We are IntechOpen, the world’s leading publisher of Open Access books
Built by scientists, for scientists

4,300
Open access books available

116,000
International authors and editors

125M
Downloads

154
Countries delivered to

TOP 1%
Our authors are among the most cited scientists

12.2%
Contributors from top 500 universities

WEB OF SCIENCE™
Selection of our books indexed in the Book Citation Index in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com
Chapter 4

Genetic Diversity in Bananas and Plantains (Musa spp.)

G. Manzo-Sánchez, M.T. Buenrostro-Nava, S. Guzmán-González, M. Orozco-Santos, Muhammad Youssef and Rosa Maria Escobedó-Gracia Medrano

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/59421

1. Introduction

Bananas and plantains belong to the family Musaceae and are cultivated throughout the humid tropics and sub-tropics. This crop is perennial with a faster relative growth rate compared to other fruit crops, while producing fruit all year round. Because of their nutritional value, bananas and plantains are considered the fourth most important crop worldwide after rice, wheat and corn. In many countries of Africa, bananas are considered an important part of the diet; the population in Uganda consumes per capita an average of 191 kg per year [1]. This crop also represents an important source of income for many rural families that work directly or indirectly in this industry. The edible Musa spp. originate from two wild species, Musa acuminata Colla and M. balbisiana Colla, with the A and B genomes, respectively, as well as their hybrids and polyploids.

The genus Musa is of great importance worldwide due to the commercial and nutritional value of cultivated varieties. Morphological data have suggested that Musa is diverse, with well-defined characteristics giving a number of indicators of the genome constitution. However, phenotyping for many physiological characteristics, including biotic and abiotic stress tolerance, particularly under controlled, contained and reproducible conditions, is difficult because of the size of the plants and their long life cycle.

DNA marker technologies have been widely used in banana genetics and diversity analysis, e.g., in taxonomy, cultivar true-to-type assessment and genetic linkage map development. Currently, proteomic analysis is giving rise to new trends in genetic diversity and plant system biology analyses. These approaches will yield detailed insights into the Musa genome and...
provide important genetic data for *Musa* breeders. In this chapter, we discuss the contribution of different DNA and protein-based markers to understanding the genetic diversity of the *Musaceae* family.

2. *Musa* genome

The *Musaceae* family genome is complex, and there are four genomes present, corresponding to the genetic constitutions of the four wild *Musa* species, i.e., *M. acuminata* (A-genome, 2n=2x=22), *M. balbisiana* (B-genome, 2n=2x=22), *M. schizocarpa* (S genome, 2n=2x=22) and *M. textillis* (T genome, 2n=2x=20). Modern bananas and plantain varieties are polyploids, and have a complex genetic structure; because most cultivated bananas are parthenocarpic (generally seedless) they have to be vegetatively propagated [2]. The traits of inflorescence in some accessions of *Musa* spp. are presented in Figure 1.

With the advent of modern DNA sequencing technologies and powerful bioinformatics tools, the sequencing and assembly of genomes for economically important crops and their relatives is becoming more common [3]. Knowing and understanding the genetic make-up of these crops represents a great opportunity to not only elucidate the function of genes of interest, but also to detect regions in the genome that could present polymorphism associated with agronomic traits [4]. This genetic variability is extremely valuable in plant breeding programmes where the selection of individuals with desirable characteristics is carried out.

Significant progress has being made towards gaining a better understanding of the *Musa* genome. Recently, [5] published the first draft of a 523-megabase *M. acuminata* genome and showed 36,542 protein-coding gene models. Transposable elements accounted for almost 50 % of the genome. More recently, a draft of the *M. balbisiana* genome sequence was published [6]. The data in that study showed a great divergence between the A and B genome, which is useful in terms of increasing genetic diversity. The *M. balbisiana* genome was shown to be 79 % smaller than that of *M. acuminata*, but with a highly similar number of predicted functional gene sequences: 36,638. Genomic information sheds light on polymorphisms that can be used in plant breeding programmes, such as the use of single nucleotide polymorphism (SNPs) where there is a high degree of heterozygosity between the A and B genome of one in every 55.9 base pairs. SNP polymorphisms have been successfully used with *Musa* spp. for mapping the genes *phytoene synthase* and *lycopene β-cyclase*, both of which are involved in β-carotene biosynthesis. Genomic information will bring about a strong and significant improvement in the use of DNA-based molecular markers, and make banana and plantain breeding programmes more efficient in improving the crop’s traits.

A large dataset of genomic information is publicly available through different databases, such as The Banana Genome Hub (http://banana-genome.cirad.fr/home), resources from the Global Musa Genomics Consortium (GMGC; http://musagenomics.org/), and the National Center for Biotechnology Information (NCBI; http://www.ncbi.nlm.nih.gov/), among many others.
3. DNA markers

In *Musa* breeding programmes, the use of morphological and cytogenetic markers has played an important role in identifying genes that control discrete traits with simple Mendelian inheritance, and has helped to estimate genome size and diversity [9]. However, a more detailed insight into the genomic polymorphism associated with agronomic traits is required. Molecular DNA-based markers are powerful tools for gaining insights into individual genetic characteristics, and for determining allele frequency. DNA markers were developed first for humans, then applied in plants, and subsequently for the analysis of the banana genome (Figure 2). This feature allows plant breeders to select only those individuals with desirable characteristics and significantly reduce the selection time. In this chapter, we discuss the development and applications of molecular marker technology to improve some of the commercially available *Musa* cultivars and to assess their genetic diversity. The more important advantages and disadvantages of some DNA markers are presented in Table 1.

3.1. Restriction Fragment Length Polymorphism (RFLP)

RFLPs markers are widely used to detect variations in DNA fragment length banding patterns of electrophoresed restriction digests of DNA samples [10]. These variations are mainly due
to the presence of a restriction enzyme cleavage site at one site in the genome of one individual, and the absence of the site in another individual. It can also detect changes in fragment size due to insertions or deletions between the restriction fragments. RFLP is a codominant marker, meaning that it is able to distinguish between homozygotes and heterozygotes. RFLP is robust, easily transferred between laboratories, and requires no prior sequence information for its use.

RFLPs have been found to be useful in *Musa* for constructing genetic linkage maps, characterizing germplasm, phylogenetic analysis [11-15] and analysis of variation in the chloroplast genome [16, 17], and most recently have been linked to polymorphisms in resistance gene analogues [18]. These markers were useful for detecting genetic variations in Indian wild *Musa balbisiana* populations associated with morphotaxonomic characterization clustering of most of the test types of bananas. However, they fail for some specific clusters [19], suggesting the need for more specific types of markers.

![Figure 2. Schematic representation describing the development of DNA marker techniques and the first report on studies of the genetic diversity in *Musa* species.](image)

RFLPs make locus-specific estimations of conserved synteny possible; however, some of the disadvantages are that it is expensive to develop, requires large amounts of DNA, is not possible to automate, unlike other DNA markers [AFLP, diversity array technology (DArT), or variable number tandem repeats (VNTR)], needs a suitable probe library, may require...
radioactive labelling, is laborious and time consuming. The relatively high cost and technically demanding nature of this technique make it inappropriate for routine breeding applications [20]. The use of more specific, PCR-based types of markers overcomes most of the disadvantages associated with RFLPs.

3.2. Random Amplified Polymorphic DNA (RAPD)

With the development of the polymerase chain reaction (PCR) technique, amplifying specific regions of an individual genome became possible, and identifying polymorphisms is made more precise by detecting small nucleotide changes compared to RFLPs. Random amplified polymorphic DNA (RAPD) depends on the PCR and is used as a very fast way to obtain information about genetic variation with a relatively low cost [21]. Some other characteristics are the fact that no prior knowledge of the genome sequence is required; low amounts of DNA template are used; and the advantage of technical simplicity. RAPD assays have proven to be powerful and efficient means of assisting introgression and backcross breeding [22]. But reproducibility is sometimes limited, and reliability depends on the skills of the operator, which is a dominant feature of the marker system.

RAPD has been widely used to distinguish diverse Musa germplasms [19, 23-26], for identification of duplications among accessions in tissue culture germplasm banks and somaclonal variation [27-29], and differentiation of irradiated banana genotypes [30, 31]. In addition, a molecular linkage map has also been developed using a variety of marker systems including RAPD [32]. Specific RAPD markers for the A and B genome of Musa have been identified [33, 34] and full-sib hybrids in plantain breeding populations [35]. These reports clearly demonstrate the potential value of this technique for germplasm characterization and cultivar identification, but give little insight into the value of the assay for molecular breeding. Despite the criticism of the technique, it is still being reported in recent publications [36-38].

Kaemmer [23] was first to report the use of RAPDs for fingerprinting of wild species and cultivars of banana (Musa spp.) by using simple sequence repeats (SSRs) as primers labelled with $^{32}$P. Fingerprinting analysis detected enough genetic variation to help discriminate between most of the 15 banana clones tested. In this study, polymorphic bands were unique to most of the clones tested and helped to distinguish between most of the wild type M. balbisiana (BB) and the M. acuminata (AAA) clones, as well as their hybrids for plantains (AAB) and cooking bananas (ABB); however, this technique generally fails to explain some of the variation within the species. Later, [24] used RAPD markers by adopting a set of nine primers that were shorter (N 10) than the SSRs (N 16) reported by [23], but the authors were able to find enough polymorphism that was unique to each of the nine genotypes representing the AA, AAA, AAB, ABB and BB genomes, and multivariate analysis showed a strong correlation between the polymorphism obtained and the morphological characters used to classify each of the groups.

The use of RAPDs was reported to identify 57 cultivars by using 60 10-mer random primers, where only 49 primers gave consistent results, and the primer OPC-15 (5’-GACGGATCAG-3’) helped to distinguish 55 of the cultivars by producing 24 bands of all tested primers [25]. These markers failed to properly characterize the clones that Gros Michel and Venkel had previously
thought belonged to the *Acuminata* group (AAA). However, chloroplast polymorphism was shown to be identical to *M. balbisiana*. This showed the potential of using RAPDs markers for proper cultivar identification and germplasm classification. Similar applications, but with a variant, were reported later by [17], who used a single primer (5’-TATAGTTAC-CAAGTGGTG-3’) designed from a human *Alu* sequence. This sequence is a member of well-conserved short interspersed nuclear elements in the primate’s genome. Its use in fingerprinting the *Musa* genome produced bands that ranged between 300 bp and 3 kb, indicating that *Alu* related sequences are inverted repeats that are relatively close to each other.

### 3.3. Variable Number Tandem Repeats (VNTR)

VNTR are generated by highly specific PCR amplification and, therefore, should not suffer from the reproducibility problems experienced with RAPD analysis. VNTR are regions of short, tandemly repeated DNA motifs (generally less than or equal to 4 bp), with an overall length in the order of tens of base pairs [39]. VNTR have been reported to be highly abundant and randomly dispersed throughout the genomes of many plant species. Variation in the number of times the motif is repeated is thought to arise through slippage errors during DNA replication.

Furthermore, the isolation of VNTR is becoming increasingly routine with the availability of automated DNA sequencing facilities, along with improved techniques for the construction of genomic libraries enriched for VNTR and improved techniques for the screening of appropriate clones [40], bacterial artificial chromosome (BAC) end-sequences and, recently, the availability of genome sequencing facilities attributed to the discovery of VNTRs [41].

The development and utilization of VNTR in *Musa* research was first reported by [35, 40, 42] and [35]. VNTR have been considered optimum markers in other systems due to their abundance, polymorphism and reliability. VNTR analysis has been shown to detect a high level of polymorphism between individuals of *Musa* breeding populations [35], and used for development of a linkage map [18]. Nevertheless, several hundred VNTR markers have been generated in *Musa* [40, 42, 43]. New VNTR loci were discovered in the *M. acuminata* Calcutta 4 BAC end-sequence [41] from five fully sequenced consensi datasets, with validation for polymorphism conducted on genotypes contrasting in host plant resistance to Sigatoka disease [44]. Recently, [45] used a transcriptome database to design primers that were able to distinguish 32 VNTR and 119 target region amplified polymorphism (TRAP) alleles in 14 diploid *Musa* accessions.

### 3.4. Inter Simple Sequence Repeats (ISSR)

The ISSR technique developed by [46] does not require the knowledge of flanking sequences and has wide applications for all organisms, regardless of the availability of information about their genome sequence. They have also proved to be simple, fast, cost effective and versatile sets of markers for repeatable amplification of DNA sequences using single primers. As for the disadvantages, the homology of the bands is uncertain, and because they are dominant
markers, they do not allow the calculation of certain parameters, requiring that heterozygous be distinguished from homozygous dominance.

ISSR and RAPD were used to determine genetic stability of three economically important micropropagated banana (Musa spp.) cultivars. The results showed that ISSR detected more polymorphism than RAPD [27]. Similarly, ISSR were used for detection of genetic uniformity of micropropagated plantlets [47] and for screening in vitro mutagenesis and variance [37]. Another study reported the use of ISSR to assess the genetic diversity and classification of 27 wild banana accessions collected in Guangxi, China. The results showed that the collected germplasm was derived from diverse origins and evolutionary paths of banana in Guangxi [48]. ISSR were employed for molecular assessment of genetic identity and genetic stability in banana cultivars [49]. Recently, ISSR were used to analyse the pattern of genetic variation and differentiation in 32 individuals along with two reference samples of wild Musa, which corresponded to three populations across the biodiversity-rich hot-spot of the southern Western Ghats of India [50].

3.5. Sequence-Related Amplified Polymorphism (SRAP)

The sequence-related amplified polymorphism (SRAP) technique is a simple and efficient marker system that can be adapted for a variety of purposes, including map construction, gene tagging, genomic and cDNA fingerprinting, and map-based cloning. It has several advantages over other molecular marker systems, such as simplicity, reasonable throughput rate, disclosure of numerous codominant markers, ease of isolation of polymorphic bands for sequencing, and most importantly, the targeting of open reading frames (ORFs) [51].

The SRAP marker has been adopted recently for the assessment of genetic diversity and relationships in Musa. In this regard, the study of Thailand’s wild landraces and cultivars of M. acuminata (A genome), M. balbisiana (B genome) and plantains using SRAP and RAPD has shown that the former marker was more efficient for detecting differences among closer cultivars in the same group, and the BB banana accessions were clustered separately from the AA banana accessions [52].

In another study, SRAP and AFLP were used to study 40 Musa accessions in a core collection being established in Mexico, which includes commercial cultivars and wild species of interest for genetic enhancement [7]. In addition to its practical simplicity, SRAP exhibited approximately three times more specific and unique bands than AFLP. Furthermore, SRAP was demonstrated to be a proficient tool for discriminating among M. acuminata, M. balbisiana and M. schizocarpa in the Musa section, as well as between plantains and cooking bananas within triploid cultivars. The six dessert banana cultivars used clustered according to their subgroup, i.e., Cavendish, Ibota and Gros Michel. Moreover, unique and specific bands were clearly recognized for each of seven subspecies of the acuminata complex, i.e., microcarpa, malaccensis, zebrina, banksii, truncata and burmanica-burmanicoides [7].

Moreover, the study of genetic relationships among some banana cultivars from China analysed by SRAP showed a correlation between the cultivars and their region of origin; the cultivars closely clustered into two major clusters according to their genome composition.
Likewise, the genetic data generated by the SRAP marker were reliable in respect to the morphology and agronomic trait classification, indicating the efficiency of SRAP for estimating genetic similarity among banana cultivars and providing a scientific basis for banana genetic and breeding research [53]. More recently, the fluorescently labelled SRAP molecular marker system was used to characterize the genetic variability within 71 accessions of a core collection, including wild species and cultivars of different subgroups [8], which complements previous work from the same collection [7].

The fluorescent SRAP marker information shows that *M. acuminata* subspecies *errans* was gathered with *banksii* in one cluster, while *malaccensis* was separated in a single sub-cluster. This study also found that Pisang Batu, which is known as diploid BB, was clustered with AB hybrids, i.e., Safet Velchi, Kunnan and Kamaramasenge, which supports the previous finding [54], and that *Pisang Batu* possesses a nucleotide polymorphism pattern of AA. Thus, the study suggested that *Pisang Batu* is a mislabelled accession of AAB or AB. Moreover, the SRAP marker system was found to be useful in identifying closely related accessions in the genus *Musa*, and facilitated the recognition of duplicates to be eliminated and clarified from uncertainties or mislabelled banana accessions introduced to the collection.

### 3.6. Amplified Fragment Length Polymorphism (AFLP)

AFLP is a DNA marker based on PCR amplification of selected restricted fragments obtained from the digestion of total genomic DNA or cDNA [55, 56]. It is a robust and reliable molecular technique recently employed in many systematic plant studies. AFLP banding patterns should be treated initially as dominant markers; this makes the information content limited. However, AFLP patterns can be detected as codominant markers in a segregating population when the analysis is applied to large populations [57].

The nature of an individual, whether a homozygote or heterozygote, could be distinguished using software developed on the basis of band intensity. Moreover, AFLP results in a binary band presence–absence matrix profile. In that case, two factors may affect band detection and analysis: the first is that identical bands may correspond to different fragments (homoplasy); the second is that different fragments appear as a single band (collision). An estimation method was reported for solving the effect of these factors [58], in which AFLP was demonstrated as a sampling procedure of fragments, with lengths sampled from a distribution. This study focused on estimation of pairwise genetic similarity, defined as average fraction of common fragments. Levels of polymorphism in *Musa* were shown to be high when analysed using AFLP, indicating that the technique was effective for genetic diversity analysis [59-62]. In this regard, three subspecies were suspected in the *acuminata* complex based on AFLP analysis, dominated by the subspecies *microcarpa*, *malaccensis* and *burmannica* [63].

The relationship between *M. acuminata* and *M. balbisiana* and their relatedness to cultivated bananas has been reported more clearly using AFLP [64]. Furthermore, several primers were selected from the AFLP results that can be easily used to identify A and B genomes within cultivars using a simple PCR. Additionally, AFLP markers were reported as a powerful tool for evaluating genetic polymorphisms and relationships in *Musa*. They also serve for discriminating amongst species with, A, B, S and T genomes within *Musa* species, as well as between
plantains and cooking bananas [7]. Compared to other DNA markers, AFLP was shown to be a more powerful tool than RAPD for assaying genetic polymorphisms, genetic relationships and cultivar identification among the West African plantain [65]. However, SRAP markers were more informative than AFLP in giving a higher number of unique bands specific for certain genotypes [7]. AFLP markers were appropriate for demonstrating 10 markers cosegregating with the presence or absence of banana streak badnavirus infection in *Musa* hybrids [66].

Figure 3. AFLP polymorphism of parental (P, bulk of 25 plants) and somatic embryogenesis regenerated *M. acuminata* AAA, cv. Williams triploid plants (R, bulk of 50), unique bands (arrows). The AFLP primer combinations as described by [69].

On the other hand, the AFLP marker technique was shown to be a good tool for detection of genetic variation in banana organogenesis and somatic embryogenesis-derived plants [67-69]. AFLP techniques have some disadvantages compared to other PCR-based markers, in that
they are technically challenging, time consuming and relatively expensive, while requiring a
number of DNA processes including digestion, ligation and amplification, as well as a complex
staining system. Additionally, a relatively large amount of high-quality DNA is necessary for
complete digestion, which is required to reduce the presence of fake polymorphisms. How‐
erver, microsatellite markers and AFLP analysis is considered to be one of the most suitable
tools for marker-assisted breeding in Musa. DNA markers associated with fruit parthenocarpy,
dwarfism and apical dominance in banana and plantain are being identified using AFLP and
SSR techniques [58, 31].

4. Single Nucleotide Polymorphism (SNP)

Single nucleotide variations in the genome sequence of individuals of a population or species
are known as single nucleotide polymorphisms (SNPs). The development of this technique in
humans demonstrated improvements in sequencing technology and availability of an
increasing number of SNP sequences [70]; this development has made direct analysis of genetic
variation at the DNA sequence level possible in genomes from different organisms [71].
Modern high-throughput DNA sequencing technologies and bioinformatics tools have led to
the discovery that SNPs constitute the most abundant molecular markers in the plant genomes,
which has revolutionized the pace and precision of plant genetic analysis, and the discovery
that SNPs are widely distributed throughout genomes, although their occurrence and distrib‐
ution varies among species [72].

4.1. Diversity Arrays Technology (DArT)

DArTs are attractive approaches to detecting large numbers of genome-specific single
nucleotide polymorphism (SNP) markers [73] and EcoTILLING. In principle, DArT is a DNA
hybridization-based genotyping technology, which enables low-cost whole-genome profiling
of crops without prior sequence information. DArT reduces the complexity of a representa‐
tive sample (such as pooled DNA representing the diversity of Musa) using the principle that the
genomic “representation” contains two types of fragments: constant fragments, found in any
“representation” prepared from a DNA sample from an individual belonging to a given
cultivar or species, and variable (polymorphic) fragments called molecular markers, found in
some but not all of the “representations”. DArT markers are biallelic and may be dominant
(present or absent) or codominant (two doses vs. one dose or absent). However, this technology
has disadvantages when compared to other molecular markers, in that it depends on the
availability of the array, a microarray printer and scanner, and computer infrastructure to
analyse, store and manage the data produced. Despite these disadvantages, the markers are
sequence-ready and, therefore, if sequenced they can be developed for a PCR analysis using
standard electrophoresis.

Sequenced DArT markers have been used with Musa in several studies. In this regard,
approximately 1,500 DArT markers have been developed using a wide array (“metagenome”)
of Musa accessions [74]. These can be attached to BAC contigs, and can simplify the construction
of high-quality physical maps of the banana genome, which is a critical step in a sequencing project. On the other hand, the carotenoid content and genetic variability of some banana accessions were evaluated from the *Musa* germplasm collection held at Embrapa Cassava and Tropical Fruits, Brazil [75]. Forty-two samples were analysed, including diploids, triploids and tetraploids. The molecular analysis performed using 653 DArT markers showed that DArT was an efficient tool and revealed wide-ranging genetic variability in the collected accessions. Furthermore, DArT markers were used to analyse a panel of 168 *Musa* genotypes; thus, the genomic origin of the markers can help to resolve the pedigree of valuable genotypes of unknown origin. A total of 836 markers were identified and used for genotyping. Ten percent of them were specific to the A genome and enabled the targeting of this genome portion in relatedness analysis among diverse ploidy constitutions.

DArT revealed genetic relationships among *Musa* genotypes consistent with those provided by the other marker technologies, but at a significantly higher resolution and speed [76]. Likewise, DArTs were used for the *Musa* framework map that was developed at CIRAD; additionally, 380 of these markers have been used in the construction of the BORLI map, also at CIRAD [18]. In another study, DArT analysis was used to verify the genome constitution of 24 Philippine *Musa* cultivars. Results of the molecular data showed that some DArT markers were specific for the B genome; subsequently, these markers can identify cultivars with B genome regardless of the presence of the A genome. Hence, these markers can be used to establish genome identity of the *Musa* cultivars. Moreover, BB and BBB accessions were separated from AA/AAA and AAB cultivars in a dendrogram based on DArT data [77].

Recently, a molecular marker-based genetic linkage map of two related diploid banana populations with a complex pedigree was achieved [78], using 121 DArT markers accompanied with allele-specific-polymerase chain reactions (AS-PCR) and simple sequence repeats (SSR). The linkage analysis indicated the likely presence of structural rearrangements.

4.2. (Eco) Targeting Induced Local Lesions in Genomes (EcoTILLING)

EcoTILLING is a high-throughput method for the discovery and characterization of SNPs and small insertions/deletions (indels) in genomes [53, 79, 80]. It is an adaptation of the enzymatic mismatch cleavage and fluorescence detection methods originally developed for the targeting induced local lesions in genomes (TILLING) reverse-genetic strategy [79, 81]. The technique was first described for *Arabidopsis* ecotypes (it was therefore named EcoTILLING). It has been used in many organisms due to its accuracy, low-cost and high-throughput routine, and it also serves for the discovery and assessment of genetic diversity. About 700–1,600 bp gene target regions are amplified using gene-specific primers that are fluorescently labelled for EcoTILLING using enzymatic mismatch cleavage. After PCR, samples are denatured and annealed, and heteroduplexed molecules are created through the hybridization of polymorphic amplicons. Mismatched regions in otherwise double-stranded duplex are then cleaved using a crude extract of celery juice containing the single-strand specific nuclease CEL I. Cleaved products are resolved by denaturing polyacrylamide gel electrophoresis (PAGE) and observed by fluorescence detection.
The EcoTILLING method was used for the discovery and characterization of nucleotide polymorphisms in *Musa*, including diploid and polyploid accessions. Over 800 novel alleles were discovered in 80 accessions, indicating that EcoTILLING is a robust and accurate platform for the discovery of polymorphisms in homologous gene targets. The method proved to be valid in identifying two SNPs that might be deleterious for the function of an important gene in phototropism. Using a principle component analysis, it was shown that evaluation of heterozygous SNPs alone was sufficient to discriminate hybrids from non-hybrids and triploid *acuminata* plants from diploids. Thus, a rapid SNP assessment can in some cases replace flow cytometric methods used to differentiate ploidy.

Furthermore, differentiation between *acuminata* and *balbisiana* diploids was adequate to uncover an accidental miss-assignment of an AA type as a BB type by the stock centre. Moreover, a high level of nucleotide diversity in *Musa* accessions was revealed by evaluating the heterozygous polymorphism and haplotype blocks. The authors concluded that EcoTILLING was an accurate and efficient method for the detection and classification of nucleotide polymorphisms in diploid and polyploid banana. It is highly scalable and many applications can be considered, from simple measurements of heterozygosity as a selection criterion in breeding programmes to more nuanced studies of chromosomal inheritance and functional genomics analysis. This strategy can be used to develop hypotheses for inheritance patterns of nucleotide polymorphisms within and between genome types [53].

More recently, single nucleotide polymorphism (SNP) studies for marker discovery of the use of beta carotene (provitamin A) in plantains [82], and SNPs found in the partial sequence of the gene encoding the large sub-units of ADP-glucose pyrophosphorylase, a key enzyme related to starch metabolism, in banana and plantains [83], give important information for new approaches to investigating the wide range of banana germplasm biodiversity and incorporating the information in banana and plantain breeding.

5. Molecular cytogenetic

Using cytogenetics, chromosome studies of humans and several plant species have been going on for more than a century, helping to establish the typical number of chromosomes for each of the species and to assign a chromosome number according to their size and centromere position. Chromosome size and banding pattern helped to identify subchromosomal regions and were associated with some phenotypic characteristics. Chromosome number has also been important for identifying individuals among species. Even though cytogenetic tools have been improved to obtain high-resolution banding patterns for identifying deletions, insertions or translocations [84], it remains a challenge to elucidate the origins of the chromosomes that are involved in chromosome rearrangements. Molecular cytogenetics is adding a set of powerful tools to those already available for studying genome organization, evolution and recombination. This technology can help to identify small changes at the level of the gene, for which several techniques have been developed.
5.1. Fluorescent in situ Hybridization (FISH) to chromosomes

Fluorescent in situ hybridization (FISH) allows hybridization sites to be visualized directly and, moreover, several probes can be simultaneously detected with different fluorochromes, allowing the physical order of the chromosomes to be determined. FISH was used on mitotic chromosomes to localize the physical sites of 18S-5.8S-25S and 5S rRNA genes in *Musa* [85, 86]. A single major intercalary site was observed on the short arm of the nucleolar organizing chromosome in both A and B genomes. Diploid, triploid and tetraploid genotypes showed two, three and four sites, respectively. Heterogeneous *Musa* lines showed different intensity of signals that indicate variation in the number of copies of these genes. In the case of 5S rDNA, eight subterminal sites were observed in Calcutta 4 (AA), while Butohan 2 (BB) had six sites. Triploid lines showed six to nine major sites of 55 rDNA of widely varying intensity, near the limit of detection. The diploid hybrids had five to nine sites of 5S rDNA while the tetraploid hybrid had 11 sites [86].

Additionally, a dual colour FISH showed that in all studied accessions, the satellite chromosomes carrying the 18S-25S loci did not carry the 55 loci [85]. On the other hand, the telomeric sequence was detected as pairs of dots at the ends of all the chromosomes analysed, but no intercalary sequences were seen [85].

Detection of the integration of viral sequences of banana streak badnavirus (BSV) in two metaphase spreads of *Obino l’Ewai* plantain (AAB) was achieved using FISH [87]. Two different BSV sequence locations were revealed in *Obino l’Ewai* chromosomes and a complex arrangement of BSV and *Musa* sequences was shown by probing stretched DNA fibres. In another study, the *monkey* retrotransposon was identified and localized in *Musa* using FISH [88]. Several copies of *monkey* were concentrated in the nucleolar organizer regions and colocalized with rRNA genes. Other copies of *monkey* appear to be dispersed throughout the genome. In addition, in order to increase the number of useful cytogenetic markers for *Musa*, low amounts of repetitive DNA sequences of BAC clones were used as probes for FISH on mitotic metaphase chromosomes [89]. Only one clone gave a single-locus signal on chromosomes of *M. acuminate* cv. Calcutta 4. The clone localized on a chromosome pair that carries a cluster of 55 rRNA genes. The remaining BAC clones gave dispersed FISH signals throughout the genome and/or failed to produce any signal. In addition, 19 BAC clones were subcloned and their ‘low-copy’ subclones were selected, to avoid the excessive hybridization of repetitive DNA sequences. Out of these, one subclone gave a specific signal in secondary constriction on one chromosome pair, and three subclones were localized into centromeric and peri-centromeric regions of all chromosomes. The nucleotide sequence analysis revealed that subclones which localized on different regions of all chromosomes contained short fragments of various repetitive DNA sequences [89].

Furthermore, a modern chromosome map technology known as high-resolution fluorescent in situ hybridization (FISH) was applied in *Musa* species using BAC clone positioning on pachytene chromosomes of Calcutta 4 (*M. acuminata, Eumusa*) and *M. velutina* (*Rodochlamys*). To make cell spread preparations appropriate for FISH, pollen mother cells were digested with pectolytic enzymes and macerated with acetic acid. BAC clones that contain markers for known resistance genes were chosen and hybridized to establish their relative positions on the
two species [90]. Centromeric retrotransposons were detected in banana chromosomes hybridized with MusA1 by FISH. Since all of the banana chromosomes are metacentric or submetacentric, signals were located around the centre of the chromosome, indicating that loci are found near or within the centromere [91]. Recently, the genomic organization and molecular diversity of two main banana DNA satellites were analysed in a set of 19 *Musa* accessions, including representatives of A, B and S genomes and their interspecific hybrids. The two DNA satellites showed a high level of sequence conservation within, and a high homology between *Musa* species. FISH with probes for the satellite DNA sequences, rRNA genes and a single-copy BAC clone 2G17 resulted in characteristic chromosome banding patterns in *M. acuminate* and *M. balbisiana*, which may aid in determining genomic constitution in interspecific hybrids. In addition, the knowledge of *Musa* satellite DNA was improved by increasing the number of cytogenetic markers and the number of individual chromosomes which can be identified in *Musa* [92].

5.2. Genomic *in situ* Hybridization (GISH)

Genomic *in situ* hybridization (GISH) is a powerful tool to differentiate alien chromosomes or chromosome segments from parental species in interspecific hybrids [93]. It has been applied to mitotic chromosomes from many plants resulting from interspecific hybridization. However, application of GISH on meiotic chromosomes is challenging, and has been reported in just a few species. In *Musa*, GISH was successfully applied to differentiate the chromosomes of different genomes on both mitotic and meiotic chromosomes.

The first study that used GISH in *Musa* was with chromosome spreads prepared from root tips. The total genomic DNA of diploid lines AA and BB was able to label the centromeric regions of all 22 chromosomes of the corresponding line. However, the two satellite chromosomes of genome B labelled strongly with genomic A DNA. GISH discriminated between A and B chromosomes in AAB and ABB cultivars. Additionally, it has immense potential for identification of chromosome origin and can be used to characterize cultivars and hybrids produced in *Musa* breeding [94]. In another study [95], GISH was used to determine the exact genome structure of interspecific cultivated clones AAB and ABB. As a notable exception, the clone ‘Pelipita’ (ABB) was found to have eight A and 25 B chromosomes instead of the predicted 11 A and 22 B. In addition, chromosome complement was determined by some clones that could not be classified by phenotypic characteristics and chromosome counts. Moreover, rDNA sites were located in *Musa* species that appeared to be frequently associated with satellites. These sites can be separated from the chromosomes providing a potential source of chromosome counting errors when conventional techniques are used. Furthermore, GISH successfully differentiated the chromosomes of the four known genomes, A, B, S and T, which correspond to the genetic constitutions of wild Eumusa species *M. acuminate*, *M. balbisiana*, *M. schizocarpa* and the Australimusa species, respectively [95, 96].

On the other hand, GISH has been used on meiotic chromosomes in plants. It is quite challenging, and the protocols used are complex and highly variable depending on the species. A method was developed to prepare chromosomes at meiosis metaphase I suitable for GISH in *Musa* [97]. The main challenge encountered was the hardness of the cell wall and the density
of the microsporocyte’s cytoplasm, which hamper the accessibility of the probes to the chromosomes and generate higher levels of background noise. It was clearly demonstrated that interspecific recombinations between *M. acuminata* and *M. balbisiana* chromosomes do occur and may be frequent in triploid hybrids. Recently, conventional cytogenetic and GISH analyses of meiotic chromosomes were used to investigate the pairing of different chromosome sets at diploid and tetraploid levels, and to reveal the chromosome constitution of hybrids derived from crosses involving allotetraploid genotype. At both ploidy levels, the analysis suggested that the newly formed allotetraploid behaves as a segmental allotetraploid. The 11 chromosomes were found as three sets in a tetrasomic pattern, three in a likely disomic pattern and the five remaining sets in an intermediate pattern. In addition, balanced and unbalanced diploid gametes were detected in progenies. The chromosome constitution was more homogenous in pollen than in ovules. The segmental inheritance pattern exhibited by the AABB allotetraploid genotype implies chromosome exchanges between *M. acuminata* and *M. balbisiana* species, and opens new horizons for reciprocal transfer of valuable alleles [97].

5.3. Flow cytometry

Flow cytometry (FCM) protocols have been applied for studying the natural variation in *Musa* nuclear genome size (DNA content) for taxonomic purposes and for checking ploidy among gene bank accessions and breeding materials [98-101]. The literature data suggest that, on average, the A genome of *M. acuminata* and clones with AA genome constitution is around 12% larger than the B genome of *M. balbisiana*, with small intraspecific variation in nuclear DNA found in a number of wild acuminata diploid and parthenocarpic bananas, whereas large variation seems to be exhibited among triploid varieties [102, 99].

The study of genomic composition of *Musa* accessions on a core collection based on ITS (internal transcribed spacer sequences of the nuclear ribosomal DNA) regions and SSR polymorphism, along with assessment of DNA content and ploidy by FCM, has given support to the hypothesis [103] of the occurrence of homologous recombination between A and B genomes, or between *M. acuminata* subspecies genomes, leading to discrepancies in the number of sets or portions from each parental genome [104]. It is worth mentioning that when the former published data sets were compared, clear differences in the genome size of *M. balbisiana* accessions were found, leaving open the question as to the origin of such variation. Basic research works of this type are of great interest to understanding the evolution and domestication of *Musa*, and for betterment of bananas using new approaches.

Research on the genetic stability/instability of *in vitro* somatic embryogenesis (SE) cultures by FCM for polyploid bananas [105, 106], and recently, by FCM and cytological analysis of embryogenic *M. acuminata* ssp. *malaccensis* cell suspension cultures, and of their somatic embryo-derived plantlets [107], adds support to the use of *in vitro* innovative cell biology tools for high-throughput production of clean banana planting materials, the rescue of hybrid true seeds by *in vitro* germination and their clonal propagation through SE, which assist in the acquisition of seedlings after interspecific hybridization and across ploidy to enhance the efficiency of genetic improvement in *Musa*.
Figure 4. Typical histogram of propidium iodide (PI)-labelled nuclei at pre-DNA synthesis (G1), synthesis (S) and post-synthesis (G2) from a triploid (2n=3x=33 chromosome) regenerated *M. acuminata* AAA, cv. Williams plant. The x axis indicates the fluorescence signal intensity (channel number), which stoichiometrically relates to DNA content. Plant material was regenerated by somatic embryogenesis [68].

<table>
<thead>
<tr>
<th>Criterion</th>
<th>RFLP</th>
<th>RAPD</th>
<th>VNTR</th>
<th>ISSR</th>
<th>SRAP</th>
<th>AFLP</th>
<th>SNPs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quantity of DNA required</td>
<td>High</td>
<td>Low</td>
<td>Low/high(^a)</td>
<td>Low</td>
<td>Low</td>
<td>Moderate</td>
<td>Low</td>
</tr>
<tr>
<td>Quantity of information</td>
<td>Low</td>
<td>High</td>
<td>High</td>
<td>High</td>
<td>High</td>
<td>High</td>
<td>Low</td>
</tr>
<tr>
<td>Replicability</td>
<td>High</td>
<td>Variable</td>
<td>High</td>
<td>High</td>
<td>High</td>
<td>High</td>
<td>Moderate</td>
</tr>
<tr>
<td>Resolution of genetic differences</td>
<td>High</td>
<td>Moderate</td>
<td>High</td>
<td>Moderate</td>
<td>High</td>
<td>High</td>
<td>High</td>
</tr>
<tr>
<td>Abundance</td>
<td>High</td>
<td>High</td>
<td>Medium</td>
<td>High</td>
<td>High</td>
<td>High</td>
<td>High</td>
</tr>
<tr>
<td>Locus specificity</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Ease of use and development</td>
<td>Difficult</td>
<td>Easy</td>
<td>Easy</td>
<td>Easy</td>
<td>Easy</td>
<td>Moderate(^b)</td>
<td>Moderate</td>
</tr>
<tr>
<td>Development time</td>
<td>Long</td>
<td>Short</td>
<td>Moderate</td>
<td>Short</td>
<td>Short</td>
<td>Moderate</td>
<td>Moderate</td>
</tr>
<tr>
<td>Development costs</td>
<td>High</td>
<td>Low</td>
<td>Moderate</td>
<td>Low</td>
<td>Low</td>
<td>Moderate</td>
<td>Low</td>
</tr>
<tr>
<td>Operational costs</td>
<td>High</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Moderate</td>
<td>Moderate / High</td>
<td></td>
</tr>
<tr>
<td>Locus specificity</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes/No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Major application</td>
<td>Physical mapping</td>
<td>Gene tagging</td>
<td>Genetic diversity</td>
<td>Gene tagging</td>
<td>Genetic diversity</td>
<td>Gene tagging SNP mapping</td>
<td></td>
</tr>
</tbody>
</table>

\(^a\)VNTR requires a high quantity of DNA when analysed by RFLP, while it needs a low quantity when analysed by PCR.\(^b\)Analysis of AFLP is easy with the help of an automated genotyper, or when using low-resolution agarose gel electrophoresis, but manual polyacrylamide electrophoresis requires a certain amount of experience.

Table 1. Comparison among different molecular markers for various criteria; adapted from [22, 51, 108].
6. Protein polymorphism

Proteomics can be defined as the systematic analysis of proteome, the protein complement of the genome, which deals with information on proteins’ abundance, their variations and modifications, and their interacting partnerships and networks, in order to understand cellular processes in biological systems. Thus, proteomics is important for understanding the molecular mechanisms involved in plant and crop biodiversity, which is a driving force behind speciation, crop domestication and improvement. This is of particular interest in bananas, which are good representatives of a complex allopolyploid and an important fruit crop.

6.1. Proteome analysis

In recent years, several proteomic studies, based on combined use of two-dimensional electrophoresis (2DE) and mass spectrometric methods, have been successfully applied to investigate the effect of osmotic stresses on banana growth and development [109], cold tolerance [110], inter-and intra-cultivar protein polymorphisms [111], the fruit proteome of banana [112, 113], and the proteomic profiling of banana roots in response to *F. oxysporum* [114]. These investigations highlight the value of new sequencing technologies for integrating the biological information of a plant system usually considered a non-model crop, and open new approaches to studying biodiversity, evolution and domestication in *Musa*. In this scenario, the proteomic analysis of shoot meristem changes in the acclimation to sucrose-mediated osmotic stress of two banana varieties uncovered several genotype-specific proteins (isoforms), enzymes of the energy metabolism (e.g., phosphoglyceraldehyde kinase, phosphoglucomutase, UDP-glucose pyrophosphorylase) and stress adaptations (e.g., OSR40-like protein, abscisic stress ripening protein-like protein, ASR) that were associated with the dehydration-tolerant variety [109].

In contrast, comparative quantitative proteomic analysis of plantain (AAB genome) response to cold stress treatments revealed that about 23.3 % of the 3477 total proteins identified were differentially expressed. The largest parts of the expressed proteins were predicted to be involved in the oxidation reduction process (including oxylipin biosynthesis), cellular process, response to stress and primary metabolic process. Interestingly, among the cold-responsive proteins involved in the oxidation reduction process, Cu/Zn SOD (superoxide dismutase), CAT 2 (catalase isozyme 2) and LOX (lipoxygenase) were found to be differentially expressed in the cold-tolerant plantain, in contrast to the cold-susceptible banana [110]. Altogether, the previous works provided clues as to the existence of inter-variety protein polymorphism related to their *acuminata* or *balbisiana* origin, and open new approaches to examining diploid and triploid bananas’ different responses to environmental stress.

Moreover, evidence from the proteome analysis of different triploid banana varieties using 2D electrophoresis revealed the following results: i) principal component analysis (PCA) showed that the principal component PC1 (which explains 39 % of the variance information) was positively correlated with the presence of the B genome; and, ii) the hierarchical clustering revealed that the first level of clustering separates the varieties of genome BBB and ABB composition from both AAB and the two AAA varieties, the second level splits both AAB
varieties from the two AAA varieties and the ABB from the BBB varieties, and the third level divides both AAA varieties and both AAB varieties. Although proteome analysis does not always correspond to the presumed genome formulae, perhaps because following polyploidization new gene copies may undergo modifications allowing functional diversification, in general, the observations at the protein level provide good indications for a more complex genome structure and genomic rearrangement in some banana varieties [111].

7. Conclusion

The limited genetic knowledge of the banana genome and the nature of the crop as a parthenocarpic fruit and a mostly triploid, sterile plant mean that many aspects of breeding and selection that have been possible in other crops cannot be applied in the banana. Several approaches to breeding and selection have been applied in numerous plant species; however, they could not be used in banana due to the unclear genetic knowledge of its genome and the natural characteristics of the crop, including parthenocarpy, ploidy and sterility. Unconventional biotechnological strategies including DNA and protein-based marker techniques have contributed considerably in providing a vast amount of information that helps in understanding the nature of *Musa* genome and its genetic diversity. DNA-based markers were developed and are being used in *Musa*, representing powerful tools to assess the genetic diversity and clarify the individual genetic characteristics and relationships. Researchers should select the best marker for a certain task; however, the recently developed molecular markers were more informative when applied in banana. For instance, SRAPs showed better assessment of the genetic diversity and increased the clarity of genetic relationships among several *Musa* species, subspecies and cultivars when compared with other traditional markers, such as RAPD and AFLP. Moreover, since SRAP is based on the amplification of ORFs, this gives another advantage to this marker over others. Involving recently developed molecular markers, such as the target region amplification polymorphism (TRAP), which is based on the amplification of target EST sequences, as well as the intron targeted amplified polymorphism (ITAP), which is based on the amplification of 3’ widely distributed intron-exon splice junction sequences, could improve the assessment of genetic diversity in *Musa* and refine the genetic relationships among the different sections, species and mislabelled accessions. In addition, despite their relatively high cost, the high-throughput technologies based on SNPs or small-scale indels are efficient alternatives to traditional markers, because of their greater abundance, high polymorphism, ease of measurement and ability to reveal hidden polymorphisms where other methods fail. SNPs also allow easy and unambiguous identification of alleles or haplotypes. A good marker system for polyploid crops should be dosage sensitive and have the ability to distinguish heterozygous genotypes with multiple haplotypes. On the other hand, molecular cytogenetics techniques have played a great role in understanding the *Musa* genome construction, determining genomic constitution of the interspecific hybrids and studying the natural variation in *Musa* nuclear genome size, as well as checking ploidy levels among gene bank accessions and breeding materials. In addition, the knowledge on *Musa* satellite DNA was improved by increasing the number of cytogenetic markers and the number of individual
chromosomes which can be identified in *Musa*. The complementation of DNA based markers with protein-based ones has completed the image of unreachable variations at DNA level. Several proteomic studies, based on combined use of two-dimensional electrophoresis (2DE) and mass spectrometric methods, have been successfully applied to investigate various approaches in banana, including protein polymorphism, and biotic and abiotic tolerance. The combination of all molecular approaches surveyed and discussed in this chapter can help in the revelation of the genetic diversity in *Musa*. High-throughput technologies based on SNPs or small-scale indels are efficient alternatives for traditional markers (RFLP, RAPD or AFLP), because of their greater abundance, high polymorphism, ease of measurement and ability to reveal hidden polymorphisms where other methods fail. SNPs also allow easy and unambiguous identification of alleles or haplotypes. A good marker system for polyploid crops should be dosage sensitive, and have the ability to distinguish heterozygous genotypes with multiple haplotypes.

**Author details**

G. Manzo-Sánchez1*, M.T. Buenrostro-Nava1, S. Guzmán-González1, M. Orozco-Santos2, Muhammad Youssef3 and Rosa Maria Escobedo-Gracia Medrano4

*Address all correspondence to: gmanzo@ucol.mx

1 School of Agronomy and Biological Sciences, University of Colima; Tecoman, Colima, Mexico
2 INIFAP-Tecoman Experimental Station, Colima, Tecoman, Colima, Mexico
3 Department of Genetics, Faculty of Agriculture, Assiut University, Assiut, Egypt
4 Plant Biochemistry and Molecular Biology Unit, Yucatan Center for Scientific Research, Chuburna de Hidalgo, Merida, Yucatan, Mexico

**References**


Imelda M, Estiati A, Hartati NS. Induction of mutation through gamma radiation in three cultivars of banana. Ann Bogorien 2001; 7 (2) 75–82.


Pillay M, Nwakanma DC, Tenkouano A. Identification of RAPD markers linked to A and B genome sequences in *Musa* L. Genome 2000; 43 (3) 763–767.


[51] Li G, Quiros CF. Sequence-related amplified polymorphism (SRAP), a new marker system based on a simple PCR reaction: Its application to mapping and gene tagging in Brassica. Theor Appl Genet 200; 103(2-3) 455–461.


[71] Soleimani VD, Baum BR, Johnson DA. Efficient validation of single nucleotide polymorphisms in plants by allele-specific PCR, with an example from barley. Plant Mol Biol Rep 2003; 21(3) 281–288.


D’Hont A. Unraveling the genome structure of polyploids using FISH and GISH; examples of sugarcane and banana. Cytogenet Genome Res 2005; 109 (1-3) 27-33.


