in: INIBAP annual report 1996. INIBAP: Montpellier (FRA), 1997. p. 24-28

Molecular marker assisted analysis of the *Musa* genome complex

Authors in alphabetical order:
Ford-Lloyd B.V., University of Birmingham, U.K.
Kaemmer D., J.W. Goethe University, Frankfurt/Main, Germany
Kahl G., J.W. Goethe University, Frankfurt/Main, Germany
Lagoda P.J.L., CIRAD-GERDAT-BIOTROP, Montpellier, France

Introduction

Since Mendel (1866) started to formalize the empirical knowledge of farmers and breeders into what later became known as genetics, the search for "markers" has increased our knowledge of the genetic basis of inheritance.

Mendel used traits such as pea shape and colour in order to trace the hereditary process of segregation and assortment. Thus, while in the past genetic traits were identified by their visible phenotype, i.e. the **genetic markers**[1] are apparent to the naked eye, modern molecular markers are not visible to the naked eye. **Molecular markers** relate directly to the plant's genotype rather than its phenotype and are based on differences in the nucleotide sequences of the DNA in different individuals. Molecular markers have greater resolution than phenotypic markers and, as they are based on physical DNA assays, require more or less sophisticated technology.

Genetic markers greatly facilitate the selection of parents and assist in the development of new improvement strategies based on marker-assisted breeding. As molecular markers can be used at early stages of plant development, they have another big advantage over most phenotypic markers.

The ability to distinguish or identify one individual or group of individuals from another has applications in many fields, and has now been made much more efficient by the use of molecular techniques. Particularly within cultivated plants, profiling of individual genotypes can assist in gene mapping, variety identification and discrimination between species of plants. Low levels of genetic variation which are induced by *in vitro* manipulation of plants can also be identified and measured. The same methods of fingerprinting can be used to study the amount of variation in germplasm of crop plants, and to make comparisons between different accessions or groups of accessions within collections to assist in the management and future use of the conserved material.

The following overview will mostly focus on DNA marker techniques that will be discussed with respect to their potential in gaining insight into the *Musa* genome.

Non-DNA molecular markers

Polyphenol diversification and isozyme polymorphism analyses (Horry 1993; Jarret & Litz 1986) were the first step into the realm of molecular markers. The chemical data supported the model of a hierarchical organization, reflecting different evolutionary levels in the *Musaceae*. The analysis of cultivar accessions suggested that domestication did occur at many evolutionary stages during differentiation of the subspecies of *M. acuminata*. Isozyme polymorphism corroborated, at the molecular level, the previously suspected reproductive isolation between *Musa* species.

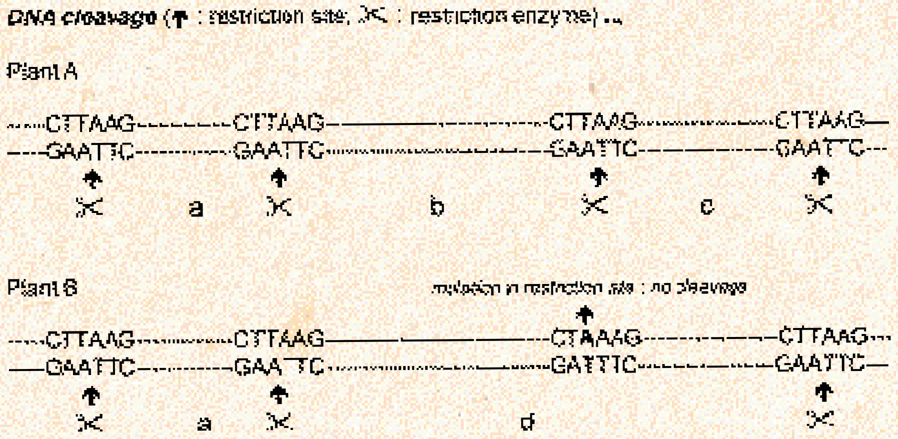
Hybridization-based DNA marker techniques

Restriction Fragment Length Polymorphism (RFLP)

The genome of any living organism is covered with very specific nucleotide motifs called "restriction sites" (or restriction enzyme recognition sites). Specific endonucleases attach to these sites and "cut" the DNA strands to produce DNA fragments. Mutations affecting these sites may destroy (or create) the recognition sequence by causing a point mutation of only one nucleotide at a given locus. The result is an altered fragment size at this locus which can be detected using DNA probes (restriction fragment length polymorphism - RFLP). Thus the analysis of a number of **RFLPs** gives a genomic profile of identified loci of the individual being studied (Figure 1).

---

[1] Editor's note: a genetic marker is a mutation in a gene of known location and effect, which facilitates the analysis of its inheritance or that of a linked gene (Elsevier's dictionary of plant genetic resources)



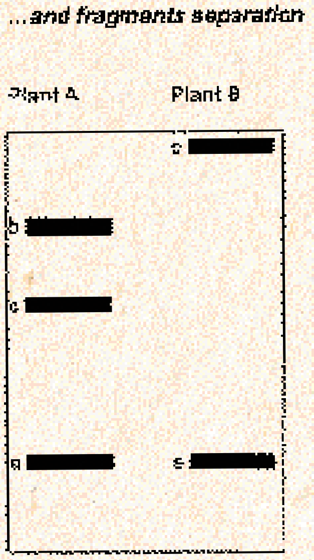


Figure 1. **RFLP**: DNA can be cleaved by restriction enzymes into fragments that can be separated by gel electrophoresis

RFLP markers, which are normally co-dominant, have been used to confirm the *Musa* classification, and to amend the genome formula and sub-species/subgroup classification of some varieties. They have been used as markers to determine the transmission of cytoplasmic DNA, and subspecies specific alleles were described, providing a powerful tool for phylogeny studies in *Musa* (Carreel *et al.* 1994). These markers are also the first anchor-markers developed for gene mapping and Quantitative trait loci identification (Faure *et al.* 1993).

RFLP analyses require an appreciable quantity of high-grade genomic DNA to be totally digested by specific endonucleases. The restriction fragments are separated according to size by gel electrophoresis prior to being fixed onto membranes by Southern blot. Finally polymorphism is revealed most often by hybridization of DNA probes to these membranes. This is a standard protocol in well equipped laboratories, but the rather complex procedure is difficult to transfer to less sophisticated laboratories.

Oligonucleotide fingerprinting

In 1985, a new class of DNA markers was detected, variously designated as "**minisatellites**", "variable number of tandem repeat sequences" (VNTR) or "hypervariable regions" (HVR). These sequences are constituted of short nucleotides sequences -5 to 50 base pairs (bp)- tandemly repeated. The number of repeats at a given locus is different between two genomes, which can be detected as polymorphism. An even more polymorphic class where the repeat unit is smaller than 5 bp was subsequently described as "**simple sequence repeat" (SSR),** "microsatellite", "variable number of dinucleotide repeat" (VNDR) or "short tandem repeat" (STR).

DNA profiling of *Musa* accessions is easily carried out using the repeat consensus of minisatellites and/or SSR. The technique of DNA fingerprinting is a derivative of RFLP analysis, but differs from the latter by the type of hybridization probe applied to reveal genetic polymorphisms. To obtain a typical multilocus DNA fingerprint, probes are used which create complex banding patterns by recognising multiple genomic loci simultaneously.

Each of these loci is characterized by more or less regular arrays of tandemly repeated DNA motifs that occur in different numbers at different loci (Figure 2). The category of probes used by oligonucleotide fingerprinting is exemplified by short, synthetic oligonucleotides which are complementary to SSRs.

In-gel hybridization using SSR oligonucleotides

The use of oligonucleotide probes is compatible with an in-gel hybridization approach that is faster and more efficient than conventional Southern blotting (Epplen J.T. 1992). After electrophoretic separation, the DNA fragments are hybridized to radioactively labelled SSR oligonucleotide probes, and detected by autoradiography.

In banana and plantain this method allows the identification of A- and B-genome-specific bands and the classification of *Musa* (sub) species and cultivars (Kaemmer *et al.* 1992 ; Bhat *et al.* 1995a). However, the necessity to use radioactivity will limit this method to advanced laboratories.

Nonradioactive chemiluminescence detection using SSR oligonucleotides

In order to perform nonradioactive hybridization using SSR oligonucleotides it is necessary to transfer the separated DNA fragments on a nylon membrane (Southern blotting) instead of immobilizing them in the agarose gel by gel drying (Bierwerth *et al.* 1992). DNA fragments linked to a membrane are accessible to complex molecules (digoxygenated SSR probes). At the end of a chain reaction, the positions of the probes are detected by emission of visible light.

Chemiluminescence detection can also be used to determine the molecular weight of SSR-containing PCR products amplified using digoxygenated primers, and has successfully been used for *Musa* STMS fragments (see below).

Polymerase chain reaction (PCR)-based DNA marker techniques

All PCR-based DNA scanning techniques can be divided into two major groups, those producing multi-locus DNA fingerprints that can be used as dominant inherited genetic markers (RAPD, DAF, (A)MP-PCR), and those producing mainly single-locus, codominant inherited genetic markers (STMS) (Caetano-Anolles 1996). The second group of DNA markers is most efficient in genetic mapping and population genetic studies and less efficient in cultivar identification and taxonomic studies. DNA fingerprinting techniques are more cost-effective, produce more markers per experiment, but are only partially transposable between different crosses or species.

Randomly Amplified Polymorphic DNA (RAPD) technology

RAPD is a very fast way of obtaining information about genetic variation (Williams *et al.* 1990; Welsh & McClelland 1990). The technique is centred around the polymerase chain reaction (PCR), a technical innovation that has swept through molecular biology laboratories and has been applied in a wide range of DNA-handling situations. The PCR is a method which allows a particular section of

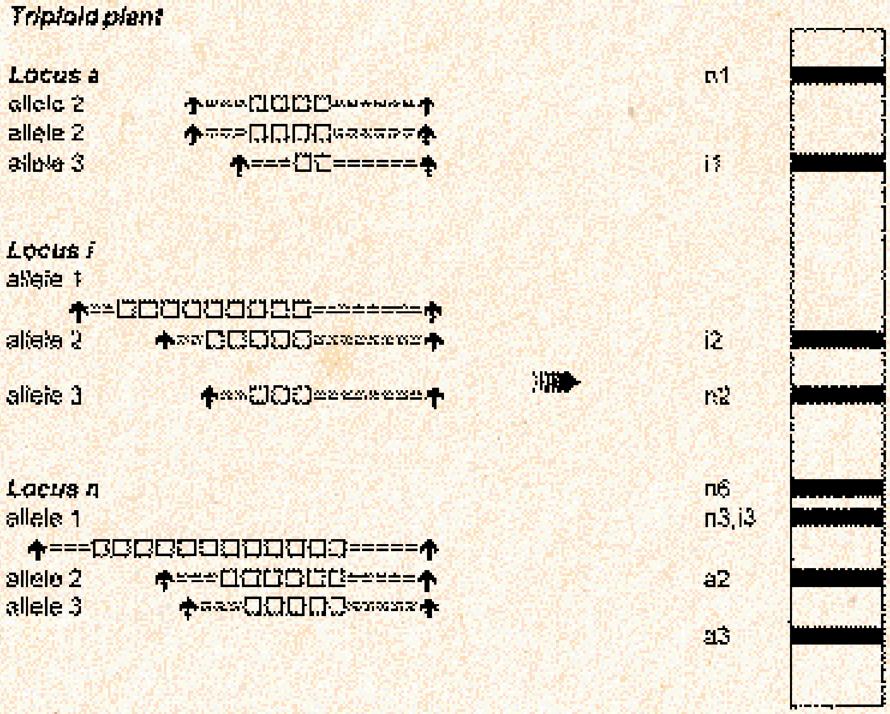


Figure 2. ***Oligonucleotide fingerprinting****: the number of repeats of a simple sequence (□) differs in the different alleles occurring at several locus. The many fragments generated through digestion by restriction enzymes are visualized by hybridization with the same sequence used as a probe after separation on the electrophoresis gel.*

the DNA of an organism to be repetitively copied (amplified) so that eventually it is present in such a high quantity that it can be simply stained by a fluorescent dye and visualized by fluorescent emission when DNA fragments are separated by electrophoresis. In order that copying can occur, short lengths of DNA (usually known as primers) are added to the PCR mix. Differences in banding patterns represent differences between DNA samples, and hence diversity. Furthermore, additional information can be obtained by using different primers (Figure 3).

Although RAPD is a good starting point for learning other molecular marker techniques such as SSR and AFLP, the technique has some disadvantages. Reproducibility in intra- and inter-laboratory comparative assays is still controversial. Recently introduced modifications of the RAPD parameters are presently used to improve the reproducibility, and separation of DNA fragments on high-resolution agarose or polyacrylamide gels with subsequent silver staining will further optimize the results of this screening technique. Nevertheless, because of the rapidity with which RAPD analyses can be carried out without any prior knowledge of the molecular genetic structure of the genome, and the small quantities of DNA which are required, large amounts of banding information can be obtained very quickly.

In *Musa* therefore, there have been a number of applications of the RAPD technique. Discrimination of the genomic composition of *Musa* triploids has been possible, together with varietal identification (Kaemmer *et al*. 1992; Howell *et al*. 1994; Bhat *et al*. 1995b). RAPD has also been used to study patterns of diversity amongst germplasm collections from different geographic regions such as the Indian subcontinent (Bhat and Jarret 1995). Radiation-induced mutants have also been detected (Kaemmer *et al*. 1992), and genetic instability monitored. For instance, markers associated with dwarf offtypes have been

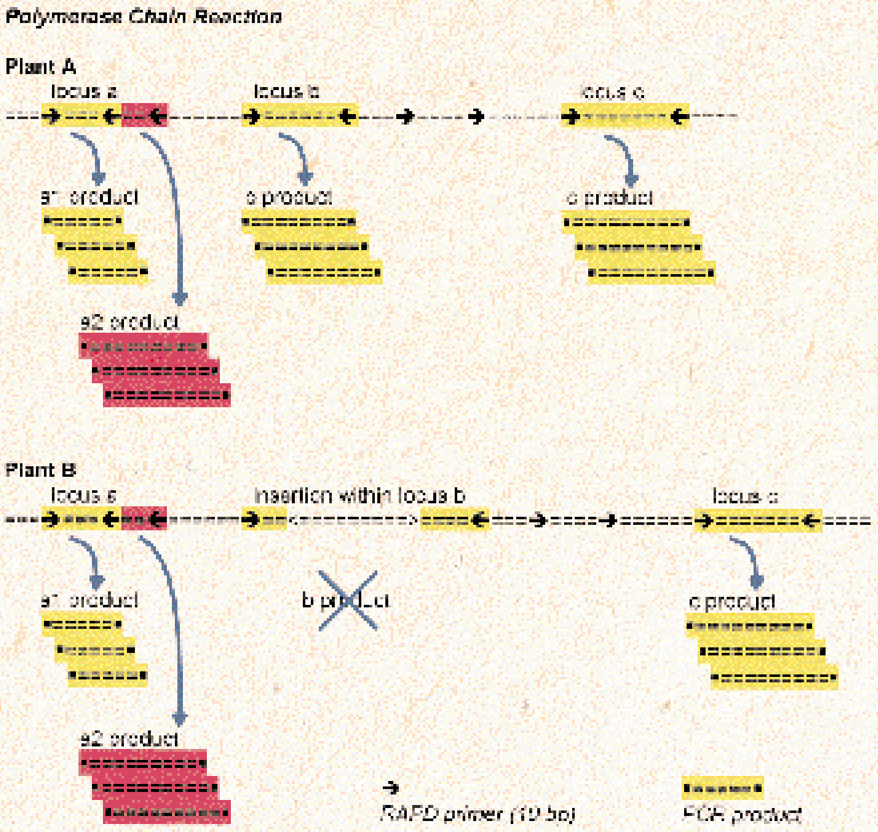


Figure 3. ***RAPD****: Short segments of DNA (primers) links with homologous sequences of the genome. When the distance is short enough, the Polymerase Chain Reaction generates a great number of copies of the different sections (PCR products) that are separated by gel electrophoresis.*

identified (Damasco *et al.* 1996) and general levels of somaclonal variation assessed in tissue culture-derived planting material of plantains (Ford-Lloyd, personal communication) using RAPD markers. Along with other types of markers, RAPD has been used for the development of linkage maps of diploid bananas (Faure *et al.* 1993)

A similar technique is DNA Amplification Fingerprinting (DAF) which differs from the RAPD technique only in the length of the primers used.

Amplified Fragment Length Polymorphism (AFLP)

The **AFLP** technique is an hybrid of RFLP and RAPD. In short, genomic DNA is cleaved by two restriction enzymes. Short double-stranded oligonucleotides (so called adaptors) are then ligated to the restriction fragments. PCR is performed with specific primers directed towards the adaptors. In order to reduce the enormous number of generated fragments, adaptors primers with one or several "anchor" nucleotides at their 3' end are employed. More than 50 fragments can be obtained in one separation. AFLPs are used for identification and genomic mapping thanks to their high polymorphism.

(Anchored)- Microsatellite-Primed PCR ((A)MP-PCR)

Summarising the PCR methods based on the use of single primers which depend on the structure of such single primers, the following methods can be distinguished: (i) 5-15 bp primer (DAF) ; (ii) 9-10 bp primer (RAPD) ; (iii) 18-32 bp primer (AP-PCR) ; (iv) 10-15 bp 5'-or 3'- anchored SSR primers (AMP-PCR). The use of the latter two primer types allows the amplification of inter-SSR regions without (MP-PCR) or including one of two SSR regions. The final population of PCR fragments will consist of the SSR primers at the ends and the inter-SSR region in between.

Since the region between two microsatellites is often variable in length, MP-PCR can be used to identify *Musa* subspecies, species and even cultivars. An additional degree of polymorphism can be introduced by anchoring the SSR primer at its 5'-end.

The PCR techniques described above have in common the ability to rapidly produce a great number of markers which can be scored. These data can be analysed statistically and used as the basis of identification systems, genetic diversity estimations and mapping.

Sequence-Tagged Microsatellite Sites (STMSs)

In contrast to multi-locus PCR, locus-specific PCR uses a pair of primers to detect **STMSs**. These microsatellite-derived markers require only simple and rapid laboratory techniques for their detection (PCR, small genomic DNA quantity and only poor DNA quality needed, possibility of automatization, possible transferability overseas "*in situ*", non radioactive detection, no hybridization, potential for automated data acquisition and analyses) (Figure 4).

Establishing microsatellite loci is expensive, thus development of microsatellite markers is currently limited in plants in general by the number of available published sequence data for most species. However, using already established STMSs is cost effective. The advantages of microsatellite markers over other PCR-based marker systems result from their potential for automated analysis and their codominant nature (as opposed to AFLP or RAPD). Thus all the advantages of PCR can be tapped for the genetic analysis of *Musa*.

Stream-lined lab procedures and protocols on microsatellite markers (i.e. PCR essay kits containing extensively tested, robust, primer pairs) are being developed as well as a high resolution, non-radioactive analysis procedure for VNTR alleles (based on silver stained polyacrylamide gels), in order to promote transfer of microsatellite markers from "biotechnology centres" in the North to "biodiversity centres" in the South (Lagoda & Noyer 1994, Lagoda *et al.* 1995).

Microsatellite markers, like RFLPs, grant direct access to the specific chromosome segments on which they are located. They widen and deepen our understanding of gene structure and expression raising very interesting fundamental questions involving genomic recombination dynamics, DNA repair mechanism and gene regulation.

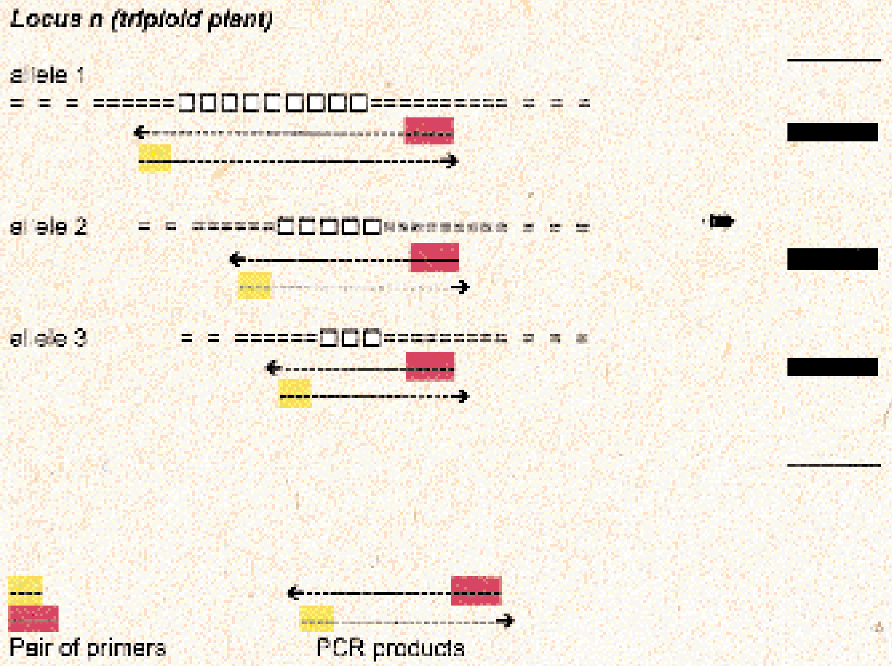


Figure 4. ***STMSs****: a locus specific pair of primers, flanking microsatellite sites (□□□), is used to perform PCR. The length of the primers, at least 20 bp, is chosen in order to guarantee specific PCR products for the studied locus.*

Future Perspectives

In future DNA marker techniques will be used to solve the following problems with substantial impact on banana and plantain breeding:

(i) The unequivocal identification of potential parents of native intraspecific *M. acuminata* hybrids from the Cavendish group (AAA).

(ii) The unequivocal identification of potential parents of native interspecific hybrids of the plantain group (AAB) and the cooking bananas (ABB).

(iii) The reconstruction of plantain phylogeny and the geographic migration of these crop plants.

(iv) The identification and classification of different resistance mechanisms working in the wild banana gene pool.

(v) The establishment of a comprehensive high density linkage map of *M. acuminata*.

(vi) Comparative mapping of *M. acuminata*, *M. balbisiana* and selected other *Musa* and *Ensete* wild species.

This paper gives an overview of the uses of molecular techniques for the analysis of the *Musa* genome. Many of these techniques are also valuable tools for "marker assisted breeding" but these applications are not covered here.

References

Bhat K.V. and R.L. Jarret. 1995. Random amplified polymorphic DNA and genetic diversity in Indian *Musa* germplasm. Genetic Resources and Crop Evolution 42(2): 107-118.

Bhat, K.V, S.R. Bhat, K.P.S Chandel, S. Lakhanpaul and S. Ali. 1995a. DNA fingerprinting of *Musa* cultivars with oligodeoxyribonucleotide probes for simple repeat motifs. Genet. Anal 12:45-51.

Bhat K.V., R.L. Jarret and R.S. Rana. 1995b. DNA profiling of banana and plantain cultivars using random amplified polymorphic DNA (RAPD) and restriction fragment length polymorphism (RFLP) markers. Electrophoresis 16:1736-1745.

Bierwerth S., G. Kahl, F. Weigand and K. Weising. 1992. Oligonucleotide fingerprinting of plant and fungal genomes: a comparison of radioactive, colorigenic and chemiluminescent detection methods. Electrophoresis 13(3): 115-122.

Caetano-Annoles G. 1996. Fingerprinting nucleic acid with arbitrary oligonucleotide primers. Agro Food Industry Hi-Tech 7 : 26-31

Carreel F., S. Fauré, D. Gonzales de León, P.J.L. Lagoda, X. Perrier, F.Bakry, H. Tézenas du Montcel, C. Lanaud and J.P. Horry. 1994. Evaluation de la diversité génétique chez les bananiers diploïdes (*Musa* sp.). Genetics Selection Evolution 26 (Supplement 1): 125s-136s

Damasco O.P., G.C. Graham, R.J. Henry, S.W. Adkins, M.K. Smith and I.D. Godwin.1996. Random amplified polymorphic DNA (RAPD) detection of dwarf off-types in micropropagated Cavendish (Musa spp AAA) bananas. Plant Cell Reports16(1-2): 118-123.

Epplen J.T. 1992. The methodology of multilocus DNA fingerprinting using radioactive or nonradioactive oligonucleotide probes specific for simple repeat. Pp 59-112 *in* Advances in Electrophoresis (Chrambach A., Dunn M.J. & Radola B.J., eds). VHC, Weinheim, Germany.

Faure S., J.L. Noyer, J.P. Horry, F. Bakry, C. Lanaud and D. Gonzalez de León. 1993. A molecular-based linkage map of diploid bananas (*Musa acuminata*). Theor. Appl. Genet. 87:517-526.

Horry J.P. 1993. Taxonomy and genetic diversity of diploid bananas. Pp 35-42 *in* Breeding Banana and Plantain for Resistance to Diseases and Pests (J. Ganry, ed.). CIRAD / INIBAP, Montpellier, France.

Jarret R.L. and R.E. Litz. 1986. Isozymes as genetic markers in bananas and plantains. Euphytica 35:539-547.

Howell E.C., H.J. Newburry, R. Swennen, L.A. Withers and B.V. Ford-Lloyd. 1994. The use of RAPD for identifying and classifying *Musa* germplasm. Genome 37: 328-332.

Kaemmer D., R. Afza, K. Weising, G. Kahl and F.J. Novak. 1992. Oligonucleotide and amplification fingerprinting of wild species and cultivars of banana (*Musa* spp.). Bio/Technology 10:1030-1035.

Lagoda P.J.L. and J.L. Noyer. 1994. Update on *Musa* genome mapping at CIRAD-AGETROP. INFO*MUSA* 3(2): 4

Lagoda P.J.L., J.L. Noyer, D. Dambier, F.C. Baurens and C. Lanaud, 1995. Abundance and distribution of SSR (simple sequence repeats) in the Musaceae family: Microsatellite markers to map the banana genome. Pp. 287-295 in Proceedings of the FAO/IAEA Symposium on Induced mutations and molecular techniques for crop improvement, IAEA-SM-340/016.

Mendel G. 1866. Versuche über Planzenhybriden. Vorhardl. Naturforsch. Ver Brünn 4(3):41.

Welsh, J. and M. McClelland 1990. Fingerprinting genomes using PCR with arbitrary primers. Nucleic Acids Research 18: 7213-7218.

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

Glossary

AFLP	*amplified fragment length polymorphism*
(A)MP-PCR	*(anchored)-microsatellite-primed PCR*
bp	*base pair*
DAF	*DNA amplification fingerprinting*
PCR	*polymerase chain reaction*
QTL	*quantitative trait loci*
RAPD	*random amplified polymorphic DNA*
RFLP	*restriction fragment length polymorphism*
SSR	*simple sequence repeat*
STMS	*sequence-tagged microsatellite sites*
VNDR	*variable number of dinucleotide repeats*
VNTR	*variable number of tandem repeats*