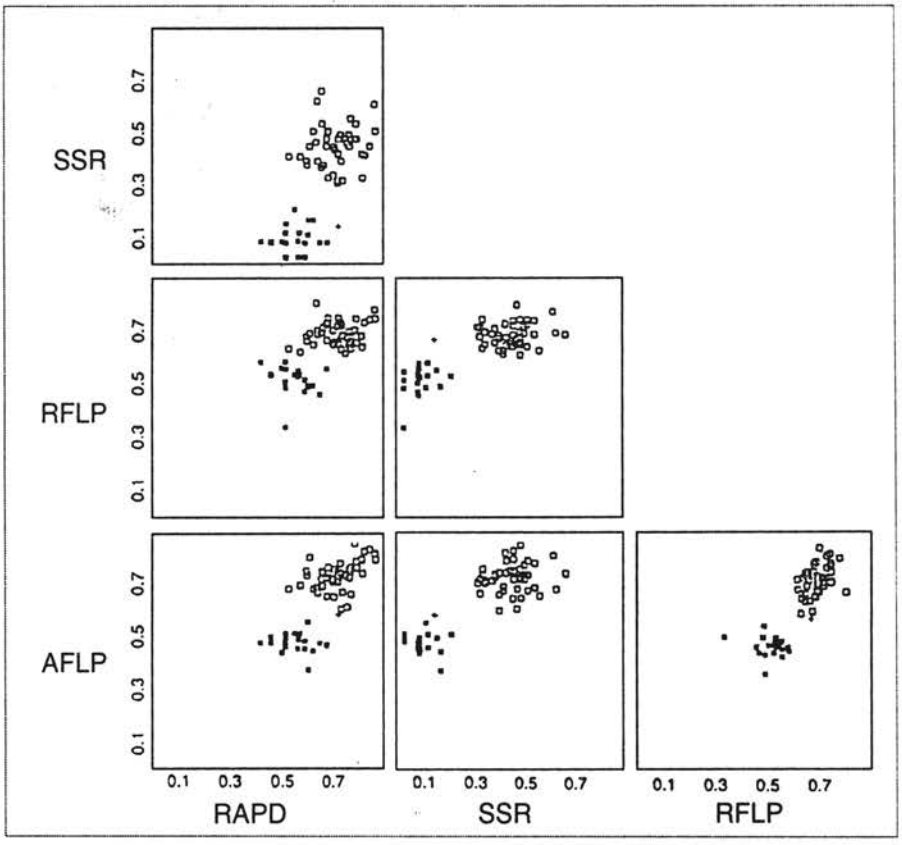


Fig. 1. Two-dimensional relations between the genetic similarities calculated from RFLP, AFLP, microsatellites (SSR), and RAPD in soybean. The white squares represent the similarities between accessions of *Glycine max*, the pluses represent similarities between accessions of *G. soja*, and the black squares represent similarities between accessions of the two species (Powell et al., 1996).

of loci analysed is most often less than 20. Messmer et al. (1991) calculated the type deviations of indexes of similarity by the jackknife method (Millier, 1974). They showed that the values of type deviations obtained with 22 isozymic loci were two to three times higher than values obtained with 144 enzyme-probe combinations of RFLP. This is also true for molecular techniques, which allow detection of more loci. Dos Santos et al. (1994) showed by the bootstrap method (Efron and Tibshirani, 1986) that the distribution of correlations between indexes calculated for 100 random samples of 56 RAPD bands taken two by two widely overlapped the distribution of correlations between indexes for 100 random samples of 56 RAPD bands and 56 RFLP bands taken two by two.

The technical quality of manipulations has a direct repercussion on the reliability of encoding of data and can thus also be a source of divergence in

the estimation of similarities. The encoding may become laborious if technical quality is not high, particularly for AFLP, which reveals several bands simultaneously. Many studies have indicated problems of reproducibility with RAPD, which necessitates repeated analyses of each sample if serious errors are to be avoided in the encoding of data (Salimath et al., 1995; Yang et al., 1996). Moreover, the profiles are difficult to reproduce from one laboratory to another, which is a handicap for studies that are part of a network (Jones et al., 1997). Finally, the competition for primers may lead in certain cases to the non-amplification of a band present in the heterozygous state, because of its considerable dilution (Hallden et al., 1996).

More alleles can be detected at a single locus with microsatellites than with other techniques, which results in an average index of similarity between individuals that is generally much lower. This may explain the low correlation between this technique and the others, especially when the individuals are not closely related (the correlation is not significant in the study of Powell et al., 1996) (Table 2).

RAPD seems sometimes to give an estimation of similarity that differs from that of other techniques, particularly when the individuals are distant (of two different species). This divergence has been observed by Thormann et al. (1994) in the genus *Brassica* and by Powell et al. (1996) in *Glycine*. Thormann et al. (1994) have shown by hybridization that, in 3 cases out of 15, bands of the same mobility are not homologous in the different species studied. RAPD may be more sensitive to this type of confusion than other markers.

The influence of evolutionary mechanisms that underlie the variation observed cannot be completely ruled out in explaining the differences between markers. N'goran et al. (1994), for example, showed that in cacao the portion of bands amplified by RAPD corresponds to those repeat sequences and that the bands amplifying repeat sequences and unique sequences do not give the same structure of genetic material.

Relationship with genealogy and morphological characters

It may be asked whether the similarities calculated from markers can be used to organize the relations between individuals into a hierarchy in the same way as genealogy or morphological characters.

The relationship between the similarity evaluated with markers and the relation coefficient of Malecot has been studied in several plants for different markers (Table 3). Overall, the correlations between the two types of information are low but significant. The low value of the correlations can be explained by the fact that the calculation of coefficients of relation is based on unrealistic hypotheses, particularly the absence of relationship between distant ancestors whose genealogy is not known. Also, genetic drift and selection are not taken into account. This caused Graner et al. (1994) to suggest

Table 3. Correlation between the Malecot coefficient of relation *f* and the similarity calculated from markers

Reference	Species	Technique	Test*	No.**	Correlation	Threshold (*P*)
Cox et al. (1985)	Soya	Isozymes	Cr	27	0.15	< 0.01
				32	0.40	< 0.01
				39	0.45	< 0.01
O'Donoughue et al. (1994)	Oat	RFLP	Tm	55	0.32	< 0.001
Graner et al. (1994)	Barley	RFLP	Cr	21	0.21	< 0.05
				17	0.42	< 0.01
Gerdes and Tracy (1994)	Maize	RFLP	Cp	42	0.54	< 0.01
		Isozymes	Cp	31	0.32	< 0.01
Tinker et al. (1993)	Barley	RAPD	Cr	27	0.51	< 0.01
Plaschke et al. (1995)	Wheat	Microsatellites	Cr	40	0.55	< 0.001

*Tm, Mantel test; Cp, Pearson coefficient of correlation; Cr, coefficient of rank correlation.
**Number of individuals used to calculate the correlation.

that the relationship calculated with molecular markers, which does not ignore these parameters, may be a more relevant tool than the *f* coefficient of Malecot.

Low but significant correlations have been observed between similarities calculated from molecular markers and distances calculated from morphological variables. For example, in *Avena sterilis*, a correlation of 0.24 ($P < 0.005$) was observed between RFLP and morphology and a correlation of 0.13 ($P < 0.005$) was found between isozymes and morphology (Beer et al., 1993). In maize, correlations of 0.167 and 0.124 ($P < 0.01$) were obtained in these two cases respectively (Gerdes and Tracy, 1994). More precisely, Dillmann et al. (1997) emphasized that the relationship between the two types of information is not linear in maize. A small distance calculated with RFLP markers corresponds always to a small morphological distance. On the other hand, a large distance calculated with markers may correspond to a large or small morphological distance (Fig. 2). This triangular relationship may be explained by the fact that, for a given morphological character, various combinations of genes may correspond to a single phenotype. Moreover, the linkage disequilibrium between the markers and the genes involved in the character can vary according to the sampling of loci. If it is low, the two types of information will be independent.

PRACTICAL IMPLEMENTATIONS OF MARKERS

Two major practical aspects of genetic marker techniques are the ease with which they can be used and their cost.

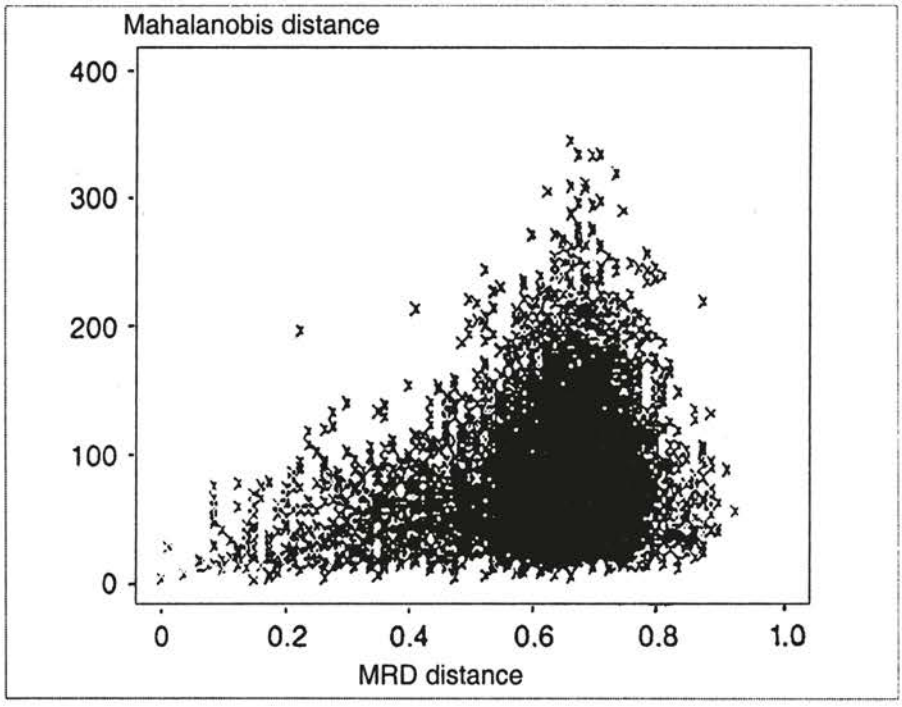


Fig. 2. Relation between the similarity (distance of Rogers, MRD) calculated from RFLP markers and the distance calculated from morphological data (Mahalanobis distance) in maize (Dillmann et al., 1997)

Use

Isozymes do not require large laboratory equipment and use only small quantities of plant material. They may be rather laborious to use on a new species. That depends on the difficulty of finding an equilibrium between the buffer for protein extraction and the electrophoresis buffer. This equilibrium is more difficult to reach for plants with organs rich in polyphenols. The main handicap of isozymes arises from the small number of loci that they can be used to detect. This number is always less than 50 and often less than 20.

Practical use of molecular markers depends on three criteria: the difficulty of using them during the start of a programme, the quantity of DNA needed to analyse each sample, and the essential laboratory equipment for routine implementation.

Genetic fingerprinting, RAPD, and AFLP can be used directly on any plant, without preliminary steps. On the other hand, RFLP and microsatellites may be very difficult to get underway unless a great deal of work has already been completed by other teams on the plant under study.

For RFLP, a source of probes must be arranged, which requires the construction and maintenance of a bank of genomic DNA or cDNA. However, in taxonomic groups that are widely studied, such as the Poaceae, it is possible to use probes that are said to be heterologous, coming from other species. A group of 180 probes of rice, barley, and oats, having a wide spectrum of hybridization over the group Poaceae, has been defined. It is distributed by Cornell University in the United States (van Deyne et al., 1998). These probes can, in principle, be used on all the species of this botanical family.

Microsatellites require a preliminary step involving screening of a bank and sequencing to define the primers in the flanking regions of the microsatellite sites, which is a difficult and costly process of molecular biology. In the most commonly studied plants, such as wheat, maize, and rice, the primer sequences of many microsatellite sites are published and this step is therefore no longer necessary, but that is not the case with most tropical plants. Moreover, sequence information often cannot be transferred from one species to another, even among taxonomically related species.

The quantity of DNA required may be a limiting factor when DNA is difficult to extract or when large samples are worked on. This quantity depends a great deal on whether the technique used relies on PCR. PCR requires low quantities of DNA. For RAPD and microsatellites, only a few nanograms of DNA are needed to complete an amplification. For AFLP, some hundreds of nanograms will do. On the other hand, 2 to 10 μg of DNA are required per track for RFLP. Moreover, microsatellites can be implemented with DNA of mediocre quality obtained with a basic extraction technique because of the high specificity of primer hybridization.

Microsatellites and RAPD can be used routinely in laboratories with rudimentary equipment. On the other hand, RFLP and AFLP rely on radioactive or biochemical marking of probes or primers and thus require greater laboratory infrastructure. AFLP may possibly be revealed with a DNA colorant such as silver nitrate (Cho et al., 1996). This technique is less constraining in terms of infrastructure but reading the profiles obtained often becomes a more delicate task.

Cost

Cost is inherently a determining factor in the choice of a molecular marker. The cost of a routine genotyping operation depends on the availability of certain molecular tools, such as probes for RFLP or primers for microsatellites, the perfection of which may lead to significant and variable costs (fabrication and maintenance of a probe bank, definition of microsatellite primers). If these costs are set aside, for a given operation, the cost of different markers can be calculated in two parts: the cost price at an elementary point (given one marker for an individual) and the time taken to obtain it. The cost price at one elementary point depends on how the laboratory gets its supplies and

its geographic location in relation to the plant material and suppliers. The size of the laboratory evaluated in terms of consumption of products or volume of results obtained for a type of marker influences this component. The time taken to obtain results is related to salary costs and costs of material depreciation. The size of the laboratory in terms of equipment and critical mass of operators (technicity) influences this second component. The cost factor is thus very difficult to extrapolate from one laboratory to another.

In the laboratory of CIMMYT (Centro Internacional de Mejoramiento de Maíz y Trigo, Mexico), Ragot and Hoisington (1993) compared the cost of two types of markers, RAPD and RFLP, distinguishing for the latter the radioactive marking of probes (rRFLP) and biochemical markers (cRFLP). They concluded that the relative cost of different techniques depends on the size of samples treated. RAPD is less expensive for small samples and, even if its cost declines, it becomes costlier in relative terms for large samples. This is mainly due to the economy of scale from reuse of membranes for RFLP markers. Ragot and Hoisington (1993) estimated that the cost begins to drop significantly from three reuses onwards. For cRFLP, it is difficult to go beyond 10 uses, routinely. For rRFLP, the number of uses can vary from 15 to 20, according to our experience. According to Ragot and Hoisington (1993), the cost of DNA extraction may represent half the cost of RAPD. To reduce this cost, Mohan et al. (1997) suggested the use of the squash technique on plant tissues (Langridge et al., 1991). The concentration of polysaccharides and other 'contaminants' of DNA present in the tissues of a good number of tropical plants often makes this technique impracticable. There is thus no universal guideline, but the positive observations of Mohan et al. (1997) show that it is very difficult to understand all the possible situations.

Risterucci (1997) compared the routine cost price of different types of marker in the Biotrop programme of CIRAD. For example, for the genotyping of 100 individuals with 100 markers, the final cost was 0.17 euros for AFLP, 0.26 euros for RAPD, 0.48 euros for RFLP, 0.49 euros for microsatelliites, and 0.61 euros for isozymes. It is clear from this comparison that the cost price of multiple allele techniques (RFLP, microsatellites, isozymes) is higher than that of genetic fingerprinting techniques. The high cost of microsatellites arises from the choice to use radioactivity to detect them. Moreover, when the final cost is considered as a function of the sample size, it appears that beyond 25 individuals studied the economies of scale are much more significant with the RFLP markers than with markers based on PCR, as was observed by Ragot and Hoisington (1993). RFLP costs almost the same as genetic fingerprinting techniques for a sample of about 200 individuals.

Even though certain trends can be inferred, it seems impractical to attempt to give a standard and universal cost for each type of marker. Rather than choose a technique on the basis of comparative cost in a necessarily highly specific context, it seems more realistic to choose a technique on the basis of the information required and then find a means of applying it as cheaply as

possible. There are ways to reduce costs that may make the choice of type of marker totally independent of cost: the subcontracting of certain steps such as occasional sequencing to service providers, agreements with other competent research laboratories in a specific field and for limited periods, and the use of kits for short periods or for occasional projects.

CONCLUSION

With biochemical and molecular markers we can obtain an image of the non-selective diversity existing in a collection. Each type of marker has advantages and disadvantages and the choice of a particular technique must be reasoned on a case-by-case basis, taking into account the objectives, available means, and state of our knowledge of the species studied. As far as we know at present, the different techniques seem to give images sufficiently similar to the genetic structure. The differences observed can be attributed to the sampling of loci, to problems of repeatability and homoplasy for certain techniques such as RAPD, and the nature of polymorphism revealed and sequences targeted by each technique. It is probable that we lack distance enough to judge on this last point.

The relationship between neutral polymorphism revealed by markers and the polymorphism of useful morphoagronomic characters is not clear. Markers reveal more or less precisely the similarities of sequences between individuals for a sample of a locus. Morphoagronomic characters measure resemblances between individuals on the basis of variables whose level of expression depends on the number of potentially epistatic genes. Markers may, however, be a useful tool to better sample genes of agronomic interest. If the species is structured in genetic groups, that signifies that the entities (populations, groups of genotypes) have had independent evolutionary histories, characterized by limited gene exchange. It is thus possible that the role of alleles determining the expression of characters of agronomic interest is different from one group to the other in response to genetic drift or to various pressures of selection. In this case, the sources of variations for characters manipulated by the selector can be better sampled if the genetic structure revealed by the markers is taken into account.

APPENDIX

Description of Molecular Techniques

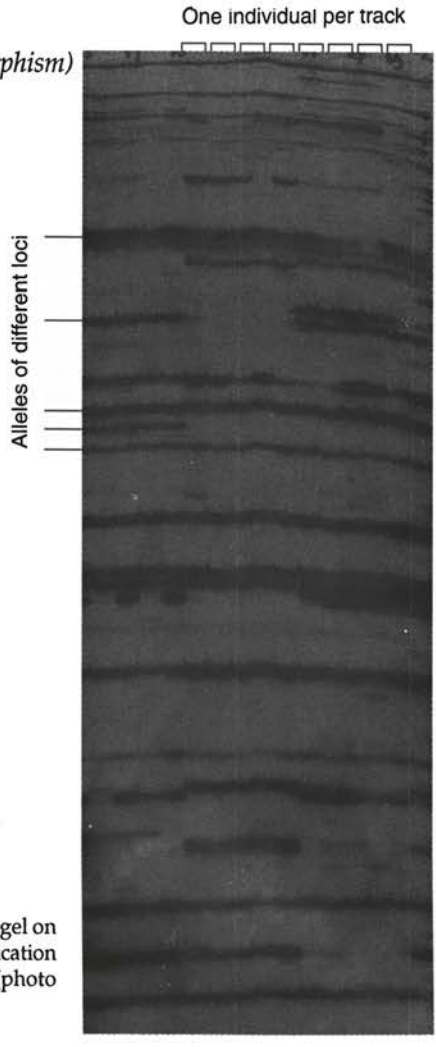
This section contains a succinct description of AFLP, RAPD, RFLP, and microsatellite techniques. For more detailed descriptions, the reader is referred to Karp et al. (1997), de Vienne and Santoni (1998), and the internet site of IPGRI (*http://www.cgie.org/ipgri/training*). Words marked by an asterisk are explained in the glossary at the end of this chapter. The internet site of the Institut Pasteur (*http://www.pasteur.fr/other/biology/francais/bio-docs-fr.html*) provides links to many dictionaries.

Isozymic techniques are not covered here, although they are competitive in terms of diversity analysis, because they are already fully described in the literature.

AFLP (amplified fragment length polymorphism)

After extraction*, DNA is hydrolysed under the action of two restriction enzymes*. Specific adaptors of two types of restriction sites* are joined (ligation*) at the tips of the fragments obtained. The DNA thus prepared is used during a PCR amplification*, which uses two primers corresponding to adaptors* to which are added one to three arbitrary bases of 3′*. One of two primers is labelled. The fragments produced by this amplification are separated by electrophoresis* on a denaturant polyacrylamide gel*.

The gel is exposed* for some days to contact with an autoradiograph film*. For each individual, several tens of fragments, or bands, are materialized. Each band is interpreted as an allele of a particular locus.

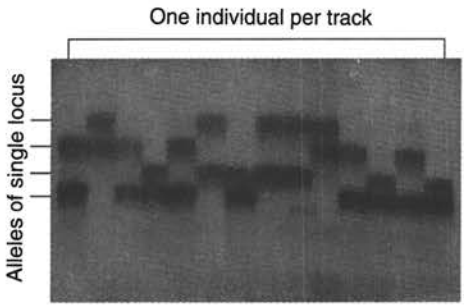


Autoradiograph of a denaturant polyacrylamide gel on which fragments from an AFLP radioactive amplification are separated. The total length of the gel is 40 cm (photo R. Purba, CIRAD).

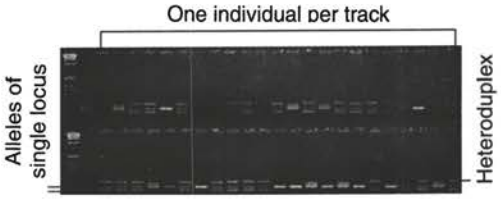
Microsatellites

After extraction*, the DNA is used in the same way as during a PCR amplification*, which uses two primers* defined in the flanking regions* after sequencing* of a microsatellite site (simple sequence repetition, SSR). The fragments produced by this amplification are separated by electrophoresis* on a resolutive agarose gel* (concentrated) or, better still, on a denaturant polyacrylamide gel*.

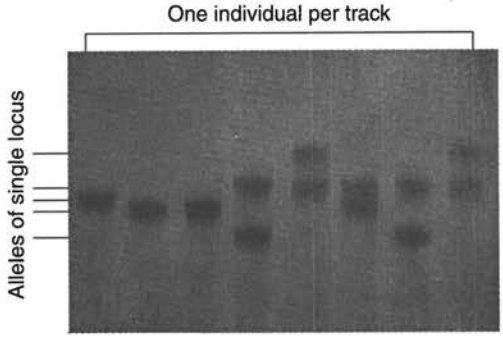
It is the variation of number of simple sequence repetitions that is revealed. The monolocus nature of microsatellites allows some flexibility as to the mode of detection of products arising from PCR amplification*. The bands observed materialize the alleles of the target locus. Each allele corresponds to a particular number of repetitions of the base motif of the microsatellite.



Autoradiograph of a long gel (40 cm) of denaturant polyacrylamide after radioactive PCR amplification of a microsatellite locus of hevea (photo M. Rodier-Goud and M. Seguin, CIRAD).

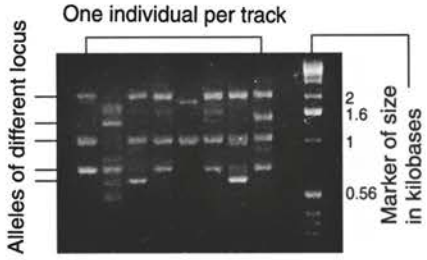


UV photograph of agarose gel. The products of PCR amplification of a microsatellite locus of hevea are stained with ethidium bromide. This type of detection is useful because it is easy to implement. The disadvantages are the lower resolutive power than those of denaturant polyacrylamide gels, as well as the appearance of supplementary bands called heteroduplex, which are the result of an artificial reassociation of alleles of a heterozygous individual at the end of PCR (photo by M. Rodier-Goud and M. Seguin, CIRAD).



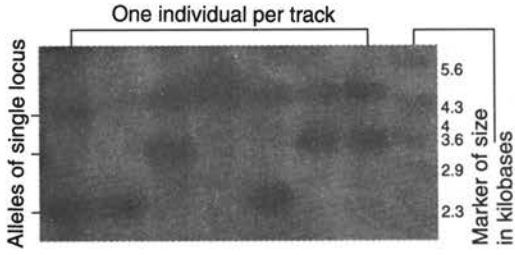
Photograph of a short gel (12 cm) of denaturant polyacrylamide. The products of PCR amplification of a microsatellite locus of banana are stained with silver nitrate (photo by J.L. Noyer, CIRAD).

RAPD (random amplified polymorphic DNA)
After extraction*, the DNA is used as during a PCR amplification*, which results in a single primer* of 10 bases defined arbitrarily. For the amplification to occur, this primer must find two inverted homologous sequences*, located at a distance such that the amplification can take place. The fragments produced by this amplification are separated by electrophoresis* on a resolutive (concentrated) agarose gel*. The result is presented in the form of a photograph of gel stained with ethidium bromide. Each band is interpreted as an allele of a particular locus.



UV photograph of 4% agarose gel. The products of RAPD amplification of DNA of 8 cacao plants are stained with ethidium bromide (photo by J. N'goran, CIRAD).

RFLP (restriction fragment length polymorphism)
After extraction*, the DNA is hydrolysed under the action of a restriction enzyme*. Fragments of variable size are then separated by electrophoresis* on an agarose gel* and then transferred on a membrane (Southern*). A labelled probe is applied on this membrane. It hybridizes on the fragments that contain all or part of a sequence* that is homologous* to it. After several washes with increasing stringency, the membrane is exposed* for a few hours or up to a few days. The result is presented in the form of an autoradiograph.. The bands observed materialize the different alleles of a target locus. Each allele corresponds to a particular configuration of restriction sites around the homologous region of the probe.



Autoradiograph of the DNA of seven diploid banana plants hybridized with a probe labelled with 32P (photo J.L. Noyer-CIRAD)

GLOSSARY

3′-5′ ends: The double strand of DNA is made up of two complementary antiparallel strands, oriented 3′OH-5′P. The synthesis of DNA occurs only from the 5′ end to the 3′ end. The bases of the primer located on the 3′ end, where the synthesis occurs, have greater significance in the accuracy of the required homology than the bases located at the other end of the primer during PCR amplifications.

Adaptor: Fragment of double-strand DNA, synthesized, the known sequence of which serves as a primer site for PCR amplifications. The adaptors are already linked to DNA fragments that are to be amplified.

Agarose gel: Support of migration in electrophoresis. The resolutive (separating) power increases with the concentration. All the same, one cannot go beyond a concentration of around 4%. The resolution does not routinely go beyond about 10 base pairs, even with high-performance agaroses.

Autoradiography: Sensitive film that can be imprinted by a radioactive or light ray. This radiation localized on a gel or on a membrane can provide a labelled primer at its tip or a marked DNA.

Denaturant polyacrylamide gel: This support of electrophoretic migration is a polyacrylamide gel in the composition of which is introduced a chaotropic (generally of urea) that keeps the denatured complementary strands dissociated. The migration technique allows the separation of DNA fragments that otherwise would not be able to penetrate the gel beyond a certain size.

Electrophoresis: A technique that allows the migration of a polarized molecule under the action of an electric field. At a constant charge density, the smallest molecules migrate further from their origin than the largest molecules. Since DNA has a negative charge in solution, one can separate fragments of different sizes.

Endonuclease: Enzyme capable of precisely catalysing the hydrolysis of DNA, which causes cleavage of two strands within the chain (as opposed to Exonucleases, which cause cleavage at the extremities of the chain).

Ethidium bromide: Intercalary agent of DNA that is fixed inside the double helix. Once fixed, it becomes fluorescent under the UV light. This characteristic is used to visualize DNA on a gel. (Ethidium bromide is a toxic product.)

Exposure: When DNA fragments have been labelled, they emit a rather luminous or radioactive glow. The radiation is detected on contact with

a sensitive radiographic film (Autoradiography). According to the intensity of the signal, the detection requires a reasonably long period of contact with a substrate (such as gel or membrane) that contains DNA.

Extraction: There are various techniques to extract DNA from plants. The high proportion of 'contaminants' that accompany DNA in this phase obliges technicians to carry out long purification steps in addition to the extraction itself. This difficulty of extraction, specific to the plant kingdom, has long delayed the development of techniques of molecular biology in comparison with those used in the animal kingdom. Even today, some techniques are difficult to use because of the sometimes mediocre quality of the DNA obtained.

Flanking regions: DNA sequences located on either side of a target sequence.

Homologue: See Homologous sequence.

Homologous sequence: Sequences identical in the series of bases that constitute them. The notion of homology can be relative and arbitrary thresholds can be fixed on it. Sequences can thus be considered homologous even if they are not totally identical, in which case one speaks of a percentage of homology (see also Hybridization).

Hybridize/Hybridization: Pairing of two complementary strands of DNA. The stringency can be controlled by temperature for a given saline charge. For a PCR primer of 10 to 30 bases, the hybridization temperature (Tm) can be calculated with precision, even if the calculation formulae have been obtained empirically. A drop of some degrees in the temperature calculated can bring the primer to hybridize on sites that are not totally homologous to it, particularly at its 5′ end. For probes used in RFLP, which are larger, the control of homology is not absolute (see also Homologous sequence).

Hydrolysis: see Endonuclease.

Ligation: Linkage of two double-strand fragments of DNA at their tips. For this, ligase is used, an enzyme that can effect this linkage no matter what kind of cut (plain or sticky ends) has produced the strands.

Marker/Marking (also *Labelling*): Technique that allows the tracking and location of a DNA molecule by addition or replacement of one or several nucleotides by nucleotides modified by an artificial radioisotope of the ^{32}P, ^{33}P, or ^{35}S type or a chemoluminescent molecule. A large fragment (more than 100-200 nucleotides) is marked by replacing a normal nucleotide by a modified nucleotide (nick translation or random priming labelling). A small fragment or oligonucleotide is marked by the addition of a base modified in terminal position (end labelling).

PCR amplification: Reaction of synthesis of DNA fragments following a series of cycles of denaturation, hybridization, and synthesis, which allows the amplification in quantity of a target DNA sequence. In theory, an amplification of n cycles gives 2n times the number of copies of fragment synthesized in the course of the first cycle. In practice, it always gives less. The discovery of heat-stable DNA polymerases, resistant to several cycles of denaturation of DNA at 95°C, has led to the spectacular development of this technique.

Primer: Fragment of single-strand DNA, synthesized, which after hybridization on a homologous sequence of a denatured strand of DNA serves as a point of anchorage to trigger the synthesis of a complementary strand. During the hybridization on genomic DNA, it is the length of the primers that defines the specificity of the target locus. In the case of RAPD, one primer of about 10 bases will hybridize on several hundreds, even thousands, of sites in each direction and allow the amplification of a small number of fragments (about 10). In the case of microsatellites, two primers of 20 bases will statistically allow the amplification of a single locus (apart from duplications).

Probe: Marked DNA fragment the sequence and function of which are not necessarily known (but may be) and that is used to locate homologous sequences among other sequences.

Restriction enzyme: Endonuclease that cuts DNA at a specific site. This site is defined by a sequence of bases. The most frequently used enzymes recognize sequences of 4 or 6 bases.

Restriction site: A sequence recognized by a restriction enzyme.

Sequence: A determined series of bases that compose a DNA strand.

Sequencing: A process that determines the sequence of a DNA fragment.

Southern (blot): After T. Southern, inventor of the technique. A technique of transfer of DNA on a membrane (blot) following a hybridization, which then allows research by means of a probe labelled with target DNA fragments. By analogy between the name of the inventor and the cardinal points, the Western blot technique was developed for the proteins and the Northern blot for RNA. The East remains to be conquered.

Stringency: Quality of a medium of hybridization or reassociation of nucleic acids. Stringency depends mainly on conditions of temperature, pH, and saline concentration. If stringency is high, all the base pairs of two nucleic acids that are to reassociate must be complementary. When the stringency drops, the reassociation is made by tolerating a growing number of base pairs with a lower percentage of homology.

Washing: In RFLP, a technique that allows the elimination of non-specific hybridization of probes from a membrane, that is, probes that do not have a sufficiently high degree of homology. As for hybridization, it is the saline charge and the temperature that determine the stringency.

REFERENCES

Beer, S.C., Goffreda, J., Phillips, T.D., Murphy, J.P., and Sorrells, M.E. 1993. Assessment of genetic variation in *Avena sterilis* using morphological traits, isozymes and RFLPs. *Crop Science*, 33: 1386-1393.

Bennett, M.D. and Leitch, I. 1995. Nuclear DNA amounts in angiosperms. *Annals of Botany*, 76: 113-176.

Bennett, M.D. and Smith, J.B. 1991. Nuclear DNA amounts in angiosperms. *Philosophical Transactions of the Royal Society of Londan*, Series B, 334: 309-345.

Caetano-Anolles, G.B. and Gresshoff, P.M. 1991. DNA amplification fingerprinting using very short arbitrary oligonucleotides primers. *Bio/Technology*, 9: 553-557.

Cho, Y.G., Blair, M.W., Panaud O., and McCouch, S. 1996. Cloning and mapping of variety-specific rice genomic DNA sequences : amplified fragment length polymorphisms (AFLP) from silver-stained polyacrylamide gels. *Genome*, 39: 373-378.

Cox, T.S., Kiang, Y.T., Gorman, M.B., and Rodgers, D.M. 1985. Relationships between coefficient of parentage and genetic similarity indices in the soybean. *Crop Science*, 25: 529-532.

de Vienne D. and Santoni, S. 1998. Les principales sources de marqueurs moléculaires. In: *Les Marqueurs Moléculaires en Génétique et Biotechnologies Végétales*. D. de Vienne, ed., paris, France, Inra, 200 p.

Dillmann, C., Bar-Hen, A., Guerin, D., Charcosset, A., and Murigneux, A. 1997. Comparison of RFLP and morphological distances between maize, *Zea mays* L., inbred lines: consequences for germplasm protection purposes. *Theoretical and Applied Genetics*, 95: 92-102.

Dolezel, J. 1991. Flow cytometric analysis of nuclear DNA content in higher plants. *Phytochemical Analysis*, 2: 143-154.

Dos Santos, J.B., Nienhuis, J., Skroch, P., Tivang, J., and Slocum, M.K. 1994. Comparison of RAPD and RFLP genetic markers in determining genetic similarity among *Brassica oleracea* L. genotypes. *Theoretical and Applied Genetics*, 87: 909-915.

Dubreuil, P. and Charcosset, A. 1998. Genetic diversity within and among maize populations: a comparison between isozyme and nuclear RFLP loci. *Theoretical and Applied Genetics*, 96: 577-587.

Efron, B. and Tibshirani, R. 1986. Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statistical Science*, 1: 54-77.

Engqvist, G.M. and Becker, H.C. 1994. Genetic diversity for allozymes, RFLPs and RAPDs in resynthesized repe. In: *Biometrics in Plant Breeding:*

Applications of Molecular Markers; IXth meetimg of the EUCARPIA section biometrics in plant breeding, J. Jansen, ed., Wageningen, Netherlands, EUCARPIA.

Gerdes, J.T. and Tracy, W.F. 1994. Diversity of historically important sweet corn inbreds as estimated by RFLPs, morphology, isozymes and pedigree. *Crop Science*, 34: 26-33.

Graner, A., Ludwig, W.F., and Melchinger, A.E. 1994. Relationships among European barley germplasm. 2. Comparison of RFLP and pedigree data. *Crop Science*, 34: 1199-1205.

Hallden, C., Hansen, M., Nilsson, N.O., Hjerdin, A., and Sall, T. 1996. Competition as a source of errors in RAPD analysis. *Theoretical and Applied Genetics*, 93: 1185-1192.

Henry, A.M. and Damerval, C. 1997. High rates of polymorphism and recombination at the *Opaque-2* locus in cultivated maize. *Molecular and General Genetics*, 256: 147-157.

Heun, M., Murpy, J.P., and Philips, T.D. 1994. A comparison of RAPD and isozyme analyses for determining the genetic relationships among *Avena sterilis* L. accessions. *Theoretical and Applied Genetics*, 87: 689-696.

Inomata, N., Shibata, H., Okuyama, E., and Yamazaki, T. 1995. Evolutionary relationships and sequence variation of α-amylase variants encoded by duplicated genes in the *Amy* locus of *Drosophila melanogaster. Genetics*, 141: 237-244.

Jarne, P. and Lagoda, J.L. 1996. Microsatellites, from molecules to populations and back. *Tree*, 11: 424-429.

Jones, C.J., Edwards, K.J., Castaglione, S., Winfield, M.O., Sala, F., van de Wile, 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., and Karp, A. 1997. Reproducibility testing of RAPD, AFLP and SSR markers in plants by a network of European laboratories. *Molecular Breeding*, 3: 381-390.

Karp, A., lsaac, P.G., and Ingram, D.S. 1998. *Molecular Tools for Screening Biodiversity*. Londres, Royaume-Uni, Chapman and Hall, 498 p.

Karp, A., Kresovich, S., Bhat, K.V., Ayad, W.G., and Hodgkin, T. 1997. Molecular tools in plant genetic resources conservation: a guide to the technologies. Rome, IPGRI, Technical Bulletin no. 2, 47 p.

Karp, A., Seberg, O., and Buiatti, M. 1996. Molecular techiques in the assessment of botanical diversity. *Annals of Botany*, 78: 143-149.

Kremer, A. 1998. Les marqueurs moléculaires en génétique des populations. In: *Les Marqueurs Moléculaires en Génétique et Biotechnologies Végétales.* D. de Vienne ed., Paris, Inra, 200 p.

Langridge, U., Schwall, M., and Langridge, P. 1991. Squashes of plant tissue as substrate for PCR. *Nucleic Acids Research*, 19: 6954.

Liu, Z. and Furnier, G.R. 1993. Comparison of allozyme, RFLP, and RAPD markers for revealing genetic variation within and between trembling aspen and bigtooth aspen. *Theoretical and Applied Genetics*, 87: 97-105.

Lu, J., Knox, M.R., Ambrose, M.J., Brown, J.K.M., and Ellis, T.H.N. 1996. Comparative analysis of genetic diversity in pea assessed by RFLP- and PCR-based methods. *Theoretical and Applied Genetics*, 93: 1103-1111.

Lynch, M. and Milligan, B.G. 1994. Analysis of population genetic structure with RAPD markers. *Molecular Ecology*, 3: 91-99.

Messmer, M.M., Melchinger, A.E., Lee, M., Woodman, W.L., Lee, E.A. and Lamkey, K.R. 1991. Genetic diversity among progenitors and elite lines from the Iowa stiff stalk synthetic (BSSS) maize population: Comparison of allozyme and RFLP data. *Theoretical and Applied Genetics*, 83: 97-107.

Miller, R.G. 1974. The jackknife: a review. *Biometrika*, 61: 1-15.

Mitchell, M.J. and Quiros, C. 1992. Genetic diversity at isozyme and RFLP loci in *Brassica campestris* as related to crop type and geographical origin. *Theoretical and Applied Genetics*, 83: 783-790.

Mohan, M., Nair, S., Bhagwat, A., Krishna, T.G., Yano, M., Bhatia, C.R., and Sasaki, T. 1997. Genome mapping, molecular markers and marker-assisted selection in crop plants. *Molecular Breeding*, 3: 87-103.

Morgante, M., Rafalski, A., Biddle, P., Tingey, S., and Olivieri, A.M. 1994. Genetic mapping and variability of seven soybean simple sequence repeat loci. *Genome*, 37: 763-769.

N'goran, J.A.K., Laurent, V., Risterucci, A.M., and Lanaud, C. 1994. Comparative genetic diversity studies of *Theobroma cacao* L. using RFLP and RAPD markers. *Heredity*, 73: 589-597.

O'Donoughue, L.S., Souza, E., Tanksley, S.D., and Sorrells, M.E. 1994. Relationships among North American oat cultivars based on restriction fragment length polymorphisms. *Crop Science*, 34: 1251-1258.

Olufowote, J.O., Xu, Y., Chen, X., Park, W.D., Beachell, H.M., Dilday, R.H., Goto, M., and McCouch, S.R. 1997. Comparative evaluation of within-cultivar variation of rice (*Oryza sativa* L.) using microsatellite and RFLP markers. *Genome*, 40: 370-378.

Orti, G., Pearse, D.E., and Avise, J.C. 1997. Phylogenetic assessment of length variation at a microsatellite locus. *Proceedings of the National Academy of Sciences of the United States of America*, 94: 10745-10749.

Peakall, R., Smouse, P.E., and Huff, D.R. 1995. Evolutionary implications of allozyme and RAPD variation in diploid populations of dioecious buffalograss *Buchloe dactyloides*. *Molecular Ecology*, 4: 135-147.

Perrier, X., Flori, A., and Bonnot, F. 1999. Les méthodes d'analyse des données. In: *Diversité Génétique des Plantes Tropicales Cultivées.* P. Hamon et al. eds., Montpellier, France, Cirad, collection Repéres, pp. 43-76.

Plaschke, J., Ganal, M.W., and Röder, M.S. 1995. Detection of genetic diversity in closely related bread wheat using microsatellites markers. *Theoretical and Applied Genetics,* 91: 1001-1007.

Powell, W., Morgante, M., Andre, C., Hanafey, M., Vogel, J., Tingey, S., and Rafalski, A. 1996. The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. *Molecular Breeding,* 2: 225-238.

Prabhu, R.R., Webb, D., Jessen, H., Luk, S., Smith, S., and Gresshoff, P.M., 1997. Genetic relatedness among soybean genotypes using DNA amplification fingerprinting (DAF), RFLP, and pedigree. *Crop Science,* 37: 1590-1595.

Ragot, M. and Hoisington, D.A. 1993. Molecular markers for plant breeding: comparisons of RFLP and RAPD genotyping costs. *Theoretical and Applied Genetics,* 86: 975-984.

Risterucci, A.M. 1997. Coût des techniques de marquage moléculaire. In: *Utilisation des Marqueurs Moléculaies en Génétique et en Amélioration des Plantes.* Training seminar. Montpellier, France, Cirad-Biotrop.

Rus-Kortekaas, W., Smulders, M.J.M., Arens, P., and 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.

Russell, J.R., Fuller, J.D., MaCaulay, M. Hatz, B.G., Jahoor, A., Powell, W., and Waugh, R., 1997. Direct comparison of levels of genetic variation among barley accessions detected by RFLPs, AFLPs, SSRs, and RAPDs. *Theoretical and Applied Genetics,* 95: 714-722.

Saghai-Maroof, M.A., Biyashev, R.M., Yang, G.P., Zhang, Q., and Allard, R.W. 1994. Extraordinarily polymorphic microsatellite DNA in barley: species diversity, chromosomal locations, and population dynamics. *Proceedings of the National Academy of Sciences of the United States of America,* 91: 5466-5470.

Salimath, S.S., de Oliveira, O.A.C., Godwin, I.D., and Bennetzen, J.L. 1995. Assessment of genome origins and genetic deversity in the genus *Eleusine* with DNA markers. *Genome,* 38: 757-763.

Santoni, S. 1996. Les marqueurs moléculaires utilisables en amélioration des plantes. *Le Sélectionneur Français,* 46: 3-18.

Taramino, G. and Tingey, S. 1996. Simple sequence repeats for germplasm analysis and mapping in maize. *Genome,* 39: 277-287.

Thormann, C.E., Ferreira, M.E., Camargo, L.E.A., Tivang, J.G., and Osborn, T.C. 1994. Comparison of RFLP and RAPD markers to estimating genetic relationships within and among cruciferous species. *Theoretical and Applied Genetics*, 88: 973-980.

Tinker, N.A., Fortin, M.G., and Mather, D.E. 1993. Random amplified polymorphic DNA and pedigree relationships in spring barley. *Theoretical and Applied Genetics*, 85: 976-984.

van Deyne, A.E., Sorello, M.E., Park, W.D., Ayres, N.M., Fu, H., Cartinhour, S.W., Paul, E., and McCouch S.R. 1998. Anchor probes for comparative mapping of grass genera. *Theoretical and Applied Genetics*, 97: 356-369.

Vos, P., Hogers, R., Bleeker, M., Reijans, M., van de Lee, T., Hornes, M., Frijters, A., Pot, J., Peleman, J., Kuiper, M., and Zabeau, M. 1995. AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Research*, 23: 4407-4414.

Weir, B.S. 1990. *Genetics Data Analysis, Methods for Discrete Population Genetic Data.* Sunderland, USA, Sinaur Associates.

Weising, K., Nybom, H., Wolff, K., and Meyer, W. 1995. DNA fingerprinting in plants and fungi. London, CRC Press.

Williams, J.G.K., Kubelik, A.R., Livak, K.J., Rafalski, J.A., and Tingey, S.V. 1990. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Research*, 18: 6531-6535.

Wu, K.S. and Tanksley, S.D. 1993. Abundance, polymorphism and genetic mapping of microsatellites in rice. *Molecular and General Genetics*, 241: 225-235.

Yang, W., de Oliveira, A.C., Godwin, I., Schertz, K., and Bennetzen, J.L. 1996. Comparisòn of DNA marker technologies in characterizing plant genome diversity: variability in Chinese sorghums. *Crop Science*, 36: 1669-1676.

Zhang, Q., Sanghi-Maroof, M.A., and Kleinhofs, A. 1993. Comparative diversity analysis of RFLPs and isozymes within and among populations of *Hordeum vulgare* subsp. *spontaneum*. *Genetics*, 134: 909-916.

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