Palacký University Olomouc
Faculty of Science
Department of Botany

and

Centre of the Region Haná for Biotechnological and Agricultural Research
Institute of Experimental Botany
Olomouc

Jana Čížková

Physical Mapping and Evolution of Banana Genome (*Musa* spp.)

Ph.D. Thesis

Supervisor: prof. Ing. Jaroslav Doležel, DrSc.

Olomouc
2013

Acknowledgements:

I would like to thank to my supervisor prof. Ing. Jaroslav Doležel, DrSc. for the opportunity to be a member of the Laboratory of Molecular Cytogenetics and Cytometry and the Centre of Plant Structural and Functional Genomics. I am very grateful for his guidance, good advice and kindness. I would like to express my thanks to my second supervisor Mgr. Eva Hřibová, Ph.D. for her continuous support and friendship. Her help and great patience at all times have been invaluable. My sincere thanks also go to my colleagues and friends in the Laboratory of Molecular Cytogenetics and Cytometry and the Centre of Plant Structural and Functional Genomics for making my Ph.D. studies an unforgettable experience. Lastly, I would like to thank to my family for all their love and encouragement.

Declaration:

I hereby declare that I elaborated this Ph.D. thesis independently under the supervision of prof. Ing. Jaroslav Doležel, DrSc. and Mgr. Eva Hřibová, Ph.D., using only information sources refered in the Literature chapter.

.....

This work was supported by the Ministry of Education, Youth and Sports of the Czech Republic and the European Regional Development Fund (Operational Programme Research and Development for Innovations No. ED0007/01/01, Ministry of Education, Youth and Sports of the Czech Republic (grant award No. LG12021), Bioversity International, Montpellier, France (LoA CfL 2010/58, LoA CfL 2011/51, LoA CfL 2012/43) and Internal Grant Agency of Palacký University (grant award No. IGA PrF/2012/001).

BIBLIOGRAPHIC IDENTIFICATION

Author's name: Jana Čížková

Title: Physical Mapping and Evolution of Banana Genome (*Musa* spp.)

Type of Thesis: Ph.D. thesis

Department: Botany

Supervisor: prof. Ing. Jaroslav Doležel, DrSc.

Second Supervisor: Mgr. Eva Hřibová, Ph.D.

The Year of Presentation: 2013

Abstract:

Bananas (*Musa* spp., family *Musaceae*) are one of the top export commodities and an essential nutrition source for millions of people, especially in developing countries of tropical and subtropical regions. Despite the unquestionable importance of this crop, there is a lack of knowledge about the genetic diversity, phylogenetic relationships and the structure, organization and evolution of banana genome. This study employed several molecular approaches with the aim to contribute to the characterization of the genetic diversity and evolution of species belonging to the family *Musaceae*. Cytogenetic mapping of specific sequences on mitotic chromosomes of bananas was used to shed more light on the structure and organization of banana genome.

In the first part of this thesis, analysis of the ITS region of nrDNA was used to study the phylogenetic relationships within the family *Musaceae*. The study provides a plausible picture of the evolution of *Musaceae* species which can be further utilized in improving the classification within the genus. The second part of the work was aimed at molecular analysis and cytogenetic mapping of two main banana DNA satellites. Their genomic organization and molecular diversity was estimated in a set of nineteen *Musa* accessions. A high potential of these satellites as cytogenetic markers was shown. Their use as probes for FISH significantly increases the number of chromosomes which can be identified in *Musa*. In the third part of this thesis, genetic diversity of a set of wild *Musa* species was analysed. The study provides a novel data on nuclear genome size and genomic distribution of rRNA genes in banana and increases the knowledge about the genome structure of wild *Musa* species. For molecular characterization of studied accessions, SSR genotyping

and analysis of the ITS region were used confirming their usefulness for assessing the genetic diversity and evolutionary relationships among banana species.
Keywords: *Musa*, banana, genome, physical cytogenetic mapping, evolution
Number of Pages/Apendices: 191/0
Language: English

BIBLIOGRAFICKÁ IDENTIFIKACE

Jméno: Jana Čížková

Téma: Fyzické mapování a evoluce genomu banánovníku (*Musa* spp.)

Typ práce: Disertační práce

Obor: Botanika

Školitel: prof. Ing. Jaroslav Doležel, DrSc.

Školitel specialista: Mgr. Eva Hřibová, Ph.D.

Rok obhajoby: 2013

Abtrakt:

Banánovník (*Musa* spp, čeleď *Musaceae*) je jednou z nejdůležitějších exportních plodin a nezastupitelný zdroj obživy pro miliony lidí žijících v rozvojových zemích tropických a subtropických oblastí. Navzdory nezpochybnitelné důležitosti této plodiny je zde nedostatek informací o genetické diverzitě, fylogenetických vztazích v rámci čeledi a struktuře, organizaci a evoluci genomu banánovníku. Tato disertační práce si kladla za cíl přispět k charakterizaci genetické diverzity a evoluce druhů čeledi *Musaceae* za pomoci molekulárních metod. Struktura a organizace genomu banánovníku byla studována pomocí cytogenetického mapování specifických sekvencí na mitotické chromozómy.

V první části práce byla pro studium fylogenetických vztahů v rámci čeledi *Musaceae* využita analýza ITS oblasti ribozomální DNA. Tato studie poskytla přesvědčivý pohled na evoluci druhů čeledi *Musaceae*, který může být dále využitelný pro zpřesnění klasifikace jednotlivých druhů. Druhá část této práce byla zaměřena na molekulární analýzu a cytogenetické mapování dvou hlavních DNA satelitů vyskytujících se v genomu banánovníku. V souboru devatenácti zástupců rodu *Musa* byla určena genomická organizace a molekulární diverzita těchto DNA satelitů. Použití těchto satelitních sekvencí jako sond pro FISH výrazně zvýšilo počet chromozómů, které mohou být genomu banánovníku identifikovány, což dokazuje jejich vysoký potenciál jakožto cytogenetických markerů. Třetí část této disertační práce se zabývala charakterizací genetické diverzity planě rostoucích druhů rodu *Musa*. Tato studie přináší nové poznatky o velikosti jaderného genomu a distribuci rRNA genů v genomu banánovníku. Zvyšuje tak znalosti o struktuře genomu planě rostoucích druhů rodu *Musa*. Pro molekulární charakterizaci studovaných položek bylo použito SSR genotypování a analýza ITS oblasti. Tento přístup se ukázal jako

užitečný a vhodný pro stanovení genetické diverzity a evolučních vztahů mezi jednotlivými druhy banánovníku.

Klíčová slova: *Musa*, banánovník, genom, fyzické cytogenetické mapování, evoluce

Počet stran/příloh: 191/0

Jazyk: anglický

CONTENT

1 INTRODUCTION

Bananas (*Musa* spp., family *Musaceae*) are giant perennial herbs growing in tropical and subtropical regions of Asia, Africa, South America and Australia. Being an essential nutrition source for millions of people in developing countries and one of the top export commodities, bananas belong to world's most important fruit crops. Despite socio-economic importance of bananas, there is a lack of knowledge about the structure, organization and evolution of banana genome.

The family *Musaceae* has been assigned to the order Zingiberales and includes three genera *Musa*, *Ensete* and *Musella*. The traditional morpho-taxonomic classification of banana is based on a set of morphological descriptors and basic chromosomes number. This conventional taxonomy has been often questioned and the phylogenetic relationships within the family *Musaceae* remains subject to debate. Different types of molecular markers have been used until now to investigate the diversity of bananas at various taxonomic levels but the taxonomy and phylogenetic relationships still remain unresolved.

The nuclear genome of banana is relatively small and divided into small and morphologically similar chromosomes. The small size and poor staining ability of banana chromosomes complicate cytogenetic studies and identification of chromosomes within a karyotype. The lack of chromosome-specific landmarks and markers complicates studies of karyotype evolution and chromosome behaviour in inter- and intraspecific hybrids.

The aim of this thesis was to shed more light on *Musa* genetic diversity, contribute to the clarification of evolutionary relationships in the family *Musaceae* and to increase the knowledge about the structure and organization of the nuclear genome of banana.

2 LITERATURE REVIEW

2.1 CHARACTERISTICS OF BANANAS

2.1.1 Morphological characteristics

Bananas are giant, monocotyledonous, herbaceous, perennial herbs. The basic structure of the banana plant is outlined in Figure 1. The plant consists of a subterranean stem that bears developing suckers, the root system, the pseudostem, leaves and the inflorescence.

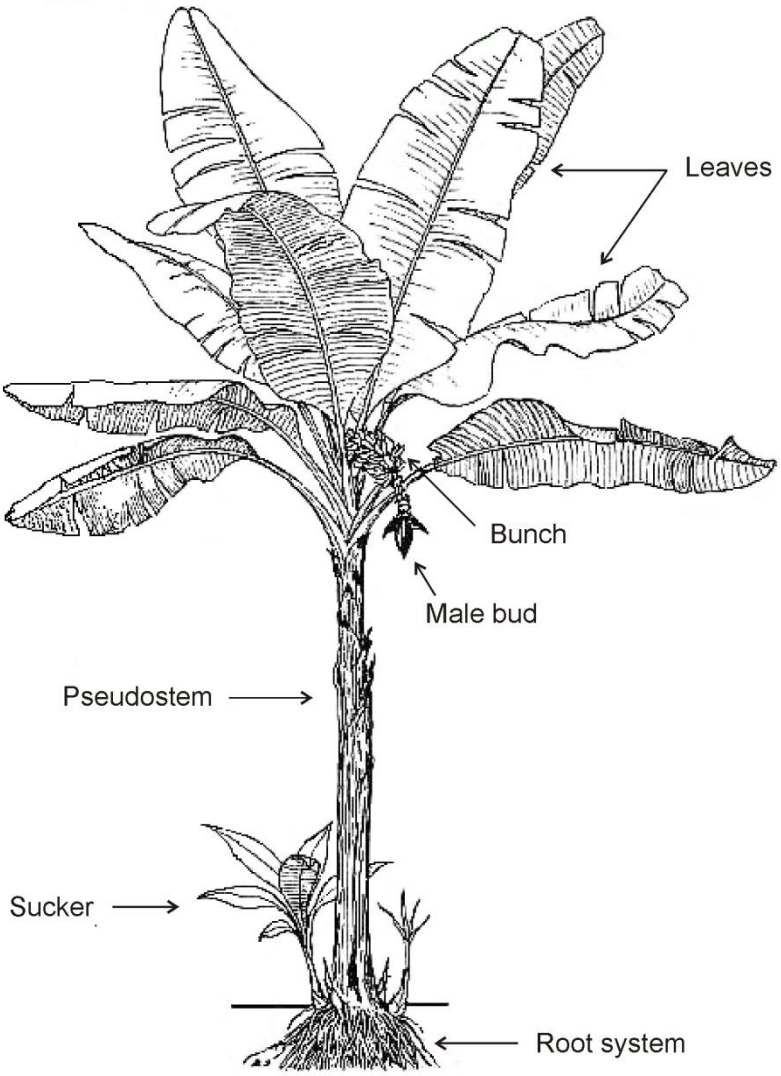


Figure 1: Basic morphology of banana plant (Dahlgren *et al.* 1985)

The true stem of the banana plant (called 'tuberous rhizome') is partly or wholly underground and has extremely short internodes. The rhizome has more or less round shape and is internally differentiated into the central cylinder and cortex. On the top of the rhizome is the apical meristem, which continuously produces new leaves. 180° opposite each leaf on the outer surface of the cortex, a vegetative bud is produced. However, only few of these buds develop into suckers which could regenerate a plant. Suckers grow successively outwards and after a short period of horizontal growth they turn upwards.

The primary seedling root of the banana dies early and is replaced by an adventitious root system immediately. In plants which are established from suckers the root system is adventitious from the beginning. Primary roots, originating from the central cylinder of the rhizome, are white when new and healthy, but later become brown and corky. From each primary root, a system of secondary a tertiary lateral roots develops. Root hairs, which are produced behind the root tip of extending primary and lateral roots, are responsible for most of the water and mineral uptake of the plant. The overall root distribution is influenced by soil type, compaction and drainage. The common horizontal extension of primary roots is 1 to 2 m but may extend up to 5 m of the plant. The vertical root zone depth is very small with about 40% of the root system in the top 10 cm and 85% in the top 30 cm of the soil (Robinson 1996). This makes banana plantations vulnerable to wind damage, and if the plants bear heavy fruit bunches, they may be blown down by strong winds.

The pseudostem can reach a height of 2-8 m in cultivated varieties and even more (10-15 m) in some wild species. The pseudostem is formed by tightly packed overlapping leaf sheaths which grow directly from the top of the rhizome. The leaf sheaths are circular, enclosing the youngest leaf sheaths and the aerial stem. With continuing growth of new leaves in the middle of the pseudostem the free margins of sheaths are forced apart. The physical strength of the pseudostem, necessary for supporting leaves and bunch of fruit is enhanced with sclerotized parenchymatous tissue present in the sheaths that are tightly packed around the aerial stem. The morphological variability of the pseudostem, including length, disposition and coloration, is useful in distinguishing among different sub-groups of bananas.

The distal end of the leaf sheaths narrow into the petiole whose color, length and shape depends on the cultivar. The petiole then becomes the leaf midrib dividing the blade into two lamina halves. New leaves (called cigar leaves) are completely

enrolled and emerge from the middle of the pseudostem. At flowering, there are 10 to 15 functional leaves on the plant, and with leaf senescence the number of leaves drops to 5 to 10 at harvest (Robinson 1996). Each newly emerging leaf is larger than the previous one with the exception of 7-11 leaves before flowering.

The inflorescence is a complex structure composed of a peduncle which bears spirally arranged flower clusters covered by a modified leaf (called bract). Closest to the base of the peduncle there are female flowers followed by hermaphrodite flowers and male flowers at the distal end. Female flowers develop into fruits (berries) which contain seeds (wild bananas), seeds and pulp, or edible pulp only (cultivated clones. The seedless fruits originate without pollination by vegetative parthenocarpy. In the bunch, each cluster of fruits is called hand and the individual fruit is called a finger. Fruit weight varies from a few grams in wild bananas to 800 g in parthenocarpic bananas. The total weight of the bunch widely differs among cultivars. The bunch of the most common export cultivar Cavendish can weigh from 15 to 70 kg (Robinson 1996). Hermaphrodite flowers do not develop into fruits and their persistence depends upon variety. The terminal male bud consists of closely overlapping bracts covering two rows of male flowers. The size, shape and number of fruits as well as the shape of bunch and the colour and shape of male bud bracts are important characters for banana classification.

2.1.2 Taxonomic classification and distribution

Bananas belong to the family *Musaceae*, a member of the order *Zingiberales*. The order *Zingiberales* is a part of the commelinid clade in the Monocots and consists of eight families (*Cannaceae*, *Costaceae*, *Heliconiaceae*, *Lowiaceae*, *Maranthaceae*, *Strelitziaceae*, *Zingiberaceae* and *Musaceae*) of tropical and subtropical perennial plants. The family *Musaceae* originally evolved in South-East Asia and surrounding tropical and subtropical regions. The family is divided into three genera *Musa, Ensete* and *Musella* (Figure 2).

The genus *Ensete* includes monocarpic plants distributed from West Africa to New Guinea. It bears no edible fruits but the starchy corm serves as food in East Africa. *Ensete ventricosum* have been domesticated especially in Ethiopia, where it became an economically important food crop (De Langhe *et al.* 2009; Hildebrand 2007). The monotypic genus *Musella* was established by Wu in 1976 for a species *Musella lasiocarpa*. This rare species grows only in the mountains of South-East

Asia. Quite recently, a new species *Musella splendida* was described in Vietnam (Valmayor and Danh 2002), but its status has not been widely accepted by other researchers.

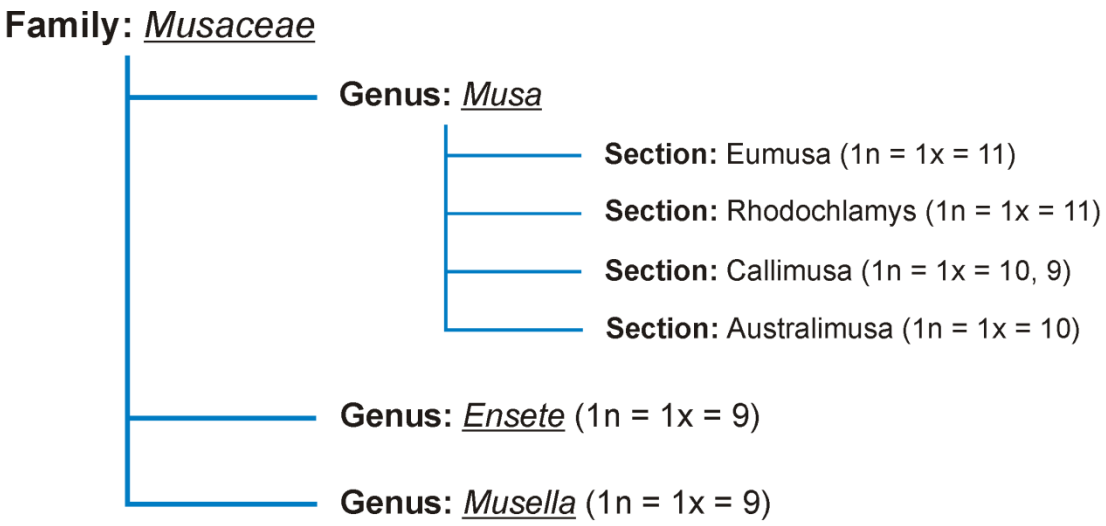


Figure 2: Taxonomic classification of the family *Musaceae*

The most important genus of the family *Musaceae* is the genus *Musa* which includes all edible bananas. This genus, comprising approximately 65 species, is naturally distributed in South-East Asia, stretching from India to Papua New Guinea and including Malaysia and Indonesia. Based on morphological characteristics and basic chromosome number, the genus has traditionally been sub-divided into four sections (Figure 2): Eumusa (2n = 2x = 22), Rhodochlamys (2n = 2x = 22), Callimusa (2n = 2x = 20 or 2n = 2x = 18) and Australimusa (2n = 2x = 20) (Cheesman 1947, Simmonds 1962) with different natural distribution (Figure 3). This classification has been widely accepted and followed by subsequent authors, but its validity has been many times questioned because of problematic classification of newly described species. For example, Argent (1976) created a new section Ingentimusa for *M. ingens* (2n = 2x = 14), which occurs only in Papua New Guinea and is described as a world's largest herb.

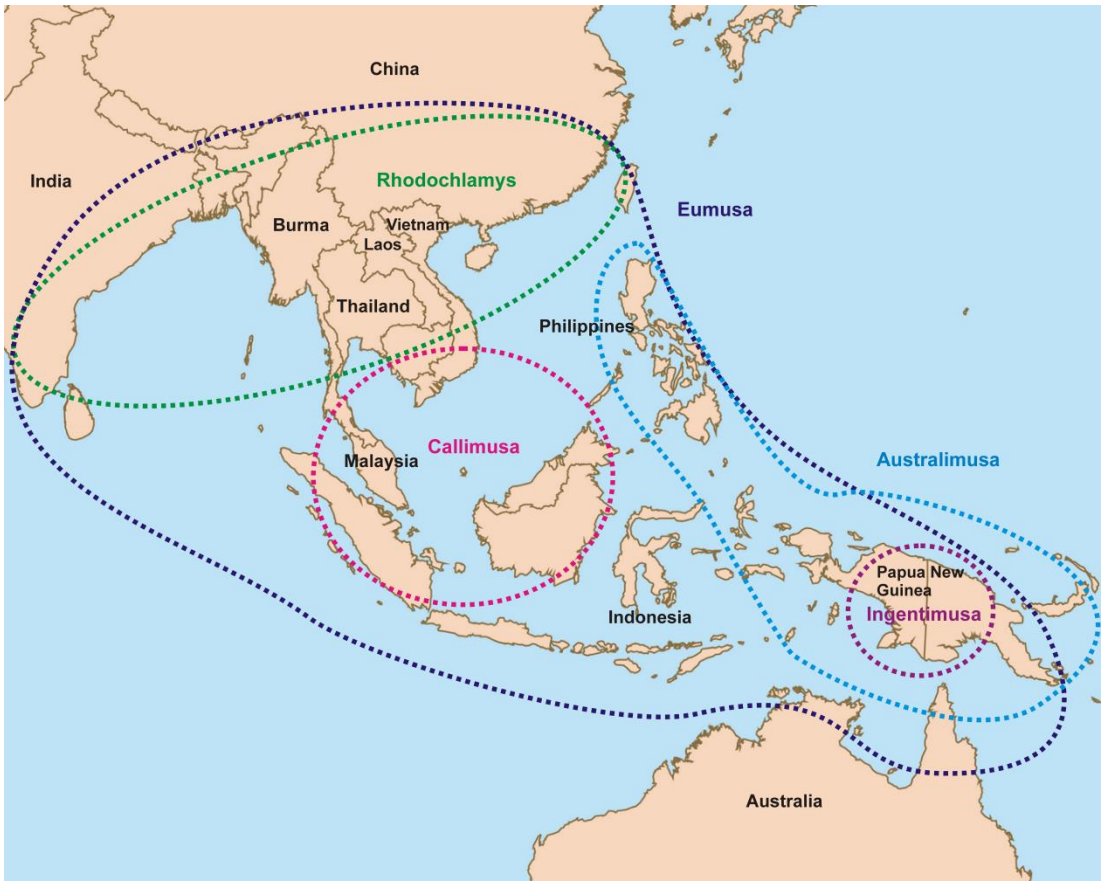


Figure 3: Geographical distribution of the sections Eumusa, Rhodochlamys, Callimusa and Australimusa (According to Simmonds 1962; De Langhe *et al.* 2009).

With the employment of molecular methods the shortcomings of current classification has been revealed even more (Gawel and Jarret 1991; Gawel *et al.* 1992). Wong *et al.* (2002) proposed joining of sections Eumusa and Rhodochlamys into one section and Callimusa and Australimusa into other one based on results of AFLP analysis. Similarity of sections Eumusa and Rhodochlamys and of sections Callimusa and Australimusa was observed by others (Nwakanma *et al.* 2003; Bartoš *et al.* 2005; Risterucci *et al.* 2009). Similarly, status of the genus *Musella* is currently uncertain, because several studies show that *Ensete* and *Musella* make up one clade and suggest joining of the two genera (Li *et al.* 2010; Liu *et al.* 2010). Clarification of the phylogeny and taxonomy of the family *Musaceae* and especially of the genus *Musa* is very important. Information on phylogenetic relationships could be valuable for collection and utilization of genetic resources necessary for further banana improvement.

Most of the currently grown banana cultivars are diploid, triploid or tetraploid clones, which originated from natural intra- and interspecific crosses between two wild diploid species of Eumusa section: *M. acuminata* (A genome) and *M. balbisiana* (B genome). The most common are triploid hybrids which are further differentiated into several groups according to a set of morphological descriptors indicating a degree of genetic inheritance of *M. acuminata* and *M. balbisiana* (Simmonds and Shepherd 1955). Cultivars derived from *M. acuminata* (AAA group) provide sweet dessert bananas, while AAB and ABB groups are usually plantains or cooking bananas characteristic by more starchy fruits (Price 1995).

Interesting difference between the diploid progenitors of common edible bananas is in their natural distribution (Figure 4). While *M. balbisiana* is located only in the northern periphery and part of the Philippines, *M. acuminata* covers almost entire Eumusa area except for the northern region occupied by *M. balbisiana* (De Langhe *et al.* 2009). Intraspecific *M. acuminata* cultivars originated in South-East Asia and the interspecific hybrids developed in the overlapping areas of their distribution.

The morphological variation among *M. acuminata* has been extensively studied for breeding purposes. Within the species, several subspecies have been discriminated based on morphological and geographical criteria. Although the number of accepted subspecies varies from six to nine the most commonly accepted subspecies are *M. acuminata* ssp. *burmannica*, *M. acuminata* ssp. *errans*, *M. acuminata* ssp. *malaccensis*, *M. acuminata* ssp. *microcarpa*, *M. acuminata* ssp. *siamea*, *M. acuminata* ssp. *truncata*, *M. acuminata* ssp. *zebrina* and *M. acuminata* ssp. *banksii* (De Langhe *et al.* 2009) (Figure 4). Contribution of individual subspecies to the origin of current edible cultivars is not clear. Although high level of genetic variability was revealed also in *M. balbisiana* (Ude *et al.* 2002), no subspecies has been defined so far.

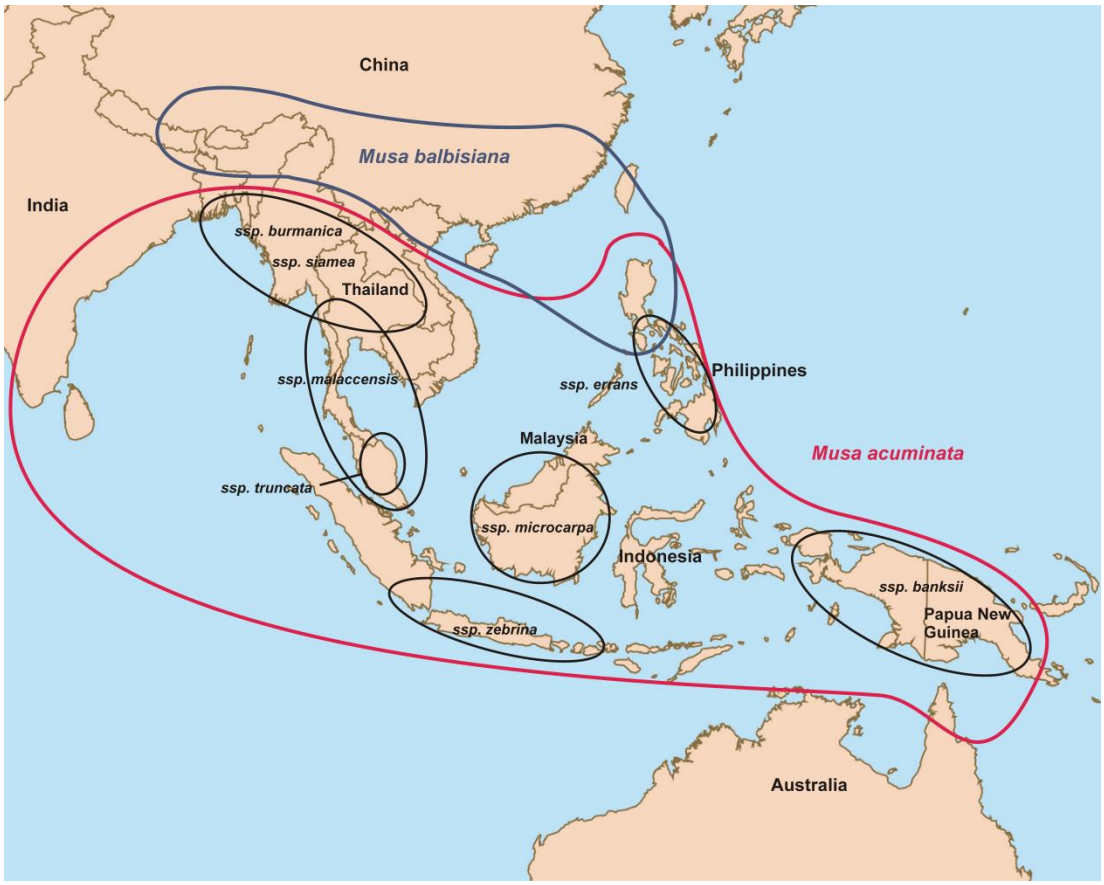


Figure 4: Geographical distribution of *M. acuminata* (and its subspecies) and *M. balbisiana* (According to De Langhe *et al.* 2009; Perrier *et al.* 2009)

Apart from edible hybrids originating from *M. acuminata* and *M. balbisiana* a few cultivars also originated from crosses between *M. acuminata* and *M. schizocarpa* (S genome) and *M. acuminata* and *M. textilis* (T genome) but their economic importance is negligible. The last group of edible bananas is represented by Fe'i bananas which are distributed throughout the Pacific region. Fe'i bananas are believed to originate in the Papua New Guinea area from ancestors belonging to the section Australimusa and they have been domesticated independently from the bananas related to *M. acuminata* and *M. balbisiana* (Ploetz *et al.* 2007). They are characteristic by their erect bunch and high levels of carotenoids.

2.1.3 Socio-economic importance

Bananas and plantains are produced in more than 130 countries and territories of tropical and subtropical regions. With the global annual production exceeding 130 million tons (faostat.fao.org), this species is one of the most important food crops.

About 90% of the production is provided by smallholder farmers and is determined for local markets. This dominating part of production is made up of a wide range of varieties, including cooking bananas, plantains, Highland bananas, sweet-acid dessert bananas and beer bananas. They represent an essential nutrition source for hundreds millions of people in developing countries. The highest consumption of bananas per person per year (200-300 kg) is reported in countries of sub-Saharan Africa, especially in Uganda, Rwanda and Gabon (faostat.fao.org). Despite the diversity of banana cultivars, the export trade (represented by 10% of the world banana production) is almost exclusively based on only one group of dessert banana – the 'Cavendish' group (AAA). These dessert bananas are the predominant group from 1960s and became a major cash crop in many countries. The main exporters of bananas can be found in Latin America (Ecuador, Costa Rica and Colombia).

Table 1: The largest banana and plantain producers in 2011 (in million metric tons) **(**faostat.fao.org**)**

Country	**Production (mT)**
India	**29.7**
China	**10.7**
Philippines	**9.2**
Ecuador	**7.4**
Brazil	**7.3**
Indonesia	**6.1**
Guatemala	**2.7**
Mexico	**2.1**
Colombia	**2.1**
Thailand	**2.0**

Bananas have been used as a food since earliest times. Sweet dessert bananas are mainly a food supplement, while cooking bananas and plantains are of higher dietary importance. They can be prepared in a wide variety of ways – boiled, roasted, fried steamed, baked or sun-dried and ground to flour (Chandler 1995). Bananas serve as an important source of carbohydrate, dietary fiber and essential vitamins such as A, B_6 and C. They also contain moderate amount of potassium, manganese and other minerals.

Except for the fruits also other parts of the banana plant can be useful. Flowers and tender parts of the plant's trunk supplement ingredients in South-East

Asian cuisine. Large, flexible and waterproof banana leaves are often used as a wrapping for food transportation or as disposable plates for serving meals. High quality textiles, ropes or paper for artistic purposes can be made from banana fibers. Industrial and medicinal alcohol has been produced from bananas in a variety of countries. Many bananas belonging to sections Callimusa and Rhodochlamys are important as ornamental flowers.

2.1.4 Diseases and breeding of resistant cultivars

The production of bananas is threatened by various diseases and pests. These problems are mainly visible in plantations of vegetatively propagated clones where the newly emerging virulent disease strains can spread quickly. Historically, the most devastating was the epidemic of Panama disease caused by the fungus *Fusarium oxysporum* f. sp. *cubense*. Panama disease destroyed over 40,000 ha of dessert banana 'Gros Michel' (AAA) in Latin America and eventually the export trade based on this cultivar. In 1950s and 1960s 'Gros Michel' was replaced by the resistant Cavendish group which dominates the commercial production since then. Currently, an extremely virulent form of Panama disease, 'Tropical Race 4', is spreading in Cavendish banana plantations in tropical countries. This new race could again cause serious devastation of dessert bananas determined for export. Infection of another fungal disease, Black sigatoga leaf spot, leads to 50% yield losses and became a striking problem in recent years. The disease agent *Mycosphaerella fijiensis* shows increasing resistance to fungicides and the used treatment is environmentally undesirable and expensive (Churchill 2011). Other banana pathogens include banana bunchy top virus causing almost complete yield losses, banana streak virus or a bacteria *Xanthomonas* which is spreading rapidly in East Africa.

Elimination of diseases by application of various chemical products is very costly, unavailable for smallholder farmers and polluting the environment. An alternative solution may come through breeding of new cultivars which would combine disease resistances, abiotic stress resistances and valued nutrition and agronomic qualities of cultivated clones. Breeding programs aiming for resistant cultivars have been carried out since 1920s in many countries including Honduras, Jamaica, Nigeria, Kenya and Brazil. Conventional cross-breeding of improved crop is complicated by the fact that cultivated bananas are almost always seedless (parthenocarpic) triploids reproduced by vegetative propagation. However, some

triploid cultivars have residual fertility and hand-pollination with diploid parents leads to production of a few seeds per bunch. Several resistant cultivars have been developed by crossing fertile diploid and tetraploid bananas, but few of them received commercial acceptance due to taste and quality issues (Heslop-Harrison and Schwarzacher 2007). Another important approach for *Musa* is breeding through mutation. There are several commercially released cultivars with desired qualities, which were derived from gamma-ray-induced mutations (Roux 2004). Development of transgenic bananas through particle bombardment (Sági *et al.* 1995) and *Agrobacterium*-mediated strategies (Rodriguez-Zapata *et al.* 2005) have been carried out.

2.1.5 Banana germplasm

The increasing need of preservation and characterization of banana diversity for the purposes of crop improvement is unquestionable. The worldwide banana germplasm, represented by accessions with a wide range of morphological variation and genome constitutions, is maintained in collections comprising between 1500 and 3000 accessions (Heslop-Harrison and Schwarzacher 2007). Field-based collections developed particularly in Asia participate in preservation of locally grown varieties. The world's largest banana germplasm collection is however maintained in tissue culture at the Bioversity International Transit Centre (ITC) (http://bananas.bioversityinternational.org/). This reference collection, hosted by the Katholieke Universiteit Leuven in Belgium, contains more than 1400 accessions and is continuously extended by various wild species and cultivars. Accessions indexed as virus negative are available for research and breeding purposes and distributed as tissue-culture plantlets.

The main objectives and means for the conservation of *Musa* diversity were formulated in The Global Conservation Strategy for *Musa* (INIBAP 2006). The main goals include establishment of agreed taxonomy, characterization of genetic diversity, conservation of the gene pool by long-term support of collections and reinforcement of the global system for the safe exchange of the germplasm. For the efficiency of collection and protection of banana diversity the unambiguous sample identification and characterization is of a great importance.

2.2 PLANT NUCLEAR GENOME

The genome can be defined as a one copy of the entire hereditary information of an organism. The majority of the DNA is stored in the nucleus and the term 'nuclear genome' is then used. The rest of DNA is stored within organelles that contain their own DNA – mitochondria and also chloroplasts, in case of photosynthetic organisms.

2.2.1 Nuclear genome size

Nuclear genome size is one of the basic characteristics of a species, useful in classification of organisms and also required in detailed analyses of genome structure and evolution. DNA content of the unreplicated haploid chromosome complement (n) is usually described with the term C-value (Swift 1950; Bennet and Smith 1976) and a nucleus in G_1 phase of the cell cycle then contains 2C DNA amount. Estimation of DNA quantity in absolute units (picograms of DNA or number of base pairs/ 1pg = $0.978 \cdot 10^9$ bp (Doležel *et al.* 2003)) has led to the discovery that C-values vary enormously among species. Eukaryotic genome sizes vary more than 200,000-fold and just within angiosperms, the range exceeds 2000-fold with the smallest known genome of 63 Mbp found in a carnivorous species *Genlisea margaretae* (Greilhuber *et al.* 2006) and the largest genome of 150 Gbp in *Paris japonica*, a rhizomatous geophyte endemic to Japan (Pellicer *et al.* 2010). It was soon discovered that the amount of nuclear DNA does not correlate with organismal complexity. This phenomenon was called "C-value paradox" (Thomas 1971) until the realization that most of the variation was caused by non-coding repetitive DNA, especially by increased number of mobile elements. Since there still remain some puzzles and unanswered questions about the genome size variation, the phenomenon is now called "C-value enigma" (Gregory 2001).

2.2.1.1 Methods for nuclear genome size estimation

Several methods have been employed to quantify nuclear genome size, but the two most widely used are Feulgen densitometry and flow cytometry. Feulgen densitometry is a precise method enabling measurement of DNA amounts in single cells. Feulgen-stained nucleus is divided into small, optically homogeneous areas and the absorption of visible monochromatic light for every spot is quantified. This

method has been more frequently used for genome size estimation until late 1990s when flow cytometry came to the fore of the field (Greilhuber *et al.* 2010; Doležel *et al.* 2007).

Flow cytometry is a convenient and reliable method, which became more popular than Feulgen densitometry due to the ease of sample preparation and high sample throughput. The method involves optical analysis of microscopic particles, constrained to flow in file within a hydrodynamically-focused fluid stream through the focus of intense light. Pulses of scattered light and the fluorescence are collected and converted to electrical pulses by photodetectors. The particles are analysed individually and at high speed. Except for nuclear genome size estimation, flow cytometry has a wide range of other applications including ploidy level screening, detection of mixoploidy and aneuploidy, assessment of the degree of polysomaty or observation of cell cycle kinetics.

To estimate DNA nuclear content, the suspension of intact nuclei is stained with a DNA-specific fluorochrome (usually propidium iodide). The amount of light emitted by the fluorochrome is then analyzed and displayed in a form of a histogram of relative fluorescence intensity (Figure 5), which represents the relative DNA content of the nucleus.

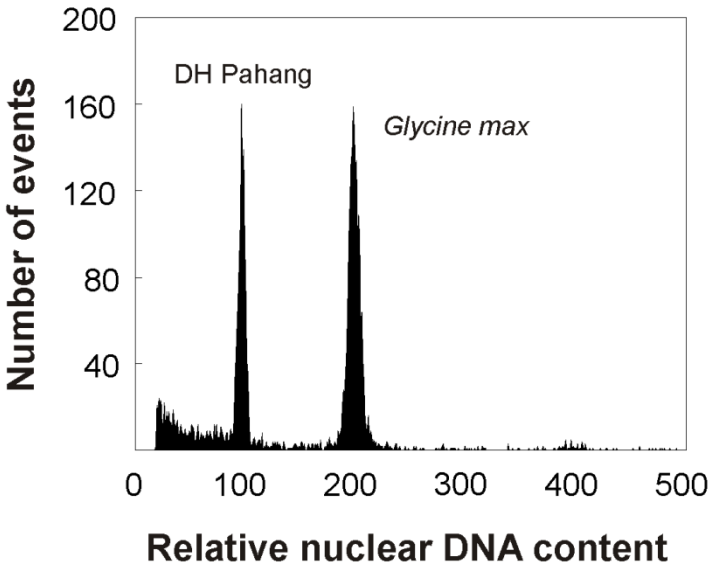


Figure 5: Histogram of relative nuclear DNA content obtained after flow cytometric analysis of propidium iodide stained nuclei of *M. acuminata* 'DH Pahang' and soybean, which were isolated, stained and analysed simultaneously. G_1 peaks of the banana and soybean (*Glycine max* 'Polanka', 2C = 2.5 pg DNA), which served as internal reference standard, are clearly visible.

Nuclei of the sample are isolated, stained and analysed simultaneously to avoid errors. 5000-20000 nuclei per sample are analysed and the absolute DNA content is then calculated from mean values of G_1 peaks following the formula:

$$\text{Sample 2C DNA content} = \frac{\text{(sample G1 peak mean) x (standard 2C DNA content)}}{\text{standard G1 peak mean}}$$

Reliable estimation of genome size of a species requires analysis of several randomly selected plants and replicate measurements of the same plant to detect variation in the procedure (Doležel and Bartoš 2005).

2.2.1.2 Nuclear genome size of banana

The first report of the nuclear genome size in *Musa* based on flow cytometric analysis was provided by Arumuganathan and Earle (1991) who estimated the genome size of 873 Mbp (2C = 1.81 pg). However, this result was significantly higher than the estimates of the genome size in following studies made by Doležel *et al.* (1994) and Lysák *et al.* (1999), in which a wider range of *M. acuminata* and *M. balbisiana* accessions and some triploid clones was used. The genome size of 600 Mbp was determined in *M. acuminata* and 550 Mbp in *M. balbisiana*. Monoploid genome size of a set of triploid clones with different genomic constitution was estimated between 566 Mbp and 621 Mbp (2C ~ 1.7-1.9 pg). These results led both Doležel *et al.* (1994) and Lysák *et al.* (1999) to assumption that the 1.81 pg 2C DNA content estimated by Arumuganathan and Earle (1991) was probably of a triploid accession.

Bartoš *et al.* (2005) extended the range of species with known nuclear DNA content by the representatives of all sections of the genus *Musa* including *M. schizocarpa* (S genome) and *M. textilis* (T genome). The differences in the genome size of the genomes A, B, S and T are summarized in Table 2.

It is evident that the sizes of A, B, S and T genomes are significantly different from each other. *M. schizocarpa* and *M. textilis* have higher DNA content than *M. acuminata* and *M. balbisiana*. *Musa* B genome is the smallest with the size ranging between 542 Mbp and 557 Mbp. *Musa* A genome is the second smallest with variation found among the subspecies and clones ranging from 599 to 619 Mbp. The difference in genome size within *M. acuminata* is supposed to be related to the geographical origin of individual accessions (Lysák *et al.* 1999).

Table 2: Nuclear DNA content and genome size of some representatives of A, B, S and T genomes

Accession name	ITC code	Genomic constitution	2C nuclear DNA content [pg]	Mean genome size [Mbp/1C]*	Reference
M. acuminata* ssp. *banksii	---	AA	1.23	601	Doležel *et al.* 1994
***M. acuminata* cv. 'Pisang Mas' (Australia)**	---	AA	1.25	611	Doležel *et al.* 1994
M. acuminata* ssp. *errans	---	AA	1.26	616	Doležel *et al.* 1994
***M. acuminata* cv. 'Pisang Mas' (Malaysia)**	---	AA	1.26	616	Doležel *et al.* 1994
Niyarma Yik	0269	AA	1.252	612	Lysák *et al.* 1999
Higa	0428	AA	1.225	599	Lysák *et al.* 1999
Pa (Rayong)	0672	AA	1.261	617	Lysák *et al.* 1999
Calcutta 4	0249	AA	1.226	599	Bartoš *et al.* 2005
Galeo	0259	AA	1.224	599	Bartoš *et al.* 2005
Pisang Mas	0653	AA	1.243	608	Bartoš *et al.* 2005
M. acuminata* ssp. *banksii	0896	AA	1.263	618	Bartoš *et al.* 2005
Guyod	0299	AA	1.266	619	Bartoš *et al.* 2005
M. balbisiana	---	BB	1.14	557	Doležel *et al.* 1994
M. balbisiana	0094	BB	1.108	542	Lysák *et al.* 1999
***M. balbisiana* 'Cameroun'**	0246	BB	1.130	553	Bartoš *et al.* 2005
Honduras	0247	BB	1.133	554	Bartoš *et al.* 2005
M. schizocarpa	0890	SS	1.377	673	Bartoš *et al.* 2005
M. textilis	0539	TT	1.435	701	Bartoš *et al.* 2005

* 1 pg DNA = $0.978 \cdot 10^9$ bp

Bartoš *et al.* (2005) also extended the knowledge of nuclear genome content by representatives of sections Rhodochlamys, Australimusa and Callimusa. Within section Rhodochlamys, 2C DNA content ranged from 1.191 to 1.299 pg (1C ~ 582 – 635 Mbp). Species belonging to Australimusa had 2C DNA content ranging from 1.435 to 1.547 pg (1C ~ 701 – 756 Mbp). The highest 2C nuclear content of 1.561 pg (1C = 763 Mbp) was found in *M. beccarii*, a representative of the section Callimusa.

2.2.2 Structure and organization of plant genomes

Despite the huge differences in the genome size, plants appear to have very similar genomic organization. The genome is divided into discrete chromosomes allowing replication, transcription and transmission of the genetic information. During interphase, chromosomes are in state of long chromatin fibers occupying specific territories within nucleus, forming loops active in gene expression. At metaphase of mitosis, the chromatin is condensed into short distinct chromosomes. Number, size and morphology of mitotic chromosomes constitute the karyotype, an important characteristics of a species. Similarly to genome size, the number of chromosomes varies widely among plants, ranging from 2n = 4 in *Haplopappus gracilis* (Jackson 1959) to 2n = 1440 in a fern *Ophioglossum reticulatum* (Khandelwal 1990). Karyotype of a single species either contains chromosomes of similar length or both large and small chromosomes can be observed (bimodal karyotype).

Morphological features, which can be identified on mitotic chromosomes, include telomeres at the ends of chromosomes, primary constriction (centromere) and secondary constriction associated with nucleolar organizing region (NOR) present on some chromosomes. Chromosome morphology is mainly defined by the position of centromere, according to which chromosomes can be classified as metacentric, submetacentric, acrocentric or telocentric.

Centromeres represent a constricted region of chromosomes on which kinetochore complex is assembled. The kinetochore is responsible for sister chromatid cohesion and essential for proper segregation of chromosomes during mitosis and meiosis. While the function of centromeres is conserved, the DNA sequence is highly diverged and variable in length (Ma *et al.* 2007). They can contain large arrays of tandemly repeated DNA (Maluszynska and Heslop-Harrison 1991; Harrison and Heslop-Harrison 1995; Nagaki *et al.* 2003), retrotransposon sequences (Miller *et al.* 1998; Gindulis *et al.* 2001) and even the presence of actively transcribed genes has been described (Yan *et al.* 2006; Mutti *et al.* 2010). On the other hand, the telomeric DNA sequence of most plant species contain a conserved 7-bp long repetitive motif $(TTTAGGG)_n$. Telomere region protects the ends of chromosome from degradation or from fusion with neighboring chromosomes being thus necessary for the key level of plant genome organization.

2.2.2.1 Composition of plant nuclear DNA

The nuclear DNA contains single-copy and low-copy coding sequences, associated regulatory sequences, promoters, introns and various classes of repetitive DNA. Investigations on the plant nuclear genome indicate that the coding DNA sequences and associated regulatory regions make up a minor part of the genome in a majority of plant species. Estimates of gene number, which are based on first genome annotation projects of *Arabidopsis*, *Medicago truncatula*, rice and poplar, varies from 25 000 to 45 000 (Lin *et al.* 1999; Yu *et al.* 2005; Tuskan *et al.* 2006) and the average size of plant gene plus its regulatory components is about 1-5 kb (Kellog and Bennetzen 2004). In plants with large genomes, most genes are clustered in gene islands, which are separated by vast expanses of regions with low gene density (gene deserts) (Feuillet and Keller 1999; Sandhu and Gill 2002). The gene clusters represent only about 12-24% of the nuclear DNA (Barakat *et al.* 1998) and their density increases from the proximal towards the distal regions of chromosome arms (Paterson *et al.* 2009; Schnable *et al.* 2009). With the growing number of sequenced plant genomes and comparative genome analysis, it has been demonstrated that gene orders are conserved among related species (Rustenholz *et al.* 2010; Freeling 2001). This colinearity decreases with the phylogenetic distance of species (Kellog and Bennetzen 2004).

Whole genome duplication (WGD) or polyploidy has played a major role in angiosperm evolution by generating fertile interspecific hybrids with multiple gene alleles at each locus. Polyploid plants can be either autopolyploids, which arise by multiplication of the genome in one plant, or allopolyploids, which are derived from a hybridization of two species with doubling of the chromosome set of one or both species involved. Duplication of genes can lead to modified gene function, adaptation and diversification of plants (Ramsey and Schemske 1998; Flagel and Wendel 2009). Among contemporary plants it has been estimated that 30-70% of species are polyploid (Wendel 2000). Recent genome-wide analysis revealed that angiosperm species have undergone one or more WGD early in their evolution and they are described as paleopolyploid (Van de Peer *et al.* 2009; Wang *et al.* 2012).

2.2.2.1.1 Repetitive DNA sequences

Repetitive DNA makes up the major fraction of most plant genomes. Based on genomic organization, repetitive DNA sequences are generally divided into two

different classes. The first group includes sequences which occur in tandem fashion forming large blocks of repeated DNA. The other group contains repeats which are dispersed throughout the genome. Both types of repetitive sequences, tandemly arranged or dispersed, are highly variable in sequence and copy number. Studying these sequence elements is important for understanding the organization and evolution of plant genomes as well as the nature and consequences of genome size variation among species.

2.2.2.1.1.1 Tandemly repeated DNA

Repetitive sequences tandemly arranged in head-to-tail fashion include microsatellites, minisatellites and satellites. Microsatellites, also called simple sequence repeats (SSR), are stretches of 2-6 bp long repeat motifs and are ubiquitous elements of plant genomes. Blocks of microsatellites are dispersed over the genome, being present in both coding and noncoding regions. They are characterized by a high degree of length polymorphism and are widely employed as genetic markers. Minisatellites with monomeric units of 6-100 bp are usually G/C-rich and organized in arrays up to several kilobases long (Tourmente *et al.* 1998; Vergnaud and Denoeud 2000). Satellite DNA sequences have typical repeat units of 160-180 bp or 320-360 bp that have been explained by the requirements of DNA length around one or two nucleosomes (Heslop-Harrison 2000). They are generally organized in several megabases long arrays and localized in constitutive heterochromatin of centromeric, pericentromeric, subtelomeric or intercalary regions (Schmidt and Heslop-Harrison 1998; Kubis *et al.* 1998). The distribution patterns are often chromosome-specific thus enabling chromosome identification using FISH (Schwarzacher 2003).

Satellite repeats are believed to be one of the most dynamic components of plant genomes. Rapid evolutionary changes of their nucleotide sequences and abundance lead to frequent occurrence of genus- or species-specific types of satellites (Macas *et al.* 2000; Tek *et al.* 2005). The mechanisms by which satellite sequences arise and evolve include unequal crossing over, gene conversion, transposition and formation of extrachromosome circular DNA (eccDNA) (Stephan 1986; Dover 2002; Navrátilová *et al.* 2008).

While the role of most tandemly repeated sequences is still not well understood, several of them have extremely important function in the genome. Except for previously mentioned centromeric and telomeric sequences, ribosomal

genes are another example of tandemly arranged clusters. Ribosomal RNA is a key component of ribosomes and essential for protein synthesis in all living organisms. High copy number and tandem arrangement of rDNA genes ensure sufficiency of ribosomal RNA in cells.

2.2.2.1.1.2 Dispersed DNA sequences

Dispersed repetitive DNA elements are scattered throughout the genome, interspersed with other sequences and distributed along all chromosomes. The majority of dispersed DNA sequences originate from mobile DNA sequence. In many plants with large complex genomes, these transposable elements make up over 70% of the nuclear DNA (SanMiguel and Bennetzen 1998). Mobile elements are divided into two major classes based on their mode of transposition.

Class I transposable elements, called retroelements, are able to amplify and move in the genome through RNA intermediates that are converted into extrachromosomal DNA by reverse transcriptase/RNaseH enzymes. Hence, they produce new copies for integration into genome without losing the original copy. They contain two or three open reading frames, which encode genes for the transcription, movement and integration proteins (Figure 6). Gag proteins are involved in intracellular packaging of the RNA transcript and the function of Pol includes the enzyme reverse transcriptase. Retroelements have been classified into viral retroelements and non-viral retroelements. Viral retroelements have long been thought to be restricted to vertebrates, but their presence has been recently confirmed also in plant genomes (Peterson-Burch *et al.* 2000). Some non-viral retroelements resemble retroviral proviruses, but they are lacking genes which encode the envelope proteins of retroviruses. Based on the structure of their DNA copies, two main subgroups of non-viral retroelements are distinguished: retrotransposons and retroposons (Kumar and Bennetzen 1999).

The most prominent feature of retrotransposons is the presence of long terminal repeats (LTR) sequences which can be several thousand bases long. They are found in all eukaryotic genomes and represent the most numerous class of retroelements in plants (Kumar *et al.* 1997). They account for 5.6-17% of the genome in plants with small genome size such as those of *Arabidospsis* or rice, and their abundance increases along with increasing genome size (McCarthy *et al.* 2002; Pereira 2004; Zuccolo *et al.* 2007). The two main subclasses of retrotransposons are

called Ty1-*copia* and Ty3-*gypsy*. They differ from each other in the position of integrase (Figure 6), the degree of reverse transcriptase sequence conservation (Xiong and Eickbush 1990) and chromosomal localization. Ty1-*copia* elements were found to be distributed along plant chromosomes. Reduced abundance of these elements was observed in certain regions, mainly in centromeres, interstitial and terminal heterochromatic regions and ribosomal DNA sites (Pearce *et al.* 1996; Heslop-Harrison *et al.* 1997). On the other hand, Ty3-*gypsy* elements were primarily detected in centromeric regions of plant chromosomes (Miller *et al.* 1998; Presting *et al.* 1998; Liu *et al.* 2008).

Retroposons do not contain long terminal repeats; therefore they are also called non-LTR retrotransposons. They are ubiquitous components of nuclear genomes in many species across the plant kingdom, but they are less abundant than LTR retrotransposons (Kumar and Bennetzen 1999). Retroposons have been subdivided into LINEs (long interspersed nuclear elements) and SINEs (short interspersed nuclear elements), their basic structure is shown in Figure 6. It has been proposed that LINEs are the most ancient class of retroelements, being precursors for LTR retrotransposons (Xiong and Eickbush 1990). The first plant LINE-like element *Cin4* was identified in the genome of maize (Schwarz-Sommer *et al.* 1987) and subsequently other LINEs were characterized in various plant genomes (Wright 1996; Noma *et al.* 1999) showing that sequence divergence and extreme heterogeneity is a typical feature of plant LINEs (Schmidt 1999). SINEs are structurally different from all the other retroelements, they have no coding capacity and use LINE-specified functions to transpose (Kajikawa and Okada 2002). SINEs are highly abundant in animals but relatively rare in plant genomes. Several SINEs have been identified in genome of rice (Mochizuki *et al.* 1992; Tsuchimoto *et al.* 2008), tobacco (Yoshioka *et al.* 1993) or *Aegilops umbellulata* (Yasui *et al.* 2001).

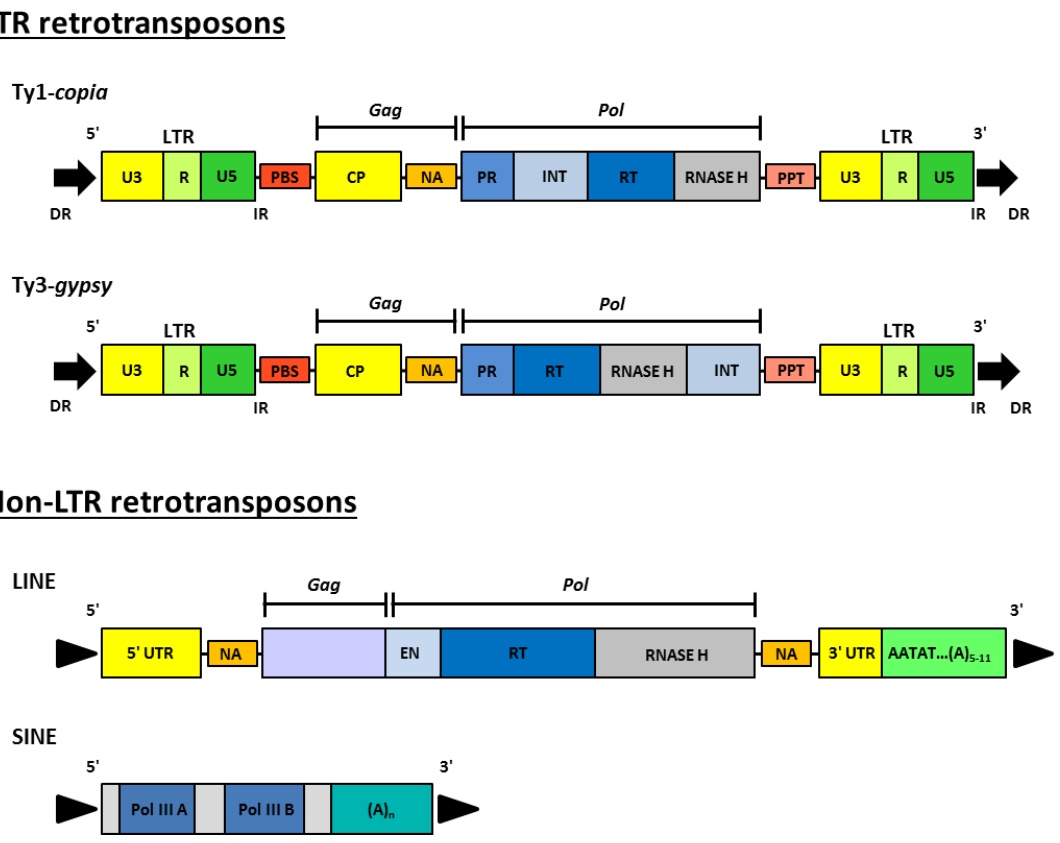


Figure 6: General structure of some retroelements. LTR retrotransposons contain long terminal repeats in direct orientation at each end. U3, R and U5 regions carry signals for initiation and termination of transciption. The genes for capsid-like protein (CP), endonuclease (EN), integrase (INT), protease (PR), reverse transcriptase (RT) and RNAse H are encoded within retrotransposons. Other sequence features include primer binding sites (PBS), polypurine tracts (PPT), nucleic acid binding moiety (NA), inverted terminal repeats (IR), flanking target direct repeats (DR), 5' untranslated region (5' UTR), 3' untranslated region (3' UTR) and Pol III A and B-promoter recognition sites for RNA polymerase III. (According to Kumar and Bennetzen 1999)

DNA transposable elements (class II transposable elements) move in the genome by utilizing single or double-stranded DNA intermediate. They are found in high density in euchromatic regions, within or in close proximity of genes (Bureau and Wessler 1992). The class II transposable elements can be further divided into three major subclasses. The best known are "cut-and-paste" transposons, which excise as double-stranded DNA and reintegrate into new genomic location. These elements, characterized by transposase gene and terminal inverted repeats (TIRs) at both ends, include *Activator*/*Dissociation* (*Ac*/*Ds*) and *Enhancer*/*Suppressor-Mutator* (*En*/*Spm*) systems first described in the genome of maize (Fedoroff *et al.* 1983). The

second subclass is represented by *Helitrons*, a recently described group of transposable elements that are predicted to amplify by a rolling-circle mechanism, which involves a single-stranded DNA intermediate (Kapitonov and Jurka 2001; Yang and Bennetzen 2009). *Mavericks*, large transposons with long TIRs and coding capacity for multiple proteins, are another class of DNA transposable elements. The mechanism of their transposition is not yet well understood, but it has been proposed that they replicate by a single-stranded DNA intermediate using a self-encoded DNA polymerase (Pritham *et al.* 2007; Feschotte and Pritham 2007).

The increasing understanding of mobile DNA elements abundance, distribution and behavior has revealed that they are the major force in the genome evolution at multiple levels. It has been shown in many studies, that they have huge impact on nuclear DNA content and that there is a positive correlation between genome size and mobile DNA elements copy number (Bennet and Leitch 2005; Paterson *et al.* 2009; Wicker *et al.* 2009). Transposable elements have high transpositional and recombinational potential and may be involved in ectopic recombination. They also conduce to the evolution of genes by contributing exonic and intronic sequences ("transposon domestication") and regulatory sequences (Miller *et al.* 1999; Volff 2006).

2.2.2.2 Structure and organization of banana genome

The relatively small genome of banana is divided into 9 – 11 chromosomes in a haploid set. Individual chromosomes are small and of similar size. Their length ranges between 1 and 2 μm when observed at mitotic metaphase. A karyotype, which is defined by the number, length and morphology of chromosomes, is an important characteristic of a species. However, the karyotype of banana has not yet been elucidated. Although the method for preparation of high quality mitotic metaphase spreads has been established (Doležel *et al.* 1998), individual chromosomes of banana have not been identified and numbered due to lack of morphological differences. Among the chromosome set, only chromosomes bearing secondary constriction can be identified. Measurements of short and long arms are also difficult because the chromosomes condense tightly. However, banana chromosomes appear to be metacentric and submetacentric. The missing karyotype data, which have proven to be useful in other species for evolutionary and phylogenetic studies

(Heslop-Harrison and Schwarzacher 2011), can only be obtained after more of chromosome-specific landmarks and markers are available.

2.2.2.2.1 Composition of nuclear DNA of banana

Like all plant genomes, the *Musa* genome consists of single copy sequences and various types of repetitive DNA. Several BAC clone libraries were developed from diploid *M. acuminata* and *M. balbisiana* species (Vilarinhos *et al.* 2003; Šafář *et al.* 2004) and the analysis of BAC sequences showed that a coding gene occurs every 6.4 to 6.9 kb in the banana nuclear genome (Aert *et al.* 2004; Cheung and Town 2007).

It has been estimated that about 55% of the banana genome is represented by repetitive and non-coding sequences (Hřibová *et al.* 2007). The most detailed study of repetitive part of the banana genome was made by Hřibová *et al.* (2010) who analyzed low-depth 454 sequencing data of *M. acuminata* cv. 'Calcutta 4' and reconstructed most of the major types of DNA repeats. LTR retrotransposons were found to be the most abundant DNA sequences. Phylogenetic analysis of their RT domains showed that Ty1-*copia* elements were represented by four distinct evolutionary lineages, while the less abundant Ty3-*gypsy* elements belonged to a single evolutionary lineage of chromoviruses. Non-LTR retrotransposons and DNA transposons were found to be relatively rare. Even less frequent seem to be tandem arranged repeats (satellites). After extensive search, Hřibová *et al.* (2010) discovered two new satellite DNA sequences, CL18 and CL33. While the satellite CL18 consisted of ~ 2 kb monomers, CL33 was characterized by ~ 130 bp monomer unit.

The recent sequencing and assembling of the genome of doubled-haploid *M. acuminata* cv. 'DH Pahang' (subspecies *malaccensis*) shed more light on banana genome structure and evolution. More than 36 000 protein-coding gene models have been identified and similarly to other plants, the gene rich regions were found to be mostly located on distal parts of chromosomes (Figure 7) (D'Hont *et al.* 2012).

Almost half of the *Musa* genome is made up of different classes of transposable elements, the majority of them being represented by LTR retrotransposons. These elements account for 37% of the genome and are particularly abundant in centromeric and pericentromeric chromosome regions. LINEs represent about 5.5% and DNA transposable elements represent only about 1.3% of the genome (D'Hont *et al.* 2012).

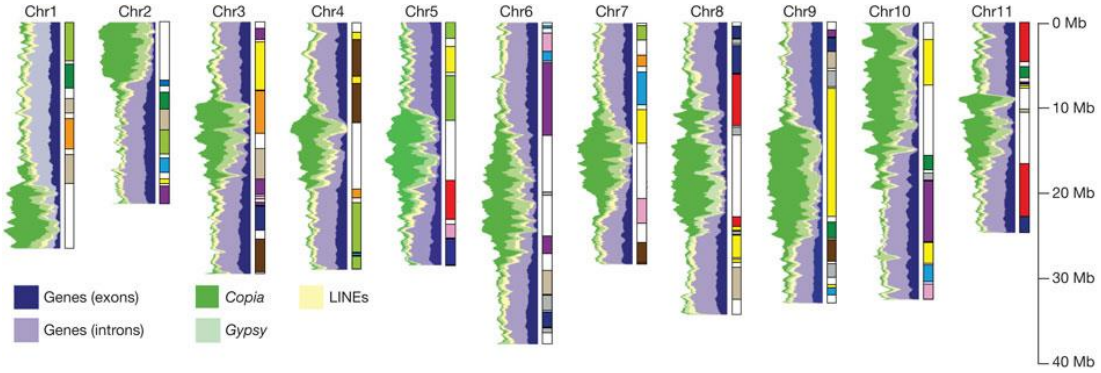


Figure 7: Chromosomal distribution of the main *M. acuminata* genome features (D'Hont *et al.* 2012).

D'Hont *et al.* (2012) detected three rounds of whole genome duplications in the *Musa* lineage, which were further followed by gene loss and chromosome rearrangements. Two WGDs were dated at a similar period around 65 Myr ago and even more ancient duplication event, occurring around 100 Myr ago, was detected. All three rounds of paleopolyploidization in the *Musa* genome occurred independently of two WGD previously reported in the *Poaceae* lineage (Tang *et al.* 2010).

2.3 PHYSICAL MAPPING OF PLANT GENOMES

Learning about the localization of genes and repetitive sequences on chromosomes is an important element leading to understanding genome organization and evolution. The map of a chromosome can be gained by two different approaches: genetic mapping and physical mapping. Genetic mapping determines the relative proximity of markers along chromosomes based on the frequency of recombination. The distance between markers is then expressed in centimorgans (cM). Genetic maps give the information about chromosomal order of markers but cannot determine the real distance of markers on chromosomes (Figure 8).

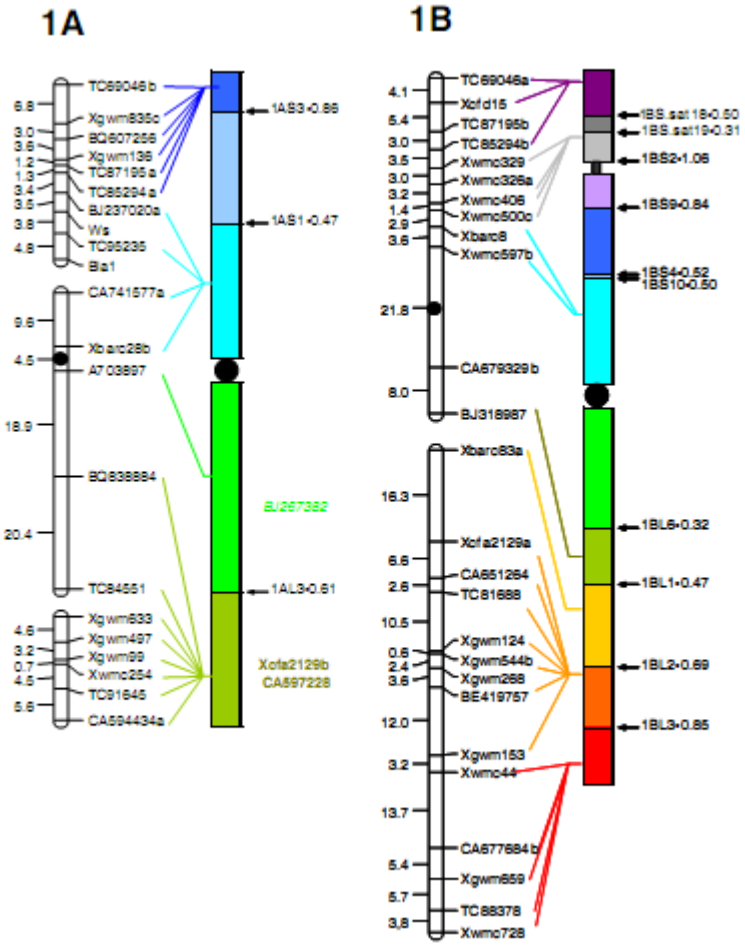


Figure 8: Comparison of genetic and physical maps of 1A and 1B chromosomes of durum wheat (Gadaleta *et al.* 2009).

On the other hand, physical mapping is based on actual localization of markers on chromosomes and the distance between markers is expressed in terms of base pairs (Figure 8). Various methods of physical mapping have been used,

including physical mapping through contig assembly with employment of large-insert DNA libraries, optical mapping or physical cytogenetic mapping.

Large-insert DNA libraries are prepared after partial digestion of genomic DNA with restriction enzymes and subsequent ligation of the fragments into high-capacity vectors. Three most common vectors are bacterial artificial chromosomes (BACs), yeast artificial chromosomes (YACs) and cosmids. The choice of vector depends on the insert size, the ease of screening and maintaining libraries, the ability to recover cloned sequences and the stability of cloned sequences in the vector (Hass-Jacobus and Jackson 2005). Physical maps are then created based on overlaps between clones. The overlaps are identified either by screening clones for presence of DNA markers or after DNA fingerprinting. In the latter, large-insert clones are digested with restriction enzymes into fragments, which are then separated and detected on a gel. Overlapping clones are identified by shared restriction fragments. Physical mapping through contig assembly is widely used for studies of plant genomes (Paux *et al.* 2008; Gu *et al.* 2009; Luo *et al.* 2010).

Optical mapping is a light microscope-based technique for creating ordered restriction maps from individual elongated chromosomes (Schwarz *et al.* 1993). In this approach, large DNA molecules are digested on open glass surfaces and individual restriction fragments are visualized by fluorescence microscopy. Ordered restriction maps are then derived from digital images of fully or partially digested molecules (Aston *et al.* 1999). This technique has been successfully used for validation of rice (Zhou *et al.* 2007) and maize (Zhou *et al.* 2009) genome sequence.

2.3.1 Cytogenetic mapping of plant genomes

Cytogenetic maps are based on ordering markers with respect to cytological landmarks of chromosomes, such as centromere and telomere. These maps are generated by *in situ* hybridization of labeled DNA directly to chromosomal DNA in microscope preparations. DNA sequences can be mapped to mitotic metaphase chromosomes (Jiang *et al.* 1995), meiotic pachytene chromosomes (de Jong *et al.* 1999, Cheng *et al.* 2001), extended DNA fibres (Fransz *et al.* 1996) or super-stretched flow-sorted chromosomes (Valárik *et al.* 2004).

The technique of *in situ* hybridization (ISH) was first described by Gall and Pardue (1969) and used radioactively labeled probes. Sites of hybridization were detected by photographic emulsion over the surface of the preparation. Since then

many improvements of the method have been introduced, the most important being the development of non-radioactive labeling. Many of the first non-radioactive ISH experiments were based on biotin-labeled probes, the affinity of biotin to avidin and hybridization site detection by colored precipitates with enzyme substrates (Rayburn and Gill 1985). The problems with difficult color differentiation on chromosomes and relatively low spatial resolution of these methods have been overcome with the application of fluorophores for probe labeling (Pinkel *et al.* 1986).

2.3.1.1 Fluorescence *in situ* hybridization

Fluorescence *in situ* hybridization (FISH) is a technique enabling detection and localization of specific DNA sequences on chromosomes. Probes for FISH experiment can be either labeled directly using various fluorophores (fluorescein, Cy3, etc.) or indirectly with biotin or a hapten (e.g., digoxigenin). Avidin and its derivatives (in case of biotin labelled probes) and anti-hapten antibodies conjugated with fluorophores are then used to detect sites of probe hybridization (Figure 9). Both direct and indirect FISH has been widely applied in plant molecular cytogenetic research (Jiang and Gill 2006).

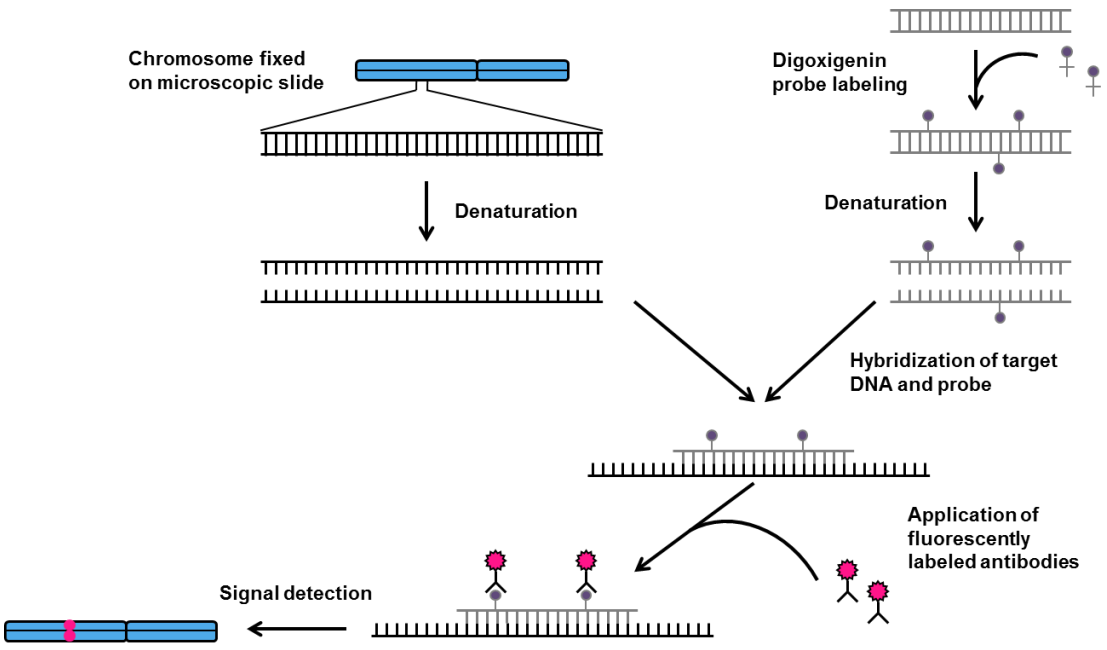


Figure 9: Principle of fluorescence *in situ* hybridization.

A probe for FISH can be a fragment of DNA or RNA depending on the requirements of the experiment. DNA *in situ* hybridization gives information about

the long-range organization, chromosomal location and genomic distribution of probes and differences in their copy number, while RNA *in situ* hybridization gives information about the location and amount of expression of particular genes. Commonly used probes include tandemly arranged gene sequences (rDNA), tandemly arranged and dispersed repetitive sequences, clones of large-insert DNA libraries or single-copy sequences and genes.

Sensitivity and resolution are the most important parameters of FISH mapping technique. They depend on preparation quality and the level of chromosome condensation. The limit of detection sensitivity ranges from 1 to 3 kb (Ohmido *et al.* 1998; Wang *et al.* 2006) but localization of probes shorter than 1 kb has been reported (Khrustaleva and Kik 2001). The resolution of FISH on mitotic metaphase chromosomes, which are the most common target for cytogenetic mapping, is about 3-8 Mb (Lamb *et al.* 2007; Danilova and Birchler 2008). On less condensed prometaphase chromosomes, the resolution increases to 1-2 Mb (Cheng *et al.* 2002). The resolving power of FISH mapping is even higher on interphase nuclei, super-stretched mitotic metaphase chromosomes and meiotic pachytene chromosomes. Position of probes separated by 60-100 kb can be determined in euchromatic regions of these targets (Jiang *et al.* 1996; Valárik *et al.* 2004; Kulikova *et al.* 2001). The highest axial-resolution of FISH mapping is, however, achieved when DNA fibres are used. Probes which are separated by only few kilobases can be detected (Fransz *et al.* 1996; Jackson *et al.* 1998).

Genomic *in situ* hybridization (GISH) is a special type of FISH, which uses the total genomic DNA of one species as a labeled probe. Unlabeled genomic DNA of another species is added in excess and serves as blocking DNA, thus increasing the specificity of probing. This method enables distinguishing parental genomes and analyzing genome organization in interspecific hybrids (Schwarzacher *et al.* 1989). GISH-based physical mapping was demonstrated in grasses *Lolium perenne* and *Festuca pratensis* (King *et al.* 2002). A diploid monosomic substitution line was produced by introgression of a single chromosome of *F. pratensis* into *L. perenne* and recombination between *F. pratensis*/*L. perenne* bivalent was visualized using the GISH method. GISH junctions represented the physical locations of recombination sites and enabled a range of recombinant chromosomes to be used for physical mapping of the introgressed *F. pratensis* chromosome (King *et al.* 2002).

FISH-based chromosome identification can also be used for karyotyping of plant genomes (Dong *et al.* 2000; Kato *et al.* 2004; Lamb *et al.* 2007). The most commonly used probes for FISH karyotyping are tandemly repeated sequences which often generate hybridization pattern specific for individual chromosomes and enables their identification (Dong *et al.* 2000). Single-gene detection was applied for karyotyping of maize chromosomes (Lamb *et al.* 2007).

2.3.1.2 Cytogenetic mapping of banana genome

As mentioned earlier, the nuclear genome of banana is relatively small (1C ~ 550-750 Mbp) and divided into small and morphologically similar chromosomes. This complicates cytogenetic studies and identification of chromosomes within karyotype. There is no reliable method to discriminate parental chromosomes in hybrid clones. In several studies, GISH has been used to distinguish the genomic constitution of interspecific hybrids (Osuji *et al.* 1997; D'Hont *et al.* 2000). However, this method is not commonly used in banana, because DNA sequences of A and B genomes are very similar and the discrimination of parental chromosomes is difficult and not reliable. Only (peri)centromeric chromosome regions are typically labelled, while distal parts of each arm remains unlabelled. This prevents the detection of potential interspecific intra-chromosomal exchanges (D'Hont *et al.* 2000).

Genomic distribution of genes for ribosomal RNA was studied in a wide range of *Musa* species using FISH. 45S rDNA was localized on one chromosome pair and the number of 5S rDNA loci varied from four to eight in representatives of section Eumusa (Figure 10, A and B) (Osuji *et al.* 1998; Doleželová *et al.* 1998; Bartoš *et al.* 2005). In sections Rhodochlamys and Australimusa, 45S rRNA genes were mapped to one or two chromosome pairs and 5S rDNA clusters were found on four or six chromosomes. *Musa beccarrii* (Callimusa section) was characterized by six 45S rDNA loci and five clusters of 5S rRNA genes (Bartoš *et al.* 2005).

Cytogenetic mapping of DNA sequence of banana streak virus (BSV) showed dispersed distribution with two main sites of integration into the genome (Harper *et al.* 1999). Copies of LTR retrotransposon *monkey* were localized in the nucleolus organizing region and colocalized with 45S rRNA genes. Other copies of *monkey* appeared to be dispersed throughout the genome (Balint-Kurti *et al.* 2000).

An important progress in cytogenetic mapping of banana genome would be the isolation of chromosome specific markers which could enable identification of individual chromosomes. Unlike other plant species (Dong *et al.* 2000; Kubaláková *et al.* 2005), most of cytogenetically mapped repetitive sequences do not show chromosome specific pattern and can not be used for chromosome identification (Valárik *et al.* 2002; Hřibová *et al.* 2010). While most transposable elements were localized along all banana chromosomes, a LINE element was found in (peri)centromeric regions of *M. acuminata* 'Calcutta 4' (Figure 10, F), thus the probe could serve as centromeric marker (Hřibová *et al.* 2010).

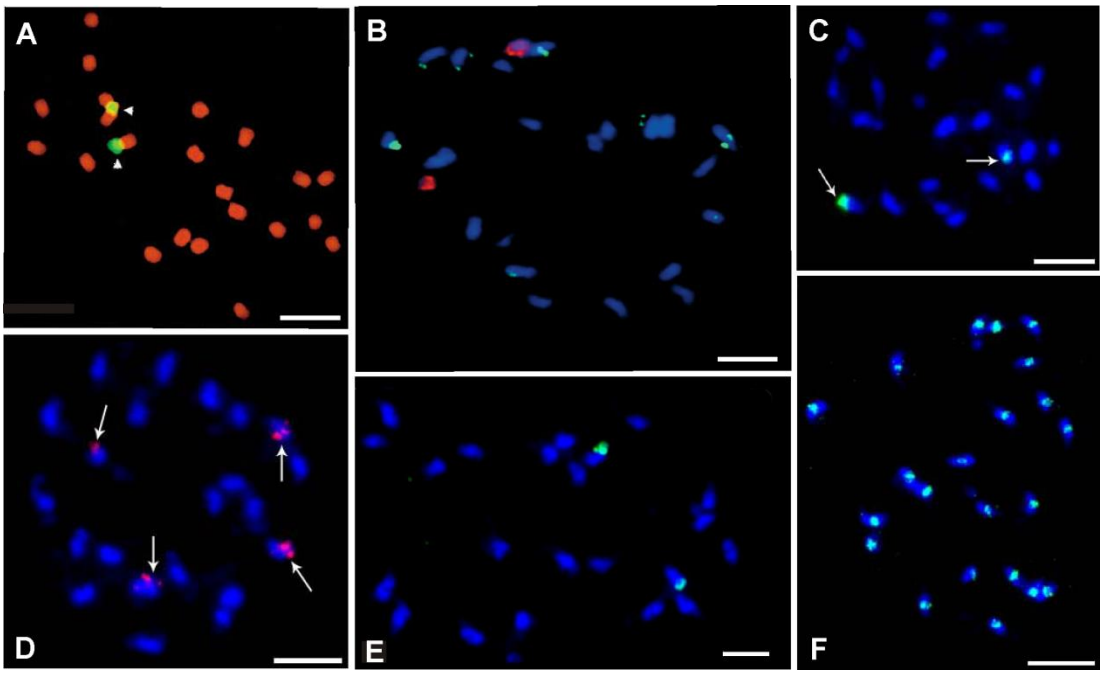


Figure 10: Cytogenetic mapping of different DNA sequences on mitotic metaphase chromosomes of *M. acuminata* 'Calcutta 4'. Chromosomes were counterstained with DAPI (red pseudocolour is used in picture A). Bar = 5 µm. (A) Clone *Radka*1 localized to secondary constriction of one chromosome pair (Valárik *et al.* 2002). (B) Genomic distribution of 45S rDNA (red) and 5S rDNA (green) (Bartoš *et al.* 2005). (C) Signals of DNA satellite CL18 were detected on two chromosomes (Hřibová *et al.* 2010). (D) DNA satellite CL33 was localized on two chromosome pairs (Hřibová *et al.* 2010). (E) BAC clone 2G17 gives single locus-specific signal on one chromosome pair (Hřibová *et al.* 2008). (F) A probe derived from LINE element was detected in centromeric regions of all chromosomes (Hřibová *et al.* 2010).

Another possibility for identification of individual chromosomes would be isolation of BAC clones from large-insert DNA libraries of banana and their

localization on mitotic chromosomes. This approach was used by Hřibová *et al.* (2008) but out of eighty clones carrying low amounts of repetitive sequences, only one clone gave a single-locus signal (Figure 10, E).

Recently, two specific satellite DNA sequences (CL18 and CL33) were identified in 454 sequencing data of *M. acuminata* 'Calcutta 4' (Hřibová *et al.* 2010). FISH on mitotic metaphase chromosomes of *M. acuminata* 'Calcutta 4' revealed clusters of signals in subtelomeric regions of one chromosome pair, when probe for CL18 was used (Figure 10, C). DNA satellite CL33 was detected on four chromosomes (Figure 10, D) (Hřibová *et al.* 2010). Potential variation in genomic distribution of these repeats in different *Musa* species could improve the knowledge of the long-range organization of their chromosomes and karyotype evolution in *Musa*.

2.4 PLANT MOLECULAR PHYLOGENETICS

The classical way of estimating the evolution and relationships between species is based on comparison of their morphological characters. Molecular phylogenetics is a branch of science which uses statistical analysis of molecular data to gain the information about evolutionary relationships among organisms. The data generated by molecular markers reflect the variability at the DNA level. This variability arises from DNA mutations which are continuously accumulated in genomes of organisms and can be used for studies of genetic diversity and for reconstruction of phylogenetic relatioships. Unlike conventional phenotype-based markers, molecular markers are not influenced by environment, usually have qualitative character and can be employed in analysis of distatnly related species.

2.4.1 Molecular DNA markers

With the development of molecular methods, molecular DNA markers have made a rapid progress and a wide range of marker techniques is now available. An ideal markar should have several important features, such as uniform distribution throughout the genome, high degree of polymorphism and a potential for high-throughput analysis. The technique should be reproducible, simple, quick and unexpensive, should not require high amounts of DNA and prior information about the genome sequence (Agarwal *et al.* 2008). However, it is extremely difficult to find a molecular marker which would meet all these requirements and the choice of a marker then depends on the type of the study to be undertaken. Molecular markers can be classified into three basic groups, based on the way they are scored: (1) molecular markers based on hybridization, (2) molecular markers based on PCR and (3) molecular markers based on sequencing.

2.4.1.1 Molecular DNA markers based on hybridization

RFLP markers (Restriction Fragment Length Polymorphism) are typical representatives of molecular markers based on hybridization. Genomic DNA is digested by restriction endonucleases, resulting fragments are separated by gel electrophoresis and transferred to a membrane via Southern blotting technique. DNA polymorphism is then detected by hybridization of a labeled probe to the membrane (Botstein *et al.* 1980). Different fragment profiles are generated by nucleotide

substitutions or DNA rearrangements. RFLP markers are highly polymorphic, reproducible and codominantly inherited. They have been successfully used in diversity and phylogenetic studies of many plant species (Tanksley *et al.* 1989; Moser and Lee 1994; Dijkhuizen *et al.* 1996). However, the technique is time consuming and requires large quantity of high quality genomic DNA and is therefore being replaced by newer approaches, such as DArT technology (Diversity Array Technology).

DArT technology is a high-throughput microarray hybridization-based technique that permits simultaneous screening of thousands of polymorphic loci distributed over a genome without any prior sequence information (Jaccoud *et al.* 2001). In this approach, representations of an organism or population of organisms are prepared by restriction enzyme digestion of genomic DNA followed by ligation of restriction fragments to adapters and subsequent amplification. Individual DNA fragments are isolated by cloning and inserts are then amplified and arrayed on a solid support (Jaccoud *et al.* 2001). Polymorphic DArT markers are revealed by hybridization of labelled genomic representations of individual genomes to the prepared discovery array. The three main advantages of this approach, the independence of DNA sequence, high-throughput and low-cost data production, have led to increasing use of this technique for genetic diversity studies in numerous species, such as barley (Wenzl *et al.* 2004), wheat (Akbari *et al.* 2006), grasses (Kopecký *et al.* 2009) or banana (Risterucci *et al.* 2009).

2.4.1.2 Molecular DNA markers based on PCR

There is a large number of approaches for generation of molecular markers based on PCR, mainly because of its simplicity and possibility to use random primers to overcome the limitation of prior sequence knowledge. RAPD markers (Random Amplified Polymorphic DNA; Williams *et al.* 1990) involve the use of short random oligonucleotide sequences (commonly ten bases long) in a PCR reaction and result in the amplification of DNA products which differ between individuals because of rearrangements or deletions at or between oligonucleotide primer binding sites in the genome (Williams *et al.* 1990). Polymorphisms detected after electrophoretic separation of amplified fragments document the genetic diversity of studied individuals. The major weakness of the method is relatively low reproducibility. It

has been shown that the profiling is highly dependent on reaction conditions and the results may substantially vary between different laboratories (Jones *et al.* 1997).

To overcome problem with reproducibility connected with RAPD markers AFLP (Amplified Fragment Length Polymorphism) method was developed (Vos *et al.* 1995). The method involves restriction digestion of the genomic DNA followed by ligation of adaptors to the ends of DNA fragments. Restriction fragments are then subjected to two subsequent PCR amplifications under highly stringent conditions. Adapter-specific primers have one to three nucleotides long extension running into the unknown sequence of restriction fragment. The amplified fragments are separated and visualized on denaturing polyacrylamide gels or using capillary electrophoresis. AFLP markers are dominant, reproducible and can be generated without prior knowledge of nucleotide sequence. They have been extensively used for studying genetic diversity in different plant species (Lu *et al.* 1996; Breyne *et al.* 1999; Steiger *et al.* 2002). Despite all the advantages, AFLP markers have several drawbacks including high requirements on the quality of genomic DNA and not completely regular distribution throughout the genome depending on restriction enzymes used for digestion.

Microsatellites, also known as simple sequence repeats (SSRs), are another example of PCR-based markers. Microsatellites are stretches of simple nucleotide motifs (1 – 6 bp long) arranged tandemly within the genomes of both prokaryotic and eukaryotic organisms. Microsatellite assays show extensive inter-individual polymorphism in the number of repeated units, which is mainly caused by polymerase slippage during DNA replication (Schlötterer and Tautz 1992). Locus-specific primers are designed to the flanking regions, previous information about nucleotide sequence is therefore necessary. Analysis of PCR products is carried out using agarose or polyacrylamide gels. The employment of fluorescently labelled primers and capillary electrophoresis has further increased the resolution and enabled high-throughput and automatisation. Microsatellites are also popular because of their high abundance, co-dominant inheritance, extent of allelic diversity and good reproducibility. They have been used for genetic diversity studies, genome mapping, marker-assisted selection and population studies in many important crops (Yang *et al.* 1994; Lu *et al.* 1996; Pessoa-Filho *et al.* 2007).

2.4.1.3 Molecular DNA markers based on sequencing

The most straightforward way of assessing genetic diversity is sequencing of defined regions of DNA, ideally sequencing and comparison of whole genomes. With the increasing accessibility of sequencing technologies, the new generation of molecular markers, such as SNPs (Single Nucleotide Polymorphism), has emerged. SNPs are sequence polymorphisms caused by a single nucleotide mutation at a specific locus in the DNA sequence. Single base transitions are the most common mutation types, but transversions, insertions and deletions also occur. SNPs are the most abundant molecular markers widely distributed throughout the entire genome, although they are more frequent in repeats, introns and pseudogenes (Balasubramanian *et al.* 2002). They are genetically stable, highly reproducible and allow high-throughput automated analysis. The disadvantage of these markers is the low level of information obtained compared with that of highly polymorphic SSR markers.

Chloroplast and mitochondrial DNA sequences, markers derived from single and low copy nuclear gene sequences and 45S rDNA locus belong to the most popular and frequent targets for phylogenetic analysis in plants. Many early plant phylogenetic studies focused on protein-coding gene sequences of chloroplast DNA such as *rbc*L (gene encoding tha large subunite of ribulose-1,5-bisphosphate carboxylase/oxygenase, RUBISCO) which was found to be suitable for eluciadation of evolutionary relationships among higher-level taxa (Chase *et al.* 1993; Källersjö *et al.* 1998; Smith *et al.* 2009). Noncoding regions of the chloroplast genome were used in lower-level studies (Taberlet *et al.* 1991; Gielly and Taberlet 1994; Small *et al.* 2005). The mitochondrial genome is characterized by a much higher rate of structural rearrangements than chloroplast genome, but is slowly evolving on the sequence level (Palmer and Herbon 1988). The slow sequence evolution of mitochondrial DNA in combination with variable occurance of introns provides information about the evolution of older land plant clades (Knoop 2004). High copy number of organellar DNA and the possibility to use universal primers simplify experimental desing. Prevalently uniparental inheritance makes the analysis of chloroplast and mitochondrial DNA suitable for identification of male and female parents of hybrids. On the other hand, this mode of inheritance limits their application in phylogenetic studies of hybrid species (Birky 1995).

Single and low copy nuclear gene sequences provide evidence for both maternal and paternal lineages. Compared to organellar DNA, nuclear genes evolve more rapidly at the sequence level (Sang 2002; Small *et al.* 2004) and provide greater resolving power for relationships clarification (Zou *et al.* 2008). Because of the existence of thousands of low copy nuclear genes in genomes, virtually limitless source of phylogenetic information appears. Major drawbacks of the use of nuclear genes for phylogenetic reconstructions are the necessity to design specific primers and the need to distiguish between ortologous and paralogous genes.

2.4.2 Ribosomal ITS sequences in molecular phylogenetics

Nuclear ribosomal RNA (rRNA) genes are the most conserved and most utilized genes in eukaryotes. They occur in hundreds to thousands of copies and are organized in head-to-tail tandem arrays in two separate genomic loci. 5S rDNA encoding 5S rRNA genes occur in hundreds of copies and are usually localized on several chromosome pairs. 45S rDNA locus is localized on one or more chromosome pairs in the secondary constriction and contains coding regions for 18S, 5.8S and 26S rRNAs which are separated by internal transcribed spacers (ITS1 and ITS2) (Figure 11). These genes are transcribed by RNA polymerase I as a single unit which is further processed to generate 18S, 5.8S and 26S rRNA molecules. 45S rDNA gene blocks are separated from each other by intergenic spacers (IGS) consisting of a 5' end and 3' end external transcribed spacers (ETS) separated by a non-transcribed region (NTS).

The ITS region (consisting of ITS1, 5.8S rDNA and ITS2) is one of the most popular genomic loci in phylogenetic studies in plants. Several unique features make the ITS region an ideal target for molecular phylogenetics. The length of the region, varying between 500 and 700 bp in angiosperms (Baldwin *et al.* 1995; Feliner and Rosselló 2007), the possibility to use universal primers designed to conserved surrounding sequences (White *et al.* 1990) and high copy number facilitate PCR amplification and subsequent sequencing. The DNA region is highly variable due to frequently occurring nucleotide polymorphisms or insertions/deletions in the sequence and is phylogenetic inference at the specific and generic levels (Baldwin 1992, Baldwin *et al.* 1995). Other advantageous features are the biparental inheritance compared to the uniparental inheritance of chloroplast and mitochondrial markers and the intragenomic sequence uniformity caused by active homogenization

of repeat units, known as concerted evolution (Zimmer *et al.* 1980; Elder and Turner 1995).

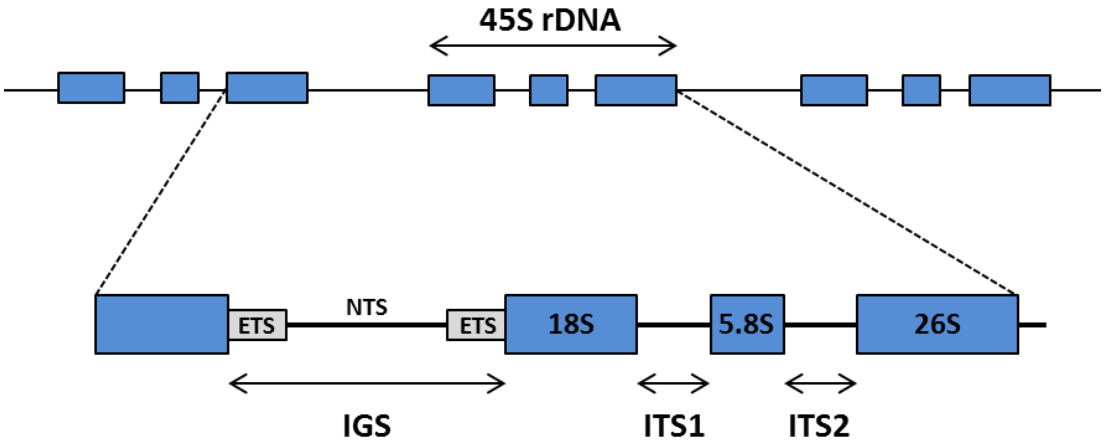


Figure 11: Schematic structure of the 45S rDNA locus. Genes coding for 18S, 5.8S and 26S rRNA are separated by internal transcribed spacers (ITS1and ITS2). Each block of 45S rDNA is separated by an intergenic spacer (IGS) which contains a non-transcribed region (NTS) and two external transcribed spacers (ETS) on both sides of NTS.

Owing to the above mentioned properties, the ITS region became one of the most widely used sequences for phylogenetic inferences. However, it is important to pay attention to some molecular processes that impact ITS sequence variation to avoid erroneous results of the analysis.

2.4.2.1 Concerted evolution

Ribosomal genes are one of the best-known examples of multigene family subjected to concerted evolution. According to this model, all copies within and among ribosomal loci evolve more or less in unison. Mutation occurring in a repeat spreads through the array via mechanisms of unequal crossing over and high-frequency gene conversion (Dover 1982; Elder and Turner 1995; Baldwin *et al.* 1995). While rRNA genes are extremely conserved among organisms, ITS1 and ITS 2 which are under lower selection pressure allow accumulation of mutations and their spreading by processes of homogenization (Álvarez and Wendel 2003).

However, concerted evolution may not be complete after processes such as hybridization or polyploidization. Various fates of rDNA loci have been observed in hybrids and allopolyploids. One possibility is that the divergent parental sequences remain conserved and evolve independently without any interaction. In this case, ITS

sequence data may be very informative and document the historical hybridization or polyploidization events. This type of evolutionary scenario was found in *Arabidopsis* (O'Kane *et al.* 1996), *Brassica* (Bennet and Smith 1991) or *Silene* (Popp and Oxelman 2001). The second possibility is that parental loci recombine, which results in chimerical ITS sequence (Barkman and Simpson 2002; Volkov *et al.* 2007). The third possible evolutionary pathway is that one repeat type dominates the rDNA repeat population within a genome. This can lead to partial or even complete homogenization of rDNA locus of the second parent (Lim *et al.* 2004; Dadejová *et al.* 2007).

2.4.2.2 Ribosomal DNA divergent paralogues and pseudogenes

The fundamental requirement for phylogenetic analysis based on nucleic acid sequences is that the compared genes are ortologous, which means that their relationship originated from organismal cladogenesis. However, multiple rDNA arrays may also include paralogous genes as a result of gene or array duplication. Paralogy can occur at many levels – among sequences within an individual, among individuals within a species, and among species. Inclusion of paralogous genes into phylogenetic analysis may then lead to phylogenetic incongruence (Wendel and Doyle 1998). Because of the possible presence of ortologous and paralogous rDNA genes in the genome, cloning and sequencing of individual ITS sequences is required.

As mentioned earlier, 18S-5.8S-26S rDNA repeats exist in thousand of copies and can occur in more than one chromosomal locus. If concerted evolution is not complete, duplicate ribosomal loci do not necessarily remain funtional and some arrays may degenerate into silenced rDNA sequences reffered to as pseudogenes. Negligence of their presence in dataset can lead to wrong inference in phylogenetic relationships (Mayol and Rosselló 2001; Álvarez and Wendel 2003). Pseudogenes can be identified "*in silico*" based on GC content, presence of large indels or changes in predicted secondary structure (Buckler *et al.* 1997).

On the other hand, several recent studies have also shown that pseudogenes can be useful in phylogenetic analysis of related species where functional copies of ITS region provide limited source of variation (Razafimandimbison *et al.* 2004; Ochieng *et al.* 2007).

2.4.2.3 Secondary structure of ITS2 and 5.8S rRNA transcripts

45S rDNA locus is transcribed as a single unit which contains 18S, 5.8S and 26S rRNA genes and both internal transcribed spacers. Although the ITS1 and ITS2 are not incorporated into the mature rRNA transcripts, they encode signals for correct processing and folding of rRNA transcripts (Hillis and Dixon 1991; Hadjiolova *et al.* 1994). It has been shown that secondary structures of functional ITS2 and 5.8S rDNA sequences are highly conserved within plants (Herskovitz and Zimmer 1996; Jobes and Thien 1997).

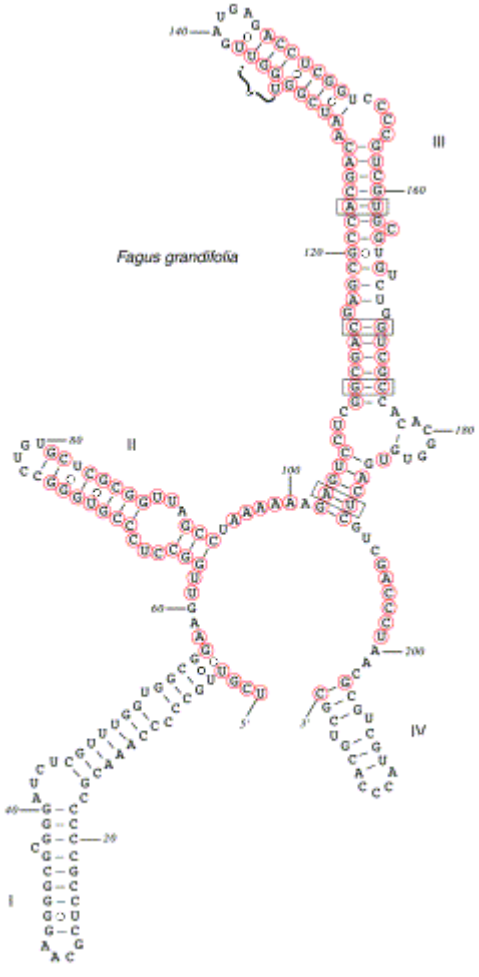


Figure 12: Diagram of secondary structure of ITS2 transcript of *Fagus grandifolia*. The four helices are labeled I-IV. Characteristic pyrimidine-pyrimidine bulge in helix II is obvious. The UGGU motif on helix III common to angiosperms is indicated by a bracket (Coleman 2003).

ITS2 spacer forms a typical four helices structure (Figure 12). Helix II is characteristic by a presence of pyrimidine-pyrimidine bulge. Helix III is the longest and contains the most conserved primary sequence on the 5' side with the UGGU motif (Mai and Coleman 1997; Schultz *et al.* 2005). Helices I and IV are more variable and only helices II and III are recognizable and common to essentially all eukaryotes (Coleman 2007). The 5.8S rDNA sequences contain three conserved motives which are neccessary for proper folding of secondary structure (Harpke and Peterson 2008). The conserved secondary structure of ITS2 and 5.8S rRNA transcripts enable identification of non-functional pseudogenes (Ochieng *et al.* 2007).

2.4.3 Diversity and phylogeny of the *Musaceae* family

The family *Musaceae* contains approximately 65 species and more than thousand cultivars and landraces which emerged during several thousands of years of banana domestication. Elucidating genetic diversity and phylogeny of the *Musaceae* family and especially of the genus *Musa* is of great importance for the collection and utilisation of genetic resources for further banana improvement.

During the past decades, a variety of molecular markers have been used to assess the genetic diversity of bananas. However, vast majority of the studies describes relatively narrow set of genotypes, which often contained locally important varieties. Wild species of Rhodochlamys, Australimusa and Callimusa sections seem to be out of focus. Grapin *et al.* (1998) used microsatellite loci to study genetic diversity among genotypes of wild or cultivated diploid *Musa acuminata* accessions. RAPD markers were employed in analysis of genetic diversity of East African highland bananas (Pillay *et al.* 2001). Microsatellite markers were used to characterize genotypes cultivated in Brazil (Creste *et al.* 2003) and plantain landraces (Noyer *et al.* 2005). Wang *et al.* (2007) assessed the genetic diversity of wild *Musa balbisiana* natural populations in China by AFLP markers.

In several studies, a broader range of accessions representing the four main sections of the genus *Musa* was used. One of these studies employed AFLP markers to analyze genetic diversity of 39 accessions and to assess sectional relationships in *Musa* (Ude *et al.* 2002). Risterucci *et al.* (2009) demonstrated the usefulness of DArT markers for the assessment of genetic diversity within *Musa* spp. However, this high-throughput approach, which is suitable for analysis of large numbers of

genotypes, is disadvantageous in cases of small sample data sets. Christelová *et al.* (2011) developed a platform for efficient genotyping in *Musa* which is based on employment of 19 microsatellite markers. This genotyping system provides a robust and reproducible approach to characterize the genetic variability of *Musa* germplasm.

The first phylogenetic studies of the *Musaceae* family used different types of molecular markers including RFLP (Jarret *et al.* 1992; Gawel *et al.* 1992) or RAPD markers (Pillay *et al.* 2001; Ruangsuttapha *et al.* 2007). Based on AFLP analysis Wong *et al.* (2002) suggested that the species of section Rhodochlamys should be combined into a single section with species of section Eumusa, and that sections Callimusa and Australimusa should be merged. These suggestions were further supported by the evidence from genome and chromosomes research (Nwakanma *et al.* 2003; Bartoš *et al.* 2005). With the rapid develoment of sequencing technologies, phylogenetic studies based on sequencing and analysis of specific regions have emerged. Chloroplast DNA sequences in combination with ITS region were used in two studies focused on molecular phylogeny of the family *Musaceae* (Liu *et al.* 2010; Li *et al.* 2010). Both studies confirmed monophyletic origin of the family *Musaceae* and further supported results showing close relationship of species from sections Eumusa and Rhodochlamys, and sections Australimusa and Callimusa. The shortcoming of these studies is the absence of sequence analysis of cloned ITS variants which would show potential heterogenity and exclude paralogous and pseudogenic sequences. More detailed knowledge of of the ITS region is needed to assess the utility of ITS region for the analysis of genetic diversity and phylogenetic relationships.

Most recently, Christelová *et al.* (2011) used gene-based markers to study the phylogeny of the family *Musaceae* and to estimate divergence times of the *Musaceae* genera and *Musa* sections. Their results confirmed close relationship of sections Australimusa and Callimusa and supported previous suggestions to merge Eumusa and Rhodochlamys sections into one. Divergence time of the family *Musaceae* was placed in the early Paleocene (69.1 Mya) and the age of *Musa* genus was estimated to be ~ 50 Mya. Speciation events within both Australimusa/Callimusa and Rhodochlamys/Eumusa were estimated to some 28 Mya and section Australimusa was shown to be evolutionarily youngest group.

2.5 REFERENCES

Aert R, Sági L, Volckaert G (2004) Gene content and density in banana (*Musa acuminata*) as revealed by genomic sequencing of BAC clones. *Theoretical and Applied Genetics* **109:** 129-139.

Agarwal M, Shrivastava N, Padh H (2008) Advances in molecular techniques and their applications in plant science. *Plant Cell Reports* **27:** 617-631.

Akbari M, Wenzl P, Caig V, Carling J, Xia L, Yang S, Uszynski G, Mohler V, Lehmensiek A, Kuchel H, Hayden MJ, Howes N, Sharp P, Vaughan P, Rathmell B, Huttner E, Kilian A (2006) Diversity arrays technology (DArT) for high-throughput profiling of hexaploid wheat genome. *Theoretical and Applied Genetics* **113(8):** 1409-1420.

Álvarez I, Wendel JF (2003) Ribosomal ITS sequences and plant phylogenetic inference. *Molecular and Phylogenetic Evolution* **29:** 417-434.

Argent GCG (1976) The wild bananas of Papua New Guinea. *Notes Royal Botanic* Garden Edinburgh **35:** 77-114.

Arumuganathan K, Earle ED (1991) Nuclear DNA content of some important plant species. *Plant Molecular Biology Reporter* **9:** 208-218.

Aston C, Mishra B, Schwarz DC (1999) Optical mapping and its potential for large-scale sequencing projects. *Trends in Biotechnology* **17(7):** 297-302.

Bailey CD, Carr TG, Harris SA, Hughes CE (2003) Characterization of angiosperm nrDNA polymorphism, paralogy, and pseudogenes. *Molecular Phylogenetics and Evolution* **29:** 435-455.

Balasubramanian S, Harrison P, Hegyi H, Bertone P, Luscombe N, Echols N, McGarvey P, Zhang Z, Gerstein M (2002) SNPs on human chromosomes 21 and 22 – analysis in terms of protein features and pseudogenes. *Pharmacogenomics* **3(3):** 393-402.

Baldwin BG (1992) Phylogenetic utility of the internal transcribed spacers of nuclear ribosomal DNA in plants: an example from the Compositae. *Molecular Phylogenetics and Evolution* **1:** 3-16.

Baldwin BG, Sanderson MJ, Porter JM, Wojcjechowski MF, Campbell CS, Donoghue MJ (1995) The ITS region of nuclear ribosomal DNA: a valuable source of evidence on angiosperm phylogeny. *Annals of the Missouri Botanical Garden* **82:** 247-277.

Balint-Kurti PJ, Clendennen SK, Doleželová M, Valárik M, Doležel J, Beetham PR, May GD (2002) *Molecular and General Genetics* **263:** 908-915.

Barakat A, Matassi G, Bernardi G (1998) Distribution of genes in the genome of Arabidopsis thaliana and its implication for the genome organization of plants. *Proceedings of the National Academy of Sciences U.S.A.* **95:** 10044-10049.

Barkman TJ, Simpson BB (2002) Hybrid origin and parentage of *Dendrochilum acuiferum* (*Orchidaceae*) inferred in a phylogenetic context using nuclear and plastid DNA sequence data. *Systematic Botany* **27(2):** 209-220.

Bartoš J, Alkhimova O., Doleželová M, De Langhe E, Doležel J (2005) Nuclear genome size and genomic distribution of ribosomal DNA in *Musa* and *Ensete* (Musaceae): taxonomic implications. *Cytogenetic and Genome Research* **109:** 50-57.

Bennett MD, Smith JB (1976) Nuclear DNA amounts in angiosperms. *Philosophical Transactions of the Royal Society in London*, Series B **274:** 227-274.

Bennett MD, Leitch IJ (2005) Nuclear DNA amounts in angiosperms: progress, problems and prospects. *Annals of Botany* **95:** 45-90.

Bennett RI, Smith AG (1991) Use of a genomic clone for ribosomal RNA from *Brassica oleracea* in RFLP analysis of *Brassica* species. *Plant Molecular Biology* **16(4):** 685-688.

Birky Jr. CW (1995) Uniparental inheritance of mitochondrial and chloroplast genes: mechanisms and evolution. *Proceedings of the National Academy of Sciences U.S.A.* **92:** 11331-11338.

Botstein D, White RL, Skolnick M, Davis RW (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *The American Journal of Human Genetics* **32:** 314-333.

Breyne P, RombautD, Van Gysel A, Van Montagu M, Gerats T (1999) AFLP analysis of genetic diversity within and between Arabidopsis thaliana ecotypes. *Molecular and General Genetics* **261:** 627-634.

Buckler ES, Ippolito A, Holtsford TP (1997) The evolution of ribosomal DNA: divergent paralogues and phylogenetic implications. *Genetics* **145:** 821-832.

Bureau TE, Wessler SR (1992) Tourist: A large family of small inverted repeat elements frequently associated with maize genes. *The Plant Cell* **4:** 1283-1294.

Chandler S (1995) The nutritional value of bananas. In: Gowen S (1995) Bananas and plantains. Chapman & Hall, London, UK.

Chase MW, Soltis DE, Olmstead RG, Morgan D, Les DH, Mishler BD, Duvall MR, Price RA, Hills HG, Qiu Y-L, Kron KA, Rettig JH, Conti E, Palmer JD, Manhart JR, Sytsma KJ, Michales HJ, Kress WJ, Karol KG, Clark WD, Hedrén M, Gaut BS, Jansen RK, Kim K-J, Wimpee CF, Smith JF, Furnier GR, Strauss SH, Xiang Q-Y, Plunkett GM, Soltis PS, Swensen SM, Williams SE, Gadek PA, Quinn CJ, Eguiarte LE, Golenberg E, Learn Jr. GH, Graham SW, Barrett SCH, Dayanandan S, Albert VA (1993) Phylogenetics of seed platns: an analysis of nucleotide sequences from tha plastid gene *rbc*L. *Annals of the Missouri Botanical Garden* **80:** 528-580.

Cheesman EE (1947) Classification of the bananas II. The genus *Musa* L. *Kew Bulletin* **2:** 106-117.

Cheng Z, Presting GG, Buell CR, Wing RA, Jiang J (2001) High-resolution pachytene chromosome mapping of bacterial artificial chromosomes anchored by genetic markers reveals the centromere location and the distribution of genetic recombination along chromosome 10 of rice. *Genetics* **157:** 1749-1757.

Cheng Z, Buell CR, Wing RA, Jiang J (2002) Resolution of fluorescence in-situ hybridization mapping on rice mitotic prometaphase chromosomes, meiotic pachytene chromosomes and extended DNA fibres. *Chromosome Research* **10:** 379-387.

Cheung F, Town CD (2007) A BAC end view of the *Musa acuminata* genome. *BMC Plant Biology* **7:** 29.

Christelová P, Valárik M, Hřibová E, De Langhe E, Doležel J (2011) A multi gene sequence-based phylogeny of the *Musaceae* (banana) family. *BMC Evolutionary Biology* **11:** 103.

Christelová P, Valárik M, Hřibová E, Van den houwe I, Channelière S, Roux N, Doležel J (2011) A platform for efficient genotyping in Musa using microsatellite markers. *AoB PLANTS* plr024 doi:10.1093/oabpla/plr024.

Churchill ACL (2011) Mycosphaerella fijiensis, the black leaf streak pathogen of banana: progress towards understanding pathogen biology and detection, disease development, and challenges of control. *Molecular Plant Pathology* **12(4):** 307-328.

Coleman AW (2003) ITS2 is a double-edged tool for eukaryote evolutionary comparisons. *Trends in Genomics* **7:** 370-375.

Coleman AW (2007) Pan-eukaryote ITS2 homologies revealed by RNA secondary structure. *Nucleic Acids Research* **35(10):** 3322-3329.

Cox DR, Burmeister M, Price ER, Kim S, Myers RM (1990) Radiation hybrid mapping: a somatic cell genetic method for constructing high-resolution maps of mammalian chromosomes. *Science* **250:** 245-250.

Creste S, Neto AT, de Oliveira Silva S, Figueira A (2003) Genetic characterization of banana cultivars (*Musa* spp.) from Brazil using microsatellite markers. *Euphytica* **132:** 259-268.

Dadejová M, Lim KY, Součková-Skalická K, Matyášek R, Grandbastien M-A, Leitch A, Kovařík A (2007) Transcription activity of rRNA genes correlates with a tendency towards intergenomic homogenization in *Nicotiana* allotetraploids. *New Phytologist* **174:** 658-668.

Dahlgren RMT, Clifford HT Yeo PF (1985) The Families of the Monocotyledons: structure, evolution and taxonomy. Springer-Verlag, Berlin, Germany.

Danilova T, Birchler J (2008) Integrated cytogenetic map of mitotic metaphase chromosome 9 of maize: resolution, sensitivity, and banding paint development. *Chromosoma* **117(4):** 345-356.

De Jong JH, Fransz P, Zabel P (1999) High resolution FISH in plants – techniques and applications. *Trends in Plant Science* **99:** 1360-1385.

De Langhe E, Vrydaghs L, De Maret P, Perrier X, Denham T (2009) Why bananas matter: An introduction to the history of banana domestication. *Ethnobotany Research & Applications* **7:** 165-177.

D'Hont A, Paget-Goy A, Escoute J, Carreel F (2000) The interspecific genome structure of cultivated banana, *Musa* spp. Revealed by genomic DNA *in situ* hybridization. *Theoretical and Applied Genetics* **100:** 177-183.

D'Hont A, Denoeud F, Aury JM, Baurens FC, Carreel F, Garsmeur O, Noel B, Bocs S, Droc G, Rouard M, Da Silva C, Jabbari K, Cardi C, Poulain J, Souquet M, Labadie K, Jourda C, Lengellé J, Rodier-Goud M, Alberti A, Bernard M, Correa M, Ayyampalayam S, Mckain MR, Leebens-Mack J, Burgess D, Freeling M, Mbéguié-A-Mbéguié D, Chabannes M, Wicker T, Panaud O, Barbosa J, Hribova E, Heslop-Harrison P, Habas R, Rivallan R, Francois P, Poiron C, Kilian A, Burthia D, Jenny C, Bakry F, Brown S, Guignon V, Kema G, Dita M, Waalwijk C, Joseph S, Dievart A, Jaillon O, Leclercq J, Argout X, Lyons E, Almeida A, Jeridi M, Dolezel J, Roux N, Risterucci AM, Weissenbach J, Ruiz M, Glaszmann JC, Quétier F, Yahiaoui N, Wincker P (2012) The banana (*Musa acuminata*) genome and the evolution of monocotyledonous plants. *Nature* **488:** 213-217.

Dijkhuzien A, Kennard WC, Havey MJ, Staub JE (1996) RFLP variation and genetic relationships in cultivated cucumber. *Euphytica* **90:** 79-87.

Doležel J, Doleželová M, Novak FJ (1994) Flow cytometric estimation of nuclear DNA amount in diploid ananas (*Musa acuminata* and *M. balbisiana*). *Biologia Plantarum* **36:** 351-357.

Doležel J, Doleželová M, Roux N, Van den houwe I (1998) A novel method to prepare slides for high resolution chromosome studies in *Musa* spp. *Infomusa* **7:** 3-4.

Doležel J, Bartoš J, Voglmayr H, Greilhuber J (2003) Nuclear DNA content and genome size of trout and human. *Cytometry A* **51:** 127-128.

Doležel J, Bartoš J (2005) Plant DNA flow cytometry and estimation of nuclear genome size. *Annals of Botany* **95:** 99-110.

Doležel J, Greilhuber J, Suda J (2007) Estimation of nuclear DNA content in plants using flow cytometry. *Nature Protocols* **2(9):** 2233-2244, doi:10.1038/nprot.2007.310.

Doleželová M, Valárik M, Swennen R, Horry JP, Doležel J (1998) Physical mapping of the 18S-25S and 5S ribosomal RNA genes in diploid bananas. *Biologia Plantarum* **41(4):** 497-505.

Dong F, Song J, Naess SK, Helgeson JP, Gebhardt C, Jiang J (2000) Developments and application of a set of chromosome-specific cytogenetic DNA markers in potato. *Theoretical and Applied Genetics* **101:** 1001-1007.

Dover G (1982) Molecular drive: a cohesive mode of speciation and evolution. *Nature* **299:** 111-117.

Dover G (2002) Molecular drive. *Trends in Genetics* **18(11):** 587-589.

Elder JF, Turner BJ (1995) Concerted evolution of repetitive DNA sequences in eukaryotes. *Quarterly Review of Biology* **70:** 297-320.

Fedoroff N, Wessler S, Shure M (1983) Isolation of the transposable maize controlling elements *Ac* and *Ds*. *Cell* **35:** 235-242.

Feliner N, Rosselló JA (2007) Better the devil you know? Guidelines for insightul utilization of nrDNA ITS in species-level evolutionary studies in plants. *Molecular Phylogenetics and Evolution* **44(2):** 911-919.

Feschotte C, Pritham EJ (2007) DNA transposons and the evolution of eukaryotic genomes. *Annual Review of Genetics* **41:** 331-368.

Feuillet C, Keller B (1999) High gene density is conserved at syntetic loci of small and large grass genomes. *Proceedings of the National Academy of Sciences U.S.A.* **96:** 8265-8270.

Flagel LE, Wendel JF (2009) Gene duplication and evolutionary novelty in plants. *New Phytologist* **183:** 557-564.

Fransz PF, Alonso-Blanco C, Liharska TB, Peeters AJM, Zabel P, de Jong JH (1996) High resolution physical mapping in *Arabidopsis thaliana* and tomato by fluorescence *in situ* hybridization to extended DNA fibres. *The Plant Journal* **9:** 421-430.

Freeling M (2001) Grasses as a single genetic system. Reassessment 2001. *Plant Physiology* **125:** 1191-1197.

Gadaleta A, Giancaspro A, Giove SL, Zacheo S, Mangini G, Someone R, Signorile A, Blanco A (2009) Genetic and physical mapping of new EST-derived SSRs on the A and B genome chromosomes of wheat. *Theoretical and Applied Genetics* **118:** 1015-1025.

Gall JG, Pardue ML (1969) Formation and detection of RNA-DA hybrid molecules in cytological preparations. *Proceedings of the National Academy of Sciences U.S.A.* **69:** 378-383.

Gao W, Chen ZJ, Yu JZ, Raska D, Kohel RJ, Womack JE, Stelly DM (2004) Wide-cross whole-genome radiation hybrid mapping of cotton (*Gossypium hirsutum* L.). *Genetics* **167:** 1317-1329.

Gawel NJ, Jarret RL (1991) Chloroplast DNA restriction fragment length polymorphisms (RFLP) in *Musa* species. *Theoretical and Applied Genetics* **81:** 783-786.

Gawel NJ, Jaret RL, Whittemore AP (1992) Restriction fragment length polymorphisms (RFLP)-based phylogenetic analysis in *Musa*. *Theoretical and Applied Genetics* **84:** 286-290.

Gielly L, Taberlet P (1994) The use of chloroplast DNA to resolve plant phylogenies: noncoding versus *rbc*L sequences. *Molecular Biology and Evolution* **11(5):** 769-777.

Gindulis F, Desel C, Galasso I, Schmidt T (2001) The large-scale organization of the centromeric region in beta species. *Genome Research* **11:** 253-265.

Grapin A, Noyer J-L, Carreel F, Dambier D, Baurens F-C, Lanaud C, Lagoda PJL (1998) Diploid *Musa acuminata* genetic diversity assayed with sequence-tagged microsatellite sites. *Electrophoresis* **19:** 1374-1380.

Gregory TR (2001) Coincidence, coevolution, or causation? DNA content, cell size, and C-value enigma. *Biological reviews* **76(1):** 65-101.

Greilhuber J, Borsch T, Muller K, Worberq A, Porembski S, Barthlott W (2006) Smallest angiosperm genomes found in lentibulariaceae, with chromosomes of bacterial size. *Plant Biology* **8(6):** 770-777.

Greilhuber J, Doležel J, Leitch IJ, Loureiro J, Suda J (2010) Genome Size. *Journal of Botany* doi:10.1155/2010/946138.

Gu YQ, Ma Y, Huo N, Vogel JP, You FM, Lazo GR, Nelson WM, Soderlund C, Dvorak J, Anderson OD, Luo MC (2009) A BAC-based physical map of *Brachypodium distachyon* and its comparative analysis with rice and wheat. *BMC Genomic* **10:** 496.

Hadjiolova KV, Norman A, Cavaillé J, Soupène E, Mazan S, Hadjiolov AA, Bachellerie JP (1994) Processing of truncated mouse or human rRNA transcribed from ribosomal minigenes transfected into mouse cells. *Molecular and Cellular Biology* **14(6):** 4044-4056.

Harper G, Osuji JO, Heslop-Harrison JS, Hull R (1999) Integration of banana streak badnavirus into the *Musa* genome: molecular and cytogenetic evidence. *Virology* **255:** 207-213.

Harpke D, Peterson A (2008) Extensive 5.8S nrDNA polymorphism in *Mammillaria* (*Cactaceae*) with special reference to the identification of pseudogenic internal transcribed spacer. *Journal of Plant Research* **121:** 261-270.

Harrison GE, Heslop-Harrison JS (1995) Centromeric repetitive DNA sequences in the genus Brassica. *Theoretical and Applied Genetics* **90:** 157-165.

Hartmann S, Nason JD, Bhattacharya D (2001) Extensive ribosomal DNA genic variation in the columnar cactus *Lophocereus*. *Journal of Molecular Evolution* **53:** 124-134.

Hass-Jacobus B, Jackson SA (2005) Physical mapping of plant chromosomes. In: Meksem K, Kahl G (2005) The handbook of plant genome mapping. WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim.

Hershkovitz MA, Zimmer EA (1996) Conservation patterns in angiosperm rDNA ITS2 sequences. *Nucleic Acids Research* **24(15):** 2857-2867.

Heslop-Harrison JS, Brandes A, Taketa S, Schmidt T, Vershinin AV, Alkhimova EG, Kamm A, Doudrick RL, Schwarzacher T, Katsiotis A, Kubis S, Kumar A, Pearce SR, Flavell AJ, Harrison GE (1997) The chromosomal distribution of Ty1-copia group retrotransposable elements in higher plants and their implications for genome evolution. *Genetica* **100:** 197-204.

Heslop-Harrison JS (2000) Comparative genome organization in plants: from sequence and markers to chromatin and chromosomes. *The Plant Cell* **12:** 617-635.

Heslop-Harrison JS, Schwarzacher T (2007) Domestication, genomics and the future for banana. *Annals of Botany* **100:** 1073-1084.

Heslop-Harrison JS, Schwarzacher T (2011) Organisation of the plant genome in chromosomes. *The Plant Journal* **66:** 18-33.

Hildebrand E (2007) A tale of two tuber crops: How attributes of Enset and Yams may have shaped prehistoric human-plant interactions in Southwest Ethiopia. In: Denham TP, Iriarte J, Vrydaghs L (2007) Rethinking agriculture: Archeological and ethnoarcheological perspectives. Left Coast Press, Walnut Creek.

Hillis DM, Dixon MT (1991) Ribosomal DNA: molecular evolution and phylogenetic inference. *The Quaterly Review of Biology* **66(6):** 410-453.

Hřibová E, Doleželová M, Town CD, Macas J, Doležel J (2007) Isolation and chracterization of the highly repeated fraction of the banana genome. *Cytogenetic and Genome Research* **119:** 268-274.

Hřibová E, Doleželová M, Doležel J (2008) Localization of BAC clones on mitotic chromosomes of *Musa acuminata* using fluorescence in situ hybridization. *Biologia Plantarum* **52(3):** 445-452.

Hřibová E, Neumann P, Matsumoto T, Roux N, Macas J, Doležel J (2010) Repetitive part of the banana (*Musa acuminata*) genome investigated by low-depth 454 sequencing. *BMC Plant Biology* **10:** 204.

http://bananas.bioversityinternational.org/

INIBAP: Global conservation strategy for *Musa* (Banana and Plantain). A consultative document prepared by INIBAP with the collaboration of numerous partners in the *Musa* research-and-development community. Montpellier: INIBAP, 2006.

Jaccoud D, Peng K, Feinstein D, Kilian A (2001) Diversity Arrays: a solid state technology for sequence information independent genotyping. *Nucleic Acids Research* **29(4):** e25.

Jackson RC (1959) A study of meiosis in Haplopappus gracilis. *American Journal of Botany* **46:** 550-554.

Jackson SA, Wang ML, Goodman HM, Jiang J (1998) Application of fiber-FISH in physical mapping of *Arabidopsis thaliana*. *Genome* **41(4):** 566-572.

Jarret RL, Gawel N, Whittemore A, Sharrock S (1992) RFLP-based phylogeny of *Musa* species in Papua New Guinea. *Theoretical and Applied Genetics* **84:** 579-584.

Jiang J, Gill BS, Wang GL, Ronald PC, Ward DC (1995) Metaphase and interphase fluorescence *in situ* hybridization mapping of the rice genome with bacterial artificial chromosomes. *Proceeding of The National Academy of Science U.S.A.* **92:** 4487-4491.

Jiang J, Gill BS (2006) Current status and the future of fluorescence in situ hybridization (FISH) in plant genome resarch. *Genome* **49:** 1057-1068.

Jiang J, Hulber SH, Gill BS, Ward DC (1996) Interphase fluorescence in situ hybridization mapping: a physical mapping strategy for plant species with large complex genomes. *Molecular and General Genetics* **252:** 497-502.

Jobes DV, Thien LB (1997) A conserved motif in the 5.8S ribosomal RNA (rRA) gene is a useful diagnostic marker for plant internal transcribed spacer (ITS) sequences. *Plant Molecular Biology Reporter* **15:** 326-334.

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

Kajikawa M, Okada N (2002) LINEs mobilize SINEs in the eel through a shared 3' sequence. *Cell* **111(3):** 433-444.

Källersjö M, Farris JS, Chase MW, Bremer B, Fay MF, Humphries CJ, Petersen G, Seberg O, Bremer K (1998) Simultaneous parsimony jackknife analysis of 2538 *rbc*L DNA sequences reveales support for major clades of green plants, land plants, seed plants and flowering plants. *Plant Systematics and Evolution* **213:** 259-287.

Kapitonov V, Jurka J (2001) Rolling-circle transposons in eukaryotes. *Proceeding of The National Academy of Science U.S.A.* **98:** 8714-8719.

Kato A, Lamb JC, Birchler JA (2004) Chromosome painting using repetitive DNA sequences as probes for somatic chromosome identification in maize. *Proceeding of The National Academy of Science U.S.A.* **101:** 13554-13559.

Kellog EA, Bennetzen JL (2004) The evolution of nuclear genome structure in seed plants. *American Journal of Botany* **91(10):** 1709-1725.

Khandelwal S (1990) Chromosome evolution in the genus *Ophioglossum* L. *Botanical Journal of the Linnean Society* **102:** 205-217.

Khrustaleva LI, Kik C (2001) Localization of single-copy T-DNA insertion in transgenic shallots (*Allium cepa*) by using ultra-sensitive FISH with tyramide signal amplification. *The Plant Journal* **25(6):** 699-707.

King J, Armstead IP, Donnison IS, Thomas HM, Jones RN, Kearsey MJ, Roberts LA, Thomas A, Morgan WG, King IP (2002) Physical and genetic mapping in the grasses *Lolium perenne* and *Fetuca pratensis*. *Genetics* **161:** 315-324.

Knoop V (2004) The mitochondrial DNA of land plants: peculiarities in phylogenetic perspective. *Current Genetics* **46:** 123-139.

Kopecký D, Bartoš J, Lukaszewski AJ, Baird JH, Černoch V, Kölliker R, Rognli OA, Blois H, Caig V, Lübberstedt T, Studer B, Shaw P, Doležel J, Kilian A (2009) Development and mapping of DArT markers within the *Festuca-Lolium* complex. *BMC Genomics* **10:** 473.

Kubaláková M, Kovářová P, Suchánková P, Čihalíková J, Bartoš J, Lucretti S, Watanabe N, Kianian SF, Doležel J (2005) Chromosome sorting in tetraploid wheat and its potential for genome analysis. *Genetics* **172(2):** 823-829.

Kubis S, Schmidt T, Heslop-Harrison JS (1998) Repetitive DNA elements as a major component of plant genomes. *Annals of Botany* **82:** 45-55.

Kulikova O, Gualtieri G, Geurts R, Kim D-J, Cook D, Huguet T, de Jong JH, Fransz PF, Bisseling T (2001) Integration of the FISH pachytene and genetic maps of *Medicago truncatula*. *The Plant Journal* **27(1):** 49-56.

Kumar A, Pearce SR, McLean K, Harrison G, Heslop-Harrison JS, Waugh R, Flavell AJ (1997) The Ty1-copia group of retrotransposons in plants: genomic organisation, evolution, and use as molecular markers. *Genetica* **100(1-3):** 205-217.

Kumar A, Bennetzen JL (1999) Plant retrotransposons. *Annual Reviews in Genetics* **33:** 479-532.

Kumar A, Simons K, Iqbal MJ, de Jiménez MM, Bassi FM, Ghavami F, Al-Azzam O, Drader T, Wang Y, Luo M-C, Gu YQ, Denton A, Lazo GR, Xu SS, Dvořák J, Kianian PMA, Kianian SF (2012) Physical mapping resources for large plant genomes: radiation hybrids for wheat D-genome progenitor *Aegilops tauschii*. *BMC Genomics* **13:** 597.

Lamb JC, Danilova T, Bauer MJ, Meer JM, Holland JJ, Jensen MD, Birchler JA (2007) Single-gene detection and karyotyping using small-target fluorescence in situ hybridization on maize somatic chromosomes. *Genetics* 175: 1047-1058.

Li HW: (1978) The Musaceae of Yunnan. *Acta Phytotaxonomica Sinica* **16:** 54-64.

Li LF, Häkkinen M, Yuan YM, Hao G, Ge XJ (2010) Molecular phylogeny and systematics of the banana family (Musaceae) inferred from multiple nuclear and chloroplast DNA fragments, with a special reference to the genus *Musa*. *Molecular Phylogenetics and Evolution* **57:** 1-10.

Lin XY, Kaul SS, Rounsley S, Shea TP, Benito MI, Town CD, Fujii CY, Mason T, Bowman CL, Barnstead M, Feldblyum TV, Buell CR, Ketchum KA, Lee J, Ronning CM, Koo HL, Moffat KS, Cronin LA, Pai G, Van Aken S, Umayam L, Tallon LJ, Gill JE, Adams MD, Carrera AJ, Creasy TH, Goodman HM, Somerville CR, Copenhaver GP, Preuss D, Nierman WC, White O, Eisen JA, Salzberg SL, Fraser CM, Venter JC (1999) Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana. *Nature* **402:** 761-768.

Liu AZ, Kress WJ, Li DZ (2010) Phylogenetic analyses of the banana family (*Musaceae*) based on nuclear ribosomal (ITS) and chloroplast (trnL-F) evidence. *Taxon* **59:** 20-28.

Liu Z, Yue W, Li D, Wang RR, Kong X, Lu K, Wang G, Dong Y, Jin W, Zhang X (2008) Structure and dynamics of retrotransposons at wheat centromeres and pericentromeres. *Chromosoma* **117(5):** 445-456.

Lu J, Knix MR, Ambrose MJ, Brown JKM, Ellis THN (1996) Comparative analysis of genetic diversity in pea assessed by RFLP- and PCR-based methods. *Theoretical and Applied Genetics* **93:** 1103-1111.

Luo M-C, Ma Y, You FM, Anderson OD, Kopecký D, Šimková H, Šafář J, Doležel J, Gill B, McGuire PE, Dvořák J (2010) Feasibility of physical map construction from fingerprinted bacterial artificial chromosome libraries of polyploid plant species. *BMC Genomics* **11:** 122.

Lysak MA, Dolezelova M, Horry JP, Swennen R, Doležel J (1999) Flow cytometric analysis of nuclear DNA content in *Musa*. *Theoretical and Applied Genetics* **98:** 1344-1350.

Ma J, Wing R, Bennetzen J, Jackson S (2007) Plant centromere organization: a dynamic structure with conserved functions. *Trends in Genetics* **23:** 134-139.

Macas J, Požárková D, Navrátilová A, Nouzová M, Neumann P (2000) Two new families of tandem repeats isolated from genus Vicia using genomic self-priming PCR. *Molecular Genetics and Genomics* **263(5):** 741-751.

Mai JC, Coleman AW (1997) The internal transcribed spacer 2 exhibits a common secondary structure in green algae and flowering plants. *Journal of Molecular Evolution* **44:** 258-271.

Maluszynska J, Heslop-Harrison JS (1991) Localization of tandemly repeated DNA sequences in *Arabidopsis thaliana*. *The Plant Journal* **1:** 159-166.

Mayol M, Rosselló JA (2001) Why nuclear ribosomal DNA spacers (ITS) tell different stories in *Quercus*. *Molecular Phylogenetics and Evolution* **19:** 167-176.

McCarthy EM, Liu J, Lizhi G, McDonald JF (2002) Long terminal repeat retrotransposons of *Oryza sativa*. *Genome Biology* **3(10):** research0053.1-0053.11.

Miller JT, Dong F, Jackson SA, Song J, Jiang J (1998) Retrotransposon-related DNA sequences in the centromeres of grass chromosomes. *Genetics* **150:** 1615-1623.

Miller WJ, McDonald JF, Nouaud D, Anxolabéhère D (1999) Molecular domestication – more than a sporadic episode in evolution. *Genetica* **107(1-3):** 197-207.

Mochizuki K, Umeda M, Ohtsubo H, Ohtsubo E (1992) Characterization of a plant SINE, p-SINE-1, in rice genomes. *Japanese Journal of Genetics* **57:** 155-166.

Moser H, Lee M (1994) RFLP variation and genealogical distance, multivariate distance, heterosis, and genetic variance in oats. *Theoretical and Applied Genetics* **87:** 947-956.

Mutti J, Sandhu D, Sidhu D, Gill K (2010) Dynamic nature of a wheat centromere with a functional gene. *Molecular Breeding* **26:** 177-187.

Nagaki K, Song J, Stupar RM, Parokonny AS, Yuan Q, Ouyang S, Liu J, Hsiao J, Jones KM, Dawe RK, Buell CR, Jiang J (2003) Molecular and cytological analysis of large tracks of centromeric DNA reveal the structure and evolutionary dynamics of maize centromeres. *Genetics* **163:** 759-770.

Navrátilová A, Koblížková A, Macas J (2008) Survey of extrachromosomal circular DNA derived from plant satellite repeats. *BMC Plant Biology* **8:** 90 doi:10.1186/1471-2229-8-90.

Noma K, Ohtsubo E, Ohtsubo H (1999) Non-LTR retrotransposons (LINEs) as ubiquitous components of plant genomes. *Molecular and General Genetics* **261:** 71-79.

Noyer JL, Causse S, Tomekpe K, Bouet A, Baurens FC (2005) A new image of plantain diversity assessed by SSR, AFLP and MSAP markers. *Genetica* **124:** 61-69.

Nwakanma DC, Pilay M, Okoli BE, Tenkouano A (2003) Sectional relationships in the genus *Musa* L. inferred from the PCR-RFLP of organelle DNA sequences. *Theoretical and Applied Genetics* **107:** 850-856.

Ochieng JW, Henry RJ, Baverstock PR, Steane DA, Shepherd M (2007) Nuclear ribosomal pseudogenes resolve a corroborated monophyly of the eucalypt genus *Corymbia* despite misleading hypotheses at functional ITS paralogs. *Molecular Phylogenetics and Evolution* **44:** 752-764.

Ohmido N, Akiyama Y, Fukui K (1998) Physical mapping of unique nucleotide sequences on identified rice chromosomes. *Plant Molecular Biology* **38:** 1043-1052.

O'Kane SL, Schaal BA, Al-Shehbaz IA (1996) The origins of *Arabidopsis suecica* (*Brassicaceae*) as indicated by nuclear rDNA sequences. *Systematic Botany* **21(4):** 559-566.

Osuji JO, Harrison G, Crouch J, Heslop-Harrison JS (1998) Identification of the genomic constitution of *Musa* L. lines (bananas, plantains and hybrids) using molecular cytogenetics. *Annals of Botany* **80:** 787-793.

Osuji JO, Crouch J, Harrison G, Heslop-Harrison JS (1998) Molecular cytogenetics of Musa species, cultivars and hybrids: location of 18S-5.8S-25S and 5S rDNA and telomere-like sequences. *Annals of Botany* **82:** 243-248.

Palmer JD, Herbon LA (1988) Plant mitochondrial DNA evolves rapidly in structure, but slowly in sequnce. *Journal of Molecular Evolution* **28:** 87-97.

Paterson AH, Bowers JE, Bruggman R, Dubchak I, Grimwood J, Gundlach H, Haberer G, Hellsten U, Mitros T, Poliakov A, Schmutz J, Spannagl M, Tang H, Wang X, Wicker T, Bharti AK, Chapman J, Feltus FA, Gowik U, Grigoriev IV, Lyons E, Maher CA, Martis M, Narechania A, Otillar RP, Penning BW, Salamov AA, Wang Y, Zhang L, Carpita NC, Freeling M, Gingle AR, Hash CT, Keller B, Klein P, Kresovich S, McCann MC, Ming R, Peterson DG, Mehboob-ur-Rahman, Ware D, Westhoff P, Mayer KF, Messing J, Rokhsar DS (2009) The Sorghum bicolor genome and the diversification of grasses. *Nature* **457:** 551-556.

Paux E, Sourdille P, Salse J, Saintenac C, Choulet F, Leroy P, Korol A, Michalak M, Kianian S, Spielmayer W, Lagudah E, Somers D, Kilian A, Alaux M, Vautrin S, Berges H, Eversole K, Appels R, Šafář J, Šimková H, Doležel J, Bernard M, Feuillet C (2008) A physical map of the 1-gigabase bread wheat chromosome 3B. *Science* **322:** 101-104.

Pearce SR, Harrison G, Li D, Heslop-Harrison JS, Kumar A, Flavell AJ (1996) The Ty1-copia group retrotransposons in Vicia species: copy number, sequence heterogeneity and chromosomal localization. *Molecular and General Genetics* **250:** 305-315.

Pellicer J, Fay MF, Leitch IJ (2010) The largest eukaryotic genome of them all? *Botanical Journal of the Linnean Society* **164:** 10-15.

Pereira V (2004) Insertion bias and purifying selection of retrotransposons in the *Arabidopsis thaliana* genome. *Genome Biology* **5:** R79.

Perrier X, Bakry F, Carreel F, Jenny C, Horry JP, Lebot V, Hippolyte I (2009) Combining biological approaches to shed light on the evolution of edible bananas. *Ethnobotany Research & Applications* **7:** 199-216.

Pessoa-Filho M, Beló A, Alcochete AAN, Rangel PHN, Ferreira ME (2007) A set of multiplex panels of microsatellite markers for rapid molecular characterization of rice accessions. *BMC Plant Biology* **7:** 23. doi:10.1186/1471-2229-7-23

Peterson-Burch BD, Wright DA, Laten HM, Voytas DF (2000) Retroviruses in plants? *Trends in Genetics* **16(4):** 151-152.

Pillay M, Ogundiwin E, Nwakanma DC, Ude G, Tenkouano A (2001) Analysis of genetic diversity and relationships in East African banana germplasm. *Theoretical and Applied Genetics* **102:** 965-970.

Pinkel D, Straume T, Gray JW (1986) Cytogenetic analysis using quantitative, high-sensitivity, fluorescence hybridization. *Proceeding of The National Academy of Science U.S.A.* **83:** 2934-2938.

Ploetz RC, Kepler AK, Daniells J, Nelson SC (2007) Bananas and plantains – an overview with emphasis on the Pacific island cultivars, ver 1. In: Elevitch CR (2007) Species Profiles for Pacific Island Agroforestry. Permanent Agriculture Resources (PAR), Holualoa, Hawaii.

Popp M, Oxelman B (2001) Inferring the history of the polyploid *Silene aegaea* (C*aryophyllaceae*) using plastid and homoeologous nuclear DNA sequences. *Molecular Phylogenetics and Evolution* **20(3):** 474-481.

Presting GG, Malysheva L, Fuchs J, Schubert I (1998) A Ty3/gypsy retrotransposon-like sequence localises to the centromeric region of cereal chromosomes. *The Plant Journal* **16:** 721-728.

Price NS (1995) The origin and development of banana and plantain cultivation. 1-13. In: Gowen S (1995) Bananas and plantains. Chapman & Hall, London, UK.

Pritham EJ, Putliwala T, Feschotte C (2007) Mavericks, a novel class of giant transposable elements widespread in eukaryotes and related to DNA viruses. *Gene* **390(1-2):** 3-17.

Ramsey J, Schemske DW (1998) Pathways, mechanisms, and rates of polyploid formation in flowering plants. *Annual Review of Ecology, Evolution, and Systematics* **29:** 467-501.

Rayburn AL, Gill BS (1985) Use of bioti-labeled probes to map specific DNA sequences on wheat chromosomes. *The Journal of Heredity* **76:** 78-81.

Razafimandimbison SG, Kellog EA, Bremer B (2004) Recent origin and phylogenetic utility of divergent ITS putative pseudogenes: a case of study from *Naucleeae* (*Rubiaceae*). *Systematic Biology* **53:** 177-192.

Riera-Lizarazu O, Vales MI, Ananiev EV, Rines HW, Phillips RL (2000) Production and characterization of maize chromosome 9 radiation hybrids derived from and oat-maize addition line. *Genetics* **156:** 327-339.

Risterucci AM, Hippolyte I, Perrier X, Xia L, Caig V, Evers M, Huttner E, Kilian A, Glaszman JC (2009) Development and assessment of Diversity Arrays

Technology for high-throughput DNA analysis in *Musa*. *Theoretical and Applied Genetics* **119:** 1093-1103.

Robinson JC (1996) Bananas and plantains. UK University Press, Cambridge.

Rodriguez-Zapata LC, Castaño E, Acereto-Escoffié POM, Chi-Manzanero BH, Echeverría S, Grijalva R, James-Kay A, González-Estrada T (2005) *Agrobacterium*-mediated transformation of *Musa acuminata* cv. 'Grand Nain' scalps by vacuum infiltration. *Scientia Horticulturae* **105:** 359-371.

Roux NS (2004) Mutation induction in *Musa* – review. In: Jain SM, Swennen R (2004) Banana improvement: cellular, molecular biology, and induced mutations. Science Publishers, Enfield, NH, USA.

Ruangsuttapha S, Eimert K, Schröder M-B, Silayoi, Denduangboripant J, Kanchanapoom K (2007) Molecular phylogeny of banana cultivars from Thailand based on HAT-RAPD markers. *Genetic Resources and Crop Evolution* doi: 10.1007/s10722-006-9169-2.

Rustenholz C, Hedley PE, Morris J, Choulet F, Feuillet C, Waugh R, Paux E (2010) Specific patterns of gene space organisation revealed in wheat by using the combination of barley and wheat genomic resources. *BMC Genomics* **11:** 714.

Šafář J, Noa-Carrazana JC, Vrána J, Bartoš J, Alkhimova O, Sabau X, Šimková H, Lheureux F, Caruana M-L, Doležel J, Piffanelli P (2004) Creation of a BAC resource to study the structure and evolution of the banana (*Musa balbisiana*) genome. *Genome* **47:** 1182-1191.

Sági L, Panis B, Remy S, Schoofs H, De Smet K, Swennen R, Cammue BPA (1995) Genetic transformation of banana and plantain (*Musa* spp.) via particle bombardment. *Bio/Technology* **13:** 481-485.

Sandhu D, Gill KS (2002) Gene-containing regions of wheat and other grass genomes. *Plant Physiology* **128:** 803-811.

Sang T (2002) Utility o low-copy nuclear gene sequences in plant phylogenetics. *Critical Reviews in Biochemistry and Molecular Biology* **37(3):** 121-147.

SanMiguel P, Bennetzen JL (1998) Evidence that a recent increase in maize genome size was caused by the massive amplification of intergene retrotransposons. *Annals of Botany* **82:** 37-44.

Schlöterrer C, Tautz D (1992) Slippage synthesis of simple sequence DNA. *Nucleic Acids Research* **20:** 211-215.

Schmidt T (1999) LINEs, SINEs and repetitive DNA: non-LTR retrotransposons in plant genomes. *Plant Molecular Biology* **40:** 903-310.

Schmidt T, Heslop-Harrison JS (1998) Genomes, genes and junk: the large-scale organization of plant genomes. *Trends in Plant Science* **3:** 195-199.

Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, Pastemak S, Liang C, Zhang J, Fulton L, Graves TA, Minx P, Reily AD, Courtney L, Kruchowski SS, Tomlinson C, Strong C, Delehaunty K, Fronick C, Courtney B, Rock SM, Belter E, Du F, Kim K, Abbott RM, Cotton M, Levy A, Marchetto P, Ochoa K, Jackson SM, Gillam B, Chen W, Yan L, Higginbotham J, Cardenas M, Waligorski J, Applebaum E, Phelps L, Falcone J, Kanchi K, Thane T, Scimone A, Thane N, Henke J, Wang T, Ruppert J, Shah N, Rotter K, Hodges J, Ingenthron E, Cordes M, Kohlberg S, Sgro J, Delgado B, Mead K, Chinwalla A, Leonard S, Crouse K, Collura K, Kudma D, Currie J, He R, Angelova A, Rajasekar S, Mueller T, Lomeli R, Scara G, Ko A, Delaney K, Wissotski M, Lopez G, Campos D, Braidotti M, Ashley E, Golser W, Kim H, Lee S, Lin J, Dujmic Z, Kim W, Talag J, Zuccolo A, Fan C, Sebastian A, Kramer M, Spiegel L, Nascimento L, Zutavern T, Miller B, Ambroise C, Muller S, Spooner W, Narchania A, Ren L, Wei S, Kumari S, Faga B, Levy MJ, McMahan L, Van Buren P, Vaughn MW, Ying K, Yeh C-T, Emrich SJ, Jia Y, Kalyanaraman A, Hsia A-P, Barbazuk WB, Baucom RS, Brutnell TP, Carpita NC, Chaparro C, Chia J-M, Deragon J-M, Estill JC, Fu Y, Jeddeloh JA, Han Y, Lee H, Li P, Lisch DR, Liu S, Liu Z, Nagel DH, McCann MC, SanMiguel P, Myers AM, Nettleton D, Nguyen J, Penning BW, Ponnala L, Schneider KL, Schwartz DC, Sharma A, Soderlund C, Springer NM, Sun Q, Wang H, Waterman M, Westerman R, Wolfgruber TK, Yang L, Yu Y, Zhang L, Zhou S, Zhu Q, Bennetzen JL, Dawe RK, Jiang J, Jiang N, Presting GG, Wessler SR, Aluru S, Martienssen RA, Clifton SW, McCombie WR, Wing RA, Wilson RK (2009) The B73 maize genome: complexity, diversity and dynamics. *Science* **326:** 1112-1115.

Schultz J, Maisel S, Gerlach D, Müller T, Wolf M (2005) A common core of secondary structure of the internal transcribed spacer 2 (ITS2) throughout Eukaryota. *RNA* **11:** 361-364.

Schwarz DC, Li X, Hernandez LI, Ramnarain SP, Huff EJ, Wang YK (1993) Ordered restriction maps of Saccharomyces cerevisiae chromosomes constructed by optical mapping. *Science* **262:** 110-114.

Schwarzacher T (2003) DNA, chromosomes, and in situ hybridization. *Genome* **46:** 953-962.

Schwarzacher T, Leitch AR, Bennett MD, Heslop-Harrison JS (1989) In situ localization of parental genomes in a wide hybrid. *Annals of Botany* **64:** 315-324.

Schwarz-Sommer Z, Leclercq L, Göbel E, Saedler H (1987) *Cin4*, an insert altering the structure of the *A1* gene in *Zea mays*, exhibits properties of nonviral retrotransposons *EMBO Journal* **6:** 3873-3880.

Simmonds NW (1962) The evolution of the bananas. Longmans, London, UK.

Simmonds NW, Shepherd K (1955) Taxonomy and origin of cultivated bananas. *Botanical Journal od Linnean Society* **55:** 302-312.

Small RL, Cronn RC, Wendel JF (2004) Use of nuclear genes for phylogeny reconstruction in plants. *Australian Systematic Botany* **17:** 145-170.

Small RL, Lickey EB, Shaw J, Hauk WD (2005) Amplification of noncoding chloroplast DNA for phylogenetic studies in lycophytes and monilophytes with a comparative example of relative phylogenetic utility from *Ophioglossaceae*. *Molecular Phylogenetics and evolution* **36:** 509-522.

Smith SA, Beaulieu JM, Donoghue MJ (2009) Mega-phylogeny approach for comparative biology: an alternative tu supertree and supermatrix approaches. *BMC Evolutionary Biology* **9:** 37.

Steiger DL, Nagai C, Moore PH, Morden CW, Osgood RV, Ming R (2002) AFLP analysis of genetic diversity within and among *Coffea arabica* cultivars. *Theoretical and Applied Genetics* **105:** 209-215.

Stephan W (1986) Recombination and the evolution of satellite DNA. *Genetic Research* **47(3):** 167-174.

Swift HH (1950) The constancy of desoxyribose nucleic acid in plant nuclei. *Proceedings of the National Academy of Science U.S.A.* **36:** 643-654.

Taberlet P, Gielly L, Pautou G, Bouvet J (1991) Universal primers for amplification of three non-coding regions of chloroplast DNA. *Plant Molecular Biology* **17:** 1105-1109.

Tang H, Bowers JE, Wang X, Paterson AH (2010) Angiosperm genome comparisons reveal early polyploidy in the monocot lineage. *Proceeding of the National Academy of Science U.S.A.* **107(1):** 472-477.

Tanksley SD, Young ND, Paterson AH, Boniebale MW (1989) RFLP mapping in plant breeding: new tools for an old science. *Nature Biotechnology* **7:** 257-264.

Tek AL, Song JQ, Macas J, Jiang JM (2005) Sobo, a recently amplified satellite repeat of potato, and its implications for the origin of tandemly repeated sequences. *Genetics* **170(3):** 1231-1238.

Thomas CA jn. (1971) The genetic organisation of chromosomes. *Annual Review of Genetics* **5:** 237-256.

Tiwari VK, Riera-Lizarazu O, Gunn HL, Lopez K, Iqbal MJ, Kianian SF, Leonard JM (2012) Endosperm tolerance of paternal aneuploidy allows rediation hybrid mapping of the wheat D-genome and a measure of γ ray-induced chromosome breaks. *PLoS ONE* **7(11):** e48815. Doi:10.1371/journal.pone.0048815.

Tourmente S, Lazreg A, Lafleuriel J, Tutois S, Cuvillier C, Espagnol MC, Picard G (1998) Identification of new minisatellites loci in Arabidospis thaliana. *Journal of Experimental Botany* **49(318)**: 21-25.

Tuskan GA, DiFazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U, Putnam N, Ralph S, Rombauts S, Salamov A, Schein J, Sterck L, Aerts A, Bhalerao RR, Bhalerao RP, Blaudez D, Boerjan W, Brun A, Brunner A, Busov V, Campbell

M, Carlson J, Chalot M, Chapman J, Chen GL, Cooper D, Coutinho PM, Couturier J, Covert S, Cronk Q, Cunningham R, Davis J, Degroeve S, Dejardin A, dePamphilis C, Detter J, Dirks B, Dubchak I, Duplessis S, Ehlting J, Ellis B, Gendler K, Goodstein D, Gribskov M, Grimwood J, Groover A, Gunter L, Hamberger B, Heinze B, Helariutta Y, Henrissat B, Holligan D, Holt R, Huang W, Islam-Faridi N, Jones S, Jones-Rhoades M, Jorgensen R, Joshi C, Kangasjarvi J, Karlsson J, Kelleher C, Kirkpatrick R, Kirst M, Kohler A, Kalluri U, Larimer F, Leebens-Mack J, Leple JC, Locascio P, Lou Y, Lucas S, Martin F, Montanini B, Napoli C, Nelson DR, Nelson C, Nieminen K, Nilsson O, Peter G, Philippe R, Pilate G, Poliakov A, Razumovskaya J, Richardson P, Rinaldi C, Ritland K, Rouze P, Ryaboy D, Schmutz J, Schrader J, Segerman , Shin H, Siddiqui A, Sterky F, Terry A, Tsai C, Uberbacher E, Unneberg P, Vahala J, Wall K, Wessler S, Yang G, Yin T, Douglas C, Marra M, Sandberg G, Van der peer Y, Rokhsar D (2006) The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray ex Brayshaw). *Science* **313:** 1596-1604.

Tsuchimoto S, Hirao Y, Ohtsubo E, Ohtsubo H (2008) New SINE families from rice, OsSN, with poly(A) at the 3' ends. *Genes and Genetic Systems* **83(3):** 227-236.

Ude G, Pillay M, Nwakanma D, Tenkouano A (2002) Analysis of genetic diversity and sectional relationships in *Musa* using AFLP markers. *Theoretical and Applied Genetics* **104:** 1239-1245.

Valárik M, Šimková H, Hřibová E, Šafář J, Doleželová M, Doležel J (2002) Isolation, characterization and chromosome localization of repetitive DNA sequences in bananas (*Musa* spp.). *Chromosome Research* **10:** 89-100.

Valárik M, Bartoš J, Kovářová P, Kubaláková M, de Jong JH, Doležel J (2004) High-resolution FISH on super-stretched flow-sorted chromosomes. *The Plant Journal* **37:** 940-950.

Valmayor RV, Le Dinh Danh (2002) Classification and characterization of *Musella splendida* sp. nov. *The Philippine Agricultural Scientist* **85(2):** 204-209.

Van de Peer Y, Fawcett JA, Proost S, Sterck L, Vandepoele K (2009) The flowering world: a tale of duplications. *Trends in Plant Science* **14(12):** 680-688.

Vergnaud G, Denoeud F (2000) Minisatellites: mutability and genome architecture. *Genome Research* **10(7):** 899-907.

Vilarinhos AD, Piffanelli P, Lagoda P, Thibivilliers S, Sabau X, Carreel F, D'Hont A (2003) Construction and characterization of a bacterial artificial chromosome library of banana (*Musa acuminata* Colla). *Theoretical and Applied Genetics* **106:** 1102-1106.

Volf JN (2006) Turning junk into gold: domestication of transposable elements and the creation of new genes in eukaryotes. *BioEssays* **28(9):** 913-922.

Volkov RA, Komarova NY, Hemleben V (2007) Ribosomal DNA in plant hybrids: inheritance, rearrangement, expression. *Systematics and Biodiversity* **5:** 261-276.

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

Wang C-J, Harper L, Cande WZ (2006) High-resolution single-copy gene fluorescence *in situ* hybridization and its use in the construction of a cytogenetic map of maize chromosome 9. *The Plant Cell* **18:** 529-544.

Wang X-L, Chiang T-Y, Roux N, Hao G, Ge X-J (2007) Genetic diversity of wild banana (*Musa balbisiana* Colla) in China as revealed by AFLP markers. *Genetic Resources and Crop Evolution* **54:** 1125-1132.

Wang Y, Wang X, Paterson AH (2012) Genome and gene duplications and gene expression divergence: a view from plants. *Annals of the New York Academy of Sciences* **1256:** 1-14.

Wendel JF, Doyle JJ (1998) Phylogenetic incongruence: window into genome history and molecular evolution. In: Soltis P, Soltis D, Doyle J (1998) Molecular systematics of plants II. Kluwer Academic Publishers, Massachusetts, USA.

Wendel JF (2000) Genome evolution in polyploids. *Plant Molecular Biology* **42:** 225-249.

Wenzl P, Carling J, Kudrna D, Jaccoud D, Huttner E, Kleinhofs A, Kilian A (2004) Diversity arrays technology (DArT) for whole-genome profiling of barley. *Proceeding of The National Academy of Science U.S.A.* **101:** 9915-9920.

White TJ, Bruns T, Lee S, Taylor J (1990) Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In: Innis MA, Gelfand DH, Sninsky JJ, White TJ (1990) PCR protocols: A guide to methods and applications. Academic Press, New York, USA.

Wicker T, Taudien S, Houben A, Keller B, Graner A, Platzer M, Stein M (2009) A whole-genome snapshot of 454 sequences exposes the composition of the barley genome size in wheat and barley. *Plant Journal* **59:** 712-722.

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

Wong C, Kiew R, Argent GCG, Set O, Lee SK, Gan YY (2002) Assessment of the validity of the sections in *Musa* (*Musaceae*) using AFLP. *Annals of Botany* **90:** 231-238.

Wright DA, Ke N, Smalle J, Hauge BM, Goodman HM, Voytas DF (1996) Multiple non-LTR retrotransposons in the genome of Arabidopsis thaliana. *Genetics* **142:** 569-578.

Wu CY (1976) *Musella lasiocarpa*. *Iconographia Cormophytorum Sinicorum* 580-582.

Xiong Y, Eickbush TH (1990) Origin and evolution of retroelements based upon their reverse transcriptase seuqences. *The EMBO Journal* **9(10):** 3353-3362.

Yan H, Ito H, Nobuta K, Ouyang S, Jin W, Tian S, Lu C, Venu RC, Wang G, Green PJ, Wing RA, Buell CR, Meyers BC, Jiang J (2006) Genomic and genetic characterization of rice Cen3 reveals extensive transcription and evolutionary implications of a complex centromere. *The Plant Cell* **18:** 2123-2133.

Yang GP, Maroof MA, Xu CG, Zhang Q, Biyashev RM (1994) Comparative analysis of microsatellite DNA polymorphism in landraces and cultivars of rice. *Molecular Genetics and Genomics* **245(2):** 187-194.

Yang L, Bennetzen JL (2009) Structure-based discovery and description of plant and animal Helitrons. *Proceeding of The National Academy of Science U.S.A.* **106:** 12832-12837.

Yasui Y, Nasuda S, Matsuoka Y, Kawahara T (2001) The Au family, a novel short interspersed element (SINE) from *Aegilops umbellulata*. *Theoretical and Applied Genetics* **102:** 463-470.

Yoshioka Y, Matsumoto S, Kojima S, Ohsima K, Okada N, Machida Y (1993) Molecular characterization of a short interspersed repetitive element from tobacco that exhibits sequence homology to specific tRNAs. *Proceeding of The National Academy of Science U.S.A.* **90:** 6562-6566.

Yu J, Wang J, Lin W, Li S, Li H, Zhou J, Ni P, Dong W, Hu S, Zeng C, Zhang J, Thang Y, Li R, Xu Z, Li S, Li X, Zheng H, Cong L, Lin L, Yin J, Geng J, Li G, Shi J, Liu J, Lv H, Li J, Wang J, Deng Y, Ran L, Shi X, Wang X, Wu Q, Li C, Ren X, Wang J, Wang X, Li D, Liu D, Zhang X, Ji Z, Zhao W, Sun Y, Zhang Z, Bao J, Han Y, Dong L, Ji J, Chen P, Wu S, Liu J, Xiao Y, Bu D, Tan J, Yang L, Ye C, Zhang J, Xu J, Zhou Y, Yu Y, Zhang B, Zhuang S, Wei H, Liu B, Lei M, Yu H, Li Y, Xu H, Wei S, He X, Fang L, Zhang Z, Zhang Y, Huang X, Su Z, Tong W, Li J, Tong Z, Li S, Ye J? Wang L, Fang L, Lei T, Chen C, Chen H, Xu Z, Li H, Huang H, Zhang F, Xu H, Li N, Zhao C, Li S, Dong L, Huang Y, Li L, Xi Y, Qi Q, Li W, Zhang B, Hu W, Zhang Y, Tian X, Jiao Y, Liang X, Jin J, Gao L, Zheng W, Hao B, Liu S, Wang W, Yuan L, Cao M, McDermott J, Samudrala R, Wang J, Wong GKS, Yang H (2005) The genomes of *Oryza sativa*: A history of Duplications. *PLoS Biology* **3(2):** e38.

Zhou S, Bechner MC, Place M, Churas CP, Pape L, Leong SA, Runnheim R, Forrest DK, Goldstein S, Livny M, Schwartz DC (2007) Validation of rice genome sequence by optical mapping. *BMC Genomics* **8:** 278.

Zhou S, Wei F, Nguyen J, Bechner M, Potamousis K, Goldstein S, Pape L, Mehan MR, Churas C, Pasternak S, Forrest DK, Wise R, Ware D, Wing RA, Waterman MS, Livny M, Schwartz DC (2009) A single molecular scaffold for the maize genome. *PLoS Genetics* **5(11):** e1000711. Doi: 10.1371/journaLpgen.1000711.

Zimmer EA, Martin SL, Beverley SM, Kan YW, Wilson AC (1980) Rapid duplication and loss of genes coding for the alphachains of hemoglobin. *Proceeding of The National Academy of Science U.S.A.* **77:** 2158-2162.

Zou X-H, Zhang F-M, Zhang J-G, Zang L-L, Tang L, Wang J, Sang T, Ge S (2008) Analysis of 142 genes resolves the rapid diversification of the rice genus. *Genome Biology* **9:** R49.

Zuccolo A, Sebastian A, Talag J, Yu Y, Kim H, Collura K, Kudrna D, Wing RA (2007) Transposable element distribution, abundance and role in genome size variation in the genus *Oryza*. **7:** 152 doi:10.1186/1471-2148-7-152.

3 AIMS OF THE THESIS

I. Clarification of phylogenetic relationships within the family *Musaceae*

The first aim of this Ph.D. thesis was to contribute to the clarification of phylogenetic relationships within the family *Musaceae* based on the analysis of the ITS1-5.8S-ITS2 sequence region and to provide detailed information about the structure and diversity of this region among representatives of the family *Musaceae*.

II. Cytogenetic mapping of major DNA satellites in *Musa*

The second aim of this thesis was to characterize genomic organization of major DNA satellites and to expand the number of individual chromosomes which can be identified in *Musa*.

III. Characterization of *Musa* genetic diversity

The third aim of this work was to contribute to characterization of banana genetic diversity in terms of cytogenetic as well as molecular characters.

4 RESULTS

4.1 ORIGINAL PAPERS

4.1.1 The ITS1-5.8S-ITS2 Sequence Region in the *Musaceae*: Structure, Diversity and Use in Molecular Phylogeny

Hřibová E, Čížková J, Christelová P, Taudien S, De Langhe E, Doležel J

PLoS ONE 6(3): e17863, 2011.

[EH, JD concieved and designed the experiments. EH, JČ, PC, ST performed the experiments. EH analyzed the data. EdL, EH, ST contributed reagents/materials/analysis tools. EH, JD, EdL wrote the paper.]

OPEN ACCESS Freely available online

PLoS one

The ITS1-5.8S-ITS2 Sequence Region in the Musaceae: Structure, Diversity and Use in Molecular Phylogeny

Eva Hřibová[1], Jana Čížková[1], Pavla Christelová[1], Stefan Taudien[2], Edmond de Langhe[3], Jaroslav Doležel[1]*

1 Centre of the Region Haná for Biotechnological and Agricultural Research, Institute of Experimental Botany, Olomouc, Czech Republic, **2** Genome Analysis, Leibniz Institute for Age Research, Fritz Lipmann Institute (FLI), Jena, Germany, **3** Laboratory of Tropical Crop Improvement, Katholieke Universiteit Leuven, Leuven, Belgium

Abstract

Genes coding for 45S ribosomal RNA are organized in tandem arrays of up to several thousand copies and contain 18S, 5.8S and 26S rRNA units separated by internal transcribed spacers ITS1 and ITS2. While the rRNA units are evolutionary conserved, ITS show high level of interspecific divergence and have been used frequently in genetic diversity and phylogenetic studies. In this work we report on the structure and diversity of the ITS region in 87 representatives of the family Musaceae. We provide the first detailed information on ITS sequence diversity in the genus *Musa* and describe the presence of more than one type of ITS sequence within individual species. Both Sanger sequencing of amplified ITS regions and whole genome 454 sequencing lead to similar phylogenetic inferences. We show that it is necessary to identify putative pseudogenic ITS sequences, which may have negative effect on phylogenetic reconstruction at lower taxonomic levels. Phylogenetic reconstruction based on ITS sequence showed that the genus *Musa* is divided into two distinct clades – *Callimusa* and *Australimusa* and *Eumusa* and *Rhodochlamys*. Most of the intraspecific banana hybrids analyzed contain conserved parental ITS sequences, indicating incomplete concerted evolution of rDNA loci. Independent evolution of parental rDNA in hybrids enables determination of genomic constitution of hybrids using ITS. The observation of only one type of ITS sequence in some of the presumed interspecific hybrid clones warrants further study to confirm their hybrid origin and to unravel processes leading to evolution of their genomes.

Citation: Hřibová E, Čížková J, Christelová P, Taudien S, de Langhe E, et al. (2011) The ITS1-5.8S-ITS2 Sequence Region in the Musaceae: Structure, Diversity and Use in Molecular Phylogeny. PLoS ONE 6(3): e17863. doi:10.1371/journal.pone.0017863

Editor: Carles Lalueza-Fox, Institut de Biologia Evolutiva - Universitat Pompeu Fabra, Spain

Received November 30, 2010; **Accepted** February 11, 2011; **Published** March 22, 2011

Funding: This work has been supported by the Grant Agency of the Academy of Sciences of the Czech Republic [grant awards no. IAA600380703 and KJB 500380901] and was performed in collaboration with the International Atomic Energy Agency [Research Agreement no. 13192]. The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: dolezel@ueb.cas.cz

Introduction

Ribosomal ribonucleic acid (rRNA) is the central component of ribosomes - small intracellular particles, which convert the information carried in the genetic code into protein molecules [1]. In plant genomes, similar to other eukaryotes, the nuclear ribosomal RNA (rRNA) genes occur in thousands of copies and are organized in tandem arrays, typically clustered in two separate genomic loci. While the 5S rDNA locus encodes the 5S rRNA genes, the 45S rDNA locus contains genes for 18S, 5.8S and 26S rRNA, which are transcribed as a single unit and then spliced. The 18S, 5.8S and 26S rRNA genes are separated by internal transcribed spacers ITS1 and ITS2, and the 45S transcription units themselves are separated from each other by non-transcribed intergenic spacers (IGS), extending from the 3′ end of 26S rRNA to 5′ end of the 18S rRNA coding regions [2].

Although the ITS are not incorporated into the mature rRNA, they encode signals for proper processing of the rRNA transcripts, which itself depends on the secondary structure of ITS RNA [3]. In contrast to the ITS1 spacer, secondary structures of functional ITS2 spacer as well as 5.8S rDNA sequences are highly conserved within plants [4,5]. ITS2 spacer forms a four helices structure with typical pyrimidine-pyrimidine bulge in helix II and the most conserved primary sequence includes the TGGT in the helix III [6–8]. The 5.8 S rDNA sequences contain three conserved motives in their nucleotide sequences that are essential for the correct folding of secondary structure [9,10]. The conserved secondary structure of ITS2 and 5.8S sequences makes it possible to identify divergent ITS paralogs with non-functional pseudogenes [10–12].

The ribosomal genes arranged in tandem arrays are characterized by low intra-genomic diversity as a consequence of concerted evolution [13–15]. Individual copies of rRNA genes evolve more or less in unison and unequal crossing-over and high frequency of gene conversion are thought to be responsible for this [13–18]. The intraspecific and intra-population stability of rRNA locus on one hand and relatively fast evolution of ITS1 and ITS2, which are under lower selective pressure, made the ITS one of the most popular marker in phylogenetic studies [17,19–21]. The analysis of the ITS was also employed to identify progenitors of hybrid species and to study the origin of polyploid species [e.g. 22–24].

However, the evolution of rRNA loci in hybrids and in allopolyploids in particular, may be complex and several evolutionary scenarios were observed which differ in the way the parental rRNA loci interact. In some species, such as *Arabidopsis*, *Brassica* or *Silene*, the divergent parental rDNA sequences remain

conserved and evolve independently in hybrids without any interaction [25–27]. In other cases, such as in allopolyploids of *Nicotiana* or *Dendrochilum*, parental rDNA loci seem to recombine as judged from the occurrence of chimerical ITS sequences [28,29]. The third evolutionary pathway is a dominance of one type of rDNA sequence that may lead to partial or even complete homogenization of rDNA locus of the second parent [18,30–32]. The dominant rDNA sequence may represent one of the parental rDNA sequence when one rDNA type is lost or eliminated or may be a chimeric sequence resulting from intergenomic recombination [16–33]. Obviously, if the concerted evolution is not complete, different types of paralogous and orthologous rDNA sequences may be present in the genome, often including silenced and non-functional rDNA sequences referred to as pseudogenes. The presence of pseudogenes in the phylogenetic data set may result in altered phylogenetic inference [34].

Banana (*Musa* spp.) is an important cash crop and staple food for millions people living in the humid tropics. Most of edible banana cultivars are seed-sterile diploid, triploid and tetraploid intra- and inter-specific hybrids, which contain various combinations of the A and B genomes originating from seed bearing progenitors *M. acuminata* Colla and *M. balbisiana* Colla, respectively [35]. A few edible hybrids originated from crosses between *M. acuminata* and *M. schizocarpa* (S genome) and *M. acuminata* and presumably *M. textilis* (T genome) as well. The genus *Musa* belongs to the family Musaceae that includes three genera, *Musa* L., *Ensete* Horan and *Musella* C.Y. Wu ex H.W. Li, and has traditionally been divided into four sections based on morphology and basic chromosome number [36]: *Eumusa* (2n = 2x = 22), *Australimusa* (2n = 2x = 20), *Rhodochlamys* (2n = 2x = 22) and *Callimusa* (2n = 2x = 18, 20). However, the traditional classification of the genus *Musa* has been questioned based on DNA analyses [37–40]. Wong *et al.* [39] proposed to compound sections *Eumusa* and *Rhodochlamys* into one and sections *Australimusa* and *Callimusa* into second one. Comparable observations on the similarity of *Eumusa* and *Rhodochlamys* on one hand and *Australimusa* and *Callimusa* on other hand were made by others [41,42]. Apart from this, the traditional taxonomy of *Musa* is also problematic for placement of newly described species [43,44].

In addition to studies which employed AFLP, SSR and DArT markers [40,45–47], only two studies on genetic diversity and phylogenetic relationships in *Musa* involved the analysis of rDNA region, namely the length and specific restriction pattern of the ITS1-5.8S-ITS2 region. Nwakanma *et al.*, [48] noted that a specific restriction pattern of the ITS region enabled differentiation between diploid AA and BB genomes as well as between the triploid AAA intraspecific hybrid clones and those that carry the B genome (AAB and ABB). More recently, the nucleotide information of the ITS region was used to study relationships of wild species of Musaceae [49,50]. Nevertheless, there has been a lack of detailed knowledge on the ITS region in *Musa* and its potential for the analysis of genetic diversity and phylogenetic relationship remains unexplored.

In this work we set out to study in detail the rDNA ITS1-5.8S-ITS2 region in a total of 87 accessions of wild diploid species and triploid hybrid clones that represent all main taxonomic groups of the family Musaceae. We analyzed DNA sequences obtained either by classical dideoxy termination chain reaction or by massively parallel 454 sequencing with the aim to (1) thoroughly describe the nucleotide structure and diversity of the ITS region in *Musa*, *Ensete* and *Musella*; (2) compare the results obtained by different sequencing approaches; (3) analyze the presence of putative pseudogenic ITS sequence types and their influence on phylogenetic reconstructions; (4) assess the utility of ITS sequences for analysis of phylogenetic relationships within Musaceae and (5) use ITS to verify the genome constitution of inter- and intra-specific banana hybrids.

Results

We have analyzed more than 2,500 nucleotide sequences from the ITS region of 87 accessions representing the family Musaceae, including 54 *Musa* diploids representing all sections of the *Musa* genus including four inter-subspecific hybrids, two *Ensete* and one *Musella* species (Table S1). In addition to the representatives of wild *Musa* species, ITS region of 30 hybrid banana clones was analyzed (Table S1). Three approaches were used to obtain ITS sequences. As the first step, the ITS region of all diploid genotypes was amplified using specific primers and four PCR products from individual PCR reactions of each accession were used for direct Sanger sequencing. Only 23 diploid genotypes produced readable ITS sequences with no polymorphism (Table S1), the remaining diploid accessions produced polymorphic ITS sequences. The ITS region of heterogeneous diploids and all hybrid clones was cloned and 15–85 DNA clones bearing ITS region were sequenced (Table S1). Finally, we have assessed the usefulness of massively parallel 454 sequencing for the analysis of the ITS loci.

Length variation, sequence diversity and GC content of the ITS region

In general, the length of ITS1 and ITS2 spacer varied from 216 to 223 bp and from 205 to 227 bp, respectively. A total length of ITS1-5.8S-ITS2 sequence region ranged from 578 bp (*M. acuminata* 'Truncata', ITC 0393) to 601 bp (*M. mannii*, ITC 1411) in all accessions, except for five representatives of section *Australimusa*: *M. maclayi* (ITC 1207), *M. maclayi* type Hung Si (ITC 0614), *M. menei* (ITC 1021), *M. textilis* (ITC 1072) and *M. maclayi* F. Muell, where the ITS region was 544 bp long. This significant length difference specific for these *Australimusa* entries was due to a 41 bp deletion in ITS1.

One of the highest sequence diversities within diploid species was observed in *M. schizocarpa* (ITC 0856, ITC 0846) (Figure S1A, B/sequence similarity matrix). The heterogenous representatives of *Australimusa* and *Rhodochlamys* sections contained more diverse ITS sequence region than the representatives of *Eumusa*. Within the hybrids, banana clones 'Cachaco' (ITC 0643) and 'Dole' (ITC 0767), representatives of Bluggoe subgroup (ABB genomes), showed the highest level of sequence divergence, while the other three Bluggoe clones showed relatively low level of sequence heterogeneity (data not shown). Sequence divergence of the entire ITS region was higher among the hybrids with genome constitution AAA, AxB and AxT than among corresponding diploid genotypes (Figure S2). On the other hand, AxS banana hybrid clones showed lower sequence divergence than the S genomes, which contained putative pseudogenes (Figure S2) GC content of ITS1 varied from 55.35 to 67.70% and was slightly lower than the GC content of ITS2 (56.11 to 70.97%).

The 5.8S rDNA sequence region had a conserved length of 155 bp or 154 bp, and its GC content varied from 49.68 to 57.48% and was significantly lower than the GC content in ITS1 and ITS2. Among all Musaceae species studied, the lowest GC content of the ITS region was identified in accession representing the section *Australimusa* (Table S2).

Secondary structure of ITS2 and 5.8S rDNA sequences in Musaceae and identification of pseudogenes

The secondary structure of ITS2 and 5.8S rDNA sequence regions was reconstructed for all accessions. ITS2 sequences

formed specific four-helices structure with typical pyrimidine-pyrimidine bulge in helix II and the most conserved primary sequence included the TGGT in the helix III (Figure 1A, B). Secondary structure of 5.8S rDNA sequence was reconstructed under specific settings for base pairing (Figure 1C, D). Moreover, 5.8S rDNA sequences were checked for the presence of three conserved motives [9]. The highly conserved sequence of 16 bp motif M1 (5′-CGATGAAGAACGTAGC-3′) is a part of two helices – helix B4 and B5 [9]. One type of ITS sequence in heterogenous diploids *M. acuminata* 'Tuu Gia' (ITC 0610) and 'Pisang Mas' (ITC 0563) had a "T" at position 11 and 16 in the motif M1. Similarly, one type of ITS in *M. acuminata* 'Truncata' (ITC 0393) and *M. schizocarpa* (ITC 0856 and ITC 0846) had an "A" at position 12. Motif M2 (5′-GAATTGCAGAATCC-3′), previously described by Jobes and Thien [5] is 14 bp long and located in the loop and 10 bp long motif M3 (5′-TTTGAACGCA-3′) is a part of the B4 and B7 helices [9] (Figure 1C, D). Some ITS types of *Musa* diploids had a changes at positions 9 and 14 in motif M2 and changes at positions 7 and 8 in the motif M3. The nucleotide changes in conserved motives of 5.8S rDNA of diploid and hybrid accessions are summarized in Table S2. The information on GC content, presence of conserved motives in the 5.8S rDNA sequence and ability of ITS2 and 5.8S rDNA sequence to fold into a conserved secondary structure allowed us to identify putative pseudogenes (Table S2).

Comparison of the ITS1-5.8S-ITS2 sequence region obtained using different sequencing approaches

In addition to Sanger sequencing, we used massively parallel sequencing (454) to study the diversity of the ITS region within six diploid Musaceae accessions (Table S1) and to compare the data obtained using the two different sequencing approaches. Reconstructed contigs of ITS regions from 454 data were characterized by high read depth (Figure 2). Analysis of the 454 data confirmed that the ITS region is highly conserved in all six accessions. Moreover, the 454 consensus ITS sequences were highly similar to those obtained using Sanger sequencing.

454 data of *M. acuminata* 'Calcutta4' resulted in one ITS sequence type that showed high similarity (99.67%) to the sequence obtained using direct Sanger sequencing - only two mismatches were found. ITS sequence region of *M. balbisiana* 'Pisang Klutug Wulung' obtained by 454 showed 99.83% similarity to the Sanger sequencing data, differing in one nucleotide insertion within the poly-G sequence region. The 454 sequencing of *M. ornata* and *E. ventricosum* resulted in two types of ITS sequence, showing 100% and 99.83% (single nucleotide substitution) similarity to the direct Sanger sequencing, output, respectively.

The highest number of 454 reads homologous to ITS region was obtained for *M. beccarii*, an observation which reflects higher copy number of 45S rRNA genes that are localized on three pairs of chromosomes in *M. beccarii* ITC 1070 [42]. The remaining species analyzed by 454 sequencing have only one pair of chromosomes bearing the 45S locus. 454 sequencing resulted in two types of ITS sequence in *M. beccarii*. The first type showed 99.65% similarity to the sequence obtained by direct Sanger sequencing (2 nucleotide changes). The second one contained 8 nucleotide substitutions showing lower similarity (98.63%) to the Sanger sequencing data and the presence of nucleotide substitutions at potential methylation sites, indicated its pseudogenic nature.

Out of the six Musaceae species evaluated using the two different sequencing approaches, only *M. textilis* gave polymorphic ITS sequence after direct Sanger sequencing. Subsequent cloning and Sanger sequencing of 15 clones carrying the ITS region resulted in two different consensus ITS sequences (0539con1 and 0539con2) (Table S1). This observation was confirmed by 454 sequencing, which resulted in two types of ITS sequence differing in three nucleotide substitutions. Both 454-ITS types showed high similarity (99.49% and 99.67%) to 0539con1 ITS sequence. Second type of ITS sequence obtained after Sanger sequencing (putative pseudogene; Table S2) was not found in the 454 dataset despite the relatively high read depth of the ITS-454 data (RD = 149.6; Figure 2).

Phylogenetic reconstruction

Position of the Musaceae clades with respect to closely related Zingiberales. We have used two different methods to reconstruct phylogenetic relationships within Musaceae considering other closely related families of the order Zingiberales - Neighbor-Joining (NJ) and Bayesian inference (BI). The best-fit model of nucleotide substitution selected by jModeltest and implemented in Bayesian analysis was GTR+I+G. If putative pseudogenic ITS types were excluded from data set, NJ and BI resulted in fully resolved trees with high NJ bootstrap support and Bayesian posterior probabilities for all main clades (Figure 3). In these analyses, ITS1-ITS2 concatenated region of all diploid Musaceae accessions, including the four inter-subspecific diploid hybrids and closely related families from Zingiberales (dataset 1) was used for NJ constructed on Jukes-Cantor distance matrix and BI analysis performed in BEAST.

The reconstructed phylogeny confirmed the monophyly of Musaceae and revealed three main clades within the family. The monophyly of *Ensete* and *Musella* was not supported as they shared the same clade. Species from the section *Australimusa* shared the same clade with the species from section *Callimusa* as did the species from sections *Eumusa* and *Rhodochlamys* (Figure 3). Moreover, the positions of the clade bearing *Ensete* and *Musella* species and that of *Australimusa* and *Callimusa* species was reverse as compared to the work of Liu *et al.* [50]. Based on sequence analysis of 19 intron containing nuclear genes Christelová *et al.* [51] created a phylogenetic tree in which the position of the clade bearing *Musella* and *Ensete* genera was sister to the *Musa* genus and the *Australimusa*/*Callimusa* sections formed a separate subcluster within the *Musa* genus clade.

In case of *Australimusa*/*Callimusa* clade, *M. coccinea* appeared as a distinct species within the clade. All accessions of cultivated Fe'i banana, except 'Utafan' (ITC 0913) clustered with the *Australimusa* species and both, BI and NJ phylogenetic analysis supported their close relationships to *M. textilis*. Moreover, the *Australimusa*/*Callimusa* clade contained one accession from the *Eumusa* section - *M. balbisiana* 10852 (ITC 0094) (Figure 3). Within the *Eumusa*/*Rhodochlamys* clade, representatives of the B genome formed strongly supported subclade separated from other *Eumusa* and *Rhodochlamys* species. However, the B genome subclade contained also one representative of Fe'i bananas - 'Utafan' (ITC 0913). Another well separated subclade was formed by representatives of section *Rhodochlamys*. However, this strongly supported subclade contained one accession of section *Australimusa* – *M. textilis* 'Née' (ITC 0563), and one representative of Fe'i bananas – 'Asupina' (ITC 1027). Moreover, two *Rhodochlamys* accessions - *M. ornata* (ITC 0637) and *M. velutina* (ITC 0627) clustered with the A genome representatives, which supports close relationships of *Eumusa* and *Rhodochlamys* species. Within the representatives of *M. acuminata*, both phylogenetic analyses supported the close relationship of *burmannica, burmannicoides, siamea* and *malaccensis* subspecies.

The BI analysis revealed a well separated and strongly supported subclade of the S genome representatives (*M. schizocarpa*)

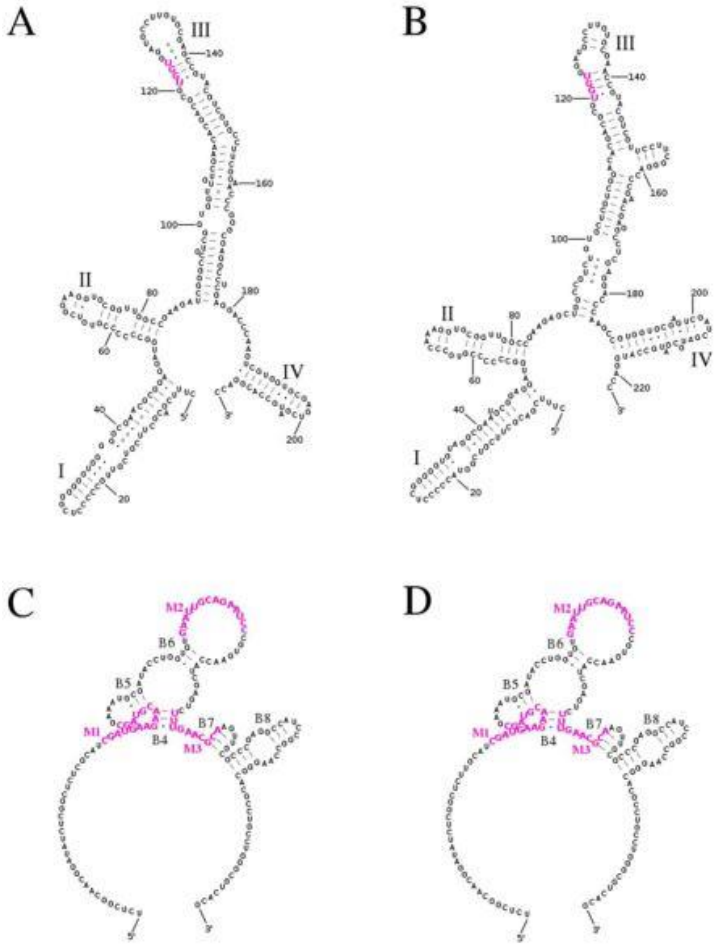


Figure 1. Examples of functional secondary structures of the ITS2 spacer and 5.8S rRNA gene. Positions within typical four-helices structures of ITS2 spacer of *M. acuminata* 'Calcutta 4' (A) and *E. gilletii* (B) are numbered every 20 nucleotides and helices are numbered I–IV. The common TGGT sequence motif in helix III is marked in red. The helices (B4–B8) in the conserved secondary structure of 5.8S rRNA gene sequences of *M. acuminata* 'Calcutta 4' (C) and *E. gilletii* (D) are numbered according to Wuyts *et al.* [82]. Conserved motives (M1–M3) are marked in violet.
doi:10.1371/journal.pone.0017863.g001

together with *M. acuminata* 'Niyarma yik' (ITC 0269). In the NJ analysis, representatives of the S genome clustered not only with *M. acuminata* 'Niyarma yik' (ITC 0269), but also with two other representatives of A genome - *M. acuminata* 'Malaccensis' (ITC 0250) and *M. acuminata* 'Galeo' (ITC 0259). BI and NJ trees differed also in the position of *M. nagensium*. In the BI tree, *M. nagensium* shared the same clade with the B genome representatives, supporting their close relationship. On the contrary, *M. nagensium* formed a highly supported separate subclade in the NJ tree, indicating its phylogenetic isolation from other *Eumusa* species (Figure 3). Finally, the analysis revealed that presence of putative pseudogenic ITS data had no significant effect on positions of main clades in the BI and NJ phylogenetic trees. Nevertheless, the positions of some accessions were changed within the main clades, especially within the A genome representatives (Figure S3).

Evolutionary relationships of banana hybrid clones. Thanks to the specific mode of the rDNA loci evolution, the ITS data can be used to analyze the genome composition and evolutionary relationships of hybrid plant species [22–24,26,27]. Here we present the first sequence data from the ITS region and its utilization for analysis of genome composition and evolutionary relationships in different banana hybrid clones. Both phylogenetic trees (BI and NJ) based on the ITS1-ITS2 concatenated region of dataset 2 (see Materials and Methods) contained cluster of ITS sequences corresponding to B genome well separated from the other genomes of *Eumusa* section. For example, banana clone 'Saba' ITC 1138, traditionally classified as ABB hybrid, contained three types of ITS sequence. One of them (1138con1) clustered with the ITS sequences corresponding to B genome and two other ITS types of 'Saba' (1138con2 and 1138con3) clustered with the A genome ITS types (Figure S4, Table S3). Similarly, *M. jackeyi* ITC 0851 which has been classified as AT hybrid contained three different types of ITS sequence region. One ITS type clustered together with the ITS corresponding to T genome (*Australimusa*/

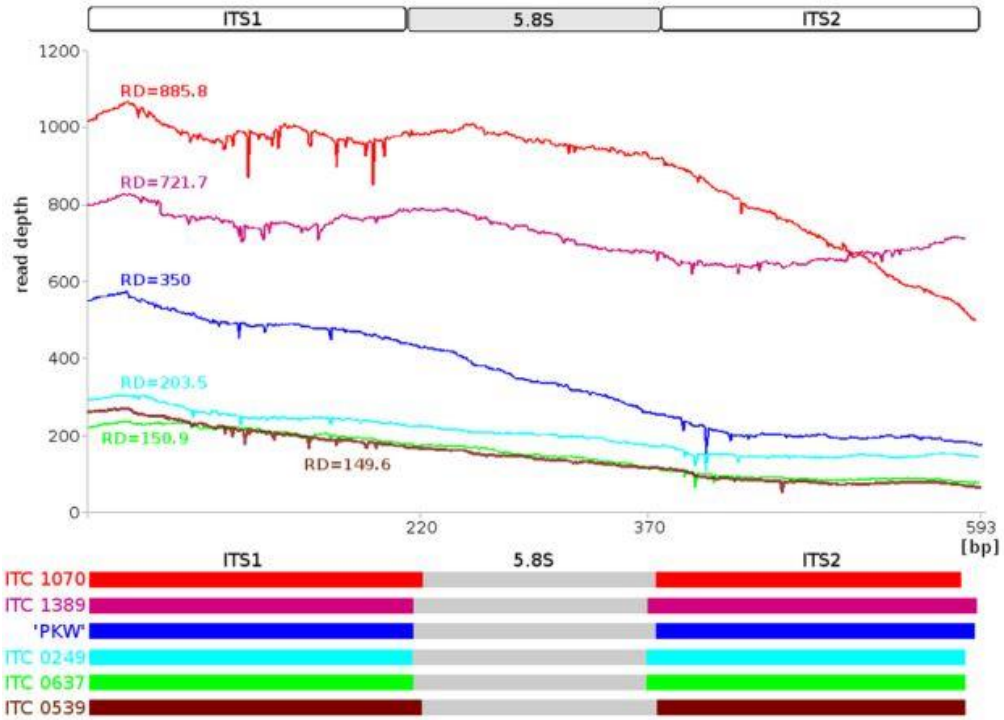


Figure 2. Reconstruction of the complete ITS1-5.8S-ITS2 region in six accessions of *Musaceae* after 454 sequencing. The graphs show the read depth (RD; number of 454 reads that were assembled over individual position) along the contig sequence. Color coding: *M. beccarii* ITC 1070 - red; *E. gilletii* ITC1389 - violet; *M. balbisiana* 'Pisang Klutug Wulung' - blue; *M. acuminata* 'Calcutta4' ITC 0249 - light blue; *M. ornata* ITC 0637 - light green and *M. textilis* ITC 0539 - brown.
doi:10.1371/journal.pone.0017863.g002

Callimusa clade) as well as the ITS sequence type (0851con3) identified as putative pseudogene. The third ITS sequence type of *M. jackeyi* (0851con2) clustered with the ITS types corresponding to A genome (Figure S4, Table S3). Similarly to dataset 1, most of the *Rhodochlamys* species formed a separated subclade. The cluster containing diploid representatives of the S genome and AS hybrids also contained ITS types from banana clones that were not classified as A×S hybrids (Figure S4). This supports close relationships of *M. acuminata* and *M. schizocarpa*. In comparison to the phylogenetic analysis using dataset 1, the position of the main clades within Musaceae in dataset 2 remained conserved in BI tree. Phylogenetic analysis using NJ of dataset 2 showed reversed position of *Ensete/Musella* and *Australimusa/Callimusa* clades (Figure S4).

BI and NJ phylogenetic study revealed that most of the banana hybrid clones contained conserved parental ITS sequence types. This information could facilitate rapid determination of genome composition of newly created hybrid clones. However, not all of the evaluated hybrid clones carried conserved parental ITS types (Figure S4, Table S3). The hybrid clones with the reported AAB genome constitution ('Maritú' ITC 0639, subgroup Iholena and '3 Hands Planty' ITC 1132, subgroup Plantain) and ABB ('Cachaco' ITC 0643, subgroup Bluggoe) did not contain ITS sequence corresponding to the B genome. These results suggest partial homogenization of the rDNA locus in the clone 'Maritú' which contained two different ITS sequence types where one of them has putative pseudogenic character. On the other hand the hybrid clones '3 Hands Planty' ITC 1132 (subgroup Plantain) and 'Cachaco' ITC 0643 (subgroup Bluggoe) contained four and six ITS sequence types, respectively. The presence of chimeric ITS sequences with higher number of putative pseudogenic sequences in these cases can be due to recombination as described in other plant hybrids [28,29]. In the Bluggoe subgroup, two other hybrid clones - 'Cachaco Enano' ITC 0632 and 'Kivuvu' ITC 0157 exhibited partial concerted evolution when the ITS sequence of A genome was not found in their genomes. Banana clone 'Butuhan' ITC 1074, reported as BB×TT hybrid contained only ITS sequence corresponding to B genome. Similarly, clone 'Tonton Kepa' ITC 0822 reported as AS hybrid contained only S genome ITS sequence (Figure S4, Table S3).

Discussion

The ITS1-5.8S-ITS2 sequence region is one of the most popular loci used in molecular phylogenetic studies [17]. Despite this, the ITS region has not been analyzed in detail in the genus *Musa* and has only been used once to examine its phylogenesis [50]. However, the study suffered from the experimental approach as it was based solely on direct sequencing of the ITS locus, and the authors did not consider sequence homogeneity/heterogeneity of the region. As the potential of the ITS analysis in the taxonomy of Musaceae remains unexplored, we chose to study the structure

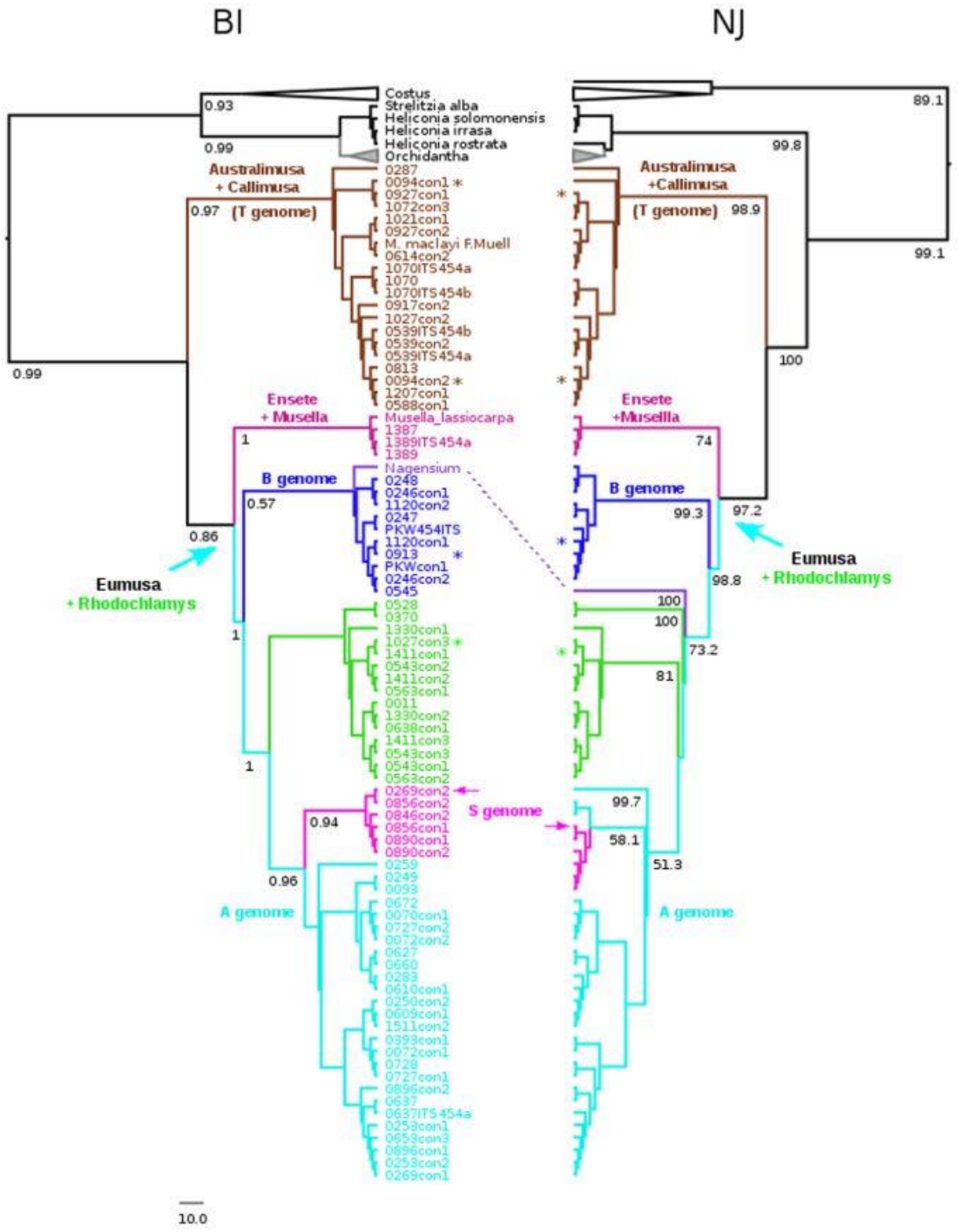


Figure 3. Phylogenetic analysis of diploid representatives of Musaceae and selected species of closely related families Strelitziaceae, Lowiaceae, Heliconiaceae and Costaceae based on the ITS1-ITS2 sequence region. Neighbor-Joining (NJ) tree constructed

from a Jukes-Cantor distance matrix of the concatenated region containing ITS1 and ITS2 spacer sequence. Bayesian inference (BI) analysis of the same data set was performed in BEAST v1.5.3 using GTR+I+G model of nucleotide substitution. Both trees are rooted at the midpoint and values below the branches indicate the bootstrap support of NJ analysis (A) and posterior probability of BI (B), respectively. Putative pseudogenes were not included in this analysis. Main clades and subclades were labeled by different colors that correspond to Figure 2. *Ensete/Musella* clade - violet; *Australimusa/Callimusa* - brown; *Rhodochlamys* - light green; B genomes - blue; A genomes - light blue and S genomes - pink.
doi:10.1371/journal.pone.0017863.g003

and diversity of the ITS1-5.8S-ITS2 region in wild species of *Musa* as well as in cultivated banana clones. We evaluated different sequencing approaches including the next generation sequencing technology and their impact on the phylogenetic reconstruction.

ITS length variation, secondary structure and identification of pseudogenes

Recent studies revealed that divergent ITS paralogs may contain non-functional pseudogenes [10–12] whose presence in phylogenetic data sets can result in altered phylogenetic inferences [34]. ITS pseudogenes were identified experimentally using transcript analysis [17]. Alternatively, *in silico* methods such as the analysis of nucleotide divergence, insertions and deletions (length of the ITS sequence region), analysis of methylation-induced substitution pattern, folding of conserved secondary structures and sequence free energy were used [52]. Due to the lack of transcript data, we evaluated ITS length, GC content and secondary structure of ITS2 and 5.8S rDNA regions. The total length of ITS regions as well as the length of putative functional ITS1 and ITS2 spacers agreed with the observations in other angiosperms [14]. However, we have identified polymorphic ITS regions in most of the *Musa* accessions and it is possible that at least some of the polymorphisms were due to the occurrence of pseudogenic sequences. In agreement with other studies [17], the putative pseudogenic ITS sequences varied in length of ITS1 spacer, had lower GC content, did not contain conserved motives in the 5.8S rDNA sequence and/or lost the ability to fold into functional ITS2 and/or 5.8S rDNA secondary structures.

Intra-individual polymorphism and evolution of ITS

Despite the concerted evolution of rDNA, ITS polymorphism within individuals is quite common [14,15,52,53] and may be due to various mechanisms. A single diploid genome contains divergent paralogs when the speciation is faster than concerted evolution or when paralogous rDNA sequences are present in non-homologous loci in the heterozygous species. Our results suggest incomplete concerted evolution in some wild diploid banana species, in which we revealed increased levels of ITS polymorphism and divergent paralogs of ITS sequences. The latter might be formed as a consequence of inter- and/or intra-genomic duplication events. Alternatively, the polymorphic character of ITS region in some wild diploids could be due to inter-subspecific hybrid origin and asexual reproduction, which bypasses meiotic recombination.

Owing to unclear origin of hybrid banana varieties and their clonal character, it is difficult to reconstruct the evolution of the 45S rDNA locus in *Musa*. Our results suggest that all three modes of evolution were involved in polyploid hybrid clones. Most of the interspecific banana hybrids presumably originating from crosses *M. acuminata*×*M. balbisiana*, *M. acuminata*×*M. schizocarpa* and *M. acuminata*×*M. textilis* contained conserved ITS sequences of both parents, indicating incomplete concerted evolution of rDNA loci. No or limited homogenization of ITS sequence was described for allopolyploid genomes of European dogroses, where most of the genomes are excluded from meiotic recombination [54]. The independent evolution of parental rDNA sequences in banana hybrid clones, which is probably due to the absence of sexual reproduction, could facilitate determination of genomic constitution of new hybrid banana cultivars. However, the fact that presence of chimeric ITS sequences and/or partial concerted evolution was observed in some hybrids may limit the potential of this approach. These hybrids show various degree of residual fertility and it is possible that their evolution involved episodes of sexual reproduction as suggested by the backcross hypothesis [55]. Our results indicate a complete concerted evolution in two banana clones: 'Butuhan' (BB×TT) and 'Tonton Kepa' (AA×SS) in which only ITS sequence of B genome and S genome was present, respectively. However, this observation may be explained either by a more complicated evolution, which included backcrossing to one of the parental species [55], or a non-hybrid origin. Clearly, the genomic constitution of both clones needs to be confirmed using additional molecular analysis.

Comparison of ITS region obtained using different sequencing approaches

Plant species and hybrid and polyploid ones in particular, often contain several rDNA loci with multiple divergent ITS sequence types. Multiple ITS sequence types may originate by different evolutionary mechanisms and their utilization in molecular phylogeny may shed light on polyploid ancestry, genome relationships, genomic constitution and evolution of hybrid species [30–32,52,56].

The ITS sequence complexity (diversity) may lead to errors in phylogenetic interference due to PCR bias and/or PCR drift [reviewed in 17]. To check the effect of ITS heterogeneity, we compared ITS sequence data obtained by classical approach (sequencing PCR products and/or PCR/cloning strategy) with the data obtained by massively parallel 454 sequencing [57]. We obtained whole genome 454 data from five *Musa* and one *Ensete* species with the length of sequencing reads sufficient to assembly individual ITS sequence types without PCR/cloning strategy. The ITS sequence data obtained using classical approach were highly similar to those obtained using 454 and resulted in the same phylogeny. However, the 454 data obtained in *M. beccarii* (ITC 1070) indicated the presence of two ITS sequence types while direct Sanger sequencing resulted in one readable ITS sequence with no polymorphism. This could be a consequence of PCR bias, when a single repeat type is preferentially amplified. Other exception was *M. textilis* (ITC 0539), in which one sequence type of ITS obtained after Sanger sequencing was not found in the 454 dataset despite the relatively high read depth. This outcome suggests some modification of the ITS sequence during the classical PCR/cloning methodology and indicate that massively parallel sequencing may be a preferable approach to study 45S rDNA sequence region as well as the other complex gene families.

Phylogenetic reconstruction using ITS

In order to analyze the position of the family Musaceae and its main clades and subclades within the order Zingiberales, we studied the ITS sequences in other families of Zingiberales (Strelitziaceae, Lowiaceae, Heliconiaceae and Costaceae). In line with the results of Liu *et al.* [50], who studied ITS and chloroplast loci, our analyses showed monophyletic origin of Musaceae. Both BI and NJ phylogenetic trees showed close relationships of species

belonging to sections *Australimusa* and *Callimusa* and sections *Eumusa* and *Rhodochlamys*, thus confirming previous findings [39,41,42,49,50]. At the same time, both trees showed reverse positions of the *Ensete/Musella* clade and *Australimusa/Callimusa* clade, probably due to the use of only one genomic locus (ITS) in the present work. Among the *Eumusa/Rhodochlamys* species, we revealed a distinct position of *M. balbisiana*, which has traditionally been placed within the *Eumusa*, and most of *Rhodochlamys* species. This suggest earlier separation of the representatives of B genome from the A genomes and *Rhodochlamys* representatives in the evolution of Musaceae and is on line with the divergence time estimates of Christelová et al. [51]. The placement of *M. acuminata* and *M. balbisiana* into two distinct subclades within *Eumusa/Rhodochlamys* clade was suggested by Li *et al.* [49] and close relationship of *M. acuminata* (section *Eumusa*) with *M. laterita* and *M. ornata* (both section *Rhodochlamys*), was also observed earlier [39,40,58]. Separation of *Eumusa* and *Rhodochlamys* sections was questioned already by Cheesman [36] and Shepherd [59] and the present as well as earlier results obtained with molecular tools indicate close relationships of the representatives from both sections [39,40,49,50].

In our study, three accessions classified as representatives of section *Australimusa* (*M. textilis* 'Née' ITC 0563 and Fe'i bananana clones 'Utafan' ITC 0913 and 'Asupina' ITC 1027) were found clustering with B genomes subclade or *Rhodochlamys* subclade. Similar situation was observed for *M. balbisiana* 10852 (ITC 0094) which grouped with *Australimusa* (see Results and Figure 3). This unexpected clustering could be due to mislabeling during their collection and/or during *in vitro* propagation in the gene bank. Alternatively, these accessions could be interspecific hybrids and/or their backcross progenies [55].

Despite a range of studies [39,40,49,50], taxonomic and phylogenetic positions of some taxa within the Musaceae are still unclear and/or controversial. One of them is *Musella*, which has been classified based on morphology as a separate genus of Musaceae [60]. The present study indicates close relationship of *Musella* and genus *Ensete* and does not confirm the monophyly of *Musella*. The same observation was made by molecular phylogenetic analysis based on the ITS and cIDNA sequences [49,50] as well as work based on genic sequences [51].

The problematic phylogenetic position was also reported for the two representatives of section *Callimusa*, which differ by basic chromosome number - *M. beccarii* (2n = 2x = 18) and *M. coccinea* (2n = 2x = 20) [42]. Our study supports close relationship of *Callimusa* with the section *Australimusa* (2n = 2x = 20), where *M. coccinea* is relatively distinct from other accessions. The distinctiveness of *M. coccinea* is supported by morphological characters [36], AFLP analysis [39] and by the ITS and cIDNA analysis of Li *et al.* [49]. In contrast to Liu *et al.* [50], our data do not support close relationship of *M. beccarii* and *M. maclayi* and indicate close relationships of cultivated Pacific Fe'i bananas to *M. textilis* and *M. maclayi*, while not confirming genetic vicinity of *M. textilis* and *M. balbisiana*.

An important outcome of this study is the identification of putative pseudogenic ITS paralogs in some banana diploid representatives and most of banana hybrid clones. Pseudogenes are expected to evolve independently at different rates than their functional counterparts and can accumulate mutations. This may lead to random clustering across phylogenetic trees and result in incorrect or incongruent phylogenies [34,61,62]. However, Ochieng *et al.* [11] found that the ITS pseudogenes can recover more corroborated phylogeny of closely related species when functional paralogs suffer selective constraints and provide too low variation. In this work we demonstrate that the presence of putative pseudogenic ITS sequences has no significant effect on the structure of main clades in the phylogenetic analysis of Musaceae. On the other hand, their presence may change the position of subclades (Figure S3). These results show that pseudogenic ITS sequences should be used with caution when analyzing phylogenesis in *Musa*. As both Liu *et al.* and Li *et al.* [49,50] used only direct Sanger sequencing, differences in the placement of some diploid species in their phylogenetic trees as compared to our results could be due to the inability to detect different ITS sequences types.

Conclusions

The present study provides the first insights into the nucleotide structure and diversity of the ITS1-5.8S-ITS2 region in Musaceae and reveal the presence of more than one type of the ITS sequence within some *Musa* species and hybrid banana clones. We show that it is necessary to use clone-based sequencing strategy for most of the species within Musaceae and identify putative pseudogenic ITS sequences that may have negative effect on phylogenetic reconstruction. We show that both Sanger and 454 sequencing lead to similar phylogenetic inferences. Phylogenetic analysis based on the ITS sequences does not support *Musella* as a distinct genus within the Musaceae and confirms close phylogenetic relationship of the species from sections *Australimusa* and *Callimusa*, and *Eumusa* and *Rhodochlamys*, where the B genome representatives form a distinct subclade. The reversed positions of the *Ensete/Musella* clade and *Australimusa/Callimusa* clade in the phylogenetic trees suggest that phylogenetic reconstruction based on a single genomic locus (ITS) may lead to inconsistent phylogenetic inference. On the other hand, the use of ITS locus in phylogenetics has proven powerful in reconstructing relationships on lower taxonomic levels (groups and subgroups). Most of the intraspecific banana hybrids analyzed in this work contain conserved parental ITS sequences, indicating incomplete concerted evolution of rDNA loci. Independent evolution of parental rDNA loci in hybrids enables verification of genomic constitution of hybrids using ITS. The observation of only one type of ITS sequence in some of the presumed interspecific hybrid clones warrants further study to confirm their hybrid origin and evolution of their genomes.

Materials and Methods

Plant material

In vitro rooted plants of most of the *Musa* and *Ensete* species used in the study as well as the hybrid clones (Table S1) were obtained from the International Transit Centre (ITC, Katholieke Universitiet, Leuven, Belgium). *M. balbisiana* 'Pisang Klutug Wulung' was obtained from Dr. François Côte (CIRAD, Guadeloupe) as rooted plants. Seeds of *M. nagensium* and *M. maclayi* F. Muell were obtained from Prof. Markku Häkkinen (Botanic Garden, University of Helsinki, Finland) and plants of *Musella lasiocarpa* were purchased from a commercial nursery. Plants and/or seeds were transferred to soil and grown in a greenhouse.

Amplification, cloning and Sanger sequencing of ITS1-5.8S-ITS2 region

Genomic DNA of all *Musa*, *Ensete* and *Musella* species used for the ITS analysis was isolated from fresh cigar leaves using the Invisorb Spin Plant Mini Kit (Invitek, Berlin, Germany) following the manufacture's recommendations. The ITS region was amplified from the genomic DNA using PCR with specific primers ITS-L and ITS-4 [48]. PCR reaction mix consisted of 10 µg genomic DNA, 1.5 mM $MgCl_2$, 0.2 mM dNTPs, 1 µM primers ITS-L and ITS-4, 1× PCR buffer and 2 U/100 of

Dynazyme™ II DNA polymerase (Finnzymes, Espoo, Finland). Amplification was performed using PTC-200 thermal cycler (BIO-RAD, Hercules, USA), with the following conditions: 94°C for 5 min (1 cycle), 94°C for 50 s, 52°C for 50 s, 72°C for 50 s (35 cycles) and 72°C for 10 min (1 cycle), and PCR products were resolved in 1.5% agarose gels.

In the first experiment, PCR products of all diploid accessions were purified using ExoSAP-IT® (USB, Cleveland, USA) according to the manufacturer's instructions and used for direct sequencing. PCR products containing polymorphic ITS sequences were cloned into TOPO vector and transformed into *E. coli* electrocompetent cells (Invitrogen Life Technologies, Carlsbad, USA). For each heterogenous accession, 15 to 85 cloned PCR products were sequenced. Sequencing was carried out using the BigDye Terminator v3.1 Cycle Sequencing kit (Applied Biosystems, Foster City, USA) according to the manufacturer's instructions and run on ABI 3730xl DNA analyzer (Applied Biosystems). Nucleotide sequences were edited using Staden Package [63].

454 sequencing and data analysis

Genomic DNA of *M. acuminata* 'Calcutta4' (ITC 0249), *M. balbisiana* 'Pisang Klutug Wulung', *M. ornata* (ITC 0637), *M. textilis* (ITC 0539) and *M. beccarii* (ITC 1070) was prepared from nuclei isolated from healthy young leaf tissues according to Zhang *et al.* [64]. Intact nuclei of *E. gilletii* (ITC 1389) were isolated by flow cytometric sorting following the protocol of Šafář *et al.* [65]. Isolated nuclei were incubated with 40 mM EDTA, 0.2% SDS and 0.25 µg/µl proteinase K for 5 hrs at 37°C, and DNA was purified by phenol/chloroform precipitation.

454 shotgun sequencing libraries were prepared by the GS Titanium library preparation kit (Roche Diagnostics, Rotkreutz, Switzerland). Single-stranded libraries were quantified by a qPCR assay [66] and processed utilizing the GS Titanium SV/LV emPCR and XLR70 sequencing kits according to the manufacturer's instructions (Roche Diagnostics). Sequencing was performed on a half 70×75 picotiter plate for each *Musa* and *Ensete* species. Sequencing reads with the similarity to ITS1-5.8S-ITS2 region and 45S rRNA genes, respectively were identified using RepeatMasker [67] and the consensus sequences of the ITS1-5.8S-ITS2 region were assembled using cap3 program [68].

The read depth (RD) of the ITS sequence region in all six accessions was estimated according to Macas *et al.* [69] using home-made perl script and plots showing RD along the ITS locus (Figure 2) were created using the CALC program (http://www.openoffice.org/product/calc.html).

ITS sequence analysis and secondary structure reconstruction

Nucleotide sequences obtained from both sequencing approaches as well as the nucleotide sequences of ITS regions of other known species from the order Zingiberales that were downloaded from the GenBank database (Table S4), were edited using Staden Package [63] and the sequence boundaries of the spacers were determined by comparison to known rice ITS sequence (GB code: AF169230). Multiple sequence alignments of the ITS regions were done using MUSCLE v3.70+fix1-2 [70] and graphically shown in SeaView v4.2.1 [71]. Multiple sequence alignment was used in accessions with polymorphic ITS sequences to identify individual ITS types, which were then used for further analysis. Nucleotide diversity (π) of ITS region was estimated in MEGA 4 [72].

GC content and sequence identity of ITS regions within heterogenous accessions was calculated using BioEdit (http://www.mbio.ncsu.edu/BioEdit/bioedit.html). Secondary structures of ITS2 region were predicted using mFOLD program version 3.2 [73,74] at the default temperature (37°C) as well as at the 25°C and XRNA program version 1.2.0 Beta (http://rna.ucsc.edu/rnacenter/xrna/xrna_download.html). No differences in folding at both temperatures were observed. The 5.8S rDNA sequences were checked for the presence of the three conserved angiosperm motifs according to Harpke and Peterson [9]. The secondary structure of 5.8S rRNA gene was reconstructed under specific settings for base pairing in XRNA program: for helix B4, F 38 98 3; helix B5, F 41 54 3; helix B6, F 62 89 3; helix B7, F 103 111 3; and for helix B8, F 112 135 4 and F 119 128 3.

Consensus sequences of the ITS regions have been deposited in the GenBank (accession numbers: FR727838–FR727974) as well as on our web site (http://olomouc.ueb.cas.cz/phylogeny-musaceae) where they are freely available.

Phylogenetic analysis

Two different datasets were generated: in order to analyze the position of the clades within Musaceae with respect to other phylogenetic groups of Zingiberales, dataset 1 comprised ITS1-ITS2 concatenated regions of all diploid Musaceae accessions as well as the known ITS1-ITS2 sequences of closely related families Strelitziaceae, Lowiaceae, Heliconiaceae and Costaceae [75,76] (Table S4). Dataset 2 comprised ITS1-ITS2 sequences of all Musaceae accessions including hybrids and was used to analyze the genome composition and evolutionary relationships of hybrid banana species. All datasets were analyzed using Neighbour joining (NJ) and Bayesian inference (BI). Neighbour joining (NJ) method was carried out using SplitsTree4 v4.1.11 [77] based on the Jukes-Cantor and uncorrected *p*-distances. Non-parametric bootstrapping with 1000 pseudoreplicates was performed to assess the nodal support.

The most appropriate model of nucleotide substitution for each dataset was determined using jModelTest v0.1.1 [78]. Likelihood calculations were carried out using integrated PhyML [79] for 11 substitution schemes (88 different models) and the model selected under Akaike information criterion (AIC) [80] was implemented in the BI settings. Bayesian inference analysis was performed in BEAST v1.5.3 [81] with four independent Markov Chain Monte Carlo (MCMC) runs, starting from randomly chosen topologies. The MCMC were run for 10,000,000 generations, data were sampled every 1,000 generations. Log-file outputs were inspected in Tracer [81] to confirm the correct convergence of the analysis. Treefiles from individual MCMC runs were subsequently combined by LogCombiner [81]. The first 25% of the generations were discarded as the burn-in, the maximum clade credibility tree and corresponding posterior probabilities were calculated using TreeAnnotator [81]. Phylogenetic trees were drawn and edited using FigTree (http://tree.bio.ed.ac.uk/software/figtree/) program.

Sequencing data analysis and statistical analysis described above were performed on LINUX systems.

Supporting Information

Figure S1 An example of similarity matrix of heterogenous ITS sequences of *M. jackeyi* (A) and *M. schizocarpa* (B).
(TIFF)

Figure S2 A box plot showing sequence diversities (π) within *Musa* representatives with different genome constitution and hybrid clones. The inter-subspecific diploids were not included in the analysis. Three diploid species which gave unexpected clustering in the phylogenetic tree (see Results and

Discussion) were included in the data set according the results of NJ and BI analysis. * Only three representatives of S genome were included in our study. Two of them contained 3 ITS types including putative pseudogenic sequences and show the highest nucleotide difference among the analyzed species.
(TIFF)

Figure S3 A BI and NJ phylogenetic trees of all ITS types including putative pseudogenes, obtained in diploid Musaceae accessions and selected species of closely related families Strelitziaceae, Lowiaceae, Heliconiaceae and Costaceae based on the ITS1-ITS2 sequence region. NJ tree was constructed from a Jukes-Cantor distance matrix and BI analysis of the same data set was performed in BEAST v1.5.3 using GTR+I+G model of nucleotide substitution. Values below the branches indicate the bootstrap support of NJ analysis and posterior probability of BI, respectively. Main clades and subclades are labeled by different colors as used in Figure 3 and putative pseudogenic types of ITS sequences are marked by asterisks.
(TIFF)

Figure S4 A BI and NJ phylogenetic analysis trees of all evaluated Musaceae accessions including banana hybrids and selected species of the closely related families Strelitziaceae, Lowiaceae, Heliconiaceae and Costaceae based on the ITS1-ITS2 sequence region. NJ tree was constructed from a Jukes-Cantor distance matrix and BI analysis of the same data set was performed in BEAST v1.5.3 using GTR+I+G model of nucleotide substitution. Values below the branches indicate the bootstrap support of NJ analysis and posterior probability of BI, respectively. Main clades and subclades are labeled by different colors as used in Figure 3 and putative pseudogenic types of ITS sequences of hybrids are marked by asterisks.
(TIFF)

Table S1 Diploid representatives of the family Musaceae and *Musa* hybrid clones used in the study. Diploid representatives that are probably inter-subspecific hybrids [83] are marked by asterisk.
(DOC)

Table S2 Sequence characteristics of ITS1-5.8S-ITS2 regions in Musaceae.
(DOC)

Table S3 Table showing genome composition of hybrid clones analyzed in this study and corresponding ITS sequence types identified after NJ and BI analysis.
(DOC)

Table S4 Representatives of the order Zingiberales.
(DOC)

Acknowledgments

We thank to Jiří Macas for providing a perl script, Ms. Radka Tušková for excellent technical assistance, and Prof. Markku Häkkinen, Dr. François Côte and Ir. Ines van den Houwe for providing the plant material.

Author Contributions

Conceived and designed the experiments: EH JD. Performed the experiments: EH JC PC ST. Analyzed the data: EH. Contributed reagents/materials/analysis tools: EdL EH ST. Wrote the paper: EH JD EdL.

References

1. Gilbert W (1986) Origin of life: the RNA world. Nature. pp 319–618.
2. Rogers SO, Bendich AJ (1987) Ribosomal RNA genes in plants: variability in copy number and in intergenic spacer. Plant Mol Biol 9: 509–520.
3. Hillis DM, Dixon MT (1991) Ribosomal DNA – molecular evolution and phylogenetic inference. Q Rev Biol 66: 410–453.
4. Hershkovitz MA, Zimmer EA (1996) Conservation patterns in angiosperm ITS2 sequences. Nucleic Acids Res 24: 2857–2867.
5. Jobes DV, Thien LB (1997) A conserved motif in the 5.8S ribosomal RNA (rRNA) gene is a useful diagnostic marker for plant internal transcribed spacer (ITS) sequence. Plant Mol Biol Rep 15: 326–334.
6. Coleman AW (2007) Pan-eukaryote ITS2 homologies revealed by RNA secondary structure. Nucleic Acids Res 35: 3322–3329.
7. Schultz J, Stefanie M, Gerlach D, Muller T, Wolf M (2007) A common core of secondary structure of the internal transcribed spacer 2 (ITS2) throughout the Eukaryota. RNA 11: 361–364.
8. Wolf M, Achtziger M, Schultz J, Dandekar T (2005) Homology modeling revealed more than 20,000 rRNA internal transcribed spacer 2 (ITS2) secondary structures. Bioinformatics 11: 1616–1623.
9. Harpke D, Peterson A (2008a) 5.8S motifs for the identification of pseudogenic ITS regions. Botany 86: 300–305.
10. Harpke D, Peterson A (2008b) Extensive 5.8S nrDNA polymorphism in *Mammillaria* (Cactaceae) with special reference to the identification of pseudogenic internal transcribed spacer region. J Plant Res 121: 261–270.
11. Ochieng JW, Henry RJ, Baverstock PR, Steane DA, Shepherd M (2007) Nuclear ribosomal pseudogenes resolve a corroborated monophyly of the eucalypt genus *Corymbia* despite misleading hypotheses at functional ITS paralogs. Mol Phylogenet Evol 44: 752–764.
12. Xiao L-Q, Möller M, Zhu H (2010) High nrDNA ITS polymorphism in the ancient extant seed plant *Cycas*: Incomplete concerted evolution and the origin of pseudogenes. Mol Phylogenet Evol 55: 168–177.
13. Dover GA, Strachan T, Coen ES (1982) Molecular drive. Science 218: 1069.
14. Baldwin BG, Sanderson MJ, Porter MJ, Wojciechowski MF, Campbell CS, et al. (1995) The ITS region of nuclear ribosomal DNA: a valuable source of evidence on angiosperm phylogeny. Ann Mo Bot Gard 82: 247–277.
15. Elder JF, Jr., Turner BJ (1995) Concerted evolution of repetitive DNA sequences in eukaryotes. Q Rev Biol 70: 297–320.
16. Wendel JF, Schnabel A, Seelanan T (1995) Bidirectional interlocus concerted evolution following allopolyploid speciation in cotton (*Gossypium*). Proc Natl Acad Sci U S A 92: 280–284.
17. Alvarez I, Wendel JF (2003) Ribosomal ITS sequences and plant phylogenetic inference. Mol Phylogenet Evol 29: 417–434.
18. Kovařík A, Matyášek R, Lim KY, Skalická K, Koukalová B, et al. (2004) Concerted evolution of 18-5.8-26S rDNA repeats in *Nicotiana* allotetraploids. Biol J Linn Soc 82: 615–625.
19. Francisco-Ortega J, Barber JC, Santos-Guerra A, Febles-Hernández R, Jansen RR (2001) Origin and evolution of the endemic genera of Gonospermiinae (Asteraceae: Anthemideae) from the Canary Islands: Evidence from nucleotide sequences of the internal transcribed spacers of the nuclear ribosomal DNA. Am J Bot 88: 161–169.
20. Renner SS, Schaefer H, Kocyan A (2007) Phylogenetics of *Cucumis* (Cucurbitaceae): Cucumber (*C. sativus*) belongs in an Asian/Australian clade far from melon (*C. melo*). BMC Evol Biol 7: 58. Available: http://www.biomedcentral.com/1471-2148/7/58. Accessed 2010 Aug 12.
21. Pettengill JB, Neel MC (2008) Phylogenetic patterns and conservation among North American members of the genus *Agalinis* (Orobanchaceae). BMC Evol Biol 8: 264. Available: http://www.biomedcentral.com/1471-2148/8/264. Accessed 2010 Aug 12.
22. Sun K, Chen X, Ma R, Li C, Wang Q, et al. (2002) Molecular phylogenetics of *Hippophae* L. (Elaegnaceae) based on the internal transcribed spacer (ITS) sequence of nrDNA. Plant Syst Evol 235: 121–134.
23. Liu QL, Ge S, Tang HB, Zhang XL, Zhu GF, et al. (2006) Phylogenetic relationships in *Elymus* (Poaceae : Triticeae) based on the nuclear ribosomal internal transcribed spacer and chloroplast trnL-F sequences. New Phytol 170: 411–4206.
24. Wang QD, Zhang T, Wang JB (2007) Phylogenetic relationships and hybrid origin of *Potamogeton* species (Potamogetonaceae) distributed in China: insight from the nuclear ribosomal internal transcribed spacer sequence (ITS). Plant Syst Evol 267: 65–78.
25. O'Kane SL, Schaal BA, Al-Shehbaz IA (1996) The origin of *Arabidopsis suecica* as indicated by nuclear rDNA sequences. Syst Bot 21: 559–566.

26. Popp M, Oxelman B (2001) Inferring the history of the polyploid *Silene aegaea* (Caryophyllaceae) using plastid and homeologous nuclear DNA sequences. Mol Phylogenet Evol 20: 474–481.
27. Bennet RI, Smith AG (1991) Use of a genomic clone for ribosomal RNA from *Brassica oleracea* in RFLP analysis of *Brassica* species. Plant Mol Biol 16: 685–688.
28. Barkman TJ, Simpson BB (2002) Hybrid origin and parentage of *Dendrochilum acuiferum* (Orchidaceae) inferred in a phylogenetic context using nuclear and plastid DNA sequence data. Syst Bot 27: 209–220.
29. Volkov RA, Komarova NY, Hemleben V (2007) Ribosomal DNA in plant hybrids: inheritance, rearrangement, expression. Syst Biodivers 5: 261–276.
30. Dadejová M, Lim KY, Souček-Skalická K, Matyášek R, Grandbastien M-A, et al. (2007) Transcription activity of rRNA genes correlates with a tendency towards intergenomic homogenization in *Nicotiana allotetraploids*. New Phytol 174: 658–668.
31. Kovařík A, Pires JC, Leitch AR, Lim KY, Sherwood AM, et al. (2005) Rapid concerted evolution of nuclear ribosomal DNA in two *Tragopogon* allopolyploids of recent and recurrent origin. Genetics 169: 931–944.
32. Lim KY, Matyášek R, Kovařík A, Leitch A (2004) Genome evolution in allotetraploid *Nicotiana*. Biol J Linn Soc 82: 599–606.
33. Volkov RA, Borisjuk NV, Panchuk II, Schweizer D, Hemleben V (1999) Elimination and rearrangement of parental rDNA in the allopolyploid *Nicotiana tabacum*. Mol Biol Evol 16: 311–320.
34. Mayol M, Rosselló JA (2001) Why nuclear ribosomal DNA spacers (ITS) tell us different stories in *Quercus*. Mol Phylogenet Evol 19: 167–176.
35. Simmonds NW, Shepherd K (1955) The taxonomy and origins of the cultivated bananas. Bot J Linn Soc 55: 302–312.
36. Cheesman EE (1947) Classification of the bananas II. The genus *Musa* L. Kew Bull 2: 106–117.
37. Gawel NJ, Jarret RL, Whittemore AP (1992) Restriction fragment length polymorphism (RFLP)-based phylogenetic analysis of *Musa*. Theor Appl Genet 84: 286–290.
38. Lanaud C, Tezenas du Montcel H, Jolivot MP, Glaszmann JC, González de León D (1992) Variation of ribosomal gene spacer length among wild and cultivated banana. Heredity 68: 147–156.
39. Wong C, Kiew R, Argent G, Set O, Lee SK, et al. (2002) Assessment of the validity of the section in *Musa* (Musaceae) using AFLP. Ann Bot 90: 231–238.
40. Ude G, Pillay M, Ogundiwin E, Tenkouano A (2002) Analysis of genetic diversity and sectional relationships in *Musa* using AFLP markers. Theor Appl Genet 104: 1239–1245.
41. Risterucci AM, Hippolyte I, Perrier X, Xia L, Caig V, et al. (2009) Development and assesment of Diversity Array technology for high-throughput DNA analysis in *Musa*. Theor Appl Genet 119: 1093–1103.
42. Bartoš J, Alkhimova O, Doleželová M, De Langhe E, Doležel J (2005) Nuclear genome size and genomic distribution of ribosomal DNA in *Musa* and *Ensete* (Musaceae): taxonomic implications. Cytogenet Genome Res 109: 50–57.
43. Hakkinen M (2004) *Musa sanasi*, a new *Musa* species from northern Borneo and discussion of the section *Callimusa* in Borneo. Acta Phytotaxon Geobot 55: 79–88.
44. Hakkinen M (2005) *Musa azizii*, a new *Musa* species (Musaceae) from northern Borneo. Acta Phytotaxon Geobot 56: 27–31.
45. Amorim EP, Vilarinhos AD, Cohen KO, Amorim VBO, dos Santos-Serejo JA, et al. (2009) Genetic diversity of carotenoid-rich bananas evaluated by Diversity Arrays Technology. Genet Mol Biol 32: 96–103.
46. Buhariwalla HK, Jarret RL, Jayashree B, Crouch JH, Ortiz R (2005) Isolation and characterization of microsatellite markers from *Musa balbisiana*. Mol Ecol Notes 5: 327–330.
47. Wang XL, Chiang TY, Roux N, Hao G, Ge XJ (2007) Genetic diversity of wild banana (*Musa balbisiana* Colla) in China as revealed by AFLP markers. Genet Resour Crop Evol 54: 1125–1132.
48. Nwakanma DC, Pillay M, Okoli BE, Tenkouano A (2003) PCR-RFLP of the ribosomal DNA internal transcribed spacer (ITS) provides markers for the A and B genomes in *Musa* L. Theor Appl Genet 108: 154–159.
49. Li L-F, Hakkinen M, Yuan Y-M, Hao G, Ge X-J (2010) Molecular phylogeny and systematics of the banana family (Musaceae) inferred from multiple nuclear and chloroplast DNA fragments, with a special reference to the genus *Musa*. Mol Phylogenet Evol 57: 1–10.
50. Liu A-Z, Kress J, Li D-Z (2010) Phylogenetic analyses of the banana family (Musaceae) based on nuclear ribosomal (ITS) and chloroplast (trnL-F) evidence. Taxon 59: 20–28.
51. Christelová P, Valárik M, Heřbová E, De Langhe E, Doležel J (2011) A multi-gene sequence based phylogeny of the Musaceae (banana) family. BMC Evol Biol (submitted).
52. Bailey CD, Carr TG, Harris SA, Hughes CE (2003) Characterization of angiosperm nrDNA polymorphism, paralogy and pseudogenes. Mol Phylogenet Evol 29: 435–455.
53. Kim ST, Sultan SE, Donoghue MJ (2008) Allopolyploid speciation in *Persicaria* (Polygonaceae): insights from a low-copy nuclear region. Proc Natl Acad Sci U S A 105: 12370–12375.
54. Ritz CM, Schmuths H, Wissemann V (2005) Evolution by reticulation: European dogroses originated by multiple hybridization across the genus Rosa. J Hered 96: 4–14.
55. De Langhe E, Heřbová E, Carpentier S, Doležel J, Swennen R (2010) Did backcrossing contribute to the origine of hybrid edible bananas? Ann Bot 106: 849–857.
56. Wissemann V (2002) Molecular evidence for allopolyploid origin of the *Rosa canina* complex (Rosaceae, Rosoideae). J Appl Bot 76: 176–178.
57. Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, et al. (2005) Genome sequencing in microfabricated high-density picolitre reactors. Nature 473: 376–380.
58. Simmond NW (1962) The evolution of the bananas. London: Longmans. 170 p.
59. Shepherd K (1988) Observation on Musa taxonomy. In: Identification of Genetic Diversity in the Genus *Musa*. Proc. Int. Workshop held at Los Banos, Philippines 5–10 September. INIBAP, Montpellier, France. pp 158–165.
60. Simmond NW (1960) Notes on banana taxonomy. Kew Bull 14: 198–212.
61. Samuel R, Bachmair A, Jobst J, Ehrendorfer F (1998) ITS sequences from nuclear rDNA suggest phylogenetic relationships between Euro-Mediterranean, East Asiatic and North American *Quercus* (Fagaceae). Plant Syst Evol 211: 129–139.
62. Manos P, Doyle JJ, Nixon KC (1999) Phylogeny, biogeography and processes of molecular differentiation in *Quercus* subgenus *Quercus* (Fagaceae). Mol Phylogenet Evol 12: 333–349.
63. Staden R (1996) The Staden sequence analysis package. Mol Biotechnol 5: 233–241.
64. Zhang HB, Zhao X, Ding X, Paterson AH, Wing RA (1995) Preparation of megabase-size DNA from plant nuclei. Plant J 7: 175–184.
65. Šafář J, Noa-Carrazana JC, Vrána J, Bartoš J, Alkhimova O, et al. (2004) Creation of a BAC resource to study the structure and evolution of the banana (*Musa balbisiana*) genome. Genome 47: 1182–1191.
66. Meyer M, Briggs AW, Maricic T, Hober B, Hoffner B, et al. (2008) From micrograms to picograms: quantitative PCR reduces the material demands of high-throughput sequencing. Nucl Acids Res 36: e5. doi:10.1093/nar/gkm1095.
67. Smit AFA, Hubley R, Green P (1996–2004) RepeatMasker Open-3.0. Available: http://www.repeatmasker.org.
68. Huan X, Madan A (1999) CAP3: A DNA sequence assembly program. Genome Res 9: 868–877.
69. Macas J, Neumann P, Navrátilová A (2007) Repetitive DNA in the pea (*Pisum sativum* L.) genome: comprehensive characterization using 454 sequencing and comparison to soybean and *Medicago truncatula*. BMC Genomics 8: 427. Available: http://www.biomedcentral.com/1471-2164/8/427. Accessed 2010 Sep 10.
70. Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32: 1792–1797.
71. Galtier N, Gouy M, Gautier C (1996) SeaView and Phylo_win, two graphic tools for sequence alignment and molecular phylogeny. Comput Appl Biosci 12: 543–548.
72. Kumar S, Tamura K, Nei M (1994) MEGA, Molecular Evolutionary Genetics Analysis software for microcomputers. Bioinformatics 10: 189–191.
73. Mathews DH, Sabina J, Zuker M, Turner DH (1999) Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. J Mol Biol 288: 911–940.
74. Zuker M (2003) Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res 31: 3406–3415.
75. Tomlinson PB (1962) Phylogeny of the Scitamineae: Morphological and anatomical considerations. Evolution 16: 192–213.
76. Kress WJ, Prince LM, Hahn WJ, Zimmer EA (2001) Unraveling the evolution of the families of the Zingiberales using morphological and molecular evidence. Syst Biol 50: 926–944.
77. Huson DH, Bryant D (2006) Application of phylogenetic networks in evolutionary studies. Mol Biol Evol 23: 254–267.
78. Posada D (2008) jModelTest: Phylogenetic model averaging. Mol Biol Evol 25: 1253–1256.
79. Guindon S, Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol 52: 696–704.
80. Akaike H (1974) A new look at the statistical model identification. IEEE Trans Autom Control 19: 716–723.
81. Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol Biol 7: 214. Available: http://www.biomedcentral.com/1471-2148/7/214. Accessed 2010 Sep 10.
82. Wuyts J, De Rijk P, Van de Peer Y, Winkelmans T, De Wachter R (2001) The European large subunit ribosomal RNA database. Nucleic Acids Res 29: 175–177.
83. Perrier X, Bakry F, Carreel F, Jenny C, Horry J-P, et al. (2009) Combining biological approaches to shed light on the evolution of edible bananas. Ethnobotany Research & Applications 7: 199–216.

4.1.1.1 SUPPORTING INFORMATION

Figure S1

Figure S1. An example of similarity matrix of heterogenous ITS sequences of M. jackeyi (A) and M. schizocarpa (B)

A) Matrix of sequence identity for ITS representatives obtained using Sanger sequencing of 18 DNA clones from diploid *M. jackeyi* ITC 0588.

B) Matrix of sequence identity for ITS representatives obtained using Sanger sequencing of 41 DNA clones from diploid *M. schizocarpa* ITC 0856

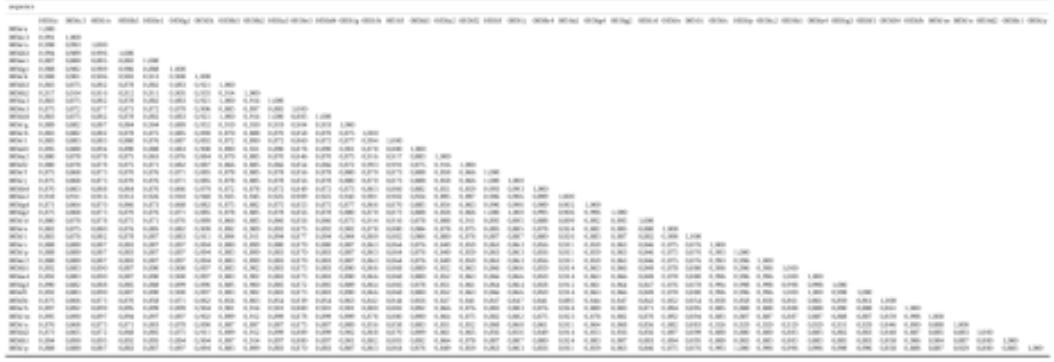


Figure S2

Figure S2. A box plot showing sequence diversities (π)

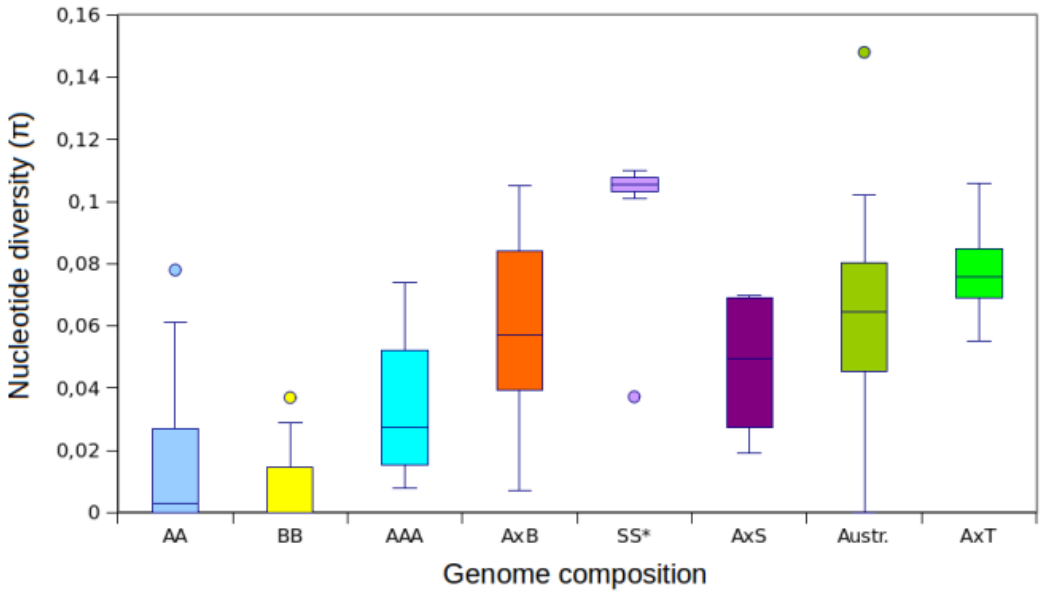


Figure S3

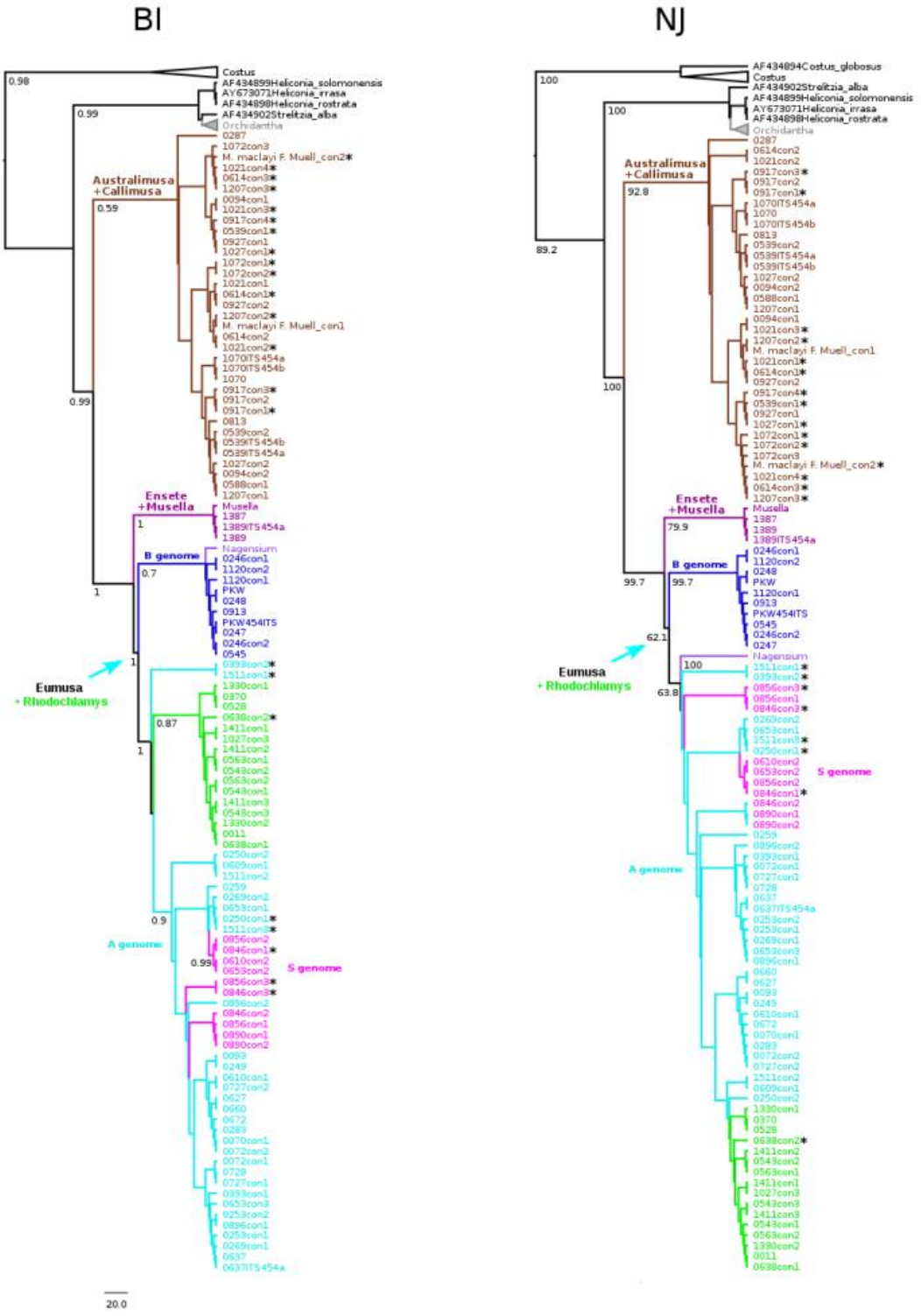

BI
NJ
Australimusa
+Callimusa
Ensete
+Musella
B genome
Eumusa
+ Rhodochlamys
A genome
S genome
20.0

Figure S4

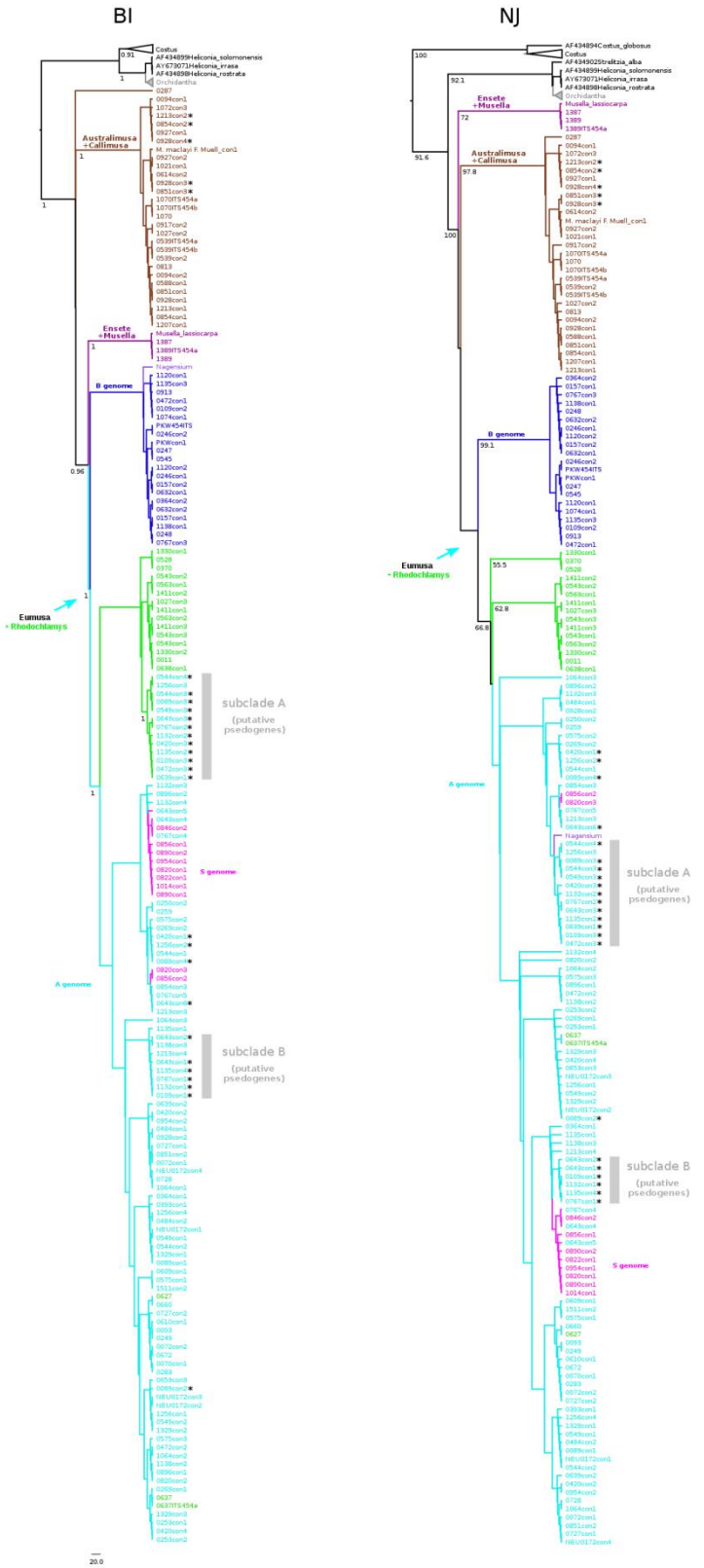

BI
NJ
Australimusa +Callimusa
Ensete +Musella
B genome
Eumusa + Rhodoclidanys
A genome
S genome
subclade A
(putative psedogenes)
subclade B
(putative psedogenes)

Table S1

Table S1: Diploid representatives of the family Musaceae

Genus	Section	Sp./Group	Ssp./Subgroup	Accession name	ITC code	Number of sequences analyzed	Note **
Musa							
	Eumusa						
		acuminata	banksii	M. acuminata ssp. banksii	0896	34	
			burmannica	Tavoy	0072	36	
			burmannica	Long Tavoy	0093	Direct	
			burmannica	Long Tavoy	0283	Direct	
			burmannicoides	Calcutta4	0249	Direct	**454**
			malaccensis	Malaccensis	0250	29	
			malaccensis	Pahang IRFA	0070	18	
			malaccensis	Pahang	0609	15	
			malaccensis	Pahang	0727	29	
			malaccensis	DH Pahang	1511	40	
			microcarpa	Borneo	0253	16	
			siamea	Khae (Phrae)	0660	Direct	
			siamea	Pa (Rayong)	0672	Direct	
			truncata	Truncata	0393	18	
			zebrina	Maia Oa	0728	Direct	
			sucrier	Pisang Mas *	0653	38	
			---	Galeo *	0259	Direct	
			---	Niyarma yik *	0269	40	
			---	Tuu Gia *	0610	15	
		schizocarpa		M. schizocarpa	0846	59	
				M. schizocarpa	0856	42	
				M. schizocarpa	0890	15	
		balbisiana		M. balbisiana (10852)	0094	37	
				Cameroun	0246	15	
				Honduras	0247	Direct	
				Singapuri	0248	Direct	
				M. balbisiana	0545	Direct	
				Tani	1120	23	
				Pisang Klutug Wulung	---	Direct	**454**
		nagensium		M. nagensium	---	15	
	Rhodochlamys	laterita		M. laterita	0627	Direct	
		ornata		M. ornata	0370	Direct	
				M. ornata	0637	Direct	**454**
				Kluai Bou	0528	Direct	
				M. ornata Red fingers	1330	35	
		sanguinea		M. mannii H. Wendl	0543	36	
				M. mannii	1411	39	
		velutina		M. velutina	0011	Direct	
				M. velutina	0638	38	
	Callimusa	beccarii		M. beccarii	1070	Direct	**454**
		coccinea		M. coccinea	0287	Direct	
	Australimusa	Fe´i		Wain	0813	Direct	
				Utafan	0913	Direct	
				Kawaputa	0927	47	
				Menei	1021	30	
				Asupina	1027	52	
		jackeyi		M. jackeyi	0588	18	
		maclayi		M. maclayi type Hung Si	0614	29	
				M. maclayi	1207	23	
				M. maclayi F.Muell	---	39	
		peekelii		M. peekelii ssp. peekelii	0917	25	
		textilis		M. textilis Née	0563	47	
				M. textilis	0539	15	**454**
				M. textilis	1072	29	
Ensete							
		ventricosum		Ensete ventricosum	1387	Direct	
		gilletii		Ensete gilletii	1389	Direct	**454**
Musella							
				Musella lasiocarpa	---	Direct	

* Diploid representatives that are probably inter-subspecific hybrids [83].
** Sequenced by 454.

Table S1: Musa hybrid clones used in the study

Genus	Genome composition	Sp./Group	Accession name	ITC code	Number of sequences analyzed
Musa					
	AAA				
		Ambon	Pisang Bakar	1064	46
		Cavendish	Grande Naine	--- (NEU0172)	51
		Gros Michel	Gros Michel	0484	63
		Red/Green Red	Red Dacca	0575	56
		Orotava	Pisang Kayu	0420	47
		---	Gran Enano	1256	59
		---	Hochuchu	0549	58
		---	Not named M. paradisiaca	0089	61
		---	Not named M. paradisiaca x	0544	54
		---	Novaria	1329	51
	AAB				
		Iholena	Maritú	0639	45
		Plantain	Obino l'Ewai	0109	57
			3 Hands Planty	1132	66
		Popoulou/Maia Maoli	Popoulou (CMR)	1135	68
	ABB				
		Bluggoe	Cachaco	0643	87
			Cachaco Enano	0632	41
			Dole	0767	73
			Kivuvu	0157	42
			Silver Bluggoe	0364	30
		Saba	Saba	1138	55
		---	Pelipita	0472	77
	AS				
		---	Ato	0820	49
		---	Tonton Kepa	0822	56
		---	Ungota	0954	46
	AxS	---	M. acuminata ssp. x M. schizocarpa	1014	22
	AxT				
		---	Karoina	0851	57
		---	Kabulupusa	0928	54
		---	Sar	1213	47
		---	Umbubu	0854	45
	BxT				
		---	Butuhan	1074	22

Table S2

Table S2. Sequence characteristics of ITS1-5.8S-ITS2 region in Musaceae

Accession name	Name of ITS type▲	GC content [%]			Position of nucleotide changes (nt-) in conserved 5.8S motives			Secondary structure of ITS2	Secondary structure of 5.8S	Note
		ITS1	5.8S	ITS2	Motif M1	Motif M2	Motif M3			
M. acuminata ssp. banksii	0896con1	62.91	57.42	70.78	conserved	conserved	conserved	four-helices	conserved	
	0896con2	62.50	56.77	68.22	conserved	conserved	conserved	four-helices	conserved	
Tavoy	0072con1	62.04	57.42	69.44	conserved	conserved	conserved	four-helices	conserved	
	0072con2	62.04	57.42	69.16	conserved	conserved	conserved	four-helices	conserved	
Long Tavoy	0093	62.04	57.42	68.69	conserved	conserved	conserved	four-helices	conserved	
Long Tavoy	0283	62.04	57.42	69.16	conserved	conserved	conserved	four-helices	conserved	
Calcutta4	0249	61.86	57.42	68.69	conserved	conserved	conserved	four-helices	conserved	
	0249ITS454-type1	61.57	57.42	69.16	conserved	conserved	conserved	four-helices	conserved	
Malaccensis	**0250con1**	**60.65**	**53.55**	**62.15**	**conserved**	**conserved**	**conserved**	**not formed**	**conserved**	**pseudogene**
	0250con2	62.96	57.42	69.16	conserved	conserved	conserved	four-helices	conserved	
Pahang IRFA	0070con1	62.04	57.42	69.16	conserved	conserved	conserved	four-helices	conserved	
Pahang	0609con1	63.59	57.42	68.37	conserved	conserved	conserved	four-helices	conserved	
Pahang	0727con1	62.04	57.42	69.77	conserved	conserved	conserved	four-helices	conserved	
	0727con2	62.04	57.42	68.69	conserved	conserved	conserved	four-helices	conserved	
DH Pahang	**1511con1**	**59.26**	**55.48**	**61.79**	**conserved**	**nt-9 "T"**	**conserved**	**not formed**	**conserved**	**pseudogene**
	1511con2	63.13	57.42	68.37	conserved	conserved	conserved	four-helices	conserved	
	1511con3	**59.72**	**53.25**	**61.68**	**conserved**	**conserved**	**conserved**	**not formed**	**not formed**	**pseudogene**
Borneo	0253con1	62.33	57.42	69.77	conserved	conserved	conserved	four-helices	conserved	
	0253con2	62.44	57.69	70.23	conserved	conserved	conserved	four-helices	conserved	
Khae (Phrae)	0660	61.57	57.42	69.16	conserved	conserved	conserved	four-helices	conserved	
Pa (Rayong)	0672	62.04	57.42	69.16	conserved	conserved	conserved	four-helices	conserved	
Truncata	0393con1	62.96	58.06	68.98	conserved	conserved	conserved	four-helices	conserved	
	0393con2	**57.41**	**52.90**	**63.77**	**nt-12 "A"**	**nt-9 "T" nt-14 "T"**	**conserved**	**not formed**	**not formed**	**pseudogene**
Maia Oa	0728	61.40	57.42	69.77	conserved	conserved	conserved	four-helices	conserved	
M. schizocarpa	**0846con1**	**55.35**	**52.26**	**62.15**	**conserved**	**conserved**	**conserved**	**not formed**	**conserved**	**pseudogene**
	0846con2	62.04	57.42	69.16	conserved	conserved	conserved	four-helices	conserved	
	0846con3	**62.50**	**54.19**	**63.08**	**nt-12 "A"**	**conserved**	**conserved**	**four-helices**	**not formed**	**pseudogene**
M. schizocarpa	0856con1	61.11	56.77	67.29	conserved	conserved	conserved	four-helices	conserved	
	0856con2	59.72	54.19	65.89	conserved	conserved	conserved	four-helices	conserved	
	0856con3	**56.94**	**50.97**	**61.68**	**nt-12 "A"**	**nt-9 "A"**	**nt-7 "T"**	**not formed**	**not formed**	**pseudogene**
M. schizocarpa	0890con1	64.06	57.42	68.22	conserved	conserved	conserved	four-helices	conserved	
	0890con2	62.96	57.42	69.44	conserved	conserved	conserved	four-helices	conserved	
M. balbisiana (10852)	0094con1	63.68	57.42	67.32	conserved	conserved	conserved	four-helices	conserved	
	0094con2	66.82	58.06	67.80	conserved	conserved	conserved	four-helices	conserved	
Cameroun	0246con1	60.47	57.42	70.09	conserved	conserved	conserved	four-helices	conserved	
	0246con2	62.33	57.42	70.97	conserved	conserved	conserved	four-helices	conserved	
Honduras	0247	62.33	57.42	70.97	conserved	conserved	conserved	four-helices	conserved	
Singapuri	0248	61.75	57.42	70.37	conserved	conserved	conserved	four-helices	conserved	
M. balbisiana	0545	62.33	57.42	70.97	conserved	conserved	conserved	four-helices	conserved	
Tani	1120con1	62.33	57.42	70.70	conserved	conserved	conserved	four-helices	conserved	
	1120con2	61.86	57.42	70.56	conserved	conserved	conserved	four-helices	conserved	
Pisang Klutug Wulung	PKW	62.33	57.42	70.83	conserved	conserved	conserved	four-helices	conserved	
	PKW454ITS	62.33	57.42	70.97	conserved	conserved	conserved	four-helices	conserved	
M. nagensium	Nagensium-con1	57.41	52.26	60.75	conserved	conserved	nt-7 "T" nt-4 "A"	four-helices	conserved	
M. laterita	0627	61.57	57.42	69.16	conserved	conserved	conserved	four-helices	conserved	
M. ornata	0370	61.57	58.06	69.77	conserved	conserved	conserved	four-helices	conserved	
M. ornata	0637	62.50	57.42	69.63	conserved	conserved	conserved	four-helices	conserved	
	0637ITS454	62.50	57.42	69.63	conserved	conserved	conserved	four-helices	conserved	
Kluai Bou	0528	61.57	58.06	69.77	conserved	conserved	conserved	four-helices	conserved	
M. ornata Red fingers	1330con1	62.21	56.77	70.09	conserved	conserved	conserved	four-helices	conserved	
	1330con2	62.39	57.42	69.48	conserved	conserved	conserved	four-helices	conserved	
M. mannii H. Wendl	0543con1	62.84	57.42	68.69	conserved	conserved	conserved	four-helices	conserved	
	0543con2	62.84	57.42	70.09	conserved	conserved	conserved	four-helices	conserved	
	0543con3	62.84	57.42	68.28	conserved	conserved	conserved	four-helices	conserved	

Table S2. Continued

Accession name	Name of ITS type▲	GC content			Position of nucleotide changes (nt-) in conserved 5.8S motives			Secondary structure of ITS2	Secondary structure of 5.8S	Note
		ITS1	5.8S	ITS2	Motif M1	Motif M2	Motif M3			
M. mannii	1411con1	63.30	57.42	69.63	conserved	conserved	conserved	four-helices	conserved	
	1411con2	62.39	56.77	69.63	conserved	conserved	conserved	four-helices	conserved	
	1411con3	63.01	57.42	68.28	conserved	conserved	conserved	four-helices	conserved	
M. velutina	0011	62.39	57.42	69.48	conserved	conserved	conserved	four-helices	conserved	
M. velutina	0638con1	62.39	57.42	69.48	conserved	conserved	conserved	four-helices	conserved	
	0638con2	**61.93**	**54.19**	**66.82**	**conserved**	**conserved**	**nt-8 "A"**	**not formed**	**conserved**	**pseudogene**
M. beccarii	1070	65.74	58.06	65.37	conserved	conserved	conserved	four-helices	conserved	
	1070ITS454a	65.32	58.06	65.37	conserved	conserved	conserved	four-helices	conserved	
	1070ITS454b	65.77	58.06	65.85	conserved	conserved	conserved	four-helices	conserved	
M. coccinea	0287	66.52	58.06	69.59	conserved	conserved	conserved	four-helices	conserved	
Wain	0813	65.47	58.06	67.65	conserved	conserved	conserved	four-helices	conserved	
Utafan	0913	62.33	57.42	70.97	conserved	conserved	conserved	four-helices	conserved	
Kawaputa	0927con1	58.30	51.61	59.02	conserved	conserved	nt-7 "T"	four-helices*	conserved	
	0927con2	60.81	53.55	64.39	conserved	conserved	nt-7 "T"	four-helices*	conserved	
Menei	1021con1	65.92	56.77	67.32	nt-11 "T"	conserved	conserved	four-helices	conserved	
	1021con2	**60.99**	**54.84**	**61.95**	**conserved**	**conserved**	**conserved**	**not formed**	**not formed**	**pseudogene**
	1021con3	**64.13**	**51.75**	**63.90**	**conserved**	**nt-6 "A"**	**conserved**	**not formed**	**not formed**	**pseudogene**
	1021con4	**59.34**	**52.26**	**61.84**	**conserved**	**conserved**	**conserved**	**not formed**	**not formed**	**pseudogene**
Asupina	**1027con1**	**58.74**	**51.61**	**58.54**	**conserved**	**conserved**	**nt-7 "T"**	**four-helices**	**not formed**	**pseudogene**
	1027con2	65.92	58.06	68.63	conserved	conserved	conserved	four-helices	conserved	
	1027con3	63.30	57.42	69.63	conserved	conserved	conserved	four-helices	conserved	
M. jackeyi	0588con1	66.82	58.06	67.80	conserved	conserved	conserved	four-helices	conserved	
M. maclayi type Hung Si	**0614con1**	**61.43**	**52.90**	**63.41**	**nt-11 "T" nt-16 "T"**	**conserved**	**conserved**	**not formed**	**not formed**	**pseudogene**
	0614con2	64.13	57.42	66.34	conserved	conserved	conserved	four-helices	conserved	
	0614con3	**59.34**	**53.55**	**61.35**	**conserved**	**nt-6 "A"**	**conserved**	**not formed**	**not formed**	**pseudogene**
M. maclayi	1207con1	67.17	58.06	67.32	conserved	conserved	conserved	four-helices	conserved	
	1207con2	**65.02**	**56.77**	**66.34**	**nt-11 "T"**	**conserved**	**conserved**	**four-helices**	**not formed**	**pseudogene**
	1207con3	**59.34**	**53.55**	**61.35**	**conserved**	**nt-6 "A"**	**conserved**	**not formed**	**not formed**	**pseudogene**
M. peekelii ssp. peekelii	**0917con1**	**64.13**	**52.90**	**64.88**	**nt-12 "A"**	**conserved**	**nt-4 "A"**	**four-helices**	**not formed**	**pseudogene**
	0917con2	62.78	52.26	64.88	nt-1 "T" nt-16 "T"	conserved	nt-4 "A"	four-helices	conserved	
	0917con3	**65.02**	**54.19**	**63.41**	**nt-11 "T"**	**nt-7 "T"**	**nt-4 "C"**	**not formed**	**not formed**	**pseudogene**
	0917con4	**57.85**	**51.61**	**63.41**	**conserved**	**conserved**	**conserved**	**four-helices**	**not formed**	**pseudogene**
M. textilis Née	0563con1	62.39	57.42	70.09	conserved	conserved	conserved	four-helices	conserved	
	0563con2	62.84	57.42	69.63	conserved	conserved	conserved	four-helices	conserved	
M. textilis	**0539con1**	**56.50**	**50.97**	**59.02**	**conserved**	**conserved**	**conserved**	**not formed**	**conserved**	**pseudogene**
	0539con2	66.82	58.06	68.78	conserved	conserved	conserved	four-helices	conserved	
	0539ITS454a	66.82	58.06	67.96	conserved	conserved	conserved	four-helices	conserved	
	0539ITS454b	66.37	58.06	68.45	conserved	conserved	conserved	four-helices	conserved	
M. textilis	**1072con1**	**63.68**	**50.97**	**61.39**	**nt-2 "A" nt-11 "T" nt-14 "T"**	**conserved**	**conserved**	**not formed**	**not formed**	**pseudogene**
	1072con2	**62.16**	**53.25**	**61.35**	**conserved**	**nt-6 "T" nt-9 "A"**	**conserved**	**four-helices**	**not formed**	**pseudogene**
	1072con3	58.79	54.19	62.32	nt-11 "T"	nt-6 "A"	conserved	four-helices	conserved	
M. maclayi F. Muell	M. maclayi F. Muell_con1	65.02	57.42	67.32	conserved	conserved	conserved	four-helices	conserved	
	M. maclayi F. Muell_con2	**58.33**	**51.61**	**59.90**	**conserved**	**conserved**	**conserved**	**not formed**	**not formed**	**pseudogene**
Ensete ventricosum	1387	61.11	57.42	66.98	conserved	conserved	conserved	four-helices	conserved	
Ensete gilletii	1389	60.19	57.42	65.77	conserved	conserved	conserved	four-helices	conserved	
	1389ITS454	60.19	57.42	65.77	conserved	conserved	conserved	four-helices	conserved	
Musella lasiocarpa	Musella	61.75	57.42	63.21	conserved	conserved	conserved	four-helices	conserved	

Table S2. Continued

Accession name	Name of ITS type▲	GC content			Position of nucleotide changes (nt-) in conserved 5.8S motives			Secondary structure of ITS2	Secondary structure of 5.8S	Note
		ITS1	5.8S	ITS2	Motif M1	Motif M2	Motif M3			
Pisang Mas	**0653con1**	**61.11**	**52.60**	**61.21**	**conserved**	**conserved**	**conserved**	**not formed**	**not formed**	**pseudogene**
	0653con2	**57.87**	**54.19**	**63.55**	**nt-11 "T" nt-16 "T"**	**nt-14 "T"**	**conserved**	**not formed**	**conserved**	**pseudogene**
	0653con3	62.50	57.42	69.91	conserved	conserved	conserved	four-helices	conserved	
Galeo	0259	61.64	55.84	69.30	conserved	conserved	conserved	four-helices	conserved	
Niyarma yik	0269con1	62.44	57.42	70.23	conserved	conserved	conserved	four-helices	conserved	
	0269con2	61.57	54.84	68.69	conserved	conserved	conserved	four-helices	conserved	
Tuu Gia	0610con1	62.04	57.42	69.16	conserved	conserved	conserved	four-helices	conserved	
	0610con2	**57.87**	**54.19**	**63.55**	**nt-11 "T" nt-16 "T"**	**nt-14 "T"**	**conserved**	**not formed**	**not formed**	**pseudogene**
Pisang Bakar	1064con1	61.57	57.42	69.77	conserved	conserved	conserved	four-helices	conserved	
	1064con2	62.91	57.42	70.78	conserved	conserved	conserved	four-helices	conserved	
	1064con3	63.59	57.42	69.77	conserved	conserved	conserved	four-helices	conserved	
Grande Naine	NEU0172con1	63.59	57.42	68.84	conserved	conserved	conserved	four-helices	conserved	
	NEU0172con2	62.50	57.42	69.44	conserved	conserved	conserved	four-helices*	conserved	
	NEU0172con3	62.50	57.42	69.72	conserved	conserved	conserved	four-helices*	conserved	
	NEU0172con4	62.04	57.42	70.05	conserved	conserved	conserved	four-helices	conserved	
Gros Michel	0484con1	62.04	57.42	68.84	conserved	conserved	conserved	four-helices	conserved	
	0484con2	63.59	57.42	68.84	conserved	conserved	conserved	four-helices	conserved	
Red Dacca	0575con1	62.67	57.42	68.37	conserved	conserved	conserved	four-helices	conserved	
	0575con2	60.19	57.42	68.37	conserved	conserved	conserved	four-helices	conserved	
	0575con3	62.91	57.42	70.78	conserved	conserved	conserved	four-helices	conserved	
Pisang Kayu	**0420con1**	**58.33**	**52.60**	**64.49**	**nt-12 "T"**	**nt-6 "A"**	**conserved**	**not formed**	**not formed**	**pseudogene**
	0420con2	62.50	57.42	69.30	conserved	conserved	conserved	four-helices	conserved	
	0420con3	**56.48**	**52.26**	**63.21**	**nt-16 "T"**	**nt-9 "T"**	**conserved**	**not formed**	**conserved**	**pseudogene**
	0420con4	62.50	57.42	70.37	conserved	conserved	conserved	four-helices	conserved	
Gran Enano	1256con1	63.59	57.42	67.59	conserved	conserved	conserved	four-helices	conserved	
	1256con2	**61.57**	**53.25**	**61.21**	**conserved**	**conserved**	**conserved**	**not formed**	**not formed**	**pseudogene**
	1256con3	59.72	55.49	62.56	conserved	nt-9 "T"	conserved	four-helices	conserved	
	1256con4	63.59	57.42	68.84	conserved	conserved	conserved	four-helices	conserved	
Hochuchu	0549con1	63.59	57.42	68.84	conserved	conserved	conserved	four-helices	conserved	
	0549con2	62.67	58.06	69.44	conserved	conserved	conserved	four-helices	conserved	
	0549con3	**57.60**	**55.48**	**61.79**	**conserved**	**nt-9 "T"**	**conserved**	**four-helices**	**conserved**	**pseudogene**
Not named M. paradisiaca	0089con1	63.59	57.42	68.84	conserved	conserved	conserved	four-helices	conserved	
	0089con2	**62.50**	**57.69**	**69.44**	**conserved**	**conserved**	**conserved**	**four-helices**	**not formed**	**pseudogene**
	0089con3	**57.41**	**54.84**	**61.79**	**conserved**	**nt-9 "T" nt-14 "T"**	**conserved**	**not formed**	**not formed**	**pseudogene**
	0089con4	**60.19**	**53.55**	**61.21**	**conserved**	**conserved**	**conserved**	**four-helices**	**not formed**	**pseudogene**
Not named M. paradisiaca x	0544con1	60.19	53.25	61.21	conserved	conserved	conserved	four-helices	conserved	
	0544con2	63.59	57.42	68.84	conserved	conserved	conserved	four-helices	conserved	
	0544con3	**57.41**	**54.84**	**61.79**	**conserved**	**nt-9 "T" nt-14 "T"**	**conserved**	**not formed**	**conserved**	**pseudogene**
	0544con4	**58.33**	**54.19**	**62.26**	**conserved**	**nt-9 "T"**	**conserved**	**not formed**	**conserved**	**pseudogene**
Novaria	1329con1	63.59	57.42	68.84	conserved	conserved	conserved	four-helices	conserved	
	1329con2	62.50	57.42	69.44	conserved	conserved	conserved	four-helice*	conserved	
	1329con3	62.50	57.42	70.09	conserved	conserved	conserved	four-helices	conserved	
Maritú	**0639con1**	**59.26**	**52.26**	**63.21**	**nt-16 "T"**	**nt-9 "T"**	**conserved**	**not formed**	**conserved**	**pseudogene**
	0639con2	62.50	57.42	70.05	conserved	conserved	conserved	four-helices	conserved	
Obino l'Ewai	**0109con1**	**56.94**	**52.26**	**64.02**	**nt-12 "A"**	**conserved**	**conserved**	**not formed**	**not formed**	**pseudogene**
	0109con2	62.33	57.42	70.97	conserved	conserved	conserved	four-helices	conserved	
	0109con3	**59.26**	**52.26**	**63.21**	**nt-16 "T"**	**nt-9 "T"**	**conserved**	**not formed**	**conserved**	**pseudogene**
3 Hands Planty	**1132con1**	**56.94**	**52.26**	**64.02**	**nt-12 "A"**	**conserved**	**conserved**	**not formed**	**not formed**	**pseudogene**
	1132con2	**59.26**	**52.26**	**63.21**	**nt-16 "T"**	**nt-9 "T"**	**conserved**	**not formed**	**conserved**	**pseudogene**
	1132con3	61.11	52.90	66.36	conserved	conserved	nt-6 "T"	four-helices	conserved	
	1132con4	62.91	57.42	70.05	conserved	conserved	conserved	four-helices	conserved	
Popoulou (CMR)	1135con1	62.50	57.42	70.23	conserved	conserved	conserved	four-helices	conserved	
	1135con2	**59.26**	**52.26**	**63.21**	**nt-16 "T"**	**nt-9 "T"**	**conserved**	**not formed**	**conserved**	**pseudogene**
	1135con3	62.33	57.42	70.97	conserved	conserved	conserved	four-helices	conserved	
	1135con4	**56.94**	**52.26**	**64.02**	**nt-12 "A"**	**conserved**	**conserved**	**not formed**	**not formed**	**pseudogene**
Cachaco Enano	0632con1	62.33	57.42	70.70	conserved	conserved	conserved	four-helices	conserved	
	0632con2	62.33	57.42	70.78	conserved	conserved	conserved	four-helices	conserved	

Table S2. Continued

Accession name	Name of ITS type▲	GC content			Position of nucleotide changes (nt-) in conserved 5.8S motives			Secondary structure of ITS2	Secondary structure of 5.8S	Note
		ITS1	5.8S	ITS2	Motif M1	Motif M2	Motif M3			
Cachaco	**0643con1**	**56.94**	**52.90.**	**64.49**	**nt-12 "A"**	**conserved**	**conserved**	**not formed**	**not formed**	**pseudogene**
	0643con2	**56.48**	**56.77**	**64.10**	**conserved**	**conserved**	**nt-9 "A"**	**not formed**	**conserved**	**pseudogene**
	0643con3	**58.80**	**52.26**	**62.26**	**nt-16 "T"**	**nt-9 "T"**	**conserved**	**not formed**	**conserved**	**pseudogene**
	0643con4	62.04	57.42	69.77	conserved	conserved	conserved	four-helices	conserved	
	0643con5	63.89	56.77	68.98	conserved	conserved	conserved	four-helices	conserved	
	0643con6	**57.87**	**54.19**	**64.09**	**nt-11 "T" nt-16 "T"**	**nt-14 "T"**	**conserved**	**not formed**	**not formed**	**pseudogene**
Dole	**0767con1**	**56.94**	**52.26**	**64.02**	**nt-12 "A"**	**conserved**	**conserved**	**not formed**	**not formed**	**pseudogene**
	0767con2	**58.80**	**52.26**	**62.26**	**conserved**	**nt-9 "T"**	**conserved**	**not formed**	**conserved**	**pseudogene**
	0767con3	62.33	57.42	70.70	conserved	conserved	conserved	four-helices	conserved	
	0767con4	62.96	57.42	69.77	conserved	conserved	conserved	four-helices	conserved	
	0767con5	60.19	56.77	67.43	conserved	conserved	conserved	four-helices	conserved	
Kivuvu	0157con1	62.33	57.42	70.70	conserved	conserved	conserved	four-helices	conserved	
	0157con2	62.33	57.42	70.70	conserved	conserved	conserved	four-helices	conserved	
Silver Bluggoe	0364con1	63.08	57.42	70.78	conserved	conserved	conserved	four-helices	conserved	
	0364con2	62.33	57.42	70.70	conserved	conserved	conserved	four-helices	conserved	
Saba	1138con1	62.33	57.42	70.70	conserved	conserved	conserved	four-helices	conserved	
	1138con2	62.91	56.49	70.64	conserved	conserved	conserved	four-helices	conserved	
	1138con3	62.96	56.13	69.30	conserved	nt-9 "T"	conserved	four-helices	conserved	
Pelipita	0472con1	62.33	57.42	71.30	conserved	conserved	conserved	four-helices	conserved	
	0472con2	62.91	57.42	70.64	conserved	conserved	conserved	four-helices	conserved	
	0472con3	**59.26**	**52.26**	**63.21**	**nt-16 "T"**	**nt-9 "T"**	**conserved**	**not formed**	**conserved**	**pseudogene**
Ato	0820con1	64.22	57.42	68.22	conserved	conserved	conserved	four-helices	conserved	
	0820con2	62.91	56.77	71.04	conserved	conserved	conserved	four-helices	conserved	
	0820con3	59.26	54.84	67.29	conserved	conserved	conserved	four-helices	conserved	
Tonton Kepa	0822con1	63.59	57.42	68.81	conserved	conserved	conserved	four-helices	conserved	
Ungota	0954con1	64.06	57.42	68.69	conserved	conserved	conserved	four-helices	conserved	
	0954con2	62.50	57.42	69.30	conserved	conserved	conserved	four-helices	conserved	
M. acuminata ssp. x M. schizocarpa	1014con1	64.06	57.42	68.22	conserved	conserved	conserved	four-helices	conserved	
M. jackeyi	0851con1	66.82	58.06	67.80	conserved	conserved	conserved	four-helices	conserved	
	0851con2	62.04	57.42	70.05	conserved	conserved	conserved	four-helices	conserved	
	0851con3	**63.23**	**56.77**	**67.80**	**nt-11 "T"**	**conserved**	**conserved**	**four-helices**	**not formed**	**pseudogene**
Kabulupusa	0928con1	66.82	58.06	67.80	conserved	conserved	conserved	four-helices	conserved	
	0928con2	62.04	57.42	70.32	conserved	conserved	conserved	four-helices	conserved	
	0928con3	**61.88**	**53.55**	**61.46**	**nt-11 "T" nt-16 "T"**	**conserved**	**conserved**	**not formed**	**not formed**	**pseudogene**
	0928con4	**56.50**	**48.39**	**56.10**	**nt-11 "T"**	**nt-7 "T" nt-9 "A"**	**nt-4 "A"**	**not formed**	**conserved**	**pseudogene**
Sar	1213con1	67.71	58.06	66.83	conserved	conserved	conserved	four-helices	conserved	
	1213con2	**57.85**	**53.55**	**66.83**	**nt-11 "T"**	**nt-7 "T"**	**conserved**	**four-helices**	**not formed**	**pseudogene**
	1213con3	**57.87**	**54.19**	**63.55**	**nt-11 "T" nt-16 "T"**	**nt-14 "T"**	**conserved**	**not formed**	**not formed**	**pseudogene**
	1213con4	62.50	57.42	68.87	conserved	conserved	conserved	four-helices*	conserved	
Umbubu	0854con1	67.71	58.06	67.32	conserved	conserved	conserved	four-helices	conserved	
	0854con2	**57.85**	**54.84**	**63.90**	**nt-11 "T"**	**conserved**	**conserved**	**not formed**	**not formed**	**pseudogene**
	0854con3	61.11	56.13	69.81	nt-16 "T"	conserved	conserved	four-helices	conserved	
Butuhan	1074con1	62.33	57.14	70.97	conserved	conserved	conserved	four-helices	conserved	

▲ Name of ITS type is based on the ITC accession number and is used in phylogenetic trees. The consensus sequences of the ITS region in diploid and triploid species with polymorphic ITS are labeled as "con + number".

* Putative pseudogenic character of the ITS sequences was estimated in silico (see Materials and Methods).

Table S3

Table S3. Previously described genome composition of hybrid clones and corresponding ITS sequence types identified in this study

Genome composition (MGIS database)	Accession name	Name of ITS type▲	ITS nucleotide sequence type	Position of putative pseudogene in the phylogenetic tree
AAA	Pisang Bakar	1064con1	A	
		1064con2	A	
		1064con3	A	
AAA	Grande Naine	NEU0172con1	A	
		NEU0172con2	A	
		NEU0172con3	A	
		NEU0172con4	A	
AAA	Gros Michel	0484con1	A	
		0484con2	A	
AAA	Red Dacca	0575con1	A	
		0575con2	A	
		0575con3	A	
AAA	Pisang Kayu	**0420con1**	**pseudogenic***	**Eumusa/Rhodochlamys clade (see Supplementary 6)**
		0420con2	A	
		0420con3	**pseudogenic***	**subclade A (see Supplementary 6)**
		0420con4	A	
AAA	Gran Enano	1256con1	A	
		1256con2	**pseudogenic***	**Eumusa/Rhodochlamys clade (see Supplementary 6)**
		1256con3	A	
		1256con4	A	
AAA	Hochuchu	0549con1	A	
		0549con2	A	
		0549con3	**pseudogenic***	**subclade A (see Supplementary 6)**
AAA	Not named M. paradisiaca	0089con1	A	
		0089con2	**pseudogenic***	**Eumusa/Rhodochlamys clade (see Supplementary 6)**
		0089con3	**pseudogenic***	**subclade A (see Supplementary 6)**
		0089con4	**pseudogenic***	**Eumusa/Rhodochlamys clade (see Supplementary 6)**
AAA	Not named M. paradisiaca x	0544con1	A	
		0544con2	A	
		0544con3	**pseudogenic***	**subclade A (see Supplementary 6)**
		0544con4	**pseudogenic***	**subclade A (see Supplementary 6)**
AAA	Novaria	1329con1	A	
		1329con2	A	
		1329con3	A	
AAB	Maritú	**0639con1**	**pseudogenic***	**subclade A (see Supplementary 6)**
		0639con2	A	
AAB	Obino l'Ewai	**0109con1**	**pseudogenic***	**subclade B (see Supplementary 6)**
		0109con2	B	
		0109con3	**pseudogenic***	**subclade A (see Supplementary 6)**
AAB	3 Hands Planty	**1132con1**	**pseudogenic***	**subclade B (see Supplementary 6)**
		1132con2	**pseudogenic***	**subclade A (see Supplementary 6)**
		1132con3	A	
		1132con4	A	
AAB	Popoulou (CMR)	1135con1	A	
		1135con2	**pseudogenic***	**subclade A (see Supplementary 6)**
		1135con3	B	
		1135con4	**pseudogenic***	**subclade B (see Supplementary 6)**
ABB	Cachaco Enano	0632con1	B	
		0632con2	B	
ABB	Cachaco	**0643con1**	**pseudogenic***	**subclade B (see Supplementary 6)**
		0643con2	**pseudogenic***	**subclade B (see Supplementary 6)**
		0643con3	**pseudogenic***	**subclade A (see Supplementary 6)**
		0643con4	A	
		0643con5	A	
		0643con6	**pseudogenic***	**Eumusa/Rhodochlamys clade (see Supplementary 6)**

Table S3. Continued

Genome composition (MGIS database)	Accession name	Name of ITS type▲	ITS nucleotide sequence type	Position of putative pseudogene in the tree
ABB	Dole	**0767con1**	**pseudogenic***	**subclade B (see Supplementary 6)**
		0767con2	**pseudogenic***	**subclade A (see Supplementary 6)**
		0767con3	B	
		0767con4	A	
		0767con5	A	
ABB	Kivuvu	0157con1	B	
		0157con2	B	
ABB	Silver Bluggoe	0364con1	A	
		0364con2	B	
ABB	Saba	1138con1	B	
		1138con2	A	
		1138con3	A	
ABB	Pelipita	0472con1	B	
		0472con2	A	
		0472con3	**pseudogenic***	**subclade A (see Supplementary 6)**
AS	Ato	0820con1	S	
		0820con2	A	
		0820con3	S	
AS	Tonton Kepa	0822con1	S	
AS	Ungota	0954con1	S	
		0954con2	A	
AxS	M. acuminata ssp. x M. schizocarpa	1014con1	S	
AxT	M. jackeyi	0851con1	T	
		0851con2	A	
		0851con3	**pseudogenic***	**Australimusa/Callimusa clade**
AxT	Kabulupusa	0928con1	T	
		0928con2	A	
		0928con3	**pseudogenic***	**Australimusa/Callimusa clade**
		0928con4	**pseudogenic***	**Australimusa/Callimusa clade**
AxT	Sar	1213con1	T	
		1213con2	**pseudogenic***	**Australimusa/Callimusa clade**
		1213con3	**pseudogenic***	
		1213con4	A	
AxT	Umbubu	0854con1	T	
		0854con2	**pseudogenic***	**Australimusa/Callimusa clade**
		0854con3	A	
BxT	Butuhan	1074con1	B	

▲ Name of ITS type is based on the ITC accession number and is used in phylogenetic trees. The consensus sequences of the ITS region in diploid and triploid species with polymorphic ITS are labeled as "con + number".

***** Putative pseudogenic character of the ITS sequences was estimated in *silico* (see Materials and Methods).

Table S4

Table S4: Representatives of the order Zingiberales

Family	Genus	Species	GenBank code	Reference
Costaceae				
	Costus	afer	AY994744	Specht 2006
		allenii	AY041043	Specht et al. 2001
		amazonicus	AY041032	Specht et al. 2001
		amazonicus krukovii	AY972879	Kay et al. 2005
		arabicus	AY041034	Specht et al. 2001
		asplundii	AY972885	Kay et al. 2005
		barbatus	AY994741	Specht 2006
		bracteatus	AY972892	Kay et al. 2005
		chartaceus	AY994719	Specht 2006
		clavigerA	AY994740	Specht 2006
		clavigerB	AY972882	Kay et al. 2005
		comosus comosus	AY972924	Kay et al. 2005
		deistelii	AY994752	Specht 2006
		dirzoi	AY972930	Kay et al. 2005
		erythrocoryne	AY972886	Kay et al. 2005
		erythrophyllus	AY972912	Kay et al. 2005
		erythrothyrsus	AY972889	Kay et al. 2005
		gabonensis	AY994747	Specht 2006
		globosus	AF434894	Johansen 2005
		guanaiensis guanaiensis	AY972883	Kay et al. 2005
		guanaiensis macrostrobilus	AY972917	Kay et al. 2005
		guanaiensis tarmicus	GQ294460	Kay and Yost 2009
		laevis	AY972922	Kay et al. 2005
		lasius	AY972893	Kay et al. 2005
		lateriflorus	AY994734	Specht 2006
		laterifolius	AY972940	Kay et al. 2005
		letestui	AY994733	Specht 2006
		letestui	AY972939	Kay et al. 2005
		lima lima	AY972926	Kay et al. 2005
		lima scabremarginatus	AY972925	Kay et al. 2005
		longebracteolatus	AY972887	Kay et al. 2005
		maculatus	AY994731	Specht 2006
		malortieanus	AY994732	Specht 2006
		megalobractea	AY994730	Specht 2006
		montanus	AY972929	Kay et al. 2005
		mosaicus	AY994728	Specht 2006
		nitidus	GQ294458	Kay and Yost 2009
		osae	AY972927	Kay et al. 2005
		phaeotrichus	AY994721	Specht 2006
		pictus	AY041033	Specht et al. 2001
		plicatus	AY994725	Specht 2006
		productus	AY972895	Kay et al. 2005
		pulverulentus	AY972897	Kay et al. 2005
		ricus	GQ294461	Kay and Yost 2009
		scaber	AY972902	Kay et al. 2005
		spicatus	AY972903	Kay et al. 2005
		spiralisA	AY972915	Kay et al. 2005
		spiralisB	AY972914	Kay et al. 2005
		stenophyllusA	AY994720	Specht 2006
		stenophyllusB	AY972931	Kay et al. 2005
		talbotii	AY972937	Kay et al. 2005
		tappenbeckianus	AY994715	Specht 2006
		vargasii	GQ294462	Kay and Yost 2009
		varzearum	AY994722	Specht 2006
		villosissimus	AY994713	Specht 2006
		vinosus	AY972923	Kay et al. 2005
		wilsonii	AY972921	Kay et al. 2005
		woodsonii	AY994712	Specht 2006
		zingiberoides	AY972910	Kay et al. 2005
Strelitziaceae				
	Strelitzia	alba	AF434902	Johansen 2005
Heliconiaceae				
	Heliconia	irrasa	AY673071	Prince and Kress 2006
		rostrata	AF434898	Johansen 2005
		solomonensis	AF434899	Johansen 2005

Table S4: Continued

Family	Genus	Species	GenBank code	Reference
Lowiaceae				
	Orchidantha	borneensisA	AF434877	Johansen 2005
		borneensisB	AF434876	Johansen 2005
		chinensis	AF434878	Johansen 2005
		fimbriata	AF434879	Johansen 2005
		grandiflora	AF434880	Johansen 2005
		holttumii	AF434881	Johansen 2005
		inouei	AF434882	Johansen 2005
		longiflora	AF434883	Johansen 2005
		maxillarioides	AF434884	Johansen 2005
		quadricolor	AF434885	Johansen 2005
		sabahensis	AF434886	Johansen 2005
		siamensis	AF434887	Johansen 2005
		suratii	AF434890	Johansen 2005

4.1.2 Molecular Analysis and Genomic Distribution of Major DNA Satellites in Banana (*Musa* spp.)

Čížková J, Hřibová E, Humplíková L, Christelová P, Suchánková P, Doležel J

PLoS ONE 8(1): e54808, 2013.

[JČ, EH, JD concieved and designed the experiments. JČ, EH, LH, PC, PS performed the experiments. JČ, EH, JD analyzed the data. JD contributed reagents/materials/analysis tools. JČ, EH, JD wrote the paper.]

OPEN ACCESS Freely available online

PLOS | ONE

Molecular Analysis and Genomic Organization of Major DNA Satellites in Banana (*Musa* spp.)

Jana Čížková, Eva Hřibová, Lenka Humplíková, Pavla Christelová, Pavla Suchánková, Jaroslav Doležel*

Institute of Experimental Botany, Centre of the Region Haná for Biotechnological and Agricultural Research, Olomouc, Czech Republic

Abstract

Satellite DNA sequences consist of tandemly arranged repetitive units up to thousands nucleotides long in head-to-tail orientation. The evolutionary processes by which satellites arise and evolve include unequal crossing over, gene conversion, transposition and extra chromosomal circular DNA formation. Large blocks of satellite DNA are often observed in heterochromatic regions of chromosomes and are a typical component of centromeric and telomeric regions. Satellite-rich loci may show specific banding patterns and facilitate chromosome identification and analysis of structural chromosome changes. Unlike many other genomes, nuclear genomes of banana (*Musa* spp.) are poor in satellite DNA and the information on this class of DNA remains limited. The banana cultivars are seed sterile clones originating mostly from natural intra-specific crosses within *M. acuminata* (A genome) and inter-specific crosses between *M. acuminata* and *M. balbisiana* (B genome). Previous studies revealed the closely related nature of the A and B genomes, including similarities in repetitive DNA. In this study we focused on two main banana DNA satellites, which were previously identified *in silico*. Their genomic organization and molecular diversity was analyzed in a set of nineteen *Musa* accessions, including representatives of A, B and S (*M. schizocarpa*) genomes and their inter-specific 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. acuminata* and *M. balbisiana* which may aid in determining genomic constitution in interspecific hybrids. In addition to improving the knowledge on *Musa* satellite DNA, our study increases the number of cytogenetic markers and the number of individual chromosomes, which can be identified in *Musa*.

Citation: Čížková J, Hřibová E, Humplíková L, Christelová P, Suchánková P, et al. (2013) Molecular Analysis and Genomic Organization of Major DNA Satellites in Banana (*Musa* spp.). PLoS ONE 8(1): e54808. doi:10.1371/journal.pone.0054808

Editor: Khalil Kashkush, Ben-Gurion University, Israel

Received July 20, 2012; **Accepted** December 17, 2012; **Published** January 23, 2013

Funding: This work was supported by the Academy of Sciences of the Czech Republic (award KJB500380901), by the Bioversity International (LoA CfL 2011/51), Ministry of Education, Youth and Sports of the Czech Republic (award LG12021) and the European Regional Development Fund (Operational Programme Research and Development for Innovations No. ED0007/01/01). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: dolezel@ueb.cas.cz

Introduction

A significant part of nuclear genomes in plants, including those with small genomes such as rice, *Brachypodium* and *Arabidopsis* is occupied by various types of repetitive DNA sequences [1,2,3]. Repetitive DNA sequences are classified based on genomic organization of repetitive units as dispersed and tandem. Dispersed repeats, represented by various classes of transposable elements encode for proteins which facilitate their replication and integration into nuclear genome [4]. Tandem repeats usually contain non-coding sequences organized in tandem arrays [5,6,7,8,9]. The arrays may be large and consist of thousands or even millions of repetitive units arranged in head-to-tail orientation [10,11]. The processes by which satellites arise and evolve are not well understood, and unequal crossing over, gene conversion, transposition and formation of extra chromosome circular DNA (eccDNA) were implicated [12].

The analysis of repetitive DNA and satellite DNA in particular, has been hampered by the presence of multiple copies of the same or similar sequences arranged in tandem. However, the situation changed recently with the advent of new sequencing technologies, which enable identification and reconstruction of tandem organized units even from partial genome sequencing data [13,14,15]. On line with the analysis at DNA level, long-range organization of satellite DNA has been studied using cytogenetic methods; fluorescence *in situ* hybridization (FISH) being the most frequent. Large blocks of satellite DNA are typically observed in heterochromatic regions of chromosomes and are usually located in centromeric and telomeric regions [16,17]. Chromosome loci rich in satellite DNA usually show specific banding pattern and this makes them useful cytogenetic markers to discriminate individual chromosomes [18,19,20,21]. The presence of chromosome landmarks enables identification of individual chromosomes and is a prerequisite to study structural changes accompanying evolution and speciation and to follow chromosome behavior and transmission in interspecific hybrids [22,23].

Bananas are a staple food and important export commodity in many countries of humid tropics. Most of banana cultivars are seed sterile diploid and triploid clones originating from natural inter- and intra-specific crosses involving wild diploid species of genus *Musa*: *M. acuminata* (A genome) and *M. balbisiana* (B genome) [24]. Both species belong to section *Eumusa* ($x = 11$); other three sections are recognized within the genus *Musa* based on chromosome number and morphology: *Rhodochlamys* ($x = 11$),

Australimusa ($x = 10$) and *Callimusa* ($x = 10$ or $x = 9$). Nuclear genome of *Musa* is relatively small (1C×~550–750 Mbp) [25,26] and divided to 1–2 µm long and morphologically similar chromosomes. This makes cytogenetic studies difficult and there is no reliable method to identify all chromosomes within a karyotype and discriminate parental chromosomes in hybrids. Genomic *in situ* hybridization (GISH) has been used occasionally to identify parental chromosomes in some hybrid cultivars [27,28]. However, the method discriminates only (peri)centromeric chromosome regions. Apart from identifying parental chromosomes, the paucity of chromosome-specific landmarks and markers hampers studies on karyotype evolution and chromosome behavior in hybrids. To date, only a few DNA sequences and probes such as rDNA, some DNA repeats and BAC clones were found useful to study the organization of plant nuclear genomes at cytogenetic level [29,30,31].

Recently, a variety of repetitive DNA elements including two new satellite DNA sequences were identified *in silico* from 454 sequencing data of *M. acuminata* cv. 'Calcutta 4' [13]. Even though most of retrotransposons were found dispersed along all banana chromosomes, some of them (a LINE element and a CRM retrotransposon) were located in (peri)centromeric regions, while the DNA satellites were identified on specific chromosome loci [13]. These results indicated a potential of these sequences as cytogenetic markers.

In this work we characterized genomic organization of two main banana DNA satellites together with other DNA sequences (a LINE element, rRNA genes, a BAC clone) in nineteen accessions of *Musa*, including inter-specific hybrids. Molecular analysis of the two DNA satellites revealed their sequence conservation within and between the accessions. The mode of their genomic distribution makes them suitable as cytogenetic markers. We show that the LINE-like element is present in centromeric regions of all nineteen accessions and can be used as centromere-specific cytogenetic marker. The present work improves the knowledge of genome structure in *Musa* and expands the number of individual chromosomes which can be identified. Differences in fluorescent labeling patterns obtained after FISH with a set of probes can be used to support determination of genomic constitution in inter-specific hybrids.

Materials and Methods

Plant Material and Genomic DNA Extraction

In vitro rooted plants of 18 *Musa* accessions were obtained from the International Transit Centre (ITC, Katholieke Universitiet, Leuven, Belgium). Five rooted plants of *M. balbisiana* clone 'Pisang Klutuk Wulung' were obtained from Dr. François Côte (CIRAD, Guadeloupe). The *in vitro* plants were transferred to soil and all plants were maintained in a greenhouse. Table 1 lists all accessions used in the present study.

Genomic DNA was prepared from nuclei isolated from healthy young leaf tissues according to Zhang et al. [32]. Isolated nuclei were incubated with 40 mM EDTA, 0.2% SDS and 0.25 µg/µl proteinase K for 5 hours at 37°C; DNA was purified by phenol/chloroform precipitation.

Estimation of Genome size

Nuclear genome size was estimated according to Bartoš et al. [25]. Suspensions of cell nuclei were prepared by chopping leaf tissues with a razor blade in a glass Petri dish containing 500 µl Otto I solution (0.1 M citric acid, 0.5% v/v Tween 20). Approximately 50 mg of young *Musa* leaf and 10 mg of leaf of soybean (*Glycine max* L. cv. Polanka, 2C = 2.5 pg DNA) [26] which served as internal standard were used for sample preparation. Crude homogenate was filtered through a 50 µm nylon mesh, nuclei were pelleted (300 *g*, 5 min) and resuspended in 300 µl Otto I solution. After 1 hour incubation at room temperature, 900 µl Otto II solution (0.4 M Na_2HPO_4) [33] supplemented with 50 µg/ml RNase, 50 µg/ml propidium iodide and 3 µl/ml 2-mercaptoethanol, were added. Samples were analyzed using Partec PAS flow cytometer (Partec GmbH, Münster, Germany) equipped with 488-nm argon laser. At least 5,000 nuclei were analyzed per sample. Three individuals were analyzed in each accession, and each individual was measured three times on three different days. Nuclear DNA content was calculated from individual measurements following the formula:

$$2C\ nuclear\ DNA\ content[pg] = 2.5 \times G_1 peak\ mean\ of\ Musa / G_1 peak\ mean\ of\ Glycine$$

Mean nuclear DNA content was then calculated for each plant and converted to the number of base pairs considering 1 pg DNA equal to 0.978×10^9 bp [34].

Dot-plot

Satellites maTR_CL18 and maTR_CL33 (GenBank accessions: JX624137 and JX624137) were originally identified *in silico* after low-pass 454 sequencing genomic DNA of *M. acuminata* 'Calcutta 4' in our previous study [13] and their presence was confirmed in the recently published genome assembly of *M. acuminata* 'DH Pahang' [35]. In order to confirm tandem arrangement of both repeats prior to our experimental work, we have analyzed the sequence data using dotter [36].

Southern Hybridization

Genomic organization of DNA satellites maTR_CL18 and maTR_CL33 [13] was analyzed in all 19 *Musa* accessions. Aliquots of genomic DNA corresponding to 1×10^7 copies of monoploid (1Cx) nuclear genomes were digested using *Dra*I whose restriction site was identified in maTR_CL18 and maTR_CL33 sequences. Digested DNA was size-fractionated by 1.2% agarose gel electrophoresis and transferred to Hybond N+ nylon membranes (Amersham, Bath, UK). Probes for satellites CL18 and CL33 were prepared after PCR amplification from genomic DNA of *M. acuminata* 'Calcutta 4' using specific primers (Table 2) and labeling by biotin. Southern hybridization was done at 68°C overnight and signals were detected using BrightStar® BioDetect™ kit under manufacturer's instructions (Ambion, Austin, USA).

Serial dilutions of genomic DNA and PCR products of satellite repeats used as standards were dot-blotted onto Hybond-N+ membranes (Amersham) with the aim to estimate copy number of satellite DNA sequences in the nuclear genomes of all evaluated *Musa* accessions. PCR products of satellite DNAs labeled by biotin were used as hybridization probes. Dots of genomic DNA and standards that gave the same intensity of hybridization signals were identified after visual inspection. Copy numbers of individual probes were estimated assuming that 1 pg of genomic DNA equals 0.978×10^9 bp [34].

Sequencing and Sequence Data Analysis of Satellite DNA

Sequences homologous to the two satellites (in the present work termed CL18-like and CL33-like satellites) were amplified from genomic DNA of all 19 *Musa* species using specific primers (Table 2). PCR reaction mix contained 10 ng genomic DNA,

Table 1. List of *Musa* accessions and their genome sizes.

Species	Subspecies	Accession name	ITC code	Genomic constitution[a]	2C nuclear DNA content [pg]		Mean monoploid genome size [Mbp/1C$_x$][b]	Reference[c]
					mean	± SD		
M. acuminata	*burmannicoides*	Calcutta4	0249	AA	1.226	0.004	600	Bartoš *et al.* [25]
	burmannica	Long Tavoy	0283	AA	1.238	0.01	605	*
	zebrina	Maia Oa	0728	AA	1.325	0.02	648	*
	malaccensis	DH Pahang	1511	AA	1.214	0.027	594	*
		Tuu Gia	0610	AA	1.261	0.02	617	*
M. balbisiana		Cameroun	0246	BB	1.121	0.003	548	Lysák *et al.* [67]
		Honduras	0247	BB	1.133	0.002	554	*
		Tani	1120	BB	1.126	0.014	551	*
		Pisang Klutuk Wulung	–	BB	1.132	0.025	554	*
M. schizocarpa		*Musa schizocarpa*	0560	SS	1.373	0.015	671	*
		Musa schizocarpa	1002	SS	1.364	0.003	667	*
Hybrids		Obino l'Ewai	0109	AAB	1.777	0.004	579	Lysák *et al.* [67]
		Maritú	0639	AAB	1.847	0.005	602	Lysák *et al.* [67]
		3 Hands Planty	1132	AAB	1.786	0.019	582	*
		Pelipita	0472	ABB	1.751	0.002	571	Lysák *et al.* [67]
		Balonkawe	0473	ABB	1.814	0.034	591	*
		Ato	0820	AS	1.294	0.015	633	*
		Tonton Kepa	0822	AS	1.318	0.021	645	*
		Umbubu	0854	AT	1.397	0.009	683	*

[a]Genome constitution as described in MGIS database (http://www.crop-diversity.org/banana/#AvailableITCAccessions).
[b]1 pg = 0.978×10^9 bp [34].
[c]The study in which DNA content was estimated.
*Estimated in the present work.
doi:10.1371/journal.pone.0054808.t001

1.5 mM $MgCl_2$, 0.2 mM dNTPs, 1 µM primers, 1×PCR buffer and 2 U/100 of Dynazyme™ II DNA polymerase (Finnzymes, Espoo, Finland). Amplification was performed as follows: 94°C for 5 min (1 cycle), 94°C for 50 s, 57°C for 90 s, 72°C for 50 s (35

Table 2. Primers used for PCR amplification of satellite DNA and a LINE element.

Type of repetitive DNA	DNA sequence name	Primer name	Primer sequence	PCR product length
Tandem repeat				
	CL18	CL18-1[a]	5′-ATCATGGGCCAACACTTGAT	Ladder-like pattern
			5′-TCGTGAGAGCGGGTTAGAGT	
		CL18-2[a]	5′- ATCATGGGCCAACACTTGAT	Ladder-like pattern
			5′-TGAGAGCGGGTTAGAGTTCC	
		CL18-IN1[b]	5′-CGAATGATTTGATGTCATCTCC	491 bp
			5′-AGTGTTGGCCCATCATGTTT	
		CL18-IN2[b]	5′-GCAATGTTTCAACTCATTACCAA	179 bp
			5′-GATGCTACCGGGAAAAATTG	
	CL33	CL33-1[a, b]	5′-AATCGATCGAACCTCGACAT	130 bp
			5′-TCCCAATAAATTTGCCTTCG	
Non-LTR retrotransposon				
	LINE element	CL1SCL8/452[a]	5′-TGAAAGCAGCTTGATTTGGA	218 bp
		CL1SCL8/452[a]	5′-CAAGGCTTGCCAACATTTTT	

[a]Primers used to prepare probes for FISH.
[b]Primers were used to prepare probes for Southern hybridization.
doi:10.1371/journal.pone.0054808.t002

cycles) and 72°C for 10 min (1 cycle). PCR products were purified by PCR Rapid Kit (Invitek, Berlin, Germany), ligated into pCR-XL-TOPO vector and transformed into One Shot TOP10 electrocompetent *E. coli* (Invitrogen Life Technologies, Carlsbad, USA). For each *Musa* accession, 24 to 60 cloned PCR products were sequenced. Sequencing was carried out using the BigDye Terminator v3.1 Cycle Sequencing kit (Applied Biosystems, Foster City, USA) according to the manufacturer's instructions and run on ABI 3730×l DNA analyzer (Applied Biosystems).

Nucleotide sequences were edited using Staden Package [37] the consensus sequences of CL33-like and the two parts of CL18-like satellites were assembled using cap3 program [38]. Multiple sequence alignments were done using MAFFT program v6.717-1 (–globalpair –maxiterate 1000) [39] and graphically displayed in SeaView v4.2.1. [40]. Sequence diversity was identified using DNA Sam program [41]. Sequence logos were generated using WebLogo tool [42]. SplitsTree4 v4.1.11 [43] was used to construct cladograms based on the Jukes-Cantor, K2P and uncorrected *p*-distances. Non-parametric bootstrapping with 1000 pseudoreplicates was performed to assess the nodal support. Cladograms were drawn and edited using FigTree program (http://tree.bio.ed.ac.uk/software/figtree/).

Chromosome Preparations

Metaphase spreads were prepared according to Doleželová et al. [29]. Actively growing root tips were pre-treated in 0.05% 8-hydroxyquinoline for 3 hrs and fixed in 3:1 ethanol : acetic acid. Fixed roots were washed in a solution of 75 mM KCl and 7.5 mM EDTA (pH 4) and meristem tips were digested in a mixture of 2% (w/v) pectinase and 2% (w/v) cellulase in 75 mM KCl and 7.5 mM EDTA (pH 4) for 90 min at 30°C. Protoplast suspension was then filtered through a 150 µm nylon mesh and pelleted. The pellet was resuspended in 75 mM KCl and 7.5 mM EDTA (pH 4) and incubated for 5 min at room temperature. After pelleting, the protoplasts were washed three times with 70% ethanol, and 5 µl of suspension were dropped onto a slide. Shortly before drying out, 5 µl of 3:1 fixative were added to the drop to induce protoplast bursting. Finally, the slide was rinsed in 100% ethanol and air-dried.

Fluorescence In Situ Hybridization (FISH)

BAC clone 2G17 [30] was labeled by Dig-11-dUTP or Bio-16-dUTP Nick Translation (Roche Applied Science, Penzberg, Germany) according to manufacturer's instructions. FISH probes for 45S rDNA and 5S rDNA were obtained by labeling *Radka*1 DNA clone (45S rDNA) and *Radka*2 DNA clone (5S rDNA) [31] with digoxigenin-11-dUTP or biotin-16-dUTP (Roche Applied Science). Both probes were labeled by PCR using M13 forward and reverse primers (Invitrogen). Banana-specific LINE element and tandem repeats maTR_CL18 and maTR_CL33 [13] were labeled with digoxigenin-11-dUTP or biotin-16-dUTP (Roche) by PCR with specific primers (Table 2).

Hybridization mixture consisting of 50% formamide, 10% dextran sulfate in 1×SSC and 1 µg/ml labeled probe was added onto slides and denatured at 80°C for 3 min. The hybridization was carried out at 37°C overnight. The sites of probe hybridization were detected using anti-digoxigenin-FITC (Roche Applied Science) and streptavidin-Cy3 (Vector Laboratories, Burlingame, USA), and the chromosomes were counterstained with DAPI. The slides were examined with Olympus AX70 fluorescence microscope and the images of DAPI, FITC and Cy-3 fluorescence were acquired separately with a cooled high-resolution black and white CCD camera. The camera was interfaced to a PC running the MicroImage software (Olympus, Tokyo, Japan).

Results

We applied various molecular and cytogenetic approaches to study genomic organization of two major DNA satellites in *Musa*: maTR_CL18 and maTR_CL33. We first confirmed using dot-plot analysis that both repeats were indeed tandem organized (Figure S1). The molecular structure and variability of tandem organized regions was investigated after sequencing products obtained after PCR with specific primers. We then used Southern hybridization and fluorescence *in situ* hybridization to estimate the copy number of ma_TR_CL18 and maTR_CL33 and characterize distribution of the repeats on mitotic chromosomes, respectively.

Sequence Analysis of DNA Satellites

A set of primers specific for CL18 and CL33 satellites was used to amplify these repeats from all nineteen *Musa* accessions. The PCR products were cloned and sequenced to study diversity of repeats within and between banana accessions (Tables S1, S2, S3).

Considering the length of CL18 repetitive units (2226 bp) and heterogeneity of the sequenced regions, two parts of the CL18-like repetitive unit were amplified and assembled: CL18-part1 and CL18-part2, which corresponded to 921 bp and 825 bp regions of CL18 [13]. The sequenced region represented 1728 bp (77.6%) of CL18. In most of the *Musa* accessions, sequences representing both parts of CL18 repetitive units were obtained (Tables S1, S2, S3). The exceptions were *M. acuminata* 'Tuu Gia' and both representatives of *M. schizocarpa*, where PCR amplification followed by sequencing did not identify sequences similar to part1 of CL18. In some other clones (*M. acuminata* 'DH Pahang', 'Long Tavoy' and 'Maia Oa', *M. balbisiana* 'Honduras' and 'Tani', and hybrid clones 'Balonkawe' (ABB), 'Ato' (AS) and 'Tonton Kepa' (AS)), relatively short sequences corresponding to part1 of CL18 were obtained. On the other hand, sequences homologous to part2 of CL18 were not obtained from hybrid clone '3 Hands Planty' (AAB).

In most of the accessions, both parts of CL18-like repeat regions shared high homology to the maTR_CL18 satellite as identified in 454 data [13]. Lower sequence homology to maTR_CL18 was observed in accessions in which PCR amplification did not result in amplification of both parts of maTR_CL18 (Tables S1, S2). Sequence diversity of regions corresponding to part2 of maTR_CL18 repetitive unit was slightly lower within individual accessions as compared to part1 (Tables S1, S2). Phylogenetic analysis based on Neighbor Joining resulted in a tree in which different clades did not contain species-specific CL18-like repetitive units. Thus, we were not able to identify B genome-specific variants of CL18 suitable as B-genome specific markers (Figure 1C, D). Altogether our observations indicated that both parts homologous to CL18 satellite were highly conserved within and between the analyzed accessions (Figures 1A, B).

Sequencing PCR products obtained with primers specific for the CL33 satellite confirmed its presence in all *Musa* accessions. In several accessions ('Maia Oa', 'Tuu Gia', 'Pelipita', '3 Hands planty', 'Ato' and 'Tonton Kepa') 188 bp repetitive units were present in addition to 134 bp repetitive units of CL33. This significant length difference was due to a 54 bp insertion/deletion. Similarly, *M. acuminata* 'Long Tavoy' contained a 110 bp repetitive unit of CL33.

Most of the analyzed *Musa* accessions shared high homology to maTR_CL33 satellite [13]. Lower sequence homology to maTR_CL33 was observed in both accessions of *M. schizocarpa* and its interspecific hybrid (Table S3). In general, a low sequence diversity of sequences corresponding to maTR_CL33 repetitive

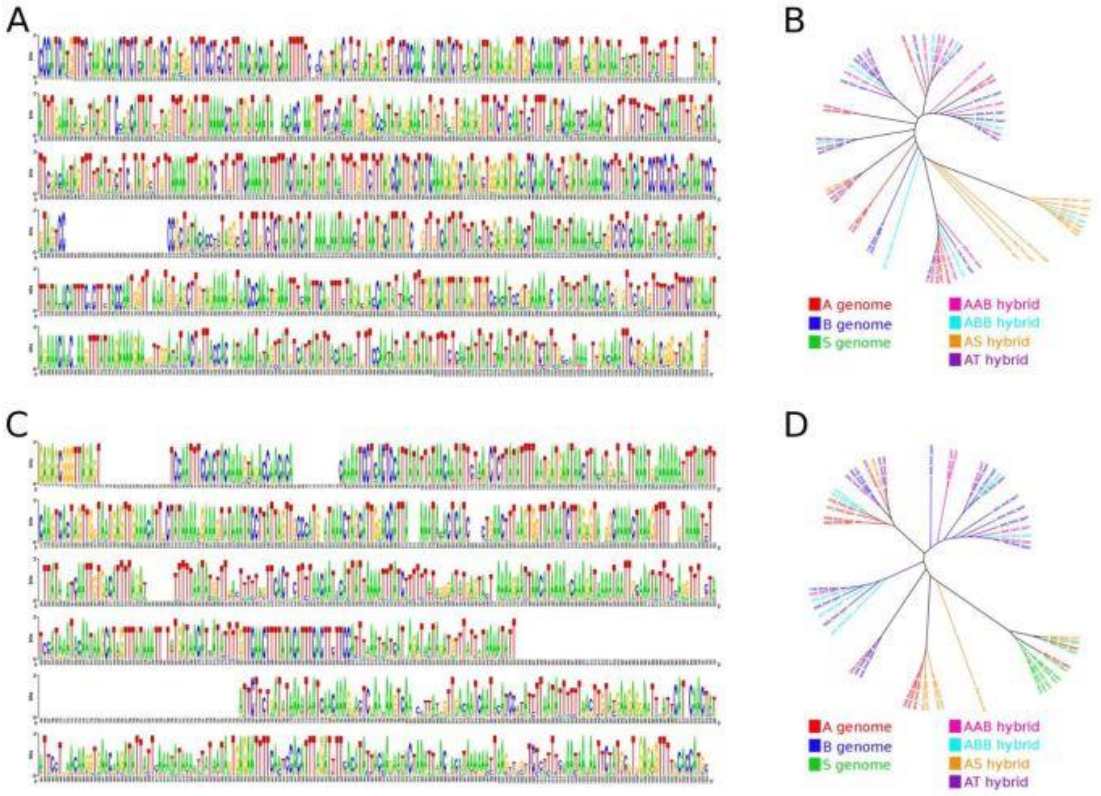


Figure 1. Sequence diversity of the two parts of CL18-like satellite sequences. Consensus sequences of both parts of CL18-like satellite were reconstructed from all analyzed accessions and are displayed as sequence logo (A: CL18-part1; C: CL18-part2). Neighbor-Joining trees constructed from a Jukes-Cantor distance matrix and rooted to the midpoint show diversity of reconstructed parts of CL18-like satellite units obtained from all studied species (B: CL18-part1; D: CL18-part2).
doi:10.1371/journal.pone.0054808.g001

unit was observed. The only exceptions were revealed in 'Balonkawe' (ABB) and one representative of *M. balbisiana* (Table S3). The high level of homology of sequences corresponding to maTR_CL33 satellite within and between all *Musa* accessions is demonstrated in Figure 2A. Similarly to the results of CL18-like repeats, cluster analysis based on Neighbor Joining showed that there were no A or B genome-specific CL33-like satellite units in the current set of *Musa* accessions (Figure 2B).

Nuclear DNA Content

The amount of nuclear DNA was estimated after flow cytometric analysis of propidium iodide-stained nuclei. All analyses resulted in histograms of relative DNA content with clearly defined peaks corresponding to G_1 nuclei of *Musa* and the reference standard (*Glycine max*) with coefficient of variation ranging from 2.5% to 4.5%. 2C nuclear DNA content ranged from 1.121 pg to 1.397 pg for diploid accessions and from 1.751 pg to 1.847 pg for triploid accessions (Table 1). Among the diploid accessions, the lowest DNA content was found in *M. balbisiana* (2C = 1.121–1.133 pg), while the highest 2C DNA content of 1.397 pg was estimated in 'Umbubu' (AT genomic constitution).

Southern Hybridization

Southern hybridization with probes for maTR_CL18 and maTR_CL33 was carried out to investigate genomic organization of the satellites. Unfortunately, a typical pattern showing individual n-mers units was not observed, most probably because *Dra*I endonuclease did not digest all repetitive units in the analyzed accessions. Both satellites gave ladder-like pattern (Figure 3A) with various hybridization signal intensities. While the signals of maTR_CL18 repeat were visible in all *Musa* accessions, the presence maTR_CL33 repeat was not confirmed in any of the four accessions of *M. balbisiana*. Accordingly, maTR_CL33 repeat resulted in weak signals in hybrid clones with ABB genome constitution (Figure 3B).

Dot-blot analysis showed a limited variation in copy number of maTR_CL18 repeat. Its abundance ranged approximately from 1×10^3 to 2×10^3 per monoploid (1Cx) genome of *M. acuminata* except of *M. acuminata* 'DH Pahang' with 3.5×10^3 to 5×10^3 copies. A similar estimate (1×10^3 to 2×10^3) was made also in both accessions of *M. schizocarpa* and in *M. balbisiana* 'Honduras'. The remaining *balbisiana* species contained of 3.5×10^3 to 5×10^3 copies of maTR_CL18 repeat. Monoploid genome of hybrid clones 'Ato' and 'Tonton Kepa' (AS) contained 1×10^3 to 2×10^3 copies of maTR_CL18 units, while hybrid clones 'Umbubu' (AT), 'Maritú'

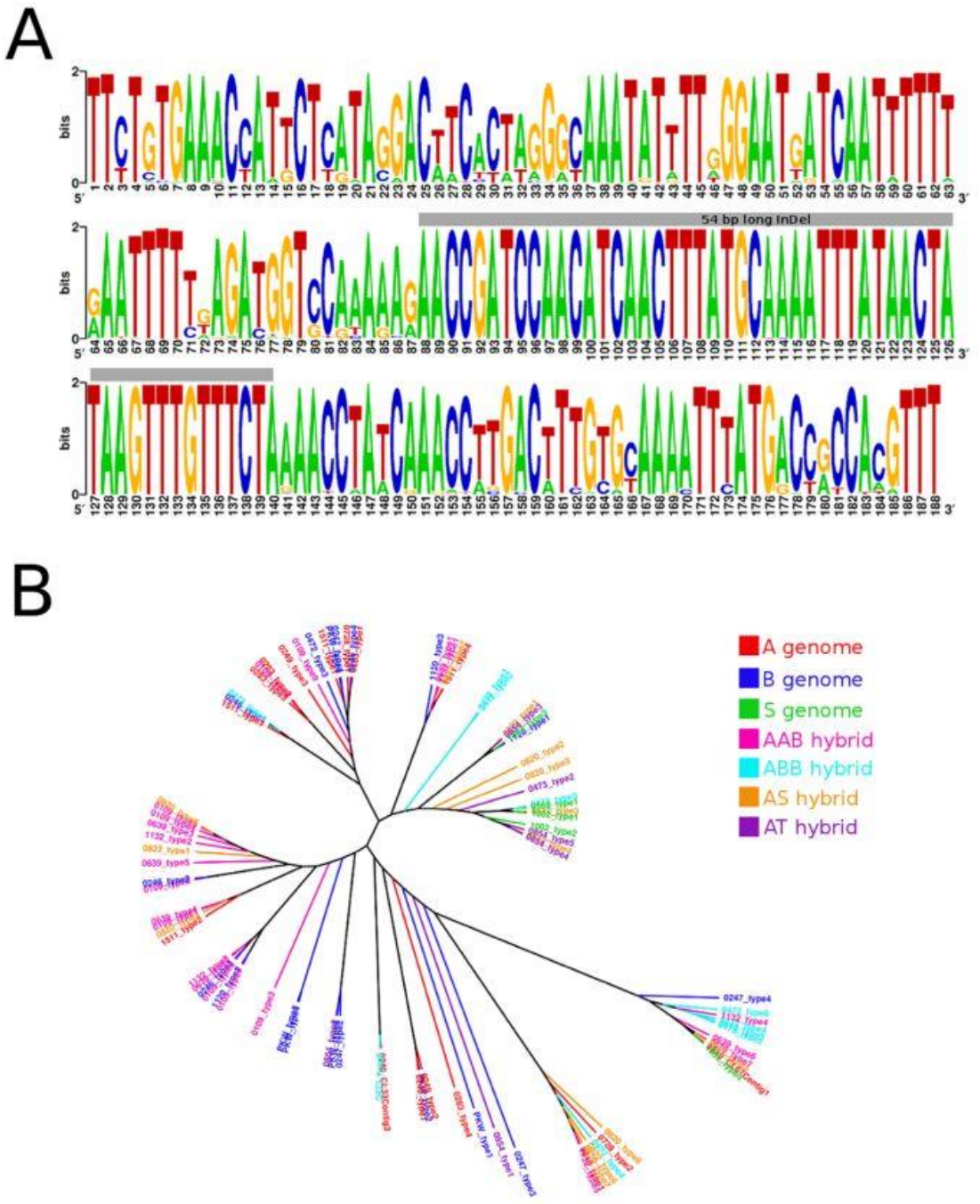


Figure 2. Diversity of CL33-like satellite sequence. Consensus sequence of CL33-like satellite was reconstructed from all obtained sequence units of all analyzed accessions and is displayed as sequence logo (A). Neighbor-Joining tree constructed from a Jukes-Cantor distance matrix and rooted to the midpoint shows diversity of individual types of CL33-like repetitive units obtained from all studied accessions (B).
doi:10.1371/journal.pone.0054808.g002

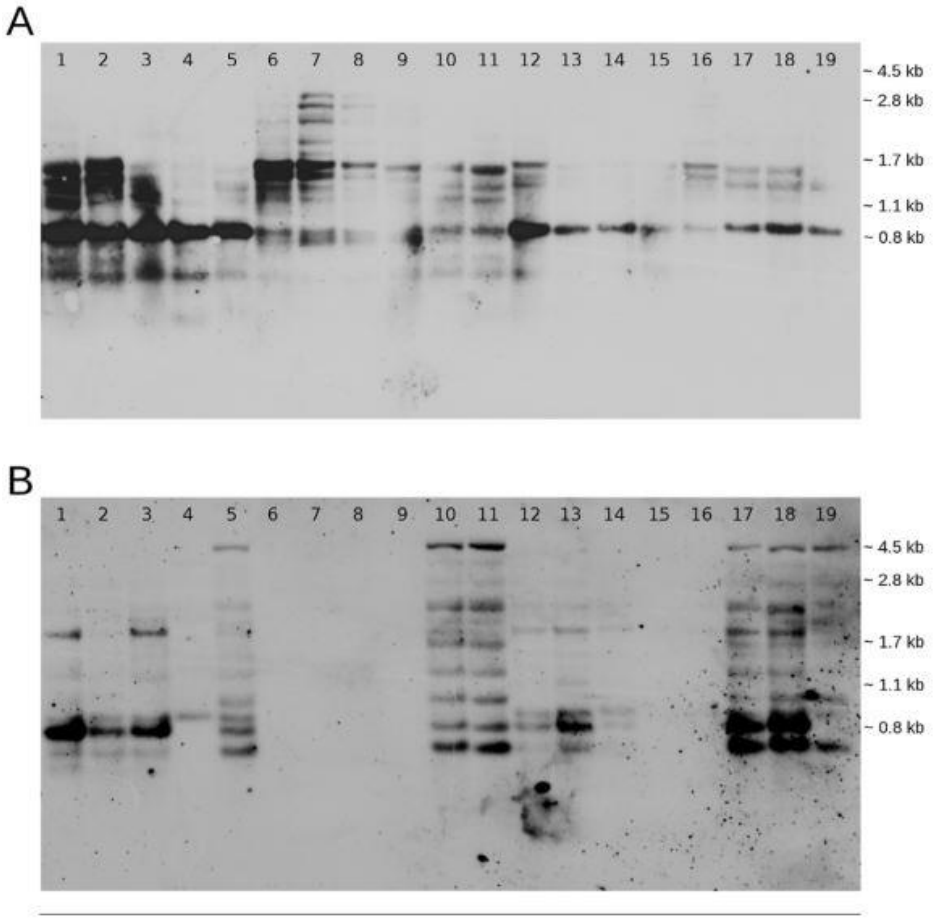


Figure 3. Southern hybridization of genomic DNA isolated from nineteen *Musa* accessions. Samples of genomic DNA corresponding to 1×10^7 copies of monoploid ($1C_x$) nuclear genomes were digested using *DraI* restriction enzyme and hybridized with probes for CL18-like satellite (A) and CL33-like satellite (B) at hybridization stringency of 85%.
doi:10.1371/journal.pone.0054808.g003

and '3 Hands Planty' contained approximately 3×10^3–4×10^3 copies of the repeat. Finally, hybrid clones 'Pelipita' and 'Balonkawe' (ABB) as well as 'Obino l'Ewai' (AAB) comprised ~5×10^3 to 6.5×10^3 copies of maTR_CL18 (Table S4).

The number of copies of maTR_CL33 repeat ranged from 2.5×10^3 to 4×10^3 per monoploid genome of *M. acuminata* 'Long Tavoy', 'Zebrina' and 'Tuu Gia' as well as in both hybrid clones 'Ato' and 'Tonton Kepa'. 5.5×10^3 to 7×10^3 copies of maTR_CL33 repetitive units were estimated in monoploid genomes of *M. acuminata* 'Calcutta4' and 'DH Pahang' and both *schizocarpa* accessions. The same amount of maTR_CL33 (5.5×10^3 to 7×10^3) was revealed in hybrid clones 'Obino l'Ewai', 'Maritú' and '3 Hands Planty' (AAB) as well as in 'Umbubu' (AT). Dot-blot analysis failed to detect presence of maTR_CL33 repeat in *M. balbisiana* species and in hybrid clones with ABB genomic constitution.

Mapping DNA Sequences on Mitotic Chromosomes

In order to study chromosome organization of the two satellites and the position of their loci in relation to other cytogenetic landmarks, multicolor FISH was done in a set of 19 *Musa* species and clones with probes for single-copy BAC clone 2G17, 45S rDNA and 5S rDNA. Moreover, taking into the account the relatively small size of banana chromosomes and difficulty in identifying primary constrictions, a probe for LINE element, which was identified and characterized previously in *M. acuminata* cv. 'Calcutta 4' [13], was used to label putative centromeric regions. The element was found in centromeric regions of all *Musa* accessions and thus was used as a centromeric marker to aid in constructing the idiograms (Figure 4, Figure S2, Figures S3A–G).

45S rRNA genes. FISH with a probe for 45S rDNA revealed its localization exclusively to secondary constriction; the only exception being a diploid hybrid clone 'Ato' (AS genome). Apart from this cultivar, in which additional 45S rDNA locus was observed on one of the nucleolar organizing chromosomes (Figure 3F), the number of 45S rDNA loci in mitotic metaphase plates corresponded to ploidy level (Table 3). 45S rDNA was detected in secondary constriction on one chromosome pair in diploid accessions, while three loci corresponding to 45S rDNA were identified in triploid accessions.

5S rRNA genes. Gene loci for 5S rRNA were observed mostly in distal chromosome regions in all accessions (Figures 4, 5, Figures S3A – G, Figure S4). Diploid chromosome sets of *M. acuminata* 'Long Tavoy', *M. acuminata* 'TuuGia' and *M. acuminata* 'Maia Oa' contained four signals corresponding to 5S rRNA genes, *M. acuminata* 'Pahang' contained six signals and eight loci of 5S rRNA gene clusters were detected in mitotic metaphase plates of *M. acuminata* 'Calcutta 4'. Diploid representatives of B genome showed presence of six 5S rDNA loci with the exception of *M. balbisiana* 'Cameroun', which contained only four 5S rDNA clusters. Both representatives of *M. schizocarpa* (SS genome) contained six chromosome pairs bearing 5S rRNA genes (Figure 4, Figure S3C, Table 3). In triploid hybrid banana clones, eight to ten 5S rDNA loci were observed (Figures S3D, E, Table 3).

BAC clone 2G17. BAC clone 2G17, which was used in this work as additional cytogenetic marker, localized to subtelomeric regions on one chromosome pair in all diploid accessions with the exception of *M. balbisiana* 'Honduras' in which four signals were observed. Only one chromosome gave FISH signal in a diploid hybrid clone 'Umbubu' (AT genome), which most probably originated from a cross between *M. acuminata* and *M. textilis* (T genome). In triploid AAB and ABB clones, three or four loci containing DNA sequences homologous to BAC 2G17 were found (Figure 5, Figures S3D, E, Figure S4, Table 3).

Satellite repeats in diploid genomes. Satellite CL18 localized to subtelomeric regions on one pair of mitotic chromosomes in all accessions of *M. acuminata* and always co-localized with BAC clone 2G17. Satellite CL33 localized to subtelomeric regions on one chromosome pair in 'Maia Oa' and on two chromosome pairs of the remaining *M. acuminata* accessions. Further FISH experiments revealed that CL33 satellite co-localized with CL18 and BAC 2G17 on one chromosome pair (Figures 4, 5, Figure S3A, Table 3) in all *M. acuminata* accessions. Similar FISH patterns were observed for CL18 and CL33 in *M. schizocarpa* (Figures 4, 5, Figure S3C, Table 3).

Three *M. balbisiana* genotypes ('Honduras', 'Tani', 'Pisang Klutuk Wulung') contained two chromosome pairs bearing satellite CL18 in their subtelomeric regions. On one of the chromosome pairs, a probe for CL18 co-localized with BAC 2G17 and on the other co-localized with 5S rDNA (Figures 4, 5, and Figure S3B). The fourth accession of *M. balbisiana* - 'Cameroun' - had only three CL18 loci as a FISH signal was missing on the chromosome with 5S rDNA locus (Figure 5, and Figure S3B). CL33 gave no visible signals on metaphase chromosomes of *M. balbisiana*.

Satellite repeats in AAB and ABB hybrids. Out of a number of the existing interspecific hybrids with a combination

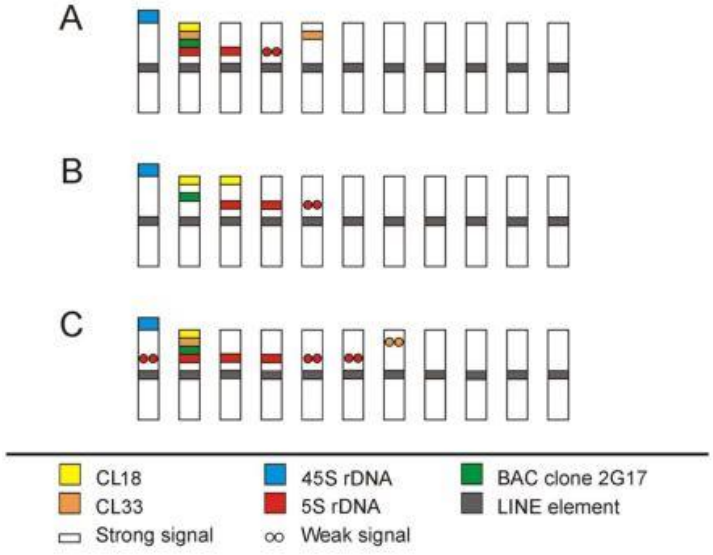


Figure 4. Idiograms of three diploid *Musa* accessions. (A) *M. acuminata* 'DH Pahang' ITC 1511; (B) *M. balbisiana* 'Pisang Klutuk Wulung'; (C) *M. schizocarpa* ITC 0560.
doi:10.1371/journal.pone.0054808.g004

Table 3. Number of loci detected using FISH with probes for CL18 and CL33 satellites, rRNA genes and BAC clone 2G17 on mitotic metaphase plates in 19 accessions of *Musa*.

Accession name	ITC* code	Genomic constitution	Chromosome number (2n)	Number of probe signals (per metaphase plate)				
				45S rDNA	5S rDNA	2G17	CL18	CL33
Calcutta 4	0249	AA	22	2	8	2	2	4
Long Tavoy	0283	AA	22	2	4	2	2	4
Maia Oa	0728	AA	22	2	4	2	2	2
DH Pahang	1511	AA	22	2	6	2	2	4
Tuu Gia	0610	AA	22	2	4	2	2	4
Cameroun	0246	BB	22	2	4	2	3	–
Honduras	0247	BB	22	2	6	2	4	–
Tani	1120	BB	22	2	6	2	4	–
Pisang Klutuk Wulung	–	BB	22	2	6	2	4	–
Musa schizocarpa	0560	SS	22	2	12	2	2	4
Musa schizocarpa	1002	SS	22	2	12	2	2	4
Obino l'Ewai	0109	AAB	33	3	9	4	4	4
Maritú	0639	AAB	33	3	8	4	4	4
3 Hands Planty	1132	AAB	33	3	9	4	4	3
Pelipita	0472	ABB	33	3	10	3	6	1
Balonkawe	0473	ABB	33	3	10	4	6	1
Ato	0820	AS	22	2	9	2	2	4
Tonton Kepa	0822	AS	22	2	9	2	2	4
Umbubu	0854	AT	21	2	4	1	1	2

*International Transit Centre (http://www.bioversityinternational.org/research/conservation/genebanks/musa_international_transit_centre.html).
doi:10.1371/journal.pone.0054808.t003

of A and B genomes, three hybrid accessions with AAB genomic constitution ('Obino l'Ewai', 'Maritú' and '3 Hands Planty') and two ABB hybrids ('Pelipita' and 'Balonkawe') were chosen for this study. All AAB interspecific hybrids and one ABB hybrid ('Balonkawe') gave signals of BAC 2G17 in subtelomeric regions of four different chromosomes. Among them, one chromosome was bearing an additional weak signal of BAC 2G17. In the second ABB hybrid 'Pelipita', BAC 2G17 localized in subtelomeric regions of three different metaphase chromosomes (Figures S3E, S4). 5S rDNA loci were found on eight metaphase chromosomes in the genome of 'Maritú', on nine chromosomes of 'Obino l'Ewai' and '3 Hands Planty', while 'Pelipita' and 'Balonkawe' contained ten clusters corresponding to 5S rDNA.

Satellite CL18 localized in subtelomeric regions on four mitotic chromosomes in AAB hybrids; FISH with satellite CL33 resulted in three visible signals in subtelomeric regions of metaphase chromosomes of '3 Hands Planty' and four signals were detected on chromosomes of 'Obino l'Ewai' and 'Maritú'. Further analysis showed that two chromosomes carried both CL18 and CL33 satellite repeats. (Figures S3D, S4, Table 3). Hybrid ABB clones 'Balonkawe' and 'Pelipita' contained six signals of satellite CL18 which were detected on different chromosomes and one signal of satellite CL33. Multicolor FISH with the combination of probes corresponding to satellite repeats and 5S rRNA genes revealed one chromosome carrying CL18, CL33, 5S rDNA and BAC 2G17. Two other chromosomes exhibiting combination of FISH signals for 2G17 and CL18 satellite repeat and three additional chromosomes bearing signals for CL18 satellite and 5S rRNA genes (Figure S3E, Table 3).

Satellite repeats in AS and AT hybrids. FISH was also used to study genomic distribution of the two satellites in diploid hybrid clones 'Ato' and 'Tonton Kepa' (AS genome) and 'Umbubu' (AT genome). Nuclear genome of AS hybrids contained two signals of CL18 and CL33 which co-localized in subtelomeric regions on two different chromosomes, while nuclear genome of AT hybrid clone 'Umbubu' contained only one signal of CL18 satellite and two signals of CL33 satellite (Figure S3F, G, Table 3).

Chromosome Identification

Multicolor FISH experiments with combinations of probes for 45S rDNA, 5S rDNA, banana-specific LINE-like element, DNA satellites CL18 and CL33 and a single copy BAC clone 2G17 facilitated identification of various numbers of chromosomes in diploid accessions of *Musa* (Figures 4, S3A–G, Table 3). Within the A genome representatives, three chromosomes could be distinguished in *M. acuminata* 'Maia Oa' and four chromosomes were identified in *M. acuminata* clones 'Long Tavoy' and 'Tuu Gia'. In *M. acuminata* 'Calcutta 4', four chromosomes exhibited a specific hybridization pattern. Although two other chromosomes carried 5S rDNA sequences in this clone, the two chromosomes could not be discriminated from each other as the probes hybridized to similar chromosome regions. Five different chromosomes could be identified in *M. acuminata* 'DH Pahang'.

Within the representatives of *M. balbisiana*, four chromosomes could be discriminated in 'Cameroun', five chromosomes were distinguished in 'Tani' and 'Pisang Klutuk Wulug' and six different chromosomes were identified in 'Honduras'. Three different chromosomes could be recognized in the two S genome representatives. Moreover, two chromosomes carrying strong

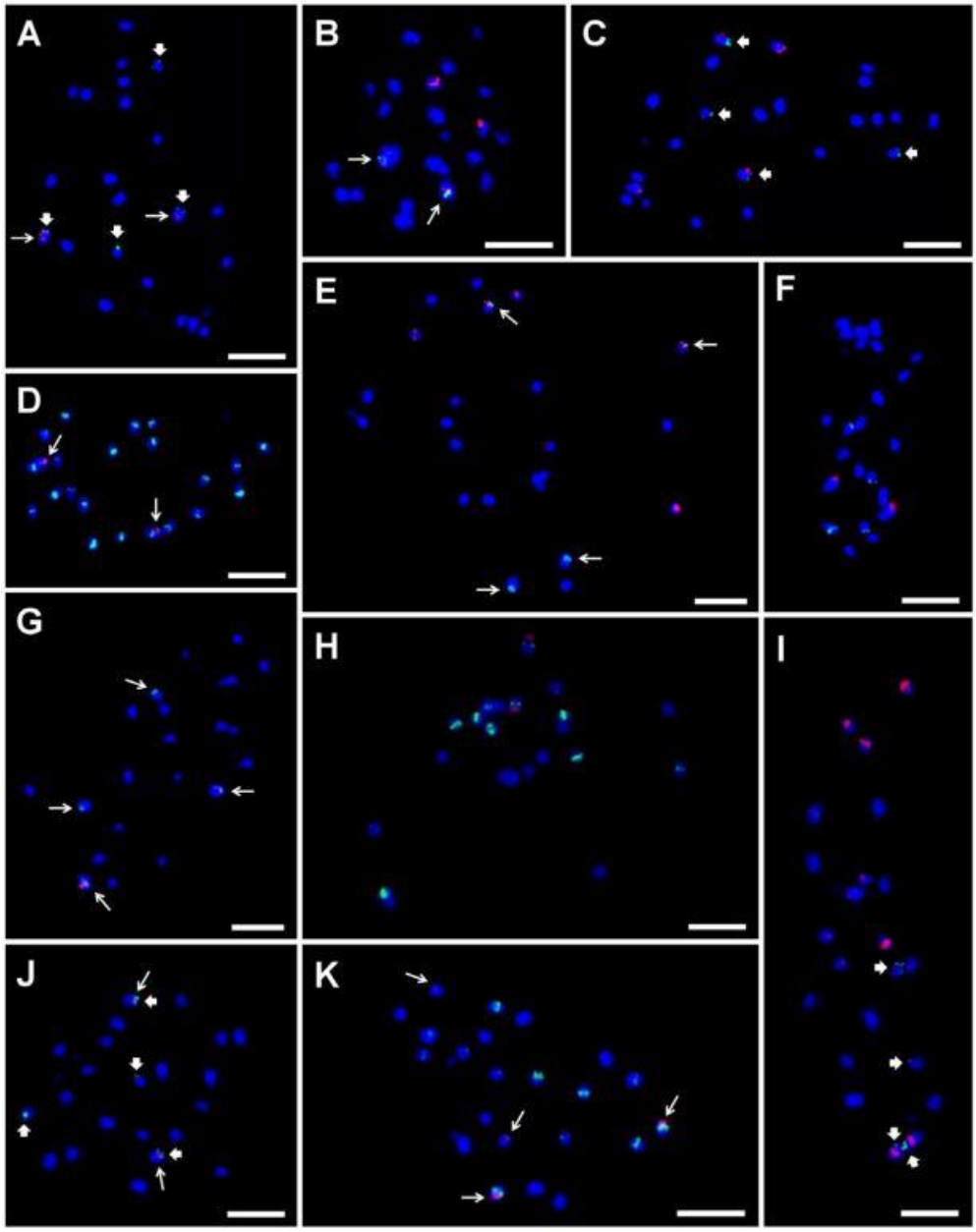


Figure 5. Examples of genomic distribution of DNA satellites as determined on mitotic metaphase chromosomes of diploid *Musa* accessions after FISH. Chromosomes were counterstained with DAPI (blue). Sites of CL18 and CL33 probe hybridization are marked by long and thick arrows, respectively. (A) CL18 (red) and CL33 (green) on chromosomes of 'Long Tavoy'. (B) 5S rDNA (red) and CL18 (green) on chromosomes of 'Maia Oa'. (C) 5S rDNA (red) and CL33 (green) on chromosomes of 'Long Tavoy'. (D) CL18 (red) and LINE element (green) on chromosomes of 'Tuu Gia'. (E) 5S rDNA (red) and CL18 (green) on chromosomes of 'Pisang Klutuk Wulung'. (F) BAC clone 2G17 (red) and 5S rDNA (green) on chromosomes of 'Cameroun'. (G) BAC clone 2G17 (red) and CL18 (green) on chromosomes of 'Tani'. (H) 45S rDNA (red) and 5S rDNA (green) on chromosomes of *M. schizocarpa* ITC 0560. (I) 5S rDNA (red) and CL33 (green) on chromosomes of *M. schizocarpa* ITC 1002. (J) CL18 (red) and CL33 (green) on chromosomes of *M. schizocarpa* ITC 1002. (K) CL18 (red) and 5S rDNA (green) on chromosomes of *M. schizocarpa* ITC 1002. Bar = 5 μm.
doi:10.1371/journal.pone.0054808.g005

signals of 5S rDNA and other two chromosomes carrying weak signals of 5S rDNA were identified in their karyotypes. Unfortunately, the probes localized to similar chromosome regions in both cases and the chromosomes could not be differentiated from each other. Due to the lack of probes specific for A, B, S and T genomes, we were not able to identify homologous chromosomes in inter-specific hybrids.

Discussion

Satellite repeats may undergo rapid changes in copy number and nucleotide sequence, which may result in the evolution of genus and species-specific repeats [18,19,20,44,45]. At the same time, their repetitive units are usually homogenized within a species due to concerted evolution [6,46,47,48]. These features make them an attractive object for evolutionary and cytogenetic studies. Although about 45% of the *Musa* genome consists of various repetitive DNA sequences [35], satellite DNA repeats represent only ~0.3% of the genome [13]. The presence of only two major satellite repeats in *Musa* and their low copy numbers may be related to relatively small genome and is on line with observations in some other small plant genomes [49,50,51].

Satellite DNA sequences are typical components of subtelomeric and centromeric chromosome regions, but may also form clusters in interstitial regions [14,18,52,53,54,55,56]. Our earlier studies [13,31,57] indicated that *Musa* does not have a typical centromeric satellite and that *Musa* centromeres are rather made of various types of retrotransposons, especially Ty3/*Gypsy*-like elements and LINE-like elements. Ty3/*Gypsy*-like elements were found in high copy numbers in centromeric regions in other plant species [58,59,60,61,62]. On the other hand, LINE elements are usually dispersed along chromosomes including their pericentromeric regions [63,64,65]. The present study confirmed preferential localization of the LINE element to centromeric regions in *Musa* and FISH with a probe for the LINE element appears a convenient way to label primary constrictions, which are not always easily visible on small and condensed mitotic metaphase chromosomes of *Musa*.

Our analyses revealed that nuclear genomes of all *Musa* accessions included in the present study contained the major DNA satellites CL33 and CL18. With a few exceptions, the satellites were found highly conserved (Figures 1A, B, 2A). Judging from the cluster analysis based on Neighbor Joining method, there are no A or B genome-specific satellite units in *Musa* (Figure 2B). These findings indicate that both satellites originated before the divergence of the section Eumusa (~28 Mya) [66] and that their sequences remained conserved during speciation. Unfortunately, our observation means that the satellites are not suitable as FISH probes to discriminate A and B genome chromosomes and hence cannot help in characterizing genomic constitution in interspecific *Musa* hybrids.

One should bear in mind that our results on the length, sequence similarity to maTR_CL18 and maTR_CL33 satellites [13] as well as the nucleotide diversity may be affected by the fact that only sequences which were obtained after PCR amplification analyzed. Our results could be supported and refined after whole genome sequencing, e.g., using next generation sequencing, and reconstruction of complete repetitive elements [13].

We were not able to detect CL33 satellite by FISH on mitotic chromosomes of *M. balbisiana*, most probably because the copy number fell below the detection limit of the method. Supported by the results of Southern hybridization, these results indicate a low copy number of this satellite in the *Musa* B genome. This corresponds to lower copy number of other DNA repeats in *M. balbisiana* [31] and may be related to smaller genome size as compared to the A genome [26,67].

Our findings extend significantly previous results of cytogenetic mapping of 45S and 5S rRNA genes and a single copy BAC clone 2G17 in *Musa* [25,28,29,30,31]. The 45S rDNA cluster localized to NORs in all accessions and the number of loci corresponded to their ploidy. The only exception was the additional locus on one satellite chromosome in diploid hybrid clone 'Ato' (AS genome).

5S rRNA genes are more diverse in the number of loci as well as in the genomic location. The largest variation in the number of 5S rDNA sites was observed among the accessions of *M. acuminata* where the number of loci ranged from 4 to 8 per mitotic metaphase plate. It is tempting to suggest different number of 5S loci in different subspecies of *M. acuminata*. Unfortunately, our results on a limited set of accessions do not support this idea. For example, 'Maia Oa' (*M. acumina* ssp. *zebrina*) and 'Long Tavoy' (*M. acuminata* ssp. *burmanica*) have the same number of 5S loci. On the other hand, cytogenetic observations may question the taxonomic classification of some accessions. Further work is needed to verify both options. The numbers of 5S rDNA loci in diploid accessions are on line with previous studies [25,28,29] except of *M. balbisiana* 'Honduras' where Bartoš et al [25] identified only four 5S rDNA loci. As the same genotype was analyzed in both studies (ITC 0247), the difference could be due to higher sensitivity of FISH in the present work. The highest number of 5S rDNA loci was found in *M. schizocarpa*. Twelve signals of 5S rDNA were detected in both representatives of this species, while Bartoš et al. [25] detected only six 5S rDNA sites in *M. schizocarpa* (ITC 0890). The discrepancy may be due to higher sensitivity of FISH. But as a different genotype was analyzed by Bartoš et al. [25] it is also possible that similarly to *M. acuminata*, there is a variation in the number of 5S loci in *M. schizocarpa*.

A BAC clone 2G17, which was originally identified as a single copy clone on mitotic chromosomes of *M. acuminata* 'Calcutta 4' [30] was localized on one chromosome pair in all diploid accessions with three exceptions. Four signals were observed on mitotic chromosomes of *M. balbisiana* 'Honduras'. This could be a consequence of the locus duplication in this genotype. On the other hand, only three signals of CL18 were found on mitotic chromosomes of *M. balbisiana* 'Cameroun' which could be explained by reduction or even loss of one locus in this clone. Only one signal of CL18 was detected in metaphase plate of hybrid clone 'Umbubu' (AT genome) suggesting that the locus diverged between the genomes of *Eumusa* and *Australimusa* species. An alternative explanation is a reduction or loss of this region during the formation and evolution of the hybrid genome.

Due to the lack of probes specific for individual chromosomes, we were not able to identify all chromosomes within the karyotypes of eleven diploid accessions and eight inter-specific hybrids of *Musa*. Nevertheless, we have succeeded in expanding significantly the number of individual chromosomes which can be identified based on specific chromosome FISH patterns (Figure S3, Table 3). All chromosomes which could be identified unambiguously occurred in pairs of homologs in all accessions of *M. acuminata*, *M. balbisiana* and *M. schizocarpa*, indicating structural homozygosity. The only exception was found in *M. balbisiana* 'Cameroun' where only three loci of CL18 were found. However, our observation cannot exclude small chromosome exchanges and concerns only a subset of chromosomes. Structural heterozygosity influences chromosome behavior in meiosis, formation gametes, their genotype, distortions from expected Mendelian segregation and may cause sterility in hybrids [68]. Differences in distribution of cytogenetic landmarks within and between the diploid *Musa* species may indicate differences in chromosome structure poten-

tially leading to aberrant meiosis and sterility in intra- and inter-specific hybrids [69,70,71].

The differences in FISH patterns between some chromosomes of *M. acuminata* and *M. balbisiana* provided information on the presence of some of their chromosomes in inter-specific hybrids and thus may contribute to the reconstruction of their genomic constitution. For example, triploid interspecific hybrid clones 'Obino l'Ewai', 'Maritú', and '3 Hands Planty' had the expected number of CL18 loci, which is on line with the reported AAB genome constitution (Table 3). Triploid ABB clones 'Pelipita' and 'Balonkawe' were characterized by weak signals of CL33 after Southern hybridization. The presence of only one chromosome on which CL33 was detected by FISH seems to support the ABB genomic constitution with only one A genome (or its part) present.

Our results indicate that NOR-bearing satellite chromosomes are maintained in all hybrid clones we have analyzed. On the other hand, variation in the number of 5S rDNA loci was observed among the triploid hybrids (Table 3). Specific numbers of 5S rDNA loci in A and B genotypes could aid in identification of the origin of hybrids. However, this would be only possible if genomic organization of 5S loci is known in a larger set of *M. acuminata* and *M. balbisiana* accessions covering the existing diversity.

FISH analysis on interspecific hybrids revealed deviations in the number of loci of BAC clone 2G17 and satellite DNA sequences from the expected numbers based on reported genomic constitution. Instead of the expected three hybridization sites of BAC 2G17, four loci were identified in three out of four triploid hybrids. Based on our observations in diploids, there should be a maximum of five CL18 loci in a triploid hybrid. However, both ABB hybrids had six loci (Table 3). As we analyzed a limited set of diploid genotypes, we cannot exclude that other diploid genotypes have different number of CL18 loci and that the hybrids originated from them. The discrepancies between the expected and observed number of loci could be also due to backcrossing of primary hybrids to one of the parental species and a loss or gain of some chromosome types. Thus, our results may provide a support to the backcross hypothesis of De Langhe et al. [72]. Clearly, further analysis is needed to clarify the issue as tandem organized sequences can increase or decrease their copy number during the interspecific hybridization and polyploidization or can be removed completely from a hybrid genome [19,73,74]. Consequently they should be used with caution when determining genomic constitution in interspecific *Musa* hybrids.

Supporting Information

Figure S1 (A) Dot-plot comparison of maTR_CL18 which was identified in 454 sequence data of *M. acuminata* 'Calcutta 4' [13] and CL18-like repeat which was identified in whole genome sequence of *M. acuminata* 'DH Pahang' [35]. Both repetitive units are more than 2.2 kb long and are organized in tandem arrays. (B) Dot-plot comparison of maTR_CL33 which was identified in 454 data of *M. acuminata* 'Calcutta 4' [13] and CL33-like repeat which was identified in whole genome sequence of *M. acuminata* 'DH Pahang' [35]. Both repetitive units are 134 bp long and are organized in tandem arrays. Sequence similarities are represented by dots and diagonal lines (A, B).
(TIFF)

Figure S2 Examples of genomic distribution of satellite DNA as determined on mitotic metaphase chromosomes of *Musa* after FISH with labeled probes for CL18 (red, labeled by arrows) and banana LINE element (green). Chromosomes were counterstained with DAPI (blue). (A) 'Maia Oa'. (B) 'Long Tavoy'. (C) 'Tani'. (D) *M. schizocarpa* ITC 0560. (E) 'Pelipita'. (F) *M. schizocarpa* ITC 1002. (G) 'Obino l'Ewai'. (H) 'Maritú'. (I) 'Ato'. (J) 'Tonton Kepa'. (K) 'Umbubu'. Bar = 5 μm.
(TIFF)

Figure S3 Idiograms of diploid (A – C) and hybrid (D – G) *Musa* accessions. (A) *M. acuminata*, (B) *M. balbisiana*, (C) *M. schizocarpa*, (D) Hybrids with AAB genomics constitution, (E) Hybrids with ABB genomic constitution, (F) Hybrids with AS genomic constitution, (G) Hybrids with AT genomic constitution. No attempt was made to identify homologs in the hybrids (D – G) and all chromosomes are shown. The chromosome sizes are only indicative.
(PDF)

Figure S4 Examples of genomic distribution of satellite DNA as determined on mitotic metaphase chromosomes of interspecific hybrids after FISH. Chromosomes were counterstained with DAPI (blue). Sites of CL18 and CL33 probe hybridization are marked by long and thick arrows, respectively. (A) CL18 (red) and CL33 (green) on chromosomes of 'Maritú'. (B) CL18 (red) and 5S rDNA (green) on chromosomes of 'Maritú'. (C) 5S rDNA (red) and CL33 (green) on chromosomes of 'Maritú'. (D) CL18 (red) and BAC clone 2G17 (green) on chromosomes of 'Pelipita'. (E) CL18 (red) and 5S rDNA on chromosomes of 'Pelipita'. (F) CL18 (red) and CL33 (green) on chromosomes of 'Ato'. (G) CL18 (red) and 5S rDNA (green) on chromosomes of 'Tonton Kepa'. (H) CL18 (red) and CL33 (green) on chromosomes of 'Umbubu'. (I) 5S rDNA (red) and CL33 (green) on chromosomes of 'Ato'. (J) CL18 (red) and CL33 (green) on chromosomes of 'Tonton Kepa'. (K) 5S rDNA (red) and CL18 (green) on chromosomes of 'Umbubu'. Bar = 5 μm.
(TIFF)

Table S1 Basic characteristics and nucleotide diversity of part1 of CL18-like repeats.
(DOC)

Table S2 Basic characteristics and nucleotide diversity of part2 of CL18-like repeats.
(DOC)

Table S3 Basic characteristics and nucleotide diversity of CL33-like repeats.
(DOC)

Table S4 Copy numbers of maTR_CL18 and maTR_CL33 satellites.
(DOC)

Acknowledgments

We thank Ir. Ines van den Houwe for providing the plant material and Ms. Radka Tušková and Ms. Eva Jahnová for excellent technical assistance.

Author Contributions

Conceived and designed the experiments: JC EH JD. Performed the experiments: JC EH LH PC PS. Analyzed the data: JC EH JD. Contributed reagents/materials/analysis tools: JD. Wrote the paper: JC EH JD.

References

1. Bennetzen JL, Kellog EA (1997) Do plants have a one-way ticket to genomic obesity? Plant Cell 9: 1509–1514.
2. Ingham LD, Hanna WW, Baier JW, Hannah LC (1993) Origin of the main class of repetitive DNA within selected *Pennisetum* species. Mol Gen Genet 238: 350–356.
3. Shapiro JA, von Sternberg R (2005) Why repetitive DNA is essential to genome function. Biol Rev 80: 227–250.
4. Kubis S, Schmidt T, Heslop-Harrison JS (1998) Repetitive DNA elements as a major component of plant genomes. Ann Bot 82: 45–55.
5. Charlesworth B, Sniegowski PE, Stephan W (1994) The evolutionary dynamics of repetitive DNA in eukaryotes. Nature 37: 215–220.
6. Elder JFJr and Turner BJ (1995) Concerted evolution of repetitive DNA sequences in eukaryotes. Q Rev Biol 70: 297–320.
7. Cheng Z, Dong F, Langdon T, Ouyang S, Buell CR, et al. (2002) Functional rice centromeres are marked by a satellite repeat and a centromere-specific retrotransposon. Plant Cell 14: 1691–1704.
8. Ellegren H (2004) Microsatellites: simple sequences with complex evolution. Nat Rev Genet 5: 435–445.
9. Tek AL, Jiang J (2004) The centromere regions of potato chromosomes contain megabase-sized tandem arrays of telomere-similar sequence. Chromosoma 113: 77–83.
10. Macas J, Požárková D, Navrátilová A, Nouzová M, Neumann P (2000) Two new families of tandem repeats isolated from genus *Vicia* using genomic self-priming PCR. Mol Gen Genet 263: 741–751.
11. Willard HF (1991) Evolution of alpha satellite. Curr Opin Genet Dev 1: 509–514.
12. Navrátilová A, Koblížková A, Macas J (2008) Survey of extrachromosomal circular DNA derived from plant satellite repeats. BMC Plant Biol 8: 90.
13. Hřibová E, Neumann P, Matsumoto T, Roux N, Macas J, et al. (2010) Repetitive part of the banana (*Musa acuminata*) genome investigated by low-depth 454 sequencing. BMC Plant Biol 10: 204.
14. Macas J, Neumann P, Navrátilová A (2007) Repetitive DNA in the pea (*Pisum sativum* L.) genome: comprehensive characterization using 454 sequencing and comparison to soybean and *Medicago truncatula*. BMC Genomics 8: 427.
15. Macas J, Kejnovský E, Neumann P, Novák P, Koblížková A, et al. (2011) Next generation sequencing-based analysis of repetitive DNA in the model dioecious plant *Silene latifolia*. PLoS One 6: e27335.
16. Schmidt T, Heslop-Harrison JS (1998) Genomes, genes and junk: the large-scale organization of plant chromosomes. Trends Plant Sci 3: 195–199.
17. Zatloukalová P, Hřibová E, Kubaláková M, Suchánková P, Šimková H, et al. (2011) Integration of genetic and physical maps of chickpea (*Cicer arietinum* L.) genome using flow-sorted chromosomes. Chromosome Res 19: 729–739.
18. Han YH, Zhang ZH, Liu JH, Lu JY, Huang SW, et al. (2008) Distribution of the tandem repeat sequences and karyotyping in cucumber (*Cucumis sativus* L.) by fluorescence in situ hybridization. Cytogenet Genome Res 122: 80–88.
19. Macas J, Navrátilová A, Koblížková A (2006) Sequence homogenization and chromosomal localization of VicTR-B satellites differ between closely related *Vicia* species. Chromosoma 115: 437–447.
20. Navrátilová A, Neumann P, Macas J (2003) Karyotype analysis of four *Vicia* species using *in situ* hybridization with repetitive sequences. Ann Bot 91: 921–926.
21. Sharma S, Raina SN (2005) Organization and evolution of highly repeated satellite DNA in plant chromosomes. Cytogenet Genome Res. 109: 15–26.
22. Kopecký D, Havránková M, Loureiro J, Castro S, Lukaszewski AJ, et al. (2010) Physical distribution of homoeologous recombination in individual chromosomes of *Festuca pratensis* in *Lolium multiflorum*. Cytogenet Genome Res. 120: 370–383.
23. Kopecký D, Lukaszewski AJ, Doležel J (2008) Cytogenetics of Festulolium (*Festuca*×*Lolium* hybrids). Cytogenet Genome Res. 120: 370–383.
24. Simmonds NW, Shepherd K (1955) The taxonomy and origins of the cultivated bananas. J Linn Soc Bot 55: 302–312.
25. Bartoš J, Alkhimova O, Doleželová M, De Langhe E, Doležel J (2005) Nuclear genome size and genomic distribution of ribosomal DNA in *Musa* and *Ensete* (Musaceae): taxonomic implications. Cytogenet Genome Res 109: 50–57.
26. Doležel J, Doleželová M, Novák FJ (1994) Flow cytometric estimation of nuclear DNA amount in diploid bananas (*Musa acuminata* and *Musa balbisiana*). Biol Plant 36: 351–357.
27. D'Hont A, Paget-Goy A, Escoute J, Carreel F (2000) The interspecific genome structure of cultivated banana, *Musa* spp. revealed by genomic DNA *in situ* hybridization. Theor Appl Genet 100: 177–183.
28. Osuji JO, Harrison G, Crouch J, Heslop-Harrison JS (1997) Identification of the genomic constitution of *Musa* L. lines (bananas, plantains and hybrids) using molecular cytogenetics. Ann Bot 80: 787–793.
29. Doleželová M, Valárik M, Swennen R, Horry JP, Doležel J (1998) Physical mapping of the 18S–25S and 5S ribosomal RNA genes in diploid bananas. Biol Plant 41: 497–505.
30. Hřibová E, Doleželová M, Doležel J (2008) Localization of BAC clones on mitotic chromosomes of *Musa acuminata* using fluorescence *in situ* hybridization. Biol Plant 52: 445–452.
31. Valárik M, Šimková H, Hřibová E, Šafář J, Doleželová M, et al. (2002) Isolation, characterization and chromosome localization of repetitive DNA sequences in bananas (*Musa* spp.). Chromosome Res 10: 89–100.
32. Zhang HB, Zhao X, Ding X, Paterson AH, Wing RA (1995) Preparation of megabase-size DNA from plant nuclei. Plant J 7: 175–184.
33. Otto F (1990) DAPI staining of fixed cells for high-resolution flow cytometry of nuclear DNA. In: Crissman HA, Darzynkiewicz Z, editors. Methods in Cell Biology. Academic Press, New York, Vol. 33. Pp. 105–110.
34. Doležel J, Bartoš J, Voglmayr H, Greilhuber J (2003) Nuclear DNA content and genome size of trout and human. Cytometry A 51: 127–128.
35. D'Hont A, Denoeud F, Aury JM, Baurens FC, Carreel F, et al. (2012) The banana (*Musa acuminata*) genome and the evolution of monocotyledonous plants. Nature 488: 213–217.
36. Sonnhammer ELL, Durbin R (1995) A dot-matrix program with dynamic threshold control suited for genomic DNA and protein sequence analysis. Gene 167:GC1-GC10.
37. Staden R (1996) The Staden sequence analysis package. Mol Biotechnol 5: 233–241.
38. Huan X, Madan A (1999) CAP3: A DNA sequence assembly program. Genome Res 9: 868–877.
39. Katoh K, Kuma K, Toh H, Miyata T (2005) MAFFT version 5: improvement in accuracy of multiple sequence alignment. Nucleic Acids Res 33: 511–518.
40. Galtier N, Gouy M, Gautier C (1996) SeaView and Phylo_win, two graphic tools for sequence alignment and molecular phylogeny. Comput Appl Biosci 12: 543–548.
41. Eckert AJ, Liechty JD, Tearse BR, Pande B, Neale DB (2010) DnaSAM: Software to perform neutrality testing for large datasets with complex null models. Mol Ecol Resour 10: 542–545.
42. Crooks GE, Hon G, Chandonia JM, Brenner SE (2004) WebLogo: a sequence logo generator. Genome Res 14: 1188–1190.
43. Huson DH, Bryant D (2006) Application of phylogenetic networks in evolutionary studies. Mol Biol Evol 23: 254–267.
44. Dechyeva D, Gindullis F, Schmidt T (2003) Divergence of satellite DNA and interspersion of dispersed repeats in the genome of the wild beet *Beta procumbens*. Chromosome Res 11: 3–21.
45. King K, Jobst J, Hemleben V (1995) Differential homogenization and amplification of two satellite DNAs in the genus *Cucurbita* (Cucurbitaceae). J Mol Evol 41: 996–1005.
46. Dover GA, Strachan T and Coen ES (1982) Molecular drive. Science 218: 1069.
47. Baldwin BG, Sanderson MJ, Porter MJ, Wojciechowski MF, Campbell CS, et al. (1995) The ITS region of nuclear ribosomal DNA: a valuable source of evidence on angiosperm phylogeny. Ann Mo Bot Gard 82: 247–277.
48. Dover G (2002) Molecular drive. Trends Genet 18: 587–589.
49. Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 408: 796–815.
50. International Brachypodium Initiative (2010) Genome sequencing and analysis of the model grass *Brachypodium distachyon*. Nature 463: 763–768.
51. International Rice Genome Sequencing Project (2005) The map-based sequence of the rice genome. Nature 436: 793–800.
52. Galasso I, Schmidt T, Pignone D, Heslop-Harrison JS (1995) The molecular cytogenetics of *Vigna unguiculata* (L.) Walp: the physical organization and characterization of 18S-5.8S-25S rRNA genes, 5S rRNA genes, telomere-like sequences, and a family of centromeric repetitive DNA sequences. Theor Appl Genet 91: 928–935.
53. Houben A, Schubert I (2003) DNA and proteins of plant centromeres. Curr Opin Plant Biol 6: 554–560.
54. Jiang J, Birchler JA, Parrott WA, Dawe RK (2003) A molecular view of plant centromeres. Trends Plant Sci 8: 570–574.
55. Macas J, Neumann P, Novák P, Jiang J (2010) Global sequence characterization of rice centromeric satellite based on oligomer frequency analysis in large-scale sequencing data. Bioinformatics 26: 2101–2108.
56. Torres GA, Gong Z, Iovene M, Hirsch CD, Buell CR, et al. (2011) Organization and evolution of subtelomeric satellite repeats in the potato genome. G3: Genes, Genomes, Genetics 1: 85–92.
57. Hřibová E, Doleželová M, Town CD, Macas J, Doležel J (2007) Isolation and characterization of the highly repeated fraction of the banana genome. Cytogenet Genome Res 119: 268–274.
58. Bao W, Zhang W, Yang Q, Zhang Y, Han B, et al. (2006) Diversity of centromeric repeats in two closely related wild rice species, *Oryza officinalis* and *Oryza rhizomatis*. Mol Genet Genomics 275: 421–430.
59. Gindullis F, Desel C, Galasso I, Schmidt T (2001) The large-scale organization of the centromeric region in *Beta* species. Genome Res 11: 253–265.
60. Kumekawa N, Ohmido N, Fukui K, Ohtsubo E, Ohtsubo H (2001) A new gypsy-type retrotransposon, RIRE7: preferential insertion into the tandem repeat sequence TrsD in pericentromeric heterochromatin regions of rice chromosomes. Mol Genet Genomics 265: 480–488.
61. Neumann P, Navrátilová A, Koblížková A, Kejnovský E, Hřibová E, et al. (2011) Plant centromeric retrotransposons: a structural and cytogenetic perspective. Mob DNA 2: 4.

62. Presting GG, Malysheva L, Fuchs J, Schubert I (1998) A TY3/GYPSY retrotransposon-like sequence localizes to the centromeric regions of cereal chromosomes. Plant J 16: 721–728.
63. Kubis SE, Heslop-Harrison JS, Desel C, Schmidt T (1998) The genomic organization of non-LTR retrotransposons (LINEs) from three *Beta* species and five other angiosperms. Plant Mol Biol 36: 821–31.
64. Kumar A and Bennetzen JL (1999) Plant retrotransposons. Annu Rev Genet 33: 479–532.
65. Schmidt T, Kubis S, Heslop-Harrison JS (1995) Analysis and chromosomal localisation of retrotransposons in sugar beet (*Beta vulgaris*) LINES and Ty1-copia- like elements as major components of the genome. Chromosome Res 3: 335–45.
66. Christelová P, Valárik M, Hřibová E, De Langhe E, Doležel J (2011) A multi gene sequence- based phylogeny of the Musaceae (banana) family. BMC Evol Biol 11: 103.
67. Lysák MA, Doleželová M, Horry JP, Swennen R, Doležel J (1999) Flow cytometric analysis of nuclear DNA content in *Musa*. Theor Appl Genet 98: 1344–1350.
68. Cai X, Xu SS (2007) Meiosis-driven genome variation in plants. Curr Genomics 8: 151–161.
69. Dodds KS, Simmonds NW (1948) Sterility and parthenocarpy in diploid hybrids of *Musa*. Heredity 2: 101–117.
70. Shepherd K (1999) Cytogenetics of the genus *Musa*. IPGRI-INIBAP. Pp.160.
71. Wilson GB (1946) Cytological studies in the Musae. II. Meiosis in some diploid clones. Genetics 31: 475–482.
72. De Langhe E, Hřibová E, Carpentier S, Doležel J, Swennen R (2010) Did backcrossing contribute to the origin of hybrid edible bananas? Ann Bot 106: 849–857.
73. Tek AL, Song JQ, Macas J, Jiang J (2005) Sobo, a recently amplified satellite repeat of potato, and its implications for the origin of tandemly repeated sequences. Genetics 170: 1231–1238.
74. Urgaković D, Plohl M (2002) Variation in satellite DNA profiles-causes and effects. EMBO J 21: 5955–5959.

4.1.2.1 SUPPORTING INFORMATION

Figure S1

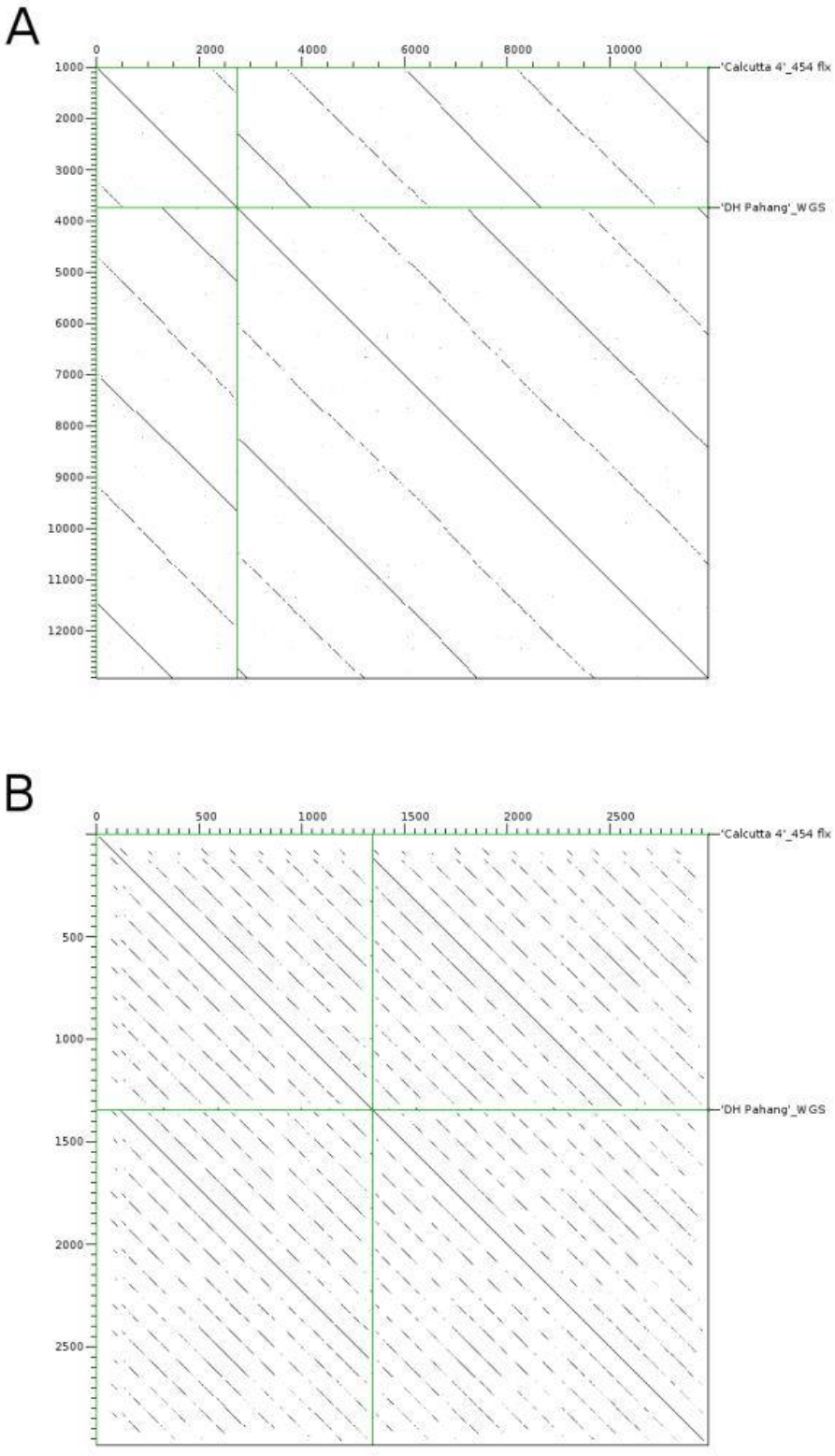


Figure S2

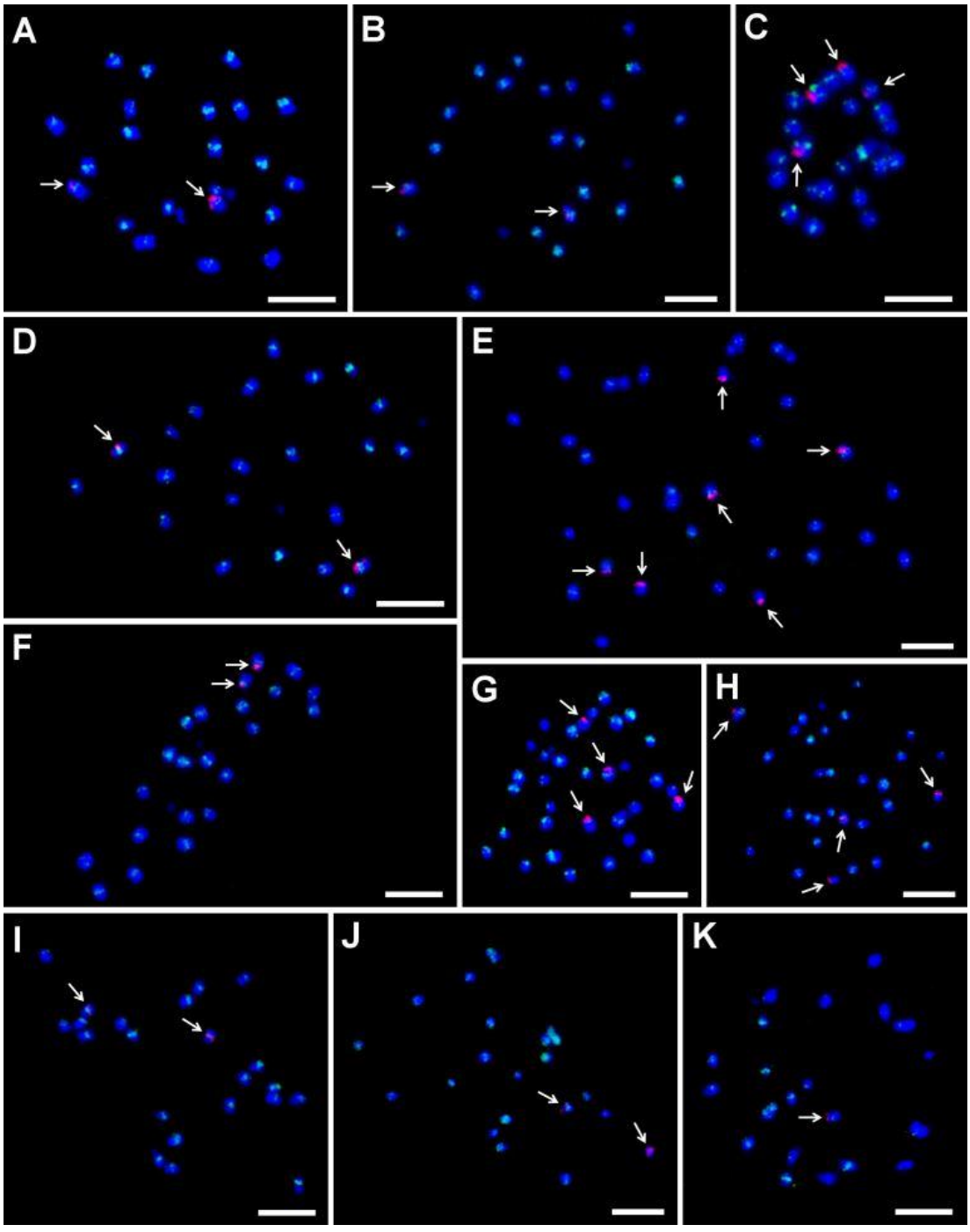
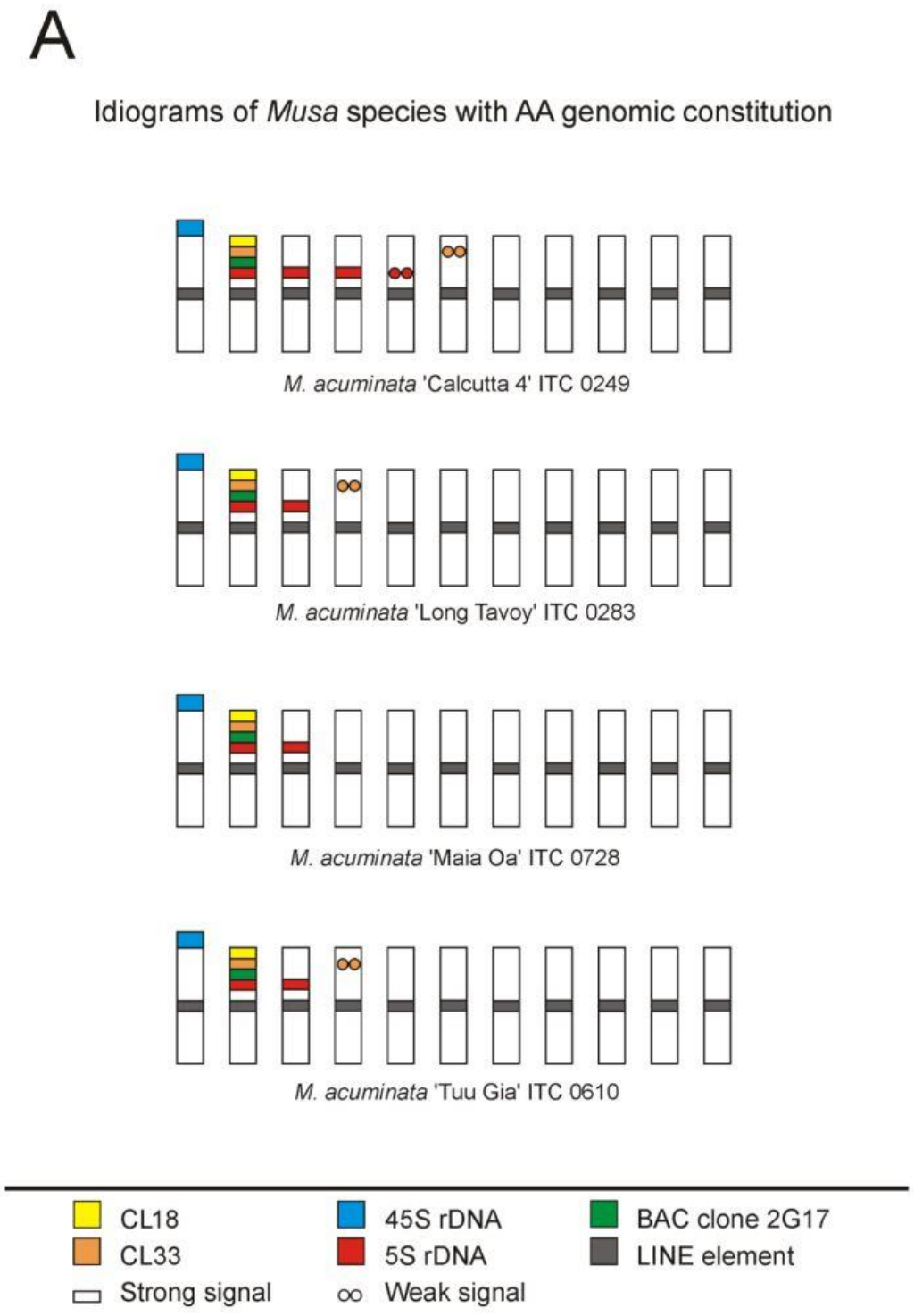
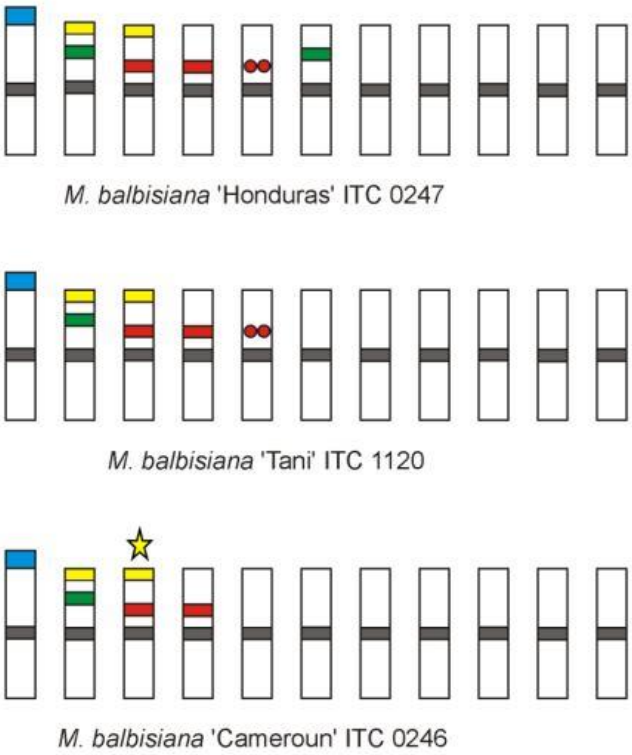


Figure S3


A
Idiograms of *Musa* species with AA genomic constitution
M. acuminata 'Calcutta 4' ITC 0249
M. acuminata 'Long Tavoy' ITC 0283
M. acuminata 'Maia Oa' ITC 0728
M. acuminata 'Tuu Gia' ITC 0610
CL18
CL33
Strong signal
45S rDNA
5S rDNA
Weak signal
BAC clone 2G17
LINE element

B

Idiograms of *Musa* species with BB genomic constitution

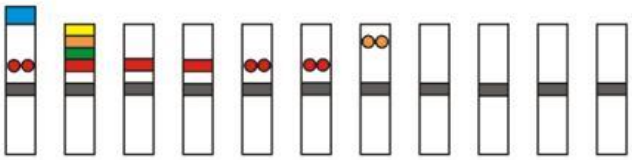


☆ Nuclear genome of *M. balbisiana* 'Cameroun' ITC 0246 contains three signals of CL18 satellite (see Results)

CL18
CL33
Strong signal
45S rDNA
5S rDNA
∞ Weak signal
BAC clone 2G17
LINE element

C

Idiogram of *Musa* species with SS genomic constitution

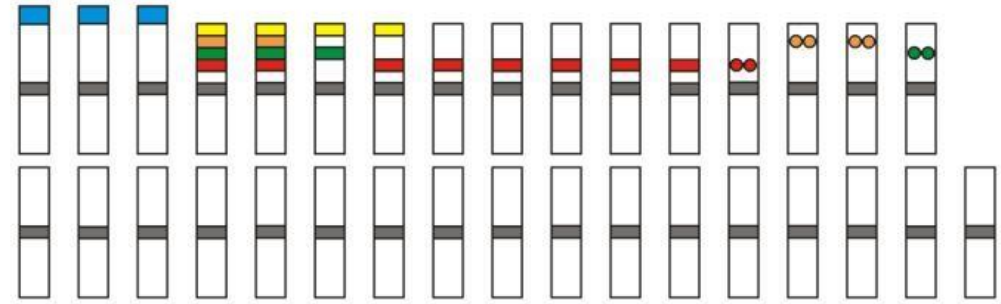


M. schizocarpa ITC 1002

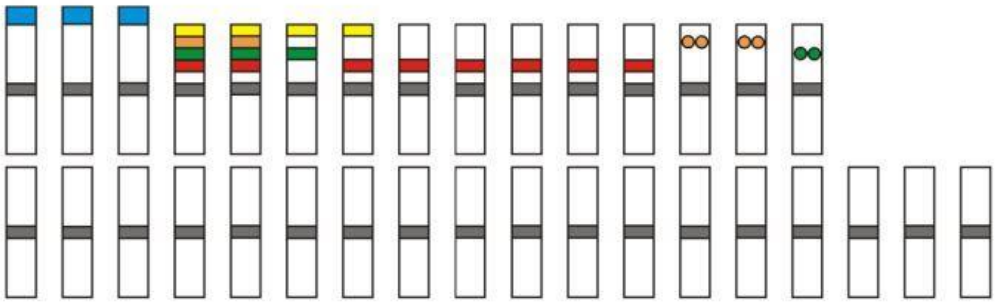
CL18
CL33
Strong signal
45S rDNA
5S rDNA
∞ Weak signal
BAC clone 2G17
LINE element

D

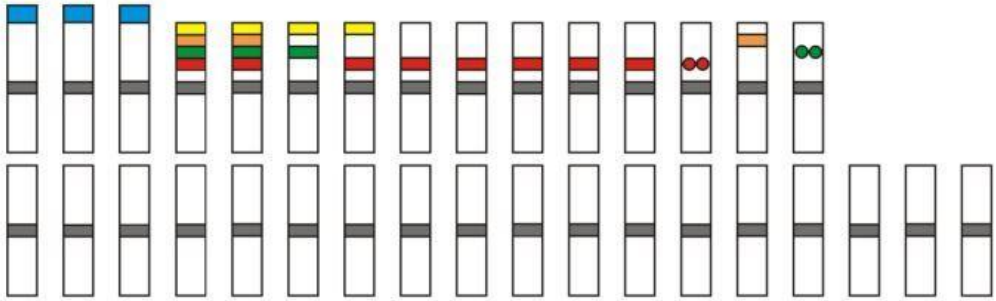
Idiograms of *Musa* hybrid clones with AAB genomic constitution



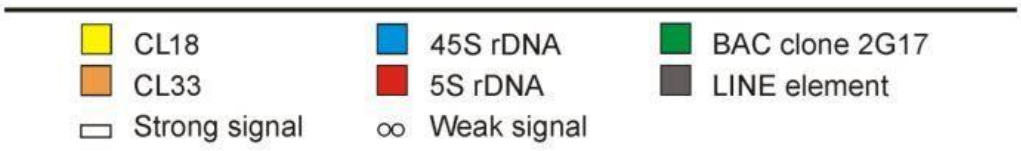
Musa hybrid clone 'Obino l'Ewai' ITC 0109



Musa hybrid clone 'Maritú' ITC 0639

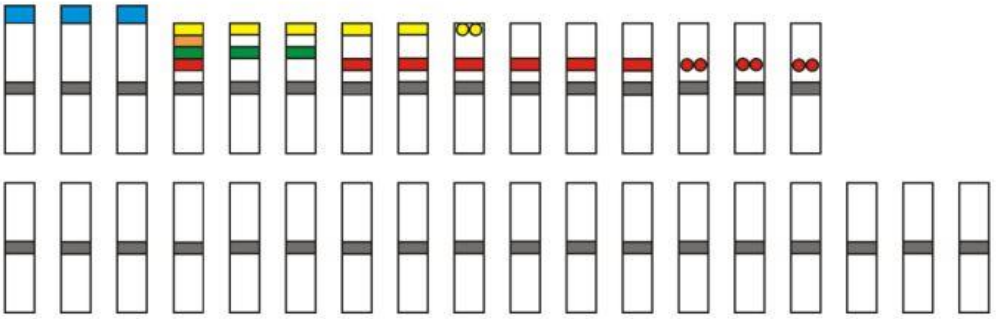


Musa hybrid clone '3 Hands Planty' ITC 1132

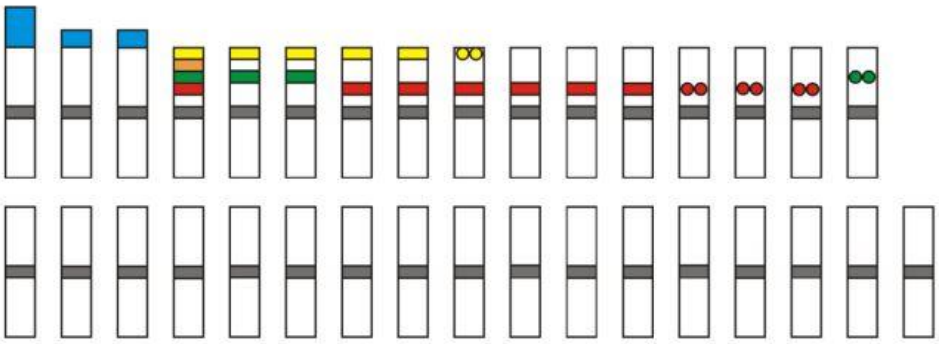


E

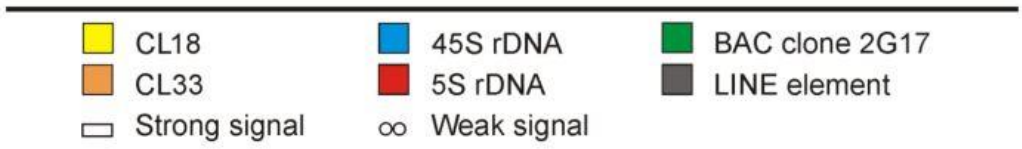
Idiograms of *Musa* hybrid clones with ABB genomic constitution



Musa hybrid clone 'Pelipita' ITC 0472

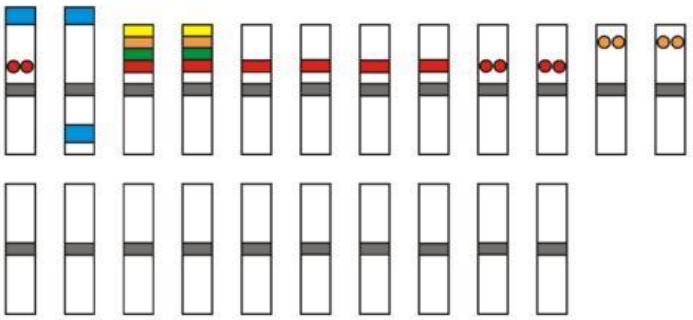


Musa hybrid clone 'Balonkawe' ITC 0473

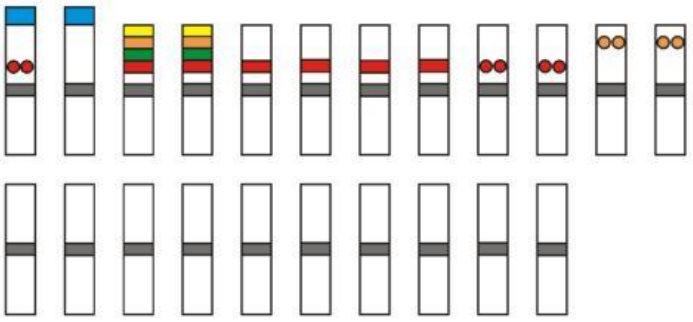


F

Idiograms of *Musa* hybrid clones with AS genomic constitution



Musa hybrid clone 'Ato' ITC 0820

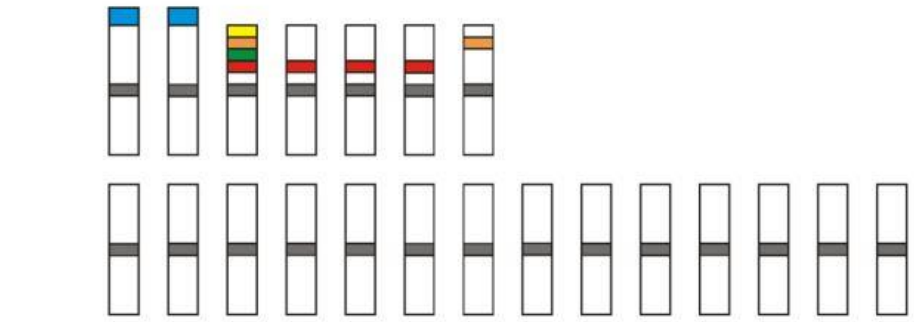


Musa hybrid clone 'Tonton Kepa' ITC 0822

CL18
CL33
Strong signal
45S rDNA
5S rDNA
∞ Weak signal
BAC clone 2G17
LINE element

G

Idiogram of *Musa* hybrid clone with AT genomic constitution



Musa hybrid clone 'Umbubu' ITC 0854

CL18
CL33
Strong signal
45S rDNA
5S rDNA
Weak signal
BAC clone 2G17
LINE element

Figure S4

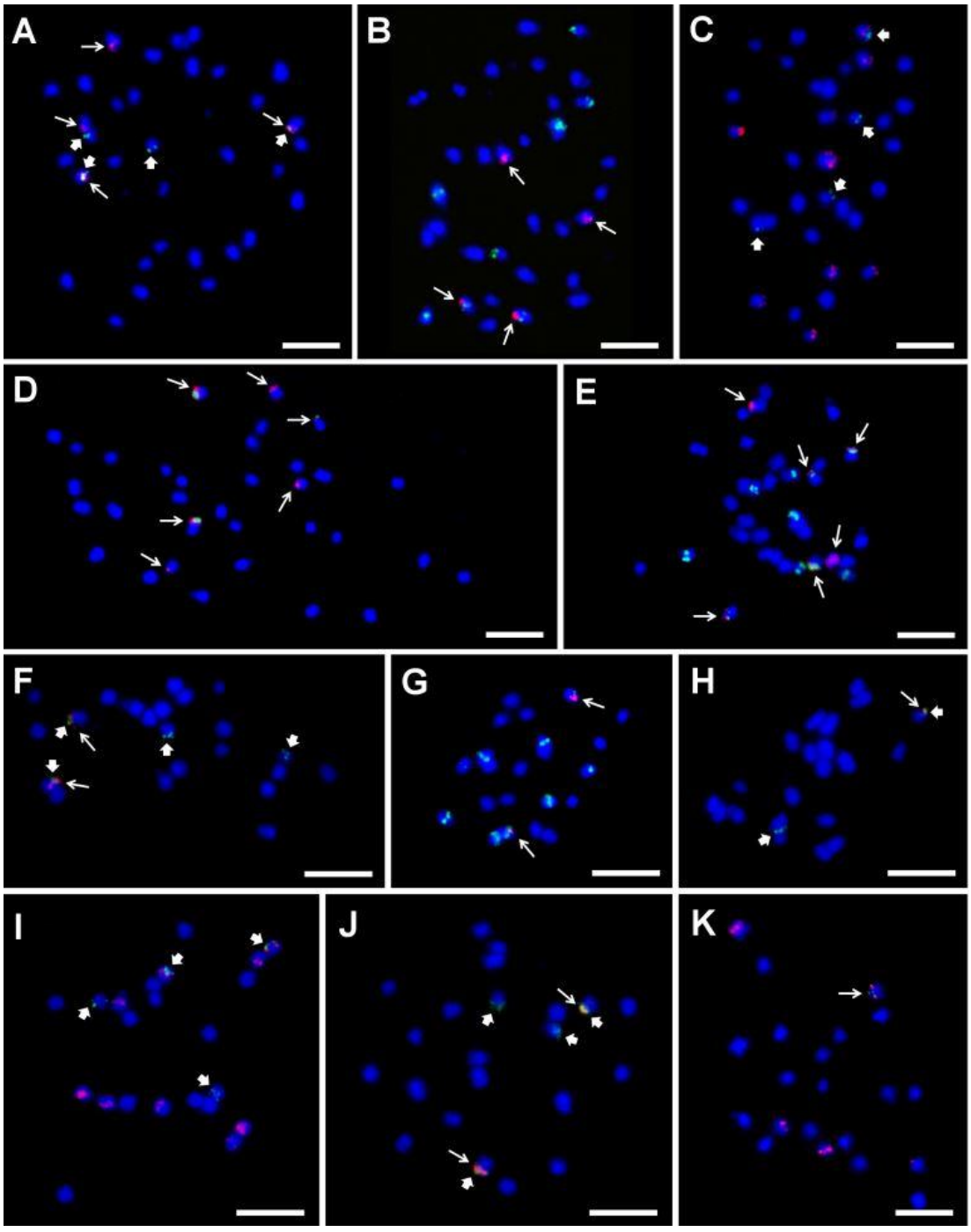


Table S1

Table S1: Basic characteristics and nucleotide diversity of part1 of CL18-like repeats

Accession code (ITC code)	Genomic constitution	Number of sequenced clones	Note		Length of sequenced region	Similarity to maTR_CL18 [%]	Nucleotide diversity 0π
0249	AA	32			921 bp	93 – 99	5.459
0283	AA	24			341 bp	95	2.353
0728	AA	24			342 bp	94	5.987
1511	AA	24			342 bp	94	6.647
0610	AA		sequences corresponding to part1 were not obtained				
0246	BB	26			934 bp	94 – 96	6.171
0247	BB	24			344 bp	94	0.000
1120	BB	30			927 bp	94 – 96	6.451
PKW	BB	26			921 bp	94 – 96	5.382
0560	SS		sequences corresponding to part1 were not obtained				
1002	SS		sequences corresponding to part1 were not obtained				
0109	AAB	24			926 bp	94 – 95	6.403
0639	AAB	24			931 bp	72 – 95	12.502
1132	AAB	24			965 bp	86 – 97	6.222
0472	ABB	28			927 bp	74 – 97	8.970
0473	ABB	24			925 bp	93 – 97	5.984
0820	AS	37	Two DNA sequences were obtained	Type 1	369 bp	59 – 74	7.256
				Type 2	367 bp	50	5.106
0822	AS	42			341 bp	96	0.185
0854	AT	24			924 bp	88 – 96	4.618

Table S2

Table S2: Basic characteristics and nucleotide diversity of part2 of CL18-like repeats

Accession code (ITC code)	Genomic constitution	Number of sequenced clones	Note		Length of sequenced region	Similarity to maTR_CL18 [%]	Nucleotide diversity 0π
0249	AA	24			738 bp	96 -99	0.667
0283	AA	27	Two DNA sequences were obtained	Type 1	728 bp	96	1.818
				Type 2	432 bp	74	0.233
0728	AA	27			727 bp	95	0.484
1511	AA	24			729 bp	95	0.094
0610	AA	58	Three DNA sequences were obtained	Type 1	259 bp	61	0.667
				Type 2	361 bp	83	1.672
				Type 3	402 bp	83	1.181
0246	BB	42			830 bp	88 – 91	9.086
0247	BB	24			823 bp	90	0.000
1120	BB	33			746 bp	88 – 94	8.698
PKW	BB	24			693 bp	90 – 95	0.122
0560	SS	24			832 bp	66 – 69	9.232
1002	SS	30			694 bp	67 – 76	8.537
0109	AAB	24			925 bp	70 – 90	6.403
0639	AAB	41	Two DNA sequences were obtained	Type 1	670 bp	72	0.000
				Type 2	742 bp	90	0.210
1132	AAB		Sequences corresponding to part2 were not obtained				
0472	ABB	42			828 bp	88 – 95	8.555
0473	ABB	29			767 bp	91 – 95	1.236
0820	AS	27			578 bp	83 – 92	8.075
0822	AS	35			401 bp	84	0. 457
0854	AT	24			826 bp	91 – 95	8.525

Table S3

Table S3: Basic characteristics and nucleotide diversity of CL33-like repeats

Accession code (ITC code)	Genomic constitution	Number of sequenced clones	Note		Length of repetitive unit	Similarity to maTR_CL33 [%]	Nucleotide diversity $\theta\pi$
0249	AA	61			134 bp	99	3.500
0283	AA	26	Two DNA sequences were obtained	Type 1	134 bp	93 – 98	3.194
				Type 2	110 bp		0.000
0728	AA	36	Two DNA sequences were obtained	Type 1	134 bp	97	2.667
				Type 2	188 bp		0.692
1511	AA	24			134 bp	98 – 99	4.248
0610	AA	24			188 bp		0.500
0246	BB	24			134 bp	96 – 98	3.225
0247	BB	24			134 bp	87 – 98	13.285
1120	BB	24			134 bp	86 – 97	4.945
PKW	BB	27			134 bp	93 – 98	5.703
0560	SS	31			134 bp	84	1.810
1002	SS	24			134 bp	83 – 86	4.862
0109	AAB	39			134 bp	95 – 98	4.825
0639	AAB	38			134 bp	85 – 98	9.387
1132	AAB	30	Two DNA sequences were obtained	Type 1	134 bp	87 – 96	4.713
				Type 2	188 bp		0.667
0472	ABB	33	Two DNA sequences were obtained	Type 1	134 bp	86 – 97	5.464
				Type 2	188 bp		0.250
0473	ABB	33			134 bp	84 – 100	15.696
0820	AS	24	Two DNA sequences were obtained	Type 1	134 bp	84 – 98	6.765
				Type 2	188 bp		0.400
0822	AS	28	Two DNA sequences were obtained	Type 1	134 bp	83 – 97	6.633
				Type 2	188 bp		0.750
0854	AT	41			134 bp	83 – 95	4.265

Table S4

Table S4: Copy numbers of maTR_CL18 and maTR_CL33 satellites

Species	Subspecies	Accession name	ITC code	maTR_CL18		maTR_CL33	
				Copy number/$1C_x$	Genome proportion [%]	Copy number/$1Cx_x$	Genome proportion [%]
M. acuminata							
	burmannicoides	Calcutta4	0249	~ $1x10^3$ - $2x10^3$	0.39 – 0.73	~ $5.5x10^3$ - $7x10^3$	0.12 – 0.16
	burmannica	Long Tavoy	0283	~ $1x10^3$ - $2x10^3$	0.39 – 0.73	~ $2.5x10^3$ - $4x10^3$	0.06 – 0.09
	zebrina	Maia Oa	0728	~ $1x10^3$ - $2x10^3$	0.34 – 0.68	~ $2.5x10^3$ - $4x10^3$	0.05 – 0.08
	malaccensis	DH Pahang	1511	~ $3.5x10^3$ - $5x10^3$	1.29 – 1.85	~ $5.5x10^3$ - $7x10^3$	0.12 – 0.16
		Tuu Gia	0610	~ $1x10^3$ - $2x10^3$	0.36 – 0.71	~ $2.5x10^3$ - $4x10^3$	0.05 – 0.09
M. balbisiana							
		Cameroun	0246	~ $3.5x10^3$ - $5x10^3$	1.41 – 2.01	below detection limit	---
		Honduras	0247	~ $1x10^3$ - $2x10^3$	0.40 – 0.79	below detection limit	---
		Tani	1120	~ $3.5x10^3$ - $5x10^3$	1.40 – 2.00	below detection limit	---
		Pisang Klutuk Wulung	---	~ $3.5x10^3$ - $5x10^3$	1.39 – 1.99	below detection limit	---
M. schizocarpa							
		Musa schizocarpa	0560	~ $1x10^3$ - $2x10^3$	0.33 – 0.66	~ $5.5x10^3$ - $7x10^3$	0.11 – 0.14
		Musa schizocarpa	1002	~ $1x10^3$ - $2x10^3$	0.33 – 0.66	~ $5.5x10^3$ - $7x10^3$	0.11 – 0.14
Hybrids							
		Obino l'Ewai	0109	~ $5x10^3$ - $6.5x10^3$	1.90 – 2.47	~ $5.5x10^3$ - $7x10^3$	0.13 – 0.16
		Maritú	0639	~ $3x10^3$ - $4x10^3$	1.10 – 1.46	~ $5.5x10^3$ - $7x10^3$	0.12 – 0.16
		3 Hands Planty	1132	~ $3x10^3$ - $4x10^3$	1.13 – 1.51	~ $5.5x10^3$ - $7x10^3$	0.13 – 0.16
		Pelipita	0472	~ $5x10^3$ - $6.5x10^3$	1.93 – 2.50	below detection limit	---
		Balonkawe	0473	~ $5x10^3$ - $6.5x10^3$	1.86 – 2.42	below detection limit	---
		Ato	0820	~ $1x10^3$ - $2x10^3$	0.35 – 0.70	~ $2.5x10^3$ - $4x10^3$	0.05 – 0.08
		Tonton Kepa	0822	~ $1x10^3$ - $2x10^3$	0.34 – 0.68	~ $2.5x10^3$ - $4x10^3$	0.05 – 0.08
		Umbubu	0854	~ $3x10^3$ - $4x10^3$	0.97 – 1.36	~ $5.5x10^3$ - $7x10^3$	0.11 – 0.14

4.1.3 Molecular and Cytogenetic Characterization of Wild *Musa* Species Newly Introduced to ITC Collection

Čížková J, Hřibová E, Christelová P, Van den Houwe I, Roux N, Doležel J

Prepared for submission.

[JČ, EH, JD concieved and designed the experiments. JČ, EH, PC performed the experiments. JČ, EH, PC analyzed the data. JD, VdH, NR contributed reagents/materials/analysis tools. JČ, EH, PN, JD wrote the paper.]

Molecular and cytogenetic characterization of wild *Musa* species newly introduced to ITC collection

Čížková J[1], Hřibová E[1], Christelová P[1], Van den Houwe I[2], Roux N[3] and Doležel J[1*]

[1] Centre of the Region Haná for Biotechnological and Agricultural Research, Institute of Experimental Botany, 78371 Olomouc, Czech Republic

[2] Bioversity International, Katholieke Universiteit Leuven, Laboratory of Tropical Crop Improvement, B-3001 Leuven, Belgium

[3] Bioversity International, 34397 Montpellier, France

[*]Corresponding author: dolezel@ueb.cas.cz

Jaroslav Doležel

Institute of Experimental Botany

Šlechtitelů 31

CZ-78371 Olomouc

Czech Republic

Phone: (+420) 585 238 703

Fax: (+420) 585 238 704

E-mail: dolezel@ueb.cas.cz

Abstract

Bananas (*Musa* spp.) are one of the major world trade commodities with unquestionable role as a staple food for millions of people in humid tropics. The production of bananas is, however, threatened by rapid spreading of various diseases and adverse environmental conditions. The preservation and characterization of banana diversity is essential for the purposes of crop improvement. The world's largest banana collection maintained at the Bioversity International Transit Centre (ITC) in Belgium is continuously extended by various edible cultivars and wild species. Detailed morphological and molecular characterization of newly introduced accessions is necessary for efficient collection and effective use of banana diversity. In this work, nuclear DNA content and genomic distribution of 45S and 5S rDNA were examined in 21 diploid accessions recently added to ITC collection, representing all four sections of the genus *Musa*. 2C DNA content is sections Eumusa and Rhodochlamys ranged from 1.217 to 1.315 pg. Species belonging to sections Australimusa and Callimusa had 2C DNA contents ranging from 1.390 to 1.772 pg. While the number of 45S rDNA loci was conserved in Eumusa and Rhodochlamys, it was highly variable in Callimusa species. 5S rRNA gene clusters were found on two to eight chromosomes per diploid cell. A set of 19 microsatellite markers was used to assess the genetic diversity and relationships of studied ITC accessions. Analysis of the ITS region suggested that some of analyzed accessions might be interspecific hybrids and/or their backcross progeny.

Key words: *Musa*; DNA content; rDNA; FISH; SSR genotyping; ITS region

Introduction

Bananas and plantains (*Musa* spp.) are one of the most important food crops with the global annual production exceeding 130 Mt (faostat.fao.org). They are grown mainly by smallholder farmers for local consumption and only about 10% of the world's production is determined for the export trade. Most of the cultivated bananas are parthenocarpic triploid (but also diploid or tetraploid) clones, which originated from natural intra- and interspecific hybridizations between various subspecies of *M. acuminata* (A genome donor) and *M. balbisiana* (B genome donor) (Simmonds and Shepherd 1955). Hybrid triploids are then classified as AAA, AAB and ABB and include a wide range of varieties of dessert bananas, cooking bananas and plantains representing an essential nutrition source for millions of people. The production of bananas is however threatened by rapid spreading of various diseases, presence of pests and adverse environmental conditions. This imposes the need for preservation and characterization of banana diversity for the purposes of crop improvement. The world's largest banana collection is maintained at the Bioversity International Transit Centre (ITC) hosted by the Katholieke Universiteit Leuven in Belgium. This reference collection contains more than 1400 accessions in tissue culture and is continuously extended by various edible cultivars, improved varieties and wild species. Since the main goal of the gene bank is long-term conservation and effective use of the entire *Musa* genepool, characterization of already deposited as well as newly introduced accessions is of great importance.

Except for *M. acuminata*, *M. balbisiana* and their hybrids, the genus *Musa* contains about 50 species. Based on a set of morphological descriptors and basic chromosome number, the genus has been traditionally subdivided into four sections: Eumusa (1n = 1x = 11), Rhodochlamys (1n = 1x = 11), Australimusa (1n = 1x = 10) and Callimusa (1n = 1x = 9, 10) (Cheesman 1974). However, this classification has been often questioned and various regroupings have been suggested (Argent 1976, Wong et al. 2002). The estimation of nuclear genome size, one of the basic characteristics of a species, previously focused mainly on edible bananas and their wild ancestors. The genome size of 600 – 650 Mbp was determined in *M. acuminata* and 550 Mbp in *M. balbisiana* (Doležel et al. 1994, Lysák et al. 1999, Čížková et al. 2013). Bartoš et al. (2005) extended the knowledge of nuclear DNA content by representatives of sections Rhodochlamys, Australimusa and Callimusa. The authors

also characterized genomic distribution of ribosomal DNA. While the number of 45S rDNA loci was found conserved within individual sections, the number of 5S rDNA loci ranged from 4 to 8 per mitotic metaphase plate and varied within sections and even within different accessions of a single species.

For molecular characterization of *Musa* germplasm, Christelová et al. (2011) established microsatellite-based genotyping platform. Microsatellites, also known as simple sequence repeats (SSRs), are 1 – 6 bp long tandemly arranged repetitive sequences. They proved to be useful in molecular genotyping of many economically important crops (This et al. 2004, Hayden et al. 2007, Pessoa-Filho et al. 2007). The platform of Christelová et al. (2011) for genotyping in *Musa* is based on 19 microsatellite loci and enables discrimination between individual species, subspecies and subgroups of banana accessions.

Sequence analysis of the internal transcribed spacer region (ITS1-5.8S-ITS2) of nuclear ribosomal DNA can provide further information about the genetic diversity and evolutionary relationships among individual accessions. The ITS region has been one of the most widely used markers in phylogenetic studies of flowering plants (Alvarez and Wendel 2003) and has recently been successfully employed in phylogenetic reconstruction in the family *Musaceae* (Hřibová et al. 2011). It has been shown that ITS locus can be used for verification of genome constitution of inter- and intraspecific banana hybrids (Hřibová et al. 2011) and for support of unambiguous results of SSR analysis (Christelová et al. 2011).

The aim of this work was to characterize a set of 21 *Musa* accessions, which were newly introduced into the ITC collection. Nuclear genome size, chromosome number and genomic distribution of rDNA were determined in individual accessions to shed light on their genome structure and evolution. The molecular characterization was performed using SSR genotyping platform and was extended by the analysis of ITS sequence region which can provide the information about the putative hybrid character of some accessions and is useful to support incongruous results of SSR analysis.

Materials and methods

Plant material and genomic DNA extraction

In vitro rooted plants of 21 *Musa* accessions (Table 1) were obtained from the International Transit Centre (ITC, Katholieke Universiteit, Leuven, Belgium). The plants were transferred to soil and maintained in a greenhouse. Fresh cigar leaves were harvested, their fragments lyophilized and maintained at room temperature until use. Genomic DNA of all *Musa* species used for the SSR genotyping and ITS analysis was isolated from lyophilized leaves using the NucleoSpin PlantII kit (Macherey-Nagel GmbH & Co. KG, Düren, Germany) following the manufacturer's recommendations.

Estimation of genome size

Nuclear genome size was estimated according to Bartoš et al. (2005). Approximately 50 mg of young *Musa* leaf and 10 mg of leaf of soybean (*Glycine max* L. cv. Polanka, 2C = 2.5 pg DNA) (Doležel et al. 1994) which served as internal standard were used for sample preparation. Suspensions of cell nuclei were prepared by simultaneous chopping of leaf tissues of *Musa* and *Glycine* in a glass Petri dish containing 500 µl Otto I solution (0.1 M citric acid, 0.5% v/v Tween 20). Crude homogenate was filtered through a 50 µm nylon mesh. Nuclei were then pelleted (300 *g*, 5 min) and resuspended in 300 µl Otto I solution. After 1 hour incubation at room temperature, 900 µl Otto II solution (0.4 M Na_2HPO_4) (Otto 1990) supplemented with 50 µg/ml RNase, 50 µg/ml propidium iodide and 3 µl/ml 2-mercaptoethanol, were added. Samples were analyzed using Partec PAS flow cytometer (Partec GmbH, Münster, Germany) equipped with 488-nm argon laser. At least 5,000 nuclei were analyzed per sample. Three individuals were analyzed for each banana accession, and each individual was measured three times on three different days. Nuclear DNA content was then calculated from individual measurements following the formula:

2C nuclear DNA content [pg] = 2.5 x G_1 peak mean of *Musa* / G_1 peak mean of *Glycine*

Mean nuclear DNA content (2C) was then calculated for each plant. Genome size (1C value) was then determined considering 1 pg DNA equal to 0.978×10^9 bp (Doležel et al. 2003).

Statistical analysis was performed using NCSS 97 statistical software (Statistical Solutions Ltd., Cork, Ireland). One-way ANOVA and a Bonferoni's (All-Pairwise) multiple comparison test were used for analysis of variation in nuclear DNA content. The significance level $\alpha = 0.01$ was used.

SSR genotyping

The standardized platform for molecular characterization of *Musa* germplasm developed by Christelová *et al.* (2011) was used to genotype all 21 *Musa* accessions. The system is based on 19 microsatellite loci that are scored after the PCR with fluorescently labeled primers and capillary electrophoretic separation with internal standard (GeneScanTM-500 LIZ size standard, Applied Biosystems). The PCR products were multiplexed prior to the separation and loaded onto the automatic 96-capillary ABI 3730xl DNA Analyzer. Electrophoretic separation and signal detection was carried out with default module settings. The resulting data were analyzed and called for alleles using GeneMarker® v1.75 (Softgenetics), manually checked and implemented into the marker panels (Christelová *et al.* 2011). The genetic similarity matrices based on Nei´s genetic distance coefficient (Nei 1973) were calculated using PowerMarker v3.25 (Liu and Muse 2005) and the unweighted pair-group method with arithmetic mean (UPGMA; Michener and Sokal 1957) was used to assess the relationship between individual genotypes. The results of UPGMA cluster analysis were visualized in a form of a tree using FigTree v1.4.0 (http://tree.bio.ed.ac.uk/software/figtree/).

Analysis of ITS1-5.8S-ITS2 region

The ITS region was amplified from the genomic DNA using PCR with specific primers ITS-L and ITS-4 (48). PCR reaction mix consisted of 10µg genomic DNA, 1.5 mM $MgCl_2$, 0.2 mM dNTPs, 1 mM primers ITS-L and ITS-4, 1x PCR buffer and 2U/100 of Dynazyme^{TM} II DNA polymerase (Finnzymes, Espoo, Finland). Amplification was performed using PTC-200 thermal cycler (BIO-RAD, Hercules, CA, USA), with the following conditions: 94°C for 5 min (1 cycle), 94°C for 50 s, 52°C for 50 s, 72°C for 50 s (35 cycles) and 72°C for 10 min (1 cycle), and PCR

products were resolved in 1.5% agarose gels.

PCR products were purified using ExoSAP-IT® (USB, Cleveland, OH, USA) according to the manufacturer's instructions and used for DNA cloning into TOPO vector and transformed into *E. coli* electrocompetent cells (Invitrogen Life Technologies, Carlsbad, USA). For each accession, at least 28 cloned PCR products were sequenced. Sequencing was carried out using the BigDye Terminator v3.1 Cycle Sequencing kit (Applied Biosystems, Foster City, USA) according to the manufacturer's instructions and run on ABI 3730xl DNA analyzer (Applied Biosystems, Foster City, USA). Nucleotide sequences were edited using Staden Package (Staden 1996). Sequence boundaries of the spacers were assessed and phylogenetic relationships analysis was performed according to Hřibová et al. (2011).

Chromosome preparations

Metaphase spreads were prepared according to Doleželová et al. (1998). Actively growing root tips were pre-treated in 0.05% (w/v) 8-hydroxyquinoline for 3 hrs at room temperature and then fixed in 3:1 ethanol : acetic acid overnight. Fixed roots were washed in a solution of 75 mM KCl and 7.5 mM EDTA (pH 4) and meristem tips were digested in a mixture of 2% (w/v) pectinase and 2% (w/v) cellulase in 75 mM KCl and 7.5 mM EDTA (pH 4) for 90 min at 30ºC. Protoplast suspension was then filtered through a 150 μm nylon mesh and pelleted. The pellet was resuspended in 75 mM KCl and 7.5 mM EDTA (pH 4) and incubated for 5 min at room temperature. After pelleting, the protoplasts were washed three times with 70% ethanol. 5 μl of suspension were dropped onto a slide and shortly before drying out, 5 μl of 3:1 fixative were added to the drop to induce protoplast bursting. Finally, the slide was rinsed in 100% ethanol and air-dried.

Fluorescence in situ hybridization (FISH)

Probes for 45S rDNA and 5S rDNA were prepared by labeling *Radka*1 DNA clone (45S rDNA) and *Radka*2 DNA clone (5S rDNA) (Valárik et al. 2002) with digoxigenin-11-dUTP or biotin-16-dUTP (Roche Applied Science). Both probes were labeled by PCR using M13 forward and reverse primers (Invitrogen).

Hybridization mixture consisting of 50% formamide, 10% dextran sulfate in

1×SSC and 1 µg/ml labeled probe was added onto slides and denatured at 80°C for 3 min. The hybridization was carried out at 37°C overnight. The sites of probe hybridization were detected using anti-digoxigenin-FITC (Roche Applied Science) and streptavidin-Cy3 (Vector Laboratories, Burlingame, USA), and the chromosomes were counterstained with DAPI. The slides were examined with Olympus AX70 fluorescence microscope and the images of DAPI, FITC and Cy-3 fluorescence were acquired separately with a cooled high-resolution black and white CCD camera. The camera was interfaced to a PC running the MicroImage software (Olympus, Tokyo, Japan).

Results

Estimation of genome size

The amount of nuclear DNA was estimated after flow cytometric analysis of propidium iodide stained nuclei. All analysis resulted in histograms of relative DNA content with two dominant peaks corresponding to G_1 nuclei of *Musa* and *Glycine* (Figure 1). 2C nuclear DNA content determined based on the ratio of G_1 peaks positions ranged from 1.217 to 1.772 pg (Table 2). The differences between the three species of Eumusa were statistically significant with 2C DNA content ranging between 1.217 and 1.311 pg. Within the section Rhodochlamys, the lowest nuclear DNA content (2C = 1.261 pg) was found in the hybrid *Musa* x *fennicae* (ITC 1522) and the highest nuclear DNA content (2C = 1.315 pg) was found in *Musa laterita* (ITC 1575). Although the differences between individual accessions were small (max. 4.3%), some of them were statistically significant. Much higher interspecific variation (27.5%) of 2C DNA content was observed within the section Callimusa (2C = 1.390 – 1.772 pg). For *Musa borneensis* (ITC 1531), the highest nuclear DNA content (2C = 1.772 pg) known for diploid *Musa* species was determined. *Musa* cf. *uranoscopos* (ITC 1532), the only representative of section Australimusa in this study, had the nuclear DNA content of 1.442 pg.

Bonferoni's (All-Pairwise) multiple comparison test revealed twelve groups based on the differences in the nuclear DNA content (Table 2). Six of these groups were represented by only one accession (*M. itinerans* var. *itinerans* (ITC 1571), *M. monticola* (ITC 1528), *M. barioensis* (ITC 1568), *M.beccarii* var. *beccarii* (ITC

1516), *M. beccarii* var. *hottana* (ITC 1529) and *M. borneensis* (ITC 1531)). Two groups comprised representatives of sections Eumusa and Rhodochlamys and one group included species belonging to sections Australimusa and Callimusa.

FISH with probes for 45S and 5S rDNA

Protoplast dropping technique was used to prepare metaphase spreads for chromosome counting and cytogenetic mapping of rRNA genes. Chromosome numbers corresponded to expected ploidy levels and section-specific chromosome numbers in all evaluated accessions (Table 2). Unfortunately, plants of *Musa* cf. *uranoscopos* (ITC 1532) and *M. campestris* var. *limbangensis* (ITC 1535) did not have actively growing roots needed for preparation of protoplast suspensions and their chromosomes could not be determined.

FISH with probes for 45S rDNA revealed distinct hybridization sites on one chromosome pair in the secondary constriction in all accession of sections Eumusa and Rhodochlamys (Table 2, Fig 2). A variable number of 45S rDNA loci was observed in Callimusa species. Four accessions contained two signals (on one chromosome pair bearing secondary constriction), *M. beccarii* var. *beccarii* (ITC 1516) was characterized by four signals, five signals were observed on chromosomes of *M. borneensis* (ITC 1531) and *M. beccarii* var. *hottana* (ITC 1529) possessed six 45S rDNA loci on three chromosome pairs (Table 2). Two of these signals were observed in the secondary constriction of one chromosome pair and the other were localized in interstitial chromosome regions.

A significantly higher variation was detected in the number of 5S rDNA loci (Table 2). Within Eumusa section, 5S rDNA sites were localized on two or three chromosome pairs. Three to seven signals were observed on chromosomes of accessions belonging to section Rhodochlamys. Callimusa species contained 5S rDNA gene clusters on two to eight chromosomes. In *M. beccarii* var. *beccarii* (ITC 1516), *M. beccarii* var. *hottana* (ITC 1529) and *M. borneensis* (ITC 1531), two of the 5S rDNA sites co-localized with interstitially localized 45S rDNA genes.

SSR genotyping

The molecular characterization of the newly introduced ITC accessions based on SSR markers included inspection of their clustering pattern within the reference set of diploid entries used in the study of Christelová et al. (2011). The UPGMA

cluster analysis based on the Nei (1973) genetic distance resulted in relatively clear grouping of genotype groups and subgroups (Figure 3). Inclusion of the new accessions did not change the overall grouping of the species as compared to the results of Christelová et al. (2011). Most of the entries under this study clustered with the reference accessions in accordance to their expected classification based on morphological description of the plants. Newly introduced accessions belonging to Eumusa and Rhodochlamys sections were grouped within the cluster A comprising species from section Rhodochlamys and several Eumusa entries (ssp. malaccensis). The two accessions described as *M. itinerans* (ITC 1526 and ITC 1571) and *M. rubinea* (ITC 1518) formed a separate clade within the cluster A. *M. yunnanensis* shared a clade with several entries of section Rhodochlamys. Accessions described as Australimusa and Callimusa species were grouped within the cluster B together with reference accessions from the sections Australimusa and Callimusa. *M.* cf. *uranoscopos* (Australimusa) was embedded among Callimusa species.

Analysis of ITS1-5.8S-ITS2 region

In general, the length of ITS1 and ITS2 spacer varied from 215 to 223 bp and from 205 to 218 bp, respectively. A total length of ITS1-5.8S-ITS2 sequence region ranged from 582 bp (*M. beccarii* var. *hottana*, ITC 1529) to 591 bp (*M. lutea*, ITC 1515) in all accessions, except for four representatives: *M. rubra* (ITC 1590), *M. rosea x siamensis* (ITC 1592), *M. yunnanensis* (ITC 1573) and *M. campestris* var. *sarawakensis* (ITC 1517) where the ITS region was shorter due to the deletions. One of the highest sequence diversities was observed in putative hybrid clones, e.g. *M. rosea x ornata* (ITC 1572) as well as in some wild diploids, e.g. *M. barionensis* (ITC 1568). GC content of ITS1 varied from 57.41 to 65.02 % and was slightly lower than the GC content of ITS2 (59.51 to 69.74 %). The 5.8S rDNA sequence region had a conserved length of 153 bp or 155 bp, except two ITS types of *M. yunnanensis* (ITC 1573) and one type of *M. rubra* (ITC 1590) and M. *rosea* x *siamensis* (ITC 1598) due to the deletion. GC content of 5.8S rDNA varied from 46.45 to 59.63% and was significantly lower than the GC content in ITS1 and ITS2.

Identification of pseudogenes and phylogenetic analysis

The secondary structure of ITS2 and 5.8S rDNA sequence regions was reconstructed for all accessions. ITS2 sequences formed specific four-helices structure with typical pyrimidine-pyrimidine bulge in helix II and the most conserved primary sequence included the TGGT in the helix III. Secondary structure of 5.8S rDNA sequence was reconstructed under specific settings for base pairing as described by Hřibová et al. (2011). Moreover, 5.8S rDNA sequences were checked for the presence of three conserved motives (Harpke and Peterson 2008). The highly conserved sequence of 16 bp motif M1 (5'-CGATGAAGAACGTAGC-3') is a part of two helices – helix B4 and B5 (Harpke and Peterson 2008). Motif M2 (5'-GAATTGCAGAATCC-3'), previously described by Jobes and Thien (1997) is 14 bp long and located in the loop and 10 bp long motif M3 (5'-TTTGAACGCA-3') is a part of the B4 and B7 helices (Harpke and Peterson 2008). The nucleotide changes in conserved motives of 5.8S rDNA of analyzed accessions, the information on GC content, presence of conserved motives in the 5.8S rDNA sequence and ability of ITS2 and 5.8S rDNA sequence to fold into a conserved secondary structure enabled identification of putative pseudogenes (Table 3).

As shown in Table 3, more than half of the analyzed accessions contained at least two types of ITS sequences. On the contrary, M. *rosea* x *siamensis* ITC 1598 which has been described as hybrid clone showed presence of only one type of ITS sequence region and its hybrid character was not confirmed. Two datasets were used to reconstruct phylogenetic analysis using BioNJ. Dataset 1 contained *Musa* ITS sequences previously described by Hřibová et al. (2011) and ITS types of 21 studied accessions, from which putative pseudogenic ITS types were excluded (Figure S1). In dataset 2, putative pseudogenes of accessions analyzed in this work were included (Figure S2). The main clades were conserved between both resulting trees (with and without pseudogenes). Pseudogenic sequences were included into the phylogenetic analysis with the aim to characterize the putative parental species of hybrid clones. Inclusion of the new accessions did not change the overall grouping of the species as compared to the results of Hřibová et al. (2011). Phylogenetic analysis of ITS region supported the clustering obtained by SSR analysis except M. *rosea* x *siamensis* ITC 1598 (Figure S1).

Discussion

Nuclear DNA content and distribution of rDNA

Most of the knowledge about the nuclear genome size and genomic distribution of rDNA loci in banana comes from the analysis of triploid cultivars and their wild ancestors (Doležel et al. 1994, Doleželová et al. 1998, Osuji et al. 1998, Lysák et al. 1999, Čížková et al. 2013). Only the study of Bartoš et al. (2005) included representatives of all *Musa* sections and provided the first complex picture of the whole genus. Our study significantly expands the number of wild *Musa* species where these key characteristics of the banana nuclear genome were estimated.

The diversity in genome size observed in Eumusa and Rhodochlamys species is in line with the previous study (Bartoš et al. 2005) and brings further evidence about the close relationships between the two sections. Similarly, the number of 45S rDNA loci was equal in accessions belonging to both sections and number of 5S rDNA loci fell within the same range. The traditional classification describes sections Eumusa and Rhodochlamys mainly based on morphological characters. The separation of these sections is however doubtful, because the morphological differences are difficult to detect in some species and occurrence of interspecific hybridization is quite frequent (Cheesman 1974, Shepherd 1999). Our findings, as well as previous analysis based on different molecular markers (Ude et al 2002, Wong et al. 2002, Nwakanma et al. 2002, Li et al. 2010), support suggestions that sections Eumusa and Rhodochlamys should be merged into one.

Section Callimusa is the only one which includes species with different basic chromosome number (1x = 9, 10). Representatives of section Callimusa analysed in our study showed high variation in the nuclear genome size as well as in the number of rDNA loci and were clearly discriminated from Eumusa and Rhodochlamys. Increased number of 45S rDNA signals was observed in the three accessions (*M. borneensis*, *M. beccarii* var. *hottana* and *M. beccarii* var. *beccarii*) with the highest nuclear DNA content. *M.* cf. *uranoscopos*, the representative of section Australimusa, was grouped together with three Callimusa species (*M. violascens*, *M. lutea* and *M. campestris* var. *limbangensis*) based on differences in nuclear DNA content (Table 2), which agrees with the previously observed similarities of Australimusa and Callimusa species (Jarret and Gawel 1995; Wong et al. 2002).

SSR and ITS analysis

The possibility to use microsatellite markers for molecular characterization of *Musa* accessions was explored by Christelová et al. (2011) who developed a standardized platform for SSR genotyping of bananas. The grouping of the 21 accessions analyzed in the present work within the diploid reference set was revealed by the UPGMA cluster analysis. In most cases, results of this grouping were consistent with the characterization based on traditional morphotaxonomic classification (Figure 3). Unexpected clustering of three accessions (*M. rubinea*, *M. yunnanensis* and *M.* cf. *uranoscopos*) with entries belonging to other sections could be explained by mislabeling of the accessions during their collection or during *in vitro* propagation in the gene bank. Hybrid character of these accessions could also explain their position within the tree.

Short length, utilization of highly conserved primers and relatively fast evolution of internal transcribed spacers ITS1 and ITS2 compared with rRNA genes, made the ITS region one of the most popular markers in phylogenetic studies. The first detailed information about the structure and diversity of ITS region in the genus *Musa* was provided by Hřibová et al. (2011) and the work of Christelová et al. (2011) showed usefulness of the ITS for further analysis of accessions which give incongruous results with SSRs.

Based on *in silico* identification of putative pseudogenes, which are expected to evolve independently at different rates than their functional counterparts and can accumulate mutations, we show that the presence of putative pseudogenic ITS data have no significant effect on the structure and positions of all main clades in the phylogenetic analysis as it was showed before (Hřibová et al. 2011).

The optimized SSR genotyping platform which was created by Christelová et al. (2011) is providing a genotyping service to the international *Musa* community under the auspices of the Bioversity International (Montpellier, FR). In this study, we confirmed the usefulness of SSR markers for molecular characterization of unknown accessions of *Musa* (Christelová et al. 2011). However, the SSR genotyping is based on scoring alleles and this might not be appropriate for diversity and phylogenetic inference estimation in case of inter-specific hybrids and their backcross progenies (Hřibová et al. 2011). Thus, we performed a detailed analysis of the nucleotide composition and structure of the ITS region and showed that some of wild diploids

contained polymorphic ITS regions. Moreover, *in silico* analysis of the ITS sequences indicated the presence of putative pseudogenic ITS types. These observations add further support to the notion that some of the analyzed accessions might be interspecific hybrids and/or their backcross progeny (Hřibová et al. 2011, De Langhe et al. 2010).

Conclusions

The present study provides novel data on nuclear genome size and genomic distribution of ribosomal genes in banana. We show a high variability in both characters, especially in section Callimusa. The data significantly improve the knowledge about the genome structure of wild *Musa* species. The genetic diversity of wild accessions was studied using the standardized platform for genotyping in banana. Cladogram based on the SSR markers together with the previously studied wild diploids showed clear clustering and enabled us to identify groups of the most related species. Moreover, phylogenetic analysis based on the ITS sequences was used to study heterozygosity or hybrid character of studied accessions. We showed that more than half of the analyzed accession were heterozygous and/or of hybrid origin. The ITS analysis also confirmed previous results of close phylogenetic relationship of the species from sections Australimusa and *Callimusa*, and *Eumusa* and *Rhodochlamys*.

Acknowledgements

We thank Jan Bartoš for valuable advices and Radka Tušková and Marie Seifertová for excellent technical assistance. This work was supported by the Ministry of Education, Youth and Sports of the Czech Republic (award INGO II No. LG12021), by the Bioversity International (LoA CfL 2010/58, LoA CfL 2011/51, LoA CfL 2012/43) and the European Regional Development Fund (Operational Programme Research and Development for Innovations No. ED0007/01/01).

References

Alvarez I, Wendel JF (2003) Ribosomal ITS sequences and plant phylogenetic inference. Molecular Phylogenetics and Evolution 29: 417-434.

Argent GCG (1976) The wild bananas of Papua New Guinea. Notes Royal Botanic Garden Edinburgh 35: 77-114.

Bartoš J, Alkhimova O., Doleželová M, De Langhe E, Doležel J (2005) Nuclear genome size and genomic distribution of ribosomal DNA in Musa and Ensete (Musaceae): taxonomic implications. Cytogenetic and Genome Research 109:50-57.

Cheesman EE (1974) Classification of the bananas. II. The genus *Musa* L. Kew Bulletin 2: 106-117.

Christelová P, Valárik M, Hřibová E, Van den houwe I, Channelière S, Roux N, Doležel J (2011) A platform for efficient genotyping in *Musa* using microsatellite markers. AoB PLANTS plr024, doi:10.1093/aobpla/plr024.

Čížková J, Hřibová E, Humplíková L, Christelová P, Suchánková P, Doležel J (2013) Molecular analysis and genomic organization of major DNA satellites in banana (*Musa* spp.). PLoS ONE 8(1): e54808. Doi: 10.1371/journal.pone.0054808.

Doležel J, Doleželová M, Novák FJ (1994) Flow cytometric estimation of nuclear DNA amount in diploid ananas (*Musa acuminata* and *M. balbisiana*). Biologia Plantarum 36: 351-357.

Doležel J, Bartoš J, Voglmayr H, Greilhuber J (2003) Nuclear DNA content and genome size of trout and human. Cytometry A 51: 127-128.

Doleželová M, Valárik M, Swennen R, Horry JP, Doležel J (1998) Physical mapping of the 18S-25S and 5S ribosomal RNA genes in diploid bananas. Biologia Plantarum 41: 497-505.

Harpke D, Peterson A (2008) 5.8S motifs for the identification of pseudogenic ITS regions. Botany 86: 300-305.

Hayden MJ, Nguyen TM, Waterman A, McMichael GL, Chalmers KJ (2007) Application of multiplex-ready PCR for fluorescence-based SSR genotyping in barley and wheat. Molecular breeding 21: 271-281.

Hřibová E, Čížková J, Christelová P, Taudien S, De Langhe E, Doležel J (2011) The ITS-5.8S-ITS2 sequence region in the *Musaceae*: structure, diversity and use in molecular phylogeny. PLoS ONE 6(3): e17863. Doi:10.1371/journal.pone.0017863.

Jarret RL, Gawel NJ (1995) Molecular markers, genetic diversity and systematis. In: Gowen S, ed. Bananas and plantains. Chapman & Hall, London, UK.

Jobes DV, Thien LB (1997) A conserved motif in the 5.8S ribosomal RNA (rRNA) gene is a useful diagnostic marker for for plant internal transcribed spacer (ITS) sequence. Plant Molecular Biology Reporter 15: 326-334.

Li L-F, Häkkinen M, Yuan Y-M, Hao G, Ge X-J (2010) Molecular phylogeny and systematics of the banana family (*Musaceae*) inferred from multiple nuclear and chloroplast DNA fragments, with a special reference to the genus *Musa*. Molecular Phylogenetics and Evolution 57: 1-10.

Liu K, Muse SV (2005) PowerMarker: integrated analysis environment for genetic marker data. Bioinformatics 21: 2128-2129.

Lysak MA, Dolezelova M, Horry JP, Swennen R, Doležel J (1999) Flow cytometric analysis of nuclear DNA content in *Musa*. Theoretical and Applied Genetics 98: 1344-1350.

Michener CD, Sokal RR (1957) A quantitative approach to a problem of classification. Evolution 11: 490-499.

Nei M (1973) Analysis of gene diversity in subdivided populations. Proceedings of the National Academy of Sciences of the USA 70: 3321-3323.

Nwakanma DC, PillayM, Okoli BE, Tenkouano A (2003) Sectional relationships in the genus *Musa* L. inferred from the PCR-RFLP of organelle DNA sequences. Theoretical and Applied Genetics 107: 850-856.

Osuji JO, Crouch J, Harrison G, Heslop-Harrison JS (1998) Molecular cytogenetics of *Musa* species, cultivars and hybrids: location of 18S-5.8S-25S and 5S rDNA and telomere-like sequences. Annals of Botany 82: 243-248.

Otto F (1990) DAPI staining of fixed cells for high-resolution flow cytometry of nuclear DNA. In: Crissman HA, Darzynkiewicz Z, ed. Methods in Cell Biology. Academic Press, New York.

Pessoa- Filho M, Beló A, Alcochete AAN, Rangel PHN, Ferreira ME (2007) A set of multiplex panels of microsatellite markers for rapid molecular characterization of rice accessions. BMC Plant Biology 7: 23.

Shepherd K (1999) Cytogenetics of the genus *Musa* (International Network for the Improvement of Banana and Plantain, Montpellier).

Simmonds NW, Shepherd K (1955) Taxonomy and origin of cultivated bananas. Botanical Journal of Linnean Society 55: 302-312.

Staden R (1996) The Staden sequence analysis package. Molecular Biotechnology 5: 233-241.

This P, Jung A, Boccaccii P, Borrego J, Botta R, Constantini L, Crespan M, Dangl GS, Eisenheld C, Ferreira-Monteiro F, Grando S, Ibanez J, Lacombe T, Laucou V, Magalhães R, Meredith CP, Milani N, Peterlung E, Regner F, Zulini L, Maul E (2004) Development of a standard set of microsatellite reference alleles for identification of grape cultivars. Theoretical and Applied Genetics 109: 1448-1458.

Ude G, Pillay M, Nwakanma D, Tenkouano A (2002) Analysis of genetic diversity and sectional relationships in *Musa* using AFLP markers. Theoretical and Applied Genetics 104: 1239-1245.

Valárik M, Šimková H, Hřibová E, Šafář J, Doleželová M, Doležel J (2002) Isolation, characterization and chromosome localization of repetitive DNA sequences in bananas (*Musa* spp.). Chromosome Research 10: 89-100.

Wong C, Kiew R, Argent GCG, Set O, Lee SK, Gan YY (2002) Assessment of the validity of the sections in *Musa* (Musaceae) using AFLP. Annals of Botany 90:231-238.

Legends to figures:

Figure 1: Histograms of relative nuclear DNA content obtained after flow cytometric analysis of propidium iodide stained nuclei isolated from Musa accessions. (A) M. itinerans var. itinerans (2C = 1.217 pg); (B) M. rosea (2C = 1.285 pg); (C) M. cf. uranoscopos (2C = 1.442 pg); (D) M. beccarii var. beccarii (2C = 1.537 pg). Nuclei isolated from soybean (Glycine max, 2C = 2.5 pg) were used as an internal reference standard.

Figure 2: Examples of genomic distribution of 45S (red, thick arrow) and 5S (green, thin arrow) rDNA as determined on mitotic metaphase chromosomes of Musa accession after FISH with labelled probes. (A) Musa yunnanensis (ITC 1573); (B) Musa mannii (ITC 1574); (C) Musa rosea x siamensis (ITC 1592); (D) Musa campestris var. sarawakensis (ITC 1517); (E) Musa monticola (ITC 1528); (F) Musa lutea (ITC 1515). Chromosomes were counterstained with DAPI. Bar = 5 μm.

Figure 3: Dendrogram showing the results of the UPGMA analysis of studied accessions within the reference set of diploid entries used in the study of Christelová et al. (2011). The main clades and subclades are distinguished by colors. The Australimusa/Callimusa clade in pink; Rhodochlamys in green; BB genotypes in blue; AA genotypes in red and SS genotypes in orange. Accessions analyzed in this study are shown in black.

Supporting information:

Figure S1: Phylogenetic analysis based on the ITS1-ITS2 sequence region according Hřibová et al. (2011). BioNJ tree constructed from a Jukes-Cantor distance matrix of the concatenated region contained ITS1 and ITS2 spacer sequence. The tree was rooted on midpoint. The main clades and subclades are distinguished by colors. The Australimusa/Callimusa clade in pink; Rhodochlamys in green; BB genotypes in blue; AA genotypes in red and SS genotypes in orange. The ITS sequences of accessions analyzed in this study are shown in black.

Figure S2: Phylogenetic analysis based on the ITS1-ITS2 sequence region according Hřibová et al. (2011). BioNJ tree constructed from a Jukes-Cantor distance matrix of the concatenated region contained ITS1 and ITS2 spacer sequence including putative pseudogenic sequences. The tree was rooted on midpoint. The main clades and subclades are distinguished by colors. The Australimusa/Callimusa clade in pink; Rhodochlamys in green; BB genotypes in blue; AA genotypes in red and SS genotypes in orange. The ITS sequences of analyzed accessions are in black color. Putative pseudogenic ITS sequence regions are marked by asterix.

Figure 1

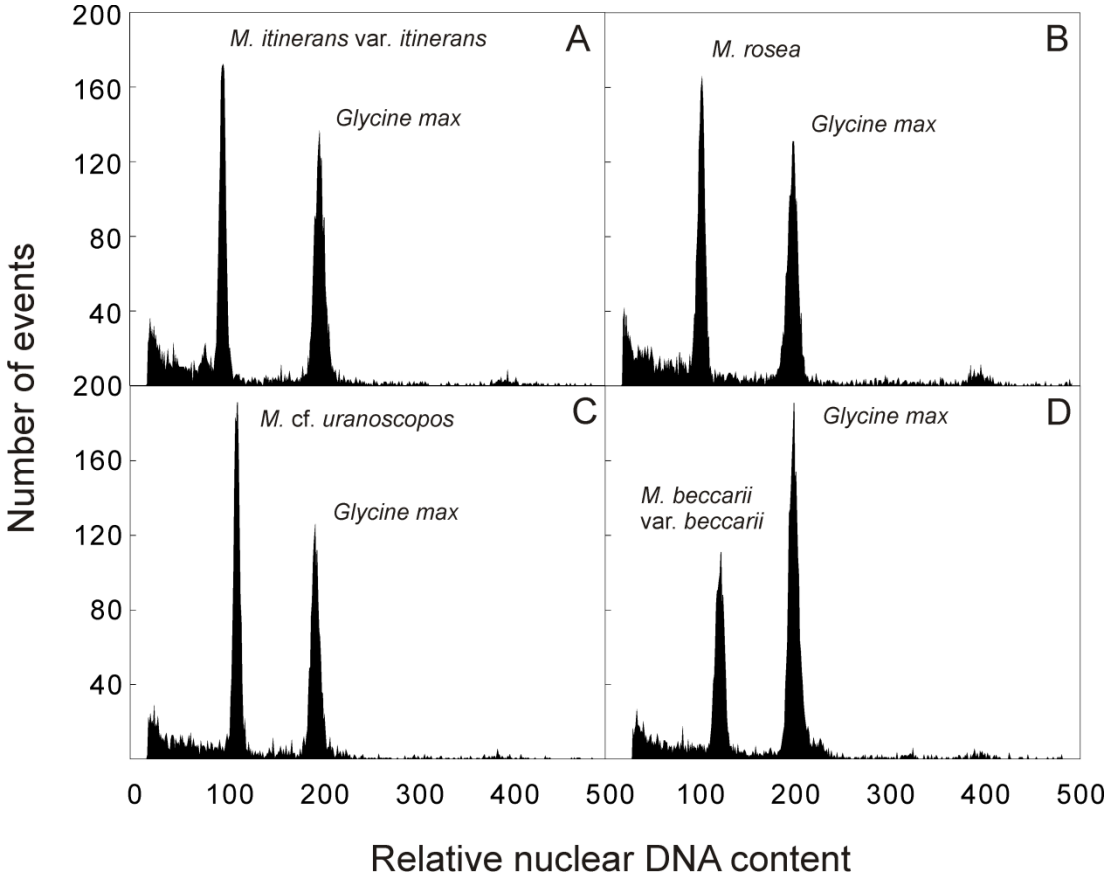

M. itinerans var. itinerans
A
Glycine max
M. rosea
B
Glycine max
M. cf. uranoscopos
C
Glycine max
Glycine max
D
M. beccarii
var. beccarii
Number of events
200
160
120
80
40
200
160
120
80
40
0
100
200
300
400
500
100
200
300
400
500
Relative nuclear DNA content

Figure 2

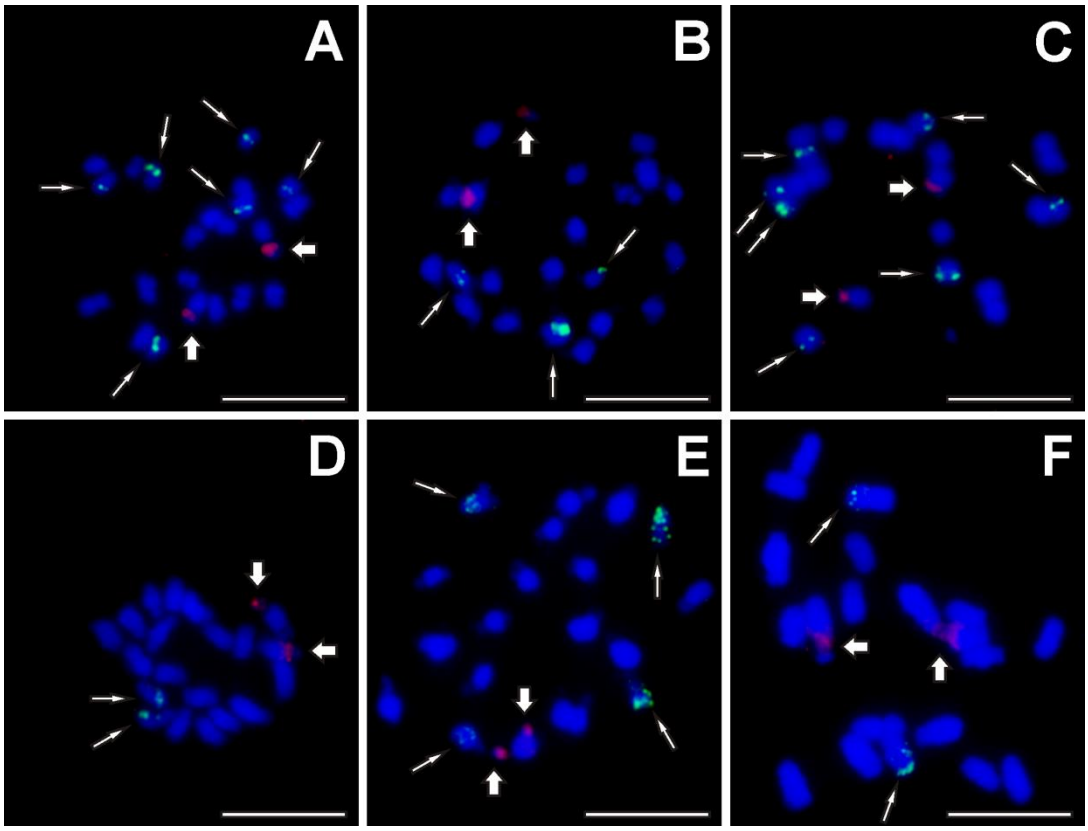


Figure 3

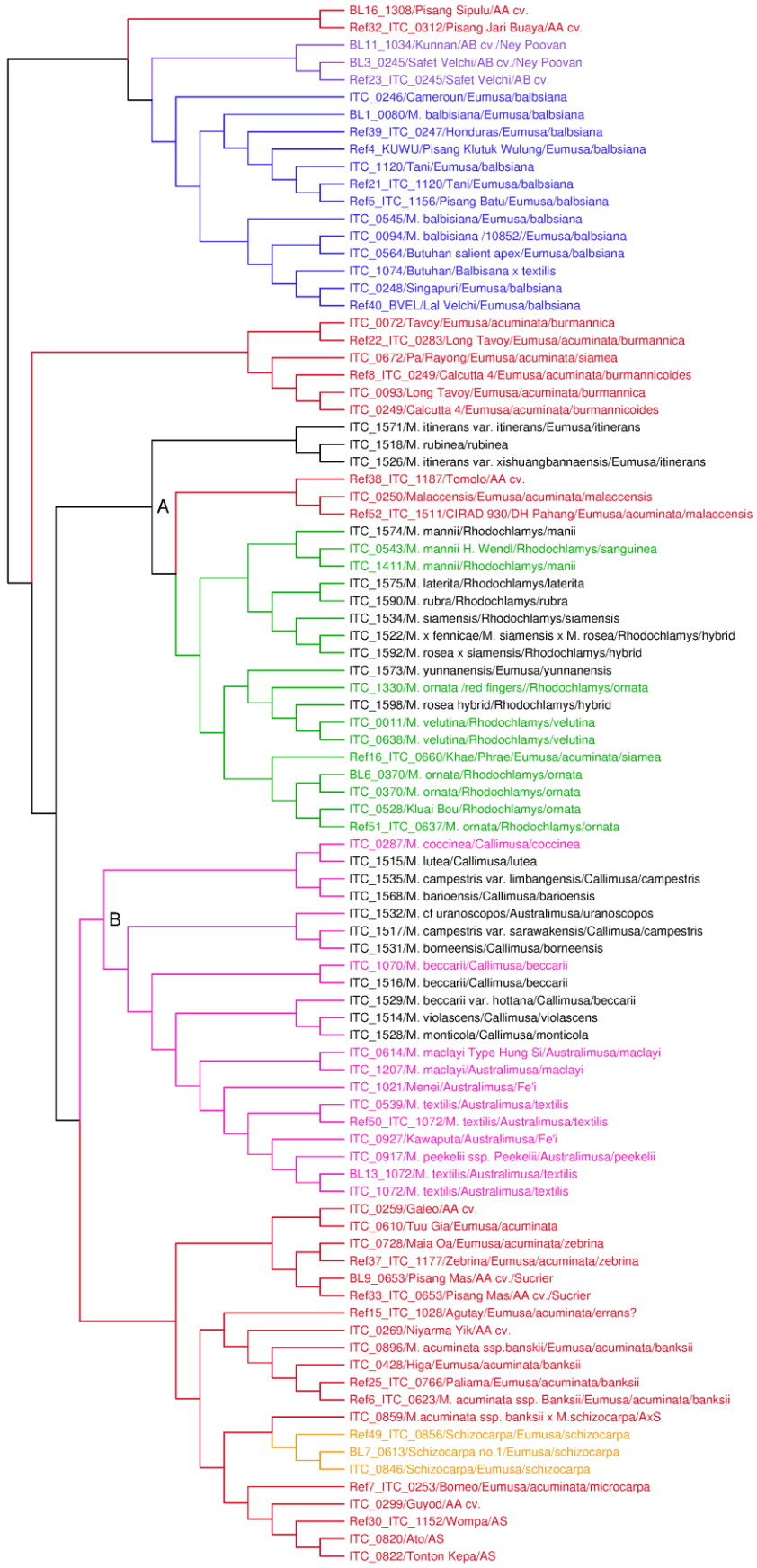

BL16_1308/Pisang Sipulu/AA cv.
Ref32_ITC_0312/Pisang Jari Buaya/AA cv.
BL11_1034/Kunnan/AB cv./Ney Poovan
BL3_0245/Safet Velchi/AB cv./Ney Poovan
Ref23_ITC_0245/Safet Velchi/AB cv.
ITC_0246/Cameroun/Eumusa/balbsiana
BL1_0080/M. balbisiana/Eumusa/balbsiana
Ref39_ITC_0247/Honduras/Eumusa/balbsiana
Ref4_KUWU/Pisang Klutuk Wulung/Eumusa/balbsiana
ITC_1120/Tani/Eumusa/balbsiana
Ref21_ITC_1120/Tani/Eumusa/balbsiana
Ref5_ITC_1156/Pisang Batu/Eumusa/balbsiana
ITC_0545/M. balbisiana/Eumusa/balbsiana
ITC_0094/M. balbisiana /10852//Eumusa/balbsiana
ITC_0564/Butuhan salient apex/Eumusa/balbsiana
ITC_1074/Butuhan/Balbisana x textilis
ITC_0248/Singapuri/Eumusa/balbsiana
Ref40_BVEL/Lal Velchi/Eumusa/balbsiana
ITC_0072/Tavoy/Eumusa/acuminata/burmannica
Ref22_ITC_0283/Long Tavoy/Eumusa/acuminata/burmannica
ITC_0672/Pa/Rayong/Eumusa/acuminata/siamea
Ref8_ITC_0249/Calcutta 4/Eumusa/acuminata/burmannicoides
ITC_0093/Long Tavoy/Eumusa/acuminata/burmannica
ITC_0249/Calcutta 4/Eumusa/acuminata/burmannicoides
ITC_1571/M. itinerans var. itinerans/Eumusa/itinerans
ITC_1518/M. rubinea/rubinea
ITC_1526/M. itinerans var. xishuangbannaensis/Eumusa/itinerans
Ref38_ITC_1187/Tomolo/AA cv.
ITC_0250/Malaccensis/Eumusa/acuminata/malaccensis
Ref52_ITC_1511/CIRAD 930/DH Pahang/Eumusa/acuminata/malaccensis
ITC_1574/M. mannii/Rhodochlamys/manii
ITC_0543/M. mannii H. Wendl/Rhodochlamys/sanguinea
ITC_1411/M. mannii/Rhodochlamys/manii
ITC_1575/M. laterita/Rhodochlamys/laterita
ITC_1590/M. rubra/Rhodochlamys/rubra
ITC_1534/M. siamensis/Rhodochlamys/siamensis
ITC_1522/M. x fennicae/M. siamensis x M. rosea/Rhodochlamys/hybrid
ITC_1592/M. rosea x siamensis/Rhodochlamys/hybrid
ITC_1573/M. yunnanensis/Eumusa/yunnanensis
ITC_1330/M. ornata /red fingers//Rhodochlamys/ornata
ITC_1598/M. rosea hybrid/Rhodochlamys/hybrid
ITC_0011/M. velutina/Rhodochlamys/velutina
ITC_0638/M. velutina/Rhodochlamys/velutina
Ref16_ITC_0660/Khae/Phrae/Eumusa/acuminata/siamea
BL6_0370/M. ornata/Rhodochlamys/ornata
ITC_0370/M. ornata/Rhodochlamys/ornata
ITC_0528/Kluai Bou/Rhodochlamys/ornata
Ref51_ITC_0637/M. ornata/Rhodochlamys/ornata
ITC_0287/M. coccinea/Callimusa/coccinea
ITC_1515/M. lutea/Callimusa/lutea
ITC_1535/M. campestris var. limbangensis/Callimusa/campestris
ITC_1568/M. barioensis/Callimusa/barioensis
ITC_1532/M. cf uranoscopos/Australimusa/uranoscopos
ITC_1517/M. campestris var. sarawakensis/Callimusa/campestris
ITC_1531/M. borneensis/Callimusa/borneensis
ITC_1070/M. beccarii/Callimusa/beccarii
ITC_1516/M. beccarii/Callimusa/beccarii
ITC_1529/M. beccarii var. hottana/Callimusa/beccarii
ITC_1514/M. violascens/Callimusa/violascens
ITC_1528/M. monticola/Callimusa/monticola
ITC_0614/M. maclayi Type Hung Si/Australimusa/maclayi
ITC_1207/M. maclayi/Australimusa/maclayi
ITC_1021/Menei/Australimusa/Fe'i
ITC_0539/M. textilis/Australimusa/textilis
Ref50_ITC_1072/M. textilis/Australimusa/textilis
ITC_0927/Kawaputa/Australimusa/Fe'i
ITC_0917/M. peekelii ssp. Peekelii/Australimusa/peekelii
BL13_1072/M. textilis/Australimusa/textilis
ITC_1072/M. textilis/Australimusa/textilis
ITC_0259/Galeo/AA cv.
ITC_0610/Tuu Gia/Eumusa/acuminata
ITC_0728/Maia Oa/Eumusa/acuminata/zebrina
Ref37_ITC_1177/Zebrina/Eumusa/acuminata/zebrina
BL9_0653/Pisang Mas/AA cv./Sucrier
Ref33_ITC_0653/Pisang Mas/AA cv./Sucrier
Ref15_ITC_1028/Agutay/Eumusa/acuminata/errans?
ITC_0269/Niyarma Yik/AA cv.
ITC_0896/M. acuminata ssp.banskii/Eumusa/acuminata/banksii
ITC_0428/Higa/Eumusa/acuminata/banksii
Ref25_ITC_0766/Paliama/Eumusa/acuminata/banksii
Ref6_ITC_0623/M. acuminata ssp. Banksii/Eumusa/acuminata/banksii
ITC_0859/M.acuminata ssp. banksii x M.schizocarpa/AxS
Ref49_ITC_0856/Schizocarpa/Eumusa/schizocarpa
BL7_0613/Schizocarpa no.1/Eumusa/schizocarpa
ITC_0846/Schizocarpa/Eumusa/schizocarpa
Ref7_ITC_0253/Borneo/Eumusa/acuminata/microcarpa
ITC_0299/Guyod/AA cv.
Ref30_ITC_1152/Wompa/AS
ITC_0820/Ato/AS
ITC_0822/Tonton Kepa/AS
A
B

Table 1: List of *Musa* accessions used in this study

Section	Accession name	ITC code
Eumusa	*Musa itinerans* var. *xishuangbannaensis*	1526
	Musa itinerans var. *itinerans*	1571
	Musa yunnanensis	1573
Rhodochlamys	*Musa rubinea*	1518
	Musa x fennicae (*Musa siamensis* (male) x *Musa rosea* (female))	1522
	Musa siamensis	1534
	Musa mannii	1574
	Musa laterita	1575
	Musa rubra	1590
	Musa rosea x *siamensis*	1592
	Musa rosea (hybrid)	1598
Australimusa	*Musa* cf. *uranoscopos*	1532
Callimusa	*Musa violascens*	1514
	Musa lutea	1515
	Musa beccarii var. *beccarii*	1516
	Musa campestris var. *sarawakensis*	1517
	Musa monticola	1528
	Musa beccarii var. *hottana*	1529
	Musa borneensis	1531
	Musa campestris var. *limbangensis*	1535
	Musa barioensis	1568

Table 2: DNA content, chromosome number and number of 45S and 5S rDNA loci

Accession name	ITC code	2C nuclear DNA content [pg]		Monoploid genome size [Mbp/1C$_x$]	Bonferoni's DNA content grouping*										Chromosome number (2n)	Number of signals**	
		Mean	± SD													45s rDNA	5S rDNA
Musa itinerans var. *xishuangbannaensis*	1526	1.311	0.017	641					E						22	2	4
Musa itinerans var. *itinerans*	1571	1.217	0.004	595	A										22	2	6
Musa yunnanensis	1573	1.259	0.011	616		B									22	2	6
Musa rubinea	1518	1.310	0.016	641					E						22	2	4
Musa x fennicae	1522	1.261	0.012	617		B	C								22	2	6
Musa siamensis	1534	1.280	0.006	626		B	C								22	2	6
Musa mannii	1574	1.282	0.008	627			C								22	2	3
Musa laterita	1575	1.315	0.006	643					E						22	2	6
Musa rubra	1590	1.306	0.01	639				D	E						22	2	6
Musa rosea x *siamensis*	1592	1.279	0.006	625			C								22	2	7
Musa rosea (hybrid)	1598	1.285	0.002	628			C	D							22	2	4
Musa cf. *uranoscopos*	1532	1.442	0.006	705							H				-	-	-
Musa violascens	1514	1.428	0.011	698						G	H				20	2	4
Musa lutea	1515	1.432	0.009	700						G	H				20	2	2
Musa beccarii var. *beccarii*	1516	1.537	0.017	752									J		18	4	4
Musa campestris var. *sarawakensis*	1517	1.417	0.008	693						G					20	2	2
Musa monticola	1528	1.390	0.016	680					F						20	2	4
Musa beccarii var. *hottana*	1529	1.673	0.02	818										K	18	6	4
Musa borneensis	1531	1.772	0.006	867										L	20	5	8
Musa campestris var. *limbangensis*	1535	1.454	0.003	711							H				-	-	-
Musa barioensis	1568	1.480	0.014	724								I			20	2	3

* Statistical analysis was performed using mean values of 2C nuclear DNA content of individual plants ($\alpha = 0.01$). DNA content within each group described by the same letter was not significantly different.
** Based on the analysis of mitotic metaphase plates (2n)

Table 3: Sequence characteristics of ITS1-5.8S-ITS2 regions in *Musa* accessions

Accession name	Label of ITS type▲	GC content [%]			Position of nucleotide changes (nt-) in conserved 5.8S motives			Secondary structure of ITS2	Secondary structure of 5.8S	Note
		ITS1	5.8S	ITS2	Motif M1	Motif M2	Motif M3			
Musa itinerans var. *xishuangbannaensis*	1526	57.67	53.55	63.76	conserved	conserved	conserved	four-helices	conserved	
Musa itinerans var. *itinerans*	1571	57.67	53.55	63.76	conserved	conserved	conserved	four-helices	conserved	
Musa yunnanensis	1573_type1	63.59	57.42	68.84	conserved	conserved	conserved	four-helices	conserved	
	1573_type2	**60.47**	**53.59**	**60.57**	**nt-2 'A'**	**conserved**	**conserved**	**not formed**	**not formed**	**pseudogene**
	1573_type3	**60.93**	**54.90**	**62.91**	**nt-2 'A'**	**conserved**	**conserved**	**four-helices**	**not formed**	**pseudogene**
	1573_type4	**57.41**	**48.94**	**64.02**	**nt-16 'T'**	**nt-9 'A'**	**nt-8 'A' nt-9 'T'**	**not formed**	**not formed**	**pseudogene**
Musa rubinea	1518_type1	57.67	53.55	63.76	conserved	conserved	conserved	four-helices	conserved	
	1518_type2	58.14	52.26	63.76	conserved	conserved	conserved	four-helices	conserved	
	1518_type3	62.96	57.42	69.55	conserved	conserved	conserved	four-helices	conserved	
Musa x fennicae (*Musa siamensis* (male) x *Musa rosea* (female))	1522_type1	62.50	57.42	69.63	conserved	conserved	conserved	four-helices	conserved	
	1522_type2	62.50	57.42	69.63	conserved	conserved	conserved	four-helices	conserved	
	1522_type3	61.57	57.42	69.16	conserved	conserved	conserved	four-helices	conserved	
	1522_type4	61.11	57.42	69.16	conserved	conserved	conserved	four-helices	conserved	
	1522_type5	62.50	57.42	69.16	conserved	conserved	conserved	four-helices	conserved	
Musa siamensis	1534_type1	58.80	52.26	62.74	nt-12 'A'	nt-9 'T'	conserved	four-helices	conserved	
	1534_type2	**57.41**	**46.45**	**62.15**	**nt-2 'A'**	**nt-6 'A' nt-9 'T' nt-14 'T'**	**nt-8 'A' nt-9 'T'**	**not formed**	**conserved**	**pseudogene**
*Musa mannii**	1574									
Musa laterita	1575_type1	58.33	55.48	62.74	conserved	nt-9 'A'	conserved	four-helices	conserved	
	1575_type2	58.33	55.48	63.21	conserved	nt-9 'A'	conserved	four-helices	conserved	
Musa rubra	1590_type1	58.33	55.48	62.74	conserved	nt-9 'A'	conserved	four-helices	conserved	
	1590_type2	58.33	55.48	63.21	conserved	nt-9 'A'	conserved	four-helices	conserved	
	1590_type3	**59.26**	**59.63**	**65.26**	**nt-11 'T'**	**not present - deletion**	**not present - deletion**	**not formed**	**not formed**	**pseudogene**
Musa rosea x *siamensis*	1592_type1	61.29	57.42	68.69	conserved	conserved	conserved	four-helices	conserved	
	1592_type2	**60.65**	**54.84**	**66.05**	**conserved**	**conserved**	**conserved**	**not formed**	**conserved**	**pseudogene**
	1592_type3	**58.80**	**55.86**	**63.51**	**conserved**	**nt1 – nt3 'deletion' nt-9 'A'**	**conserved**	**four-helices**	**not formed**	**pseudogene**
Musa rosea (hybrid)	1598	63.59	57.42	68.37	conserved	conserved	conserved	four-helices	conserved	
Musa cf. *uranoscopos*	1532_type1	63.68	58.06	66.83	conserved	conserved	conserved	four-helices	conserved	
	1532_type2	63.68	58.06	66.34	conserved	conserved	conserved	four-helices	conserved	
	1532_type3	65.02	58.06	66.83	conserved	conserved	conserved	four-helices	conserved	
Musa violascens	1514	65.02	58.06	66.83	conserved	conserved	conserved	four-helices	conserved	
Musa lutea	1515	59.28	54.19	63.26	nt-11 'T' nt-12 'A'	conserved	nt-6 'T'	four-helices	conserved	
Musa beccarii var. *beccarii**	1516									
Musa campestris var. *sarawakensis*	1517	59.64	48.95	61.46	nt-11 'T' nt-12 'A'	nt-7 'T'	conserved	four-helices	conserved	
Musa monticola	1528	65.92	58.71	68.29	conserved	conserved	nt-3 'C	four-helices	conserved	
Musa beccarii var. *hottana*	1529	65.32	58.06	65.37	conserved	conserved	conserved	four-helices	conserved	
*Musa borneensis**	1531									
Musa campestris var. *limbangensis*	1535_type1	64.13	58.06	65.85	conserved	conserved	conserved	four-helices	conserved	
	1535_type2	64.57	58.06	66.83	conserved	conserved	conserved	four-helices	conserved	
	1535_type3	**56.50**	**50.32**	**59.51**	**nt-12 'A' nt-16 'A'**	**conserved**	**conserved**	**not formed**	**conserved**	**pseudogene**
Musa barioensis	1568_type1	61.88	56.13	62.80	nt-10 'G' nt-11 'A'	nt-9 'A'	nt-9 'T'	four-helices	conserved	
	1568_type2	63.77	56.13	62.80	nt-10 'G' nt-11 'A'	nt-9 'A'	nt-9 'T'	four-helices	conserved	
	1568_type3	65.92	58.06	68.29	conserved	conserved	conserved	four-helices	conserved	
	1568_type4	**60.99**	**54.19**	**63.77**	**conserved**	**conserved**	**conserved**	**not formed**	**conserved**	**pseudogene**

▲ Label of ITS type is based on the ITC accession number and is maintained in the phylogenetic trees. The consensus sequences of the ITS regions in heterogenous diploid and triploid species are labeled by “con1” means consensus type 1.

* Data not available.

Table 4: Nucleotide diversity of ITS1-5.8S-ITS2 regions in *Musa*

Section	Accession name	ITC code	Number of sequenced clones	Nucleotide diversity θ_π
Eumusa	*Musa itinerans* var. *xishuangbannaensis*	1526	31	9.202
	Musa itinerans var. *itinerans*	1571	30	5.724
	Musa yunnanensis	1573	46	27.513
Rhodochlamys	*Musa rubinea*	1518	30	24.021
	Musa x fennicae (*Musa siamensis* (male) x *Musa rosea* (female))	1522	75	8.155
	Musa siamensis	1534	30	25.644
	Musa rosea x *ornata*	1572	Data not available	
	Musa laterita	1575	32	5.452
	Musa rubra	1590	62	15.528
	Musa rosea x *siamensis*	1592	91	13.266
	Musa rosea (hybrid)	1598	67	7.303
Australimusa	*Musa* cf. *uranoscopos*	1532	75	5.214
Callimusa	*Musa violascens*	1514	37	3.667
	Musa lutea	1515	28	0.668
	Musa beccarii var. *beccarii*	1516	Data not available	
	Musa campestris var. *sarawakensis*	1517	32	2.054
	Musa monticola	1528	37	0.963
	Musa beccarii var. *hottana*	1529	32	12.695
	Musa borneensis	1531	Data not available	
	Musa campestris var. *limbangensis*	1535	42	17.336
	Musa barioensis	1568	55	20,506

4.1.3.1 SUPPORTING INFORMATION

Figure S1:

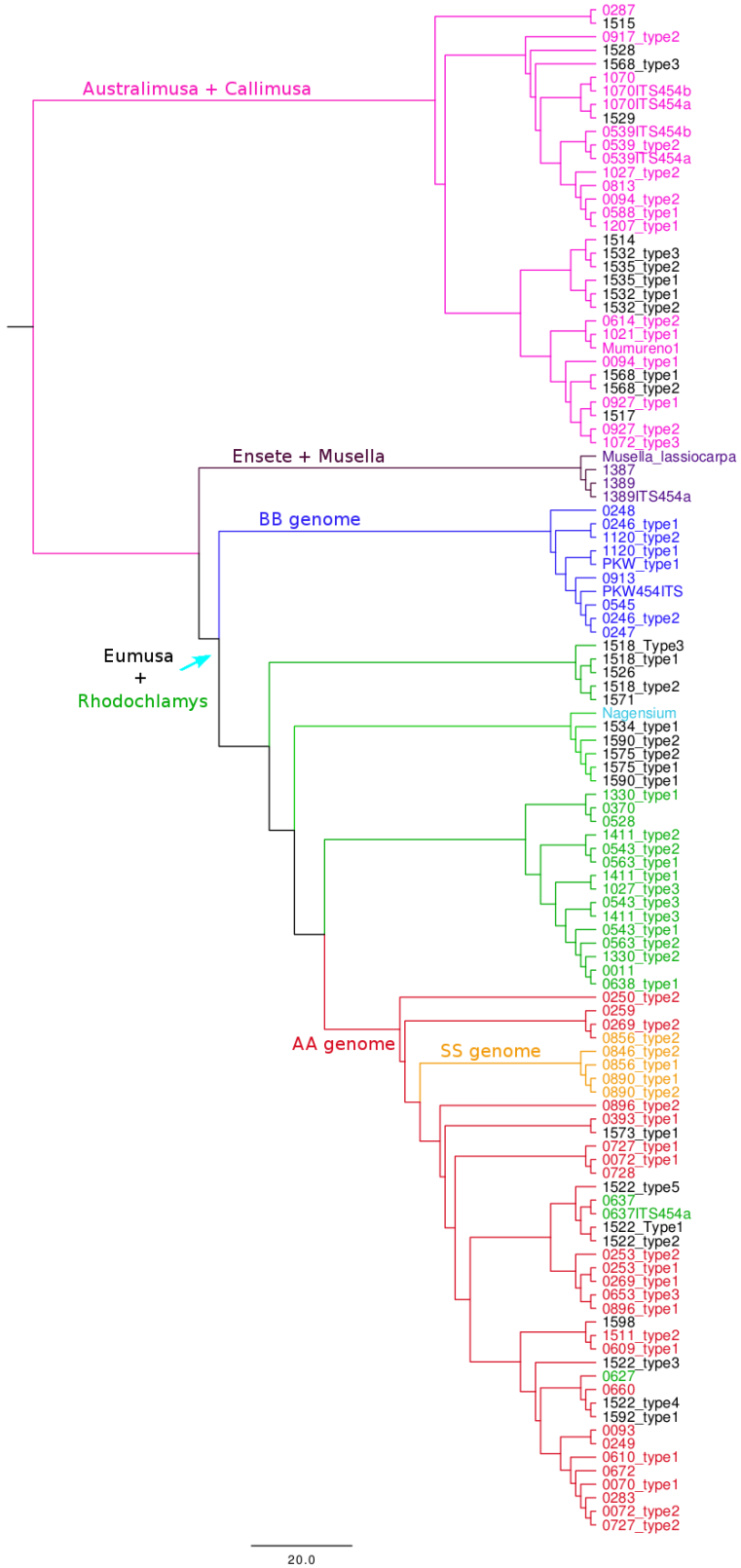
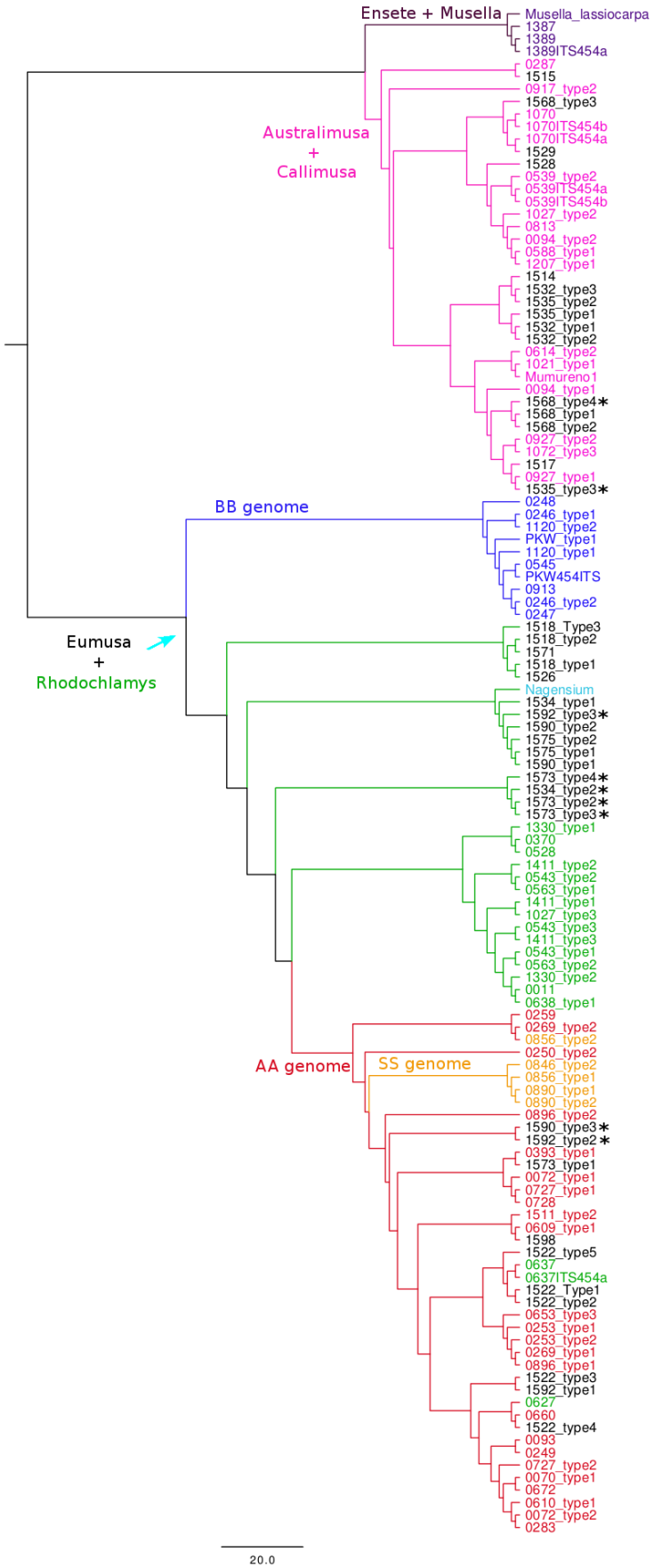


Figure S2:


Ensete + Musella
Musella_lassiocarpa
1387
1389
1389ITS454a
Australimusa
+
Callimusa
0287
1515
0917_type2
1568_type3
1070
1070ITS454b
1070ITS454a
1529
1528
0539_type2
0539ITS454a
0539ITS454b
1027_type2
0813
0094_type2
0588_type1
1207_type1
1514
1532_type3
1535_type2
1535_type1
1532_type1
1532_type2
0614_type2
1021_type1
Mumureno1
0094_type1
1568_type4*
1568_type1
1568_type2
0927_type2
1072_type3
1517
0927_type1
1535_type3*
BB genome
0248
0246_type1
1120_type2
PKW_type1
1120_type1
0545
PKW454ITS
0913
0246_type2
0247
Eumusa
+
Rhodochlamys
1518_Type3
1518_type2
1571
1518_type1
1526
Nagensium
1534_type1
1592_type3*
1590_type2
1575_type2
1575_type1
1590_type1
1573_type4*
1534_type2*
1573_type2*
1573_type3*
1330_type1
0370
0528
1411_type2
0543_type2
0563_type1
1411_type1
1027_type3
0543_type3
1411_type3
0543_type1
0563_type2
1330_type2
0011
0638_type1
0259
0269_type2
0856_type2
AA genome
SS genome
0250_type2
0846_type2
0856_type1
0890_type1
0890_type2
0896_type2
1590_type3*
1592_type2*
0393_type1
1573_type1
0072_type1
0727_type1
0728
1511_type2
0609_type1
1598
1522_type5
0637
0637ITS454a
1522_Type1
1522_type2
0653_type3
0253_type1
0253_type2
0269_type1
0896_type1
1522_type3
1592_type1
0627
0660
1522_type4
0093
0249
0727_type2
0070_type1
0672
0610_type1
0072_type2
0283
20.0

4.2 CHAPTERS IN BOOKS

4.2.1 Development of Physical Cytogenetic Maps for Bananas and Plantains

Hřibová E, Doleželová M, Němcová P, Čížková J, Schillerová L, Doležel J

In: Physical mapping technologies for identification and characterization of mutated genes to crop quality. Pp. 61-70, IAEA, Vienna, 2011.

[JČ, EH, JD concieved and designed the experiments. JČ, PN, EH, LS performed the experiments. JČ, EH, PN analyzed the data. JD, MD contributed reagents/materials/analysis tools. JČ, EH, JD wrote the paper.]

DEVELOMENT OF PHYSICAL CYTOGENETIC MAPS FOR BANANAS AND PLANTAINS

E. HRIBOVA, M. DOLEZELOVA, P. NEMCOVA,
J. CIZKOVA, L. SCHILLEROVA, J. DOLEZEL
Laboratory of Molecular Cytogenetics and Cytometry,
Institute of Experimental Botany,
Olomouc, Czech Republic

Abstract

The aim of the research project carried out under the research contract was to unravel the long-range molecular organization of banana (Musa spp.) chromosomes and develop a fine physical cytogenetic map of this crop. During the project execution, new cytogenetic markers were isolated and used to characterize the banana karyotype. The analysis of nucleotide sequence of the ITS region was used to characterize genetic diversity and establish phylogenetic relationships within the family Musaceae. Moreover, massively parellel 454 sequencing was used for low-depth sequencing of nuclear genome of M. acuminata 'Calcutta 4'. Most of all types of the repetitive elements in banana genome were characterized and a specific database of banana repetitive elements was created and used for analysis of sequence organization in already sequenced BAC clones. The 454 sequence data obtained in this work should facilitate annotation of nucleotide sequences during the ongoing banana genome sequencing project.

1. INTRODUCTION

Bananas (Musa spp.) are giant perennial herbs growing mainly in developing countries of tropics and subtropics. The so-called desert bananas are palatable when eaten raw and constitute a major export commodity second only to citrus in terms of the world fruit trade. For many developing countries, bananas provide an essential source of foreign exchange. Nevertheless, only about 13% of banana production enters the world trades. Most of bananas are grown by small farmers for local consumption and eaten raw, cooked (so-called plantains and cooking bananas) or fermented (called beer banana). In terms of production as well as its gross value, banana ranks fourth (after rice, maize and wheat) in the world. By providing the staple diet for millions, banana production plays important role in ensuring food security in many countries of Asia, Africa and Latin America. The production of bananas has been seriously threatened by an ever-increasing range of fungal, viral and insect diseases. In the absence of locally adapted resistant varieties, the diseases can only be controlled by extensive use of pesticides, which represents a considerable economic and environmental burden. Clearly, there is an urgent need for a wider diversity of genetically improved cultivars with disease resistance, increased yield, and adaptability to a wider range of environmental conditions.

Unfortunately, banana breeding has been complicated by the plant biology. While the wild bananas are diploid (2n=2x=22) and bear seeds, most of cultivated edible bananas are sterile triploid (2n=3x=33), produce fruits parthenocarpically and are multiplied vegetatively. Moreover, the origin of cultivated clones is not known. As a consequence, classical breeding endeavors achieved a limited success and no man-bred variety has been widely accepted until now. There is a great hope that mutation induction and genetic transformation will facilitate localization, isolation and transfer of genes controlling various characters, and aid in creation of new banana varieties resistant to diseases and with improved quality traits.

The objective of the research carried out under the research contract is to analyze the molecular organization of the banana (Musa spp.) genome, unravel the long-range molecular organization of its chromosomes, develop a fine physical cytogenetic map of banana and, if possible, integrate it with the existing genetic linkage map. The new information obtained, physical maps and cytogenetic markers should be used to characterize genetic diversity within the Musa genus and analyze karyotype changes in mutant banana stocks with altered chromosome numbers.

2. MATERIALS AND METHODS

2.1. Plant material

All banana plants were obtained as in vitro rooted plants from INIBAP Transit Centre (Katholieke Universiteit Leuven, Belgium). After transfer to soil, plants were maintened in the greenhouse.

2.2. Estimation of nuclear genome size

Approximately 50 mg of midrib was cut from a young Musa leaf and transferred to a glass Petri dish. About 10 mg of a young leaf of soybean (Glycine max L. cv. Polanka) with 2C = 2.5 pg DNA (Doležel et al., 1994) was added and served as an internal reference standard. The tissues were chopped simultaneously in 1 ml of Otto I buffer (0.1 M citric acid, 0.5 % v/v Tween 20; Otto, 1990). Crude suspension of isolated nuclei was filtered trough a 50 µm nylon mesh. Nuclei were then peleted (300 g, 5 min.), resuspended in 200 µl Otto I and incubated for 1hour at room temperature. Finally, 600 µl Otto II buffer (0.4 M Na_2HPO_4; Otto, 1990), supplemented with 50 µg/ml RNase and 50 µg/ml propidium iodide (PI), was added. Samples were analysed using Partec PAS flow cytometer (Partec GmbH, Münster, Germany) equipped with 488-nm argon laser. The gain of the instrument was adjusted so that peak representing soybean G_1 nuclei appeared approximately on channel 200 on histogram of relative fluorescence intensity when using 512-channel scale. About 5,000 nuclei were analysed at rate 10 – 25 nuclei/sec. Three plants were measured per accession.

2.3. BAC library screening for "low-copy" clones

Genomic BAC library of Musa acuminata cv. Calcutta 4 (C4BAM) was doubly spotted on one 22 × 22-cm Hybond N+ filter (AP Biotech) with the GeneTAC™ G3 workstation (Genomic Solutions). Putative "low-copy" BAC clones were selected based on weak Southern hybridization signals with genomic DNA of 'Calcutta 4', which was labeled with digoxigenin. Hybridization was done at 65°C overnight in 5 mL of hybridization buffer (5 × SSC, 2% blocking reagent 0.1% sodium N-lauroylsarcosin, 0.02% SDS) containing 400 ng of labeled probe. Stringency washing was done by incubation twice in 200 mL of 0.1 × SSC, 0.1% SDS buffer at 68°C. Hybridization signals were detected using anti-digoxigenin-AP (Roche Applied Science) and visualized after incubation with CDP Star chemiluminescent substrate (Roche Applied Science). BAC clones with very weak or no hybridization signals were selected for further work.

2.4. Selection of BAC clones after screening BAC pools with microsatllite markers

After growing overnight in 2YT medium supplemented with chloramphenicol (12.5 µg/mL) in 384-well plates, BAC clones from each of the 384-well plates were pooled, pelleted and resuspended in 4 mL of TE (10mM Tris, 1mM EDTA) buffer. Bacterial suspensions were lysed at 95°C for 30 min, pelleted at 3,000g for 60 min and supernatant was diluted 25-fold by deionized water for PCR reaction. Primers specific for the sequences of microsatelite markers available in GenBank (GB codes: X87258 - X87265 and X90740 - X90750, Lagoda et al., unpublished) were designed using Primer3 software (Rozen and Skaletsky 2000). The PCR reaction mix (10 µL) consisted of 2 µL template pooled BAC DNA, 1× PCR buffer, 1.5 mM $MgCl_2$, 0.2 mM dNTPs, 1.25 µM primers, and 0.5U of AmpliTaq DNA polymerase (Roche, Mannheim, Germany). PCR reaction was performed as follows: 30 cycles of 30 sec at 94°C, 30sec at 55°C, and 40 sec at 72°C, and final extension at 72°C for 10 min. Presence of PCR products was checked by electrophoresis on 1.5% agarose gel. Positive BAC clones were selected for further work.

2.5. Characterization of selected BAC clones for insert size

Individual BAC clones were cultured overnight in 1.5 mL LB medium supplemented with 12.5 µg/mL chloramphenicol. BAC DNAs were isolated using standard alkaline lysis method and digested to completion with NotI. DNA fragments were size separated by pulsed field gel electrophoresis in 1%

Gold SeaKem agarose (BMA) gel at 6V/cm, with a 1 – 40 s switch time ramp, angle 120°, for 14 hr at 14.0°C in 0.5 × TBE buffer. The inserts of BAC clones was found to range from 60 kbp to 120 kbp.

2.6. BAC subcloning

In order to eliminate the negative effect of disperse repeats on localization of BAC clones, low-copy subclones wee isolated from selected BAC clones. DNA of 'low copy' BAC clones was isolated using the Large-Construct Kit (Qiagen) and physically fragmented using a HydroShear DNA Shearing Device (GeneMachines). Fragments of 7 -10 kb were ligated using TOPO Shotgun Subcloning Kit (Invitrogen Life Technologies). Ligation mixtures were transformed into Electrocompetent TOP10 E. coli (Gibco BRL). 384 subclones from each of BAC clones were ordered in 384-well plates filled with the freezing medium (Woo et al., 1994), incubated at 37°C overnight and stored at -80°C. All BAC subclones were spotted onto a Hybond N+ filter and screened with labeled genomic DNA. Five subclones from each BAC showing weak signal after hybridization with labeled genomic DNA were randomly chosen for mapping using fluorescence in situ hybridization on mitotic chromosomes of M. acuminata 'Calcutta 4'.

2.7. Construction of a short insert DNA library enriched for repetitive DNA

Genomic DNA was extracted from fresh cigar leaves of Musa acuminata 'Calcutta 4', and sheared by sonication to fragments of 300-400 bp. In order to isolate genomic DNA fraction enriched in highly repetitive DNA, DNA fragments were denatured and reannealed to $C_0t = 0.1$ and $C_0t = 0.05$. After the reassocation, double-stranded DNA was separated using hydroxyapatite chromatography and cloned by blunt-end ligation to pBluescript II SK^+. The ligation mixture was used to transform Escherichia coli XL1-Blue MRF cells. C_0t libraries were constructed from 2,688 clones from $C_0t = 0.05$ fraction, and 4,608 clones from $C_0t = 0.1$ fraction. All clones from both C_0t libraries were spotted on two 8 × 12-cm Hybond N+ filters (AP Biotech) with the GeneTAC™ G3 workstation (Genomic Solutions). The filters were hybridized with probes for 45S and 5S ribosomal DNA (Radka1 and Radka2) and other known repetitive DNA sequences isolated from banana (Valárik et al., 2002) using the AlkPhos Direct Labeling and Detection System (Amersham). Hybridization signals were visualized after incubation with CDP Star chemiluminescent substrate (Roche Applied Science). 6,759 clones were identified as negative, indication high proportion of non-characterized and potentially repetitive DNA sequences.

2.8. Sequence analysis

DNA clones selected from C_0t libraries were sequenced at the Department of Plant Sciences, The University of Arizona (Tucson) and at The Institute for Genomic Research (Rockville). Sequence data were compared using the Dotter software (Sonnhammer and Durbin, 1995) and searched for homology to sequences in the GeneBank database using BLAST 2.0.2 (Altschul et al., 1997).

2.9. Genomic organization of tandem repeats and their copy number

Aliquots of genomic DNA of M. acuminata 'Calcutta 4' were digested using eight restriction endonucleases, RsaI, HaeIII, MseI, AluI, EcoRI, SmaI, SacI and DraI. BAC clones carrying 45S rDNA that were selected from the MA4 BAC library were digested using DraI, SacI and TaqI. Digested genomic DNA and BAC DNA were size-fractionated by 1.5% agarose gel electrophoresis, and transferred onto Hybond N+ nylon membrane (Amersham). Clones from C_0t-0.05 library containing tandemly organized sequence units (C427, 2F10, 4E2, C444 and 7D20) revealed by Dotter software were labeled using alkaline phosphatase (AlkPhos Direct Kit, Amersham) and used as hybridization probes of genomic DNA. The probes for C427, 2F10, 4E2, 18S rDNA and 26S rDNA respectively, were used for hybridization with digested DNA of BAC clones. The copy number of newly identified repetitive DNA sequences was estimated for the genomes of M. acuminata 'Calcutta 4' (AA genome) and M. balbisiana 'Tani' (BB genome). Serial dilutions of genomic DNA and PCR products of isolated repeats used as standards were dot-blotted onto Hybond-N+ membranes (Amersham). The PCR products were labeled using alkaline phosphatase and used as hybridization probes. Dots of genomic DNA and standards that gave the same intensity of hybridization signals were

identified after visual inspection. Copy numbers of individual probes were estimated assuming that 1pg of genomic DNA equals 0.978 x 10^9 bp (Doležel et al., 2003).

2.10. Sequence analysis of the 45S rDNA locus and analysis of genetic diversity in Musa

DNA region containing ITS1 and ITS2 was amplified using specific primers ITSL and ITS4 (Nwakanma et al., 2003). Primer ITSL – 5´-TCG TAA CAA GGT TTC CGT AGG TG-3´ (Hsiao et al., 1994) is complementary to 18S rDNA and primer ITS4 – 5´-TCC TCC GCT TAT TGA TAT GC-3´ (White et al. 1990) anneals to 26S rDNA. The PCR reaction mix (25 µL) consisted of 10 ng of genomic DNA, 1× PCR buffer, 1.5 mM $MgCl_2$, 0.2 mM dNTPs, 1.25 µM primers, and 0.25U of AmpliTaq DNA polymerase (Finnzymes, Finland). PCR reaction was performed as follows: 30 cycles of 50 sec at 94°C, 50sec at 52°C, and 50 sec at 72°C, and final extension at 72°C for 10 min. Presence of PCR products was checked by electrophoresis on 1.5% agarose gel. PCR products were clean up using PCR Rapid Kit (Invitek, Germany) and used for direct sequencing and for cloning of the PCR products, respectively.

In the first step, PCR products of forty-eight ITS regions from different diploid banana genotypes were used for direct sequencing using dideoxy-chain reaction. Thirty-three genotypes produced readable ITS sequence with no polymorphism and fifteen genotypes produced unreadable highly polymorphic ITS sequences. PCR products of ITS region in highly polymorphic genotypes and hybrid banana clones were ligated into TOPO vector (Invitrogene, USA) and transformed into electrocompetent cells and four recombinant clones per individual genotypes were sent for sequencing. Nucleotide sequences were assembled using the Staden Sequence Analysis Package (Staden 1996) and further analyzed in BioEdit (molecular tools: Clustal W and DNA Maximum Likelihood program, Neighbor program, http://www.mbio.ncsu.edu/BioEdit/bioedit.html). Secondary structures of ITS regions were analyzed using Mfold (Zuker 2003).

2.11. Large-scale sequencing

Intact nuclei of M. acuminata 'Calcutta 4' were isolated and used as template DNA for 454 sequencing. Sequencing was provided on GS FLX system and resulted in about 400 000 reads with average sequence length 200 bases (~ 100 MB). These amount of sequencing data represets 15% of nuclear genomes of M. acuminata 'Calcutta 4' (1C = 620 Mbp).

Duplicated sequences, specific linkers or primers used in sequencing reaction were removed from the dataset under the homology search using BLAST (Altschul et al., 1997). BLAST program was also used for identification of other known banana DNA sequences deposited in the GenBank. Sequencing reads were assembled into contigs using cap3 program. The repeat reconstruction was done using TIGR Gene Indices clustering tools (TGICL), which have been optimized for 454 dataset (Macas et al., 2007). All contigs were characterized by calculating their average read depth and genome representation, which were used for estimation of the copy number of assembled contigs (Macas et al., 2007). All repetitive clusters that represent more than 0.01% of the nuclear genome of banana will be further characterized. Tandem Repeats Finder (Benson, 1999) and TRAP (Sobreira et al., 2006) programs were used for identification of tandem organized repetitive units.

2.12. Cytogenetic mapping using FISH and PRINS

Metaphase spreads were prepared from root tips of M. acuminata 'Calcutta 4' according to Doleželová et al. (1998). For FISH, inserts of selected DNA clones were labeled with digoxigenin-11-dUTP (Roche). FISH probe for 45S rDNA was obtained by labeling a Radka1 DNA clone containing 26S rRNA gene (Valárik et al., 2002) with digoxigenin-11-dUTP or biotin-16-dUTP (Roche). 5S rDNA probe (Radka 2) was prepared from 400 bp insert of a part of the 5S rRNA gene (Valárik et al., 2002). The sites of digoxigenin- and biotin-labeled probe hybridization were detected using anti-digoxigenin fluorescein (Roche) and streptavidine conjugated to Cy3 (Sigma), respectively.

PRINS was done according to Kubaláková et al. (1997). Briefly, the reaction mixture consisted of 0.1 mM dATP, dCTP, dGTP, and 0.01 mM Alexa Fluor 488-5-dUTP, 0.017 mM dTTP, 2.5 mM $MgCl_2$ and 3U/40µl of Taq polymerase (Finnzymes) in 1 x PCR buffer. Synthetic oligonucleotides specific for tandem repeats C427, 4E8 and 2F10, respectively were designed using the Primer3 software (Rozen and Skaletsky, 2000) and were used as primers (Table 1) at 1µM concentration. Temperature profile of the reaction consisted of denaturation at 92ºC for 1 min, primer annealing at 58ºC for 50 s and extension at 72ºC for 1 min.

Following FISH or PRINS, the preparations were counterstained with DAPI (0.2 µg/ml) and mounted in Vectashield antifade solution (Vector Laboratories). The preparations were evaluated using Olympus AX70 microscope and the images of DAPI, fluorescein and Cy3 fluorescence were acquired separately with a b/w CCD camera, which was interfaced to a PC running the ISIS software (Metasystems). The images were superimposed after contrast and background optimization.

3. RESULTS AND DISCUSSION

3.1. Characterization of a set of Musa species for genome size and the number and genomic distribution of rDNA loci

Nuclear DNA content and genomic distributions of 5S and 45S rDNA were examined in nineteen diploid accessions of the genus Musa representing its four sections Eumusa, Rhodochlamys, Callimusa and Australimusa, and in Ensete gilletii, which was outgroup in this study. The results showed that 2C nuclear DNA content ranged from 1.130 to 1.561 pg in accessions representing the genus Musa. E. gilletii, which was outgroup in this work, had 2C DNA content of 1.210 pg. Within the section Eumusa, the lowest nuclear DNA content was found in both accessions of M. balbisiana (2C = 1.130 and 1.133 pg). The highest DNA content was found in M. schizocarpa (2C = 1.377 pg). An intermediate 2C DNA content (1.224 - 1.266 pg) was observed in M. acuminata. The differences between the three species of Eumusa were statistically significant. Although the differences between the accessions of M. acuminata were small (max. 3.4 %), some of them were statistically significant as well.

Smaller interspecific variation of 2C DNA content was observed within the section Rhodochlamys (1.191 – 1.299 pg) but differences between some species were still statistically significant. The smallest range of nuclear DNA content variation (7.8 %) was found between the species of Australimusa, with 2C value ranging from 1.435 to 1.547 pg. The highest 2C nuclear DNA content in this study (1.561 pg) was found in M. beccarii, the only representative of the section Callimusa in this study. Bonferroni's multiple comparison test revealed 10 groups distinguishable according to relative nuclear DNA content, three of them being represented by only one accession (M. schizocarpa, M. ornata and M. textilis). Five groups comprised representatives of at least two different sections; two of them involved accessions belonging to Musa and Ensete.

FISH with the probe for 45S rDNA revealed distinct hybridization sites on one pair of nucleolar organizing chromosomes in all accessions of Eumusa and Australimusa. In Eumusa, the sites of hybridization coincided with secondary constrictions of both chromosomes of the homologue pair. On the other hand, secondary constriction was not detectable on one of the homologues in all four accessions of Australimusa, indicating only one active nucleolar organizer region. A variable number of 45S rDNA sites were observed in the section Rhodochlamys. While three accessions possessed two sites (one chromosome pair), two accessions representing M. ornata were characterized by four 45S loci (two chromosome pairs). However, two additional loci were detected as very weak hybridization signals. They were located in the terminal position and did not coincide with secondary constrictions. M. beccarii was characterized by six sites of 45S rDNA genes (three chromosome pairs). Among the six strong hybridization clusters, only two coincided with the secondary constriction. The highest number (four pairs) of 45S rDNA loci was observed in E. gilletii. The intensity of the signals on different chromosome pairs differed, indicating a difference in the copy number of the 18S-5.8S-26S rRNA genes. A significantly larger variation was observed in the number of 5S rDNA loci. In the

Eumusa section, the number of 5S rDNA sites ranged from four to eight, five loci were observed in the seed-sterile clone 'Pisang Mas'. All Rhodochlamys accessions comprised two pairs of chromosomes bearing 5S rRNA genes except M. velutina, which had three chromosome pairs bearing 5S rDNA. In this case, two sites were major and four sites were minor, with significantly lower copy number. In contrast to a large variation in the number of 5S rDNA loci in other sections, all Australimusa accessions possessed four 5S rDNA sites. M. beccarii (Callimusa) and E. gilletii contained 5S rRNA gene clusters on five and six chromosomes, respectively. Two of the five 5S rDNA sites in M. beccarii were localized at terminal positions of chromosomes with interstitially localized signals of 45S rDNA.

3.2. Selection of "low-copy" BAC clones based on hybridization with genomic DNA

Inserts of selected BAC clones were labeled with digoxigenin and used as probes for fluorescence in situ hybridization (FISH) on mitotic chromosomes of M. acuminata 'Calcutta 4' (Doleželová et al., 1998). Despite the expectations, even the use of pre-selected BAC clones, presumably containing least amounts of repetitive DNA, did not result in single locus signals. FISH with probes prepared from inserts of the BAC clones resulted in hybridization signals along all chromosomes. Alternatively, multiple sites of hybridization were observed resembling hybridization with tandem organized repeats. These results, together with the results obtained previously during the third year of project execution indicate that the pre-screening the BAC library with the genomic DNA and choosing for the clones with weak signals is not efficient to generate chromosome-specific cytogenetic markers. In order to eliminate the negative effect of disperse repeats on localization of BAC clones, low-copy subclones were isolated from selected BAC clones. However, none of the selected subclones could be localized. This could be due to the size of inserts which was probably too small to be localized on condensed mitotic chromosomes of banana. Therefore, another strategy based on selection of marker-tagged BAC clones was tested.

3.3. Selection of BAC clones after screening BAC pools with microsatllite markers

As the strategy based on screening BAC library with the probe for the genomic DNA and selecting clones with the weak signals was not efficient to generate chromosome-specific cytogenetic markers, we tested a different strategy based on PCR screening of the BAC pools with the primers specific for the sequences of microsatelite markers available in the GenBank (GB codes: X87258 – X87265 and X90740 – X90750). The positive BAC clones were isolated and used as probes for FISH on mitotic chromosomes of M. acuminata 'Calcutta 4'. Twenty different BAC clones were isolated and used as probes for cytogenetic mapping. Out of them, six BAC clones gave weak discrete signals on one pair of mitotic chromosomes of M. acuminata 'Calcutta 4' (unpublished).

In this work, the limited spatial resolution due to small size of compact mitotic metaphase chromosomes hampered a detailed analysis of genomic distribution of DNA sequences by FISH. In order to overcome this bottleneck, we have optimized a protocol for preparation of pachytene spreads in Musa. Preliminary results confirmed a possibility to perform high-resolution FISH o these preparations. The next step will be to localize selected BAC clones by FISH on pachytene chromosomes of M. acuminata 'Calcutta 4'.

3.4. Isolation of new repetitive DNA sequences and analysis of their genomic distribution

Knowledge of repetitive DNA can facilitate mapping of important traits, phylogenetic studies and the tandem organized DNA repeats are useful cytogenetic markers. In this part of the work, we used reassociation kinetics to isolate and characterized the highly repetitive part of the banana genome. The so-called Cot-based cloning and sequencing approach was used to prepare two low-Cot libraries enriched on the highly repetitive DNA sequences. 614 of the Cot-clones were sequenced (sequences are deposited in the GenBank, GB codes: ED827164 – ED827777) and searched for homologies with the sequences deposited in the GenBank using BLASTn. Out of the 614 sequenced clones, 48% represented novel undescribed sequences. Up to 24.1% homologous sequenced clones showed

similarity to different types of retrotransposons (Fig. 1). Dot-plot analysis revealed that 14% of the sequenced clones contain tandem organized repeats. Out of these, three clones that carried different types of tandem organized units were selected for copy-number estimation and for cytogenetic localization using PRINS. One clone showed strong signal in the secondary constriction and weak signals in the centromeric regions of additional chromosomes. The other tandem organizes sequences were localized also in the secondary constriction and showed additional cluster signals on the other chromosomes of M. acuminata 'Calcutta 4' (Hřibová et al., 2007).

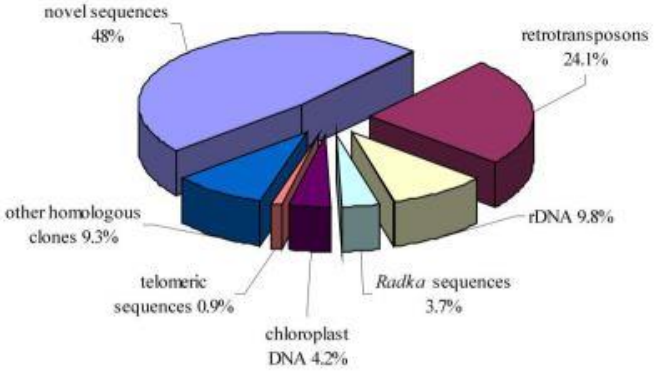


Fig. 1: Homology of DNA clones isolated from the $C_0t \leq 0.05$ library to sequences deposited in GenBank. For details see *Hřibová et al. (2007).*

3.5. Localization of the new tandem repeats with respect to 45S rDNA

To confirm that the three tandem repeats constitute a part of the 45S rDNA unit that is localized in the secondary constriction, genomic BAC library of M. acuminata 'Calcutta 4' (MA4, Vilarinhos et al. 2003) was screened was doubly spotted the Hybond N+ filter (AP Biotech) with the GeneTAC^{TM} G3 workstation (Genomic Solutions). BAC clones carrying 45S rDNA were selected based on Southern hybridization signals with probes for 18S rDNA and 26S rDNA clones, which were labeled using alkaline phosphatase (AlkPhos Direct labeling, Amersham). Three BAC clones carrying 45S rDNA units (MA4_1P13, MA4_1J14 and MA4_1B22) selected and were digested using DraI, SacI and TaqI restriction endonucleases, size fractionated on the agarose gel and blotted onto a nylon membrane (Hybond N+). The membranes with restricted BAC clone were hybridized separately with the probes for C427, 4E2, 2F10, as well as 18S rDNA and 26S rDNA clones. Southern hybridization confirmed that the clones C427, 2F10 and 4E2 hybridized with the same band of the restricted BAC clones and indicated that the clones were parts of the 45S rDNA unit localized in the secondary constriction. For details see Hřibová et al. (2007).

3.6. Karyotype analysis

Cytogenetic analysis of the plant nuclear genome is one of the basic tools used in phylogenetic analysis. In order to characterize karyotype diversity in Musa, we elected to analyze two varieties of M. beccarrii (var. beccarii and var. hottana), section Callimusa. As described previously, M. beccarii var. beccarii contain of 18 chromosomes (2x=2n=18) with the 2C nuclear DNA content 1.561 pg. We have found that the karyotype of M. beccarii var. hottana consits of 18 chromosomes (2x=2n=18) with different chromosome sizes and one chromosome pair carrying secondary constriction. As described previously, 45S rDNA is localized in the NOR and also in the centre of the two other chromosome pairs in the M. beccarii var. beccarii (Fig. 2).

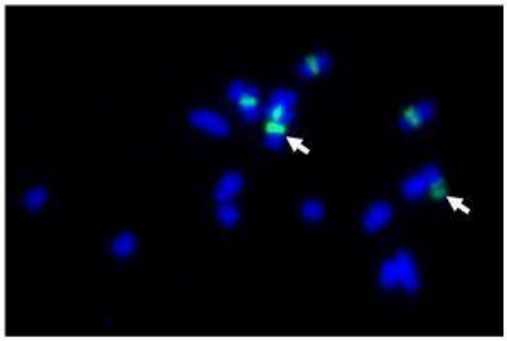


Fig. 2: Localization of 45S rDNA on mitotic chromosomes of M. beccarii var. beccarii. Probe was labelled by digoxigenin (green signals). The secondary constrctions are marked by arrows.

3.7. Evaluation of genetic diversity using repetitive DNA clones

Out of the set of 615 repetitive DNA clones that wee isolate from the High-Cot DNA library of M. acuminata Calcutta 4, a total of 576 repetitive DNA clones that represent various types of DNA repeats were selected. They were sent to the laboratory of Kornel Burg (Austrian Research Centers, Seibersdorf Research GmbH, Seibersdorf, Austria), where they are being used to assemble a boutique microarray.

3.8. Sequence analysis of the 45S rDNA locus and analysis of genetic diversity in Musa

The taxonomy and phylogenetic relationships within the genus Musa have never been fully resolved and remain a subject of debate. Recently, the internal transcribed spacer (ITS) region of the 18S-5.8S-26S nuclear rDNA has been utilized as a marker for phylogenetic analysis in many plant families. In this part of the research project, we focused on the ITS1-5.8S-ITS2 region of nuclear ribosomal DNA. The nucleotide sequences of the ITS regions were edited using BioEdit software and a phylogenetic tree was constructed based on the neighbor joining method. The tree rooted on Zingiber spp. supports the genus Musa as monophyletic group that is separated from genus Ensete and Musella. The genus Musa is devided into two distinct clades: a clade of Callimusa and Australimusa and a clade of Eumusa and Rhodochlamys, where the A genome is strictly separated from the B genome (unpublished data).

In order to obtain further insights into the structure of the 45S rDNA locus in Musa and isolate additional molecular markers suitable for analysis of genetic diversity, we selected three BAC clones from M. acuminata 'Calcutta 4', M. acuminata 'Tuu Gia' and M. balbisiana 'Pisang Klutug Wulung' carrying 45S rDNA. These clones were fully sequenced and the sequence data are now being analyzed. We expect that markers developed from the external spacer will be useful for the analysis of genetic diversity in Musa as well as to study genomic constitution in interspecific hybrids.

3.10. Large-scale sequencing

In this project, we used the 454 method for sequencing nuclear genome of M. acuminata 'Calcutta 4'. In one sequencing run provided by GS FLX system (Roche), we obtained almost 100 Mb of nucleotide sequence that represents 16% of the M. acuminata 'Calcutta 4' genome. Until now, this is the largest amount of genomic sequence data available for Musa. Based on the experience with other species (Macas et al. 2007) we expect that this amount of sequence data will make it possible to capture most of the repetitive DNA in the genome of M. acuminata. The data are currently being analyzed using different bioinformatics tools. Various types of mobile elements and new tandem organized repeats were classified and characterized for copy number and genomic distribution (Fig. 3). The 454 sequence data will be also very important for annotation of nucleotide sequences that will be obtained during the ongoing banana sequencing project.

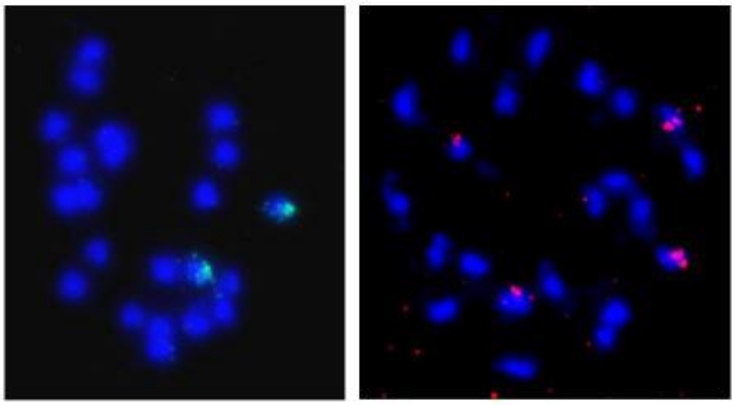


Fig. 3: Localization of tandem repeat CL18 (left) and CL33 (right) on mitotic chromosomes of M. acuminata 'Calcutta 4'.

REFERENCES

[1] ALTSCHUL, S.F., T.L. MADDEN, A.A. SCHAFFER, J.H. ZHANG, W. MILLER and D.J. LIPMAN (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acid Res 25: 3389-3402.

[2] BENSON, G. (1999). Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acid Res 27: 573-580.

[3] DOLEŽEL, J., J. BARTOŠ, H. VOGLMAYR, J. GREILHUBER (2003). Nuclear DNA content and genome size of trout and human. Cytometry 51: 127-128.

[4] DOLEŽELOVÁ, M., M. VALÁRIK, R. SWENNEN, J.P. HORRY and J. DOLEŽEL (1998). Physical mapping of the 18S-25S and 5S ribosomal RNA genes in diploid bananas. Biol. Plant. 41: 497-505.

[5] DOVER, G. (1982). Molecular drive: a cohesive mode of species evolution. Nature. 299:111-117.

[6] HSIAO, C., N.J. CHATTERTON, K.H. ASAY, K.B. JENSEN (1994). Phylogenetic relationships of 10 grass species: an assessment of phylogenetic utility of the internal transcribed spacer region in nuclear ribosomal DNA in monocots. Genome 37:112-20.

[7] KUBALÁKOVÁ, M., J. MACAS and J. DOLEŽEL (1997). Mapping of repeated DNA sequences in plant chromosomes by PRINS and C-PRINS. Theor. Appl. Genet. 94: 758-763.

[8] MACAS, J., P. NEUMANN, A. NAVRÁTILOVÁ (2007). Repetitive DNA in the pea (Pisum sativum L.) genome: comprehensive characterization using 454 sequencing and caomparison to soybean and Medicago truncatula. BMC Genomics 8: 427.

[9] NWAKANMA, D.C., M. PILLAY, B.E. OKOLI, A. TENKOUANO (2003). PCR-RFLP of the ribosomal DNA internal transcribed spacers (ITS) provides markers for the A and B genomes in Musa L.

[10] ROZEN, S. and H.J. SKALETSKY (2000). Primer3 on the WWW for general users and for biologist programmers. In: Krawetz S, Misener S (eds). Bioinformatics Methods and Protocols: Methods in Molecular Biology: 365-386.

[11] SOBREIRA, T.J., A.M. DURHAM, A. GRUBER (2006). TRAP: automated classification, quantification and annotation of tandemly repeated sequences. Bioinformatics. 22:361-362.

[12] STADEN, R. (1996). The Staden sequence analysis package. Mol Biotechnol. 5:233-

41.
[13] VALÁRIK, M., H. ŠIMKOVÁ, E. HŘIBOVÁ, J. ŠAFÁŘ, M. DOLEŽELOVÁ and J. DOLEŽEL (2002) Isolation, characterization and chromosome localization of repetitive DNA sequences in bananas (Musa spp.). Chromosome Research 10: 89-100.
[14] VILARHINOS, A.D., P. PIFFANELLI, P. LAGODA, S. THIBIVILLIERS, X. SABAU, F. CARREEL, A. D´HONT (2003). Construction and characterization of a bacterial artificial chromosome library of banana (Musa acuminta Colla). Theor Appl Genet 106: 1102-1106.
[15] WHITE, T.J., T. BRUNS, S. LEE, J. TAYLOR (1990). Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenies. In: Innis MA, Gelfand DH, Sninsky JJ, White TJ (eds) PCR protocols: a guide to methods and applications. Academic Press, San Diego, California, USA, pp 315–322.
[16] ZUKER, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res. 31:3406-3415.

4.3 PUBLISHED ABSTRACTS

4.3.1 Phylogenetic Relationships in the Family *Musaceae* Based on the Genic Sequences, Sequence of the ITS1-5.8S-ITS2 Region and DArT Markers

Němcová P, Hřibová E, Valárik M, Čížková J, Schillerová L, Kilian A, Doležel J

Global Perspectives on Asian Challenges. p. 18. International Society for Horticultural Science/ProMusa, Guangzhou, 2009.

[Oral presentation]

Phylogenetic Relationships in Family *Musaceae* Based on the Genic Sequences, Sequence of the ITS1-5.8S-ITS2 Region and DArT Markers

P. Němcová[1], E. Hřibová[1], M. Valárik[1], J. Čížková[1], L. Schillerová[1], A. Kilian[2] and J. Doležel[1]

[1]Laboratory of Molecular Cytogenetics and Cytometry, Institute of Experimental Botany, Olomouc, Czech Republic
[2] Diversity Arrays Technology, Canberra, Australia

Despite the socio-economic importance of banana (*Musa* spp.), phylogenetic relationships within the family *Musaceae* as well as the classification of subspecies and clones remain subject to debate. Different types of molecular markers have been used until now with various degrees of success and a detailed picture is still lacking. An unexplored opportunity is to compare DNA sequences of a large set of genes across the *Musaceae*. In this work, we searched banana ESTs for homology to single-copy rice genes and hybridized candidates with banana genomic DNA. During the first phase of the project, sequences of 24 candidate genes from 14 carefully selected species (ITC collection, Belgium) representing maximal diversity within the *Musaceae* were analyzed. The sequence analysis confirmed previously reported observations on clear divergence of *Musa, Ensete* and *Musella* as monophyletic groups. Moreover, within the *Musa* genus clade, section *Eumusa* clustered with *Rhodochlamys*, clearly separated from the cluster of *Australimusa* and *Callimusa*. These findings are in concordance with the previously reported close relation of these sections. The significance of our results was further supported by comparison with the results of nucleotide sequence analysis of ITS1-5.8S-ITS2 region in selected *Musaceae* species as well as analysis with DArT markers. Although each of the methods used is based on a different principle, data that were collected and analyzed supported the same concept of evolutionary relationship within the *Musaceae* species. We expect that a detailed analysis of the sequence data from higher number of genes and extension to other members of the family will clearly reveal phylogenetic relationships within the *Musaceae* and provide markers for unambiguous identification of subspecies and clones. This work has been supported by Grant Agency of the Academy of Sciences of the Czech Republic (grant award no. IAA600380703).

4.3.2 Physical Mapping of the Banana (*Musa* spp.) Genome Using Microdissected Chromosomes

Čížková J, Černohorská H, D'Hont A, Hřibová E, Rubeš J, Doležel J

Abstracts of the International Conference „Society for Experimental Biology Annual Meeting". P. 275. Prague, 2010.

[Poster presentation]

Physical Mapping of the Banana (*Musa* spp.) Genome Using Microdissected Chromosomes

J. Čížková[1], H. Černohornská[2], A. D'Hont[3], E. Hřibová[1], J. Rubeš[2], J. Doležel[1]

[1] Laboratory of Molecular Cytogenetics and Cytometry, Institute of Experimental Botany, Sokolovska 6, CZ-77200 Olomouc, Czech Republic
[2] Laboratory of Genetics and Reproduction, Veterinary Research Institute, Hudcova 70, CZ-62132 Brno, Czech Republic
[3] CIRAD, UMR 1098 DAP, Avenue Agropolis, 34398 Montpellier cedex 5, France

Bananas and plantains are giant perennial tropical plants within the genus *Musa*. Most of cultivated bananas are seed sterile diploid, triploid or tetraploid clones derived from interspecific and intersub-specific crosses between two wild diploid (2n = 22) species *M. acuminata* and *M. balbisiana*. Despite the socio-economic importance of this crop and the need of improved varieties, there is a lack of knowledge on genome structure and shortage of reliable genetic maps. Observation of chromosome pairing in meiosis suggests that banana sub-species and thus banana cultivars differ by translocations and inversions. The chromosomal rearrangements in parents of mapping populations cause segregation distortion, result in pseudo-linkages and hamper ordering of markers into linkage groups. Together with difficulties in establishing mapping population, these features hamper genetic mapping in *Musa*. One possibility to verify the position of markers in the genome is to locate them using PCR on isolated chromosomes. As it is not possible to sort *Musa* chromosomes by flow cytometry, we have used laser microbeam microdissection and catapulting to isolate single chromosomes from double haploid *M. acuminata* 'Pahang'. DNA of isolated chromosomes was amplified using linker adapter PCR and preliminary results indicate representative amplification even if only one copy of chromosome is used as a template. The work is in progress to verify linkage groups of the 'Pahang' genetic map using a set of SSR and DArT markers and to study chromosomal translocations among different subspecies and clones of *Musa*.

PHYSICAL MAPPING OF THE BANANA (*MUSA* SPP.) GENOME USING MICRODISSECTED CHROMOSOMES

JANA ČÍŽKOVÁ[1], HALINA ČERNOHORSKÁ[2], ANGELIQUE D'HONT[3], EVA HŘIBOVÁ[1], JIŘÍ RUBEŠ[2], JAROSLAV DOLEŽEL[1]

[1] Laboratory of Molecular Cytogenetics nad Cytometry, Institute of Experimental Botany, Sokolovská 6, CZ-77200 Olomouc, Czech Republic
[2] Laboratory of Genetics and Reproduction, Veterinary Research Institute, Hudcova 70, CZ-62132 Brno, Czech Republic
[3] CIRAD, UMR 1098 DAP, Avenue Agropolis, 34398 Montpellier cedex 5, France

INTRODUCTION

- Bananas are giant perennial tropical plants within the genus *Musa*
- Most of cultivated bananas are seed sterile diploid, triploid or tetraploid clones derived from interspecific and intersub-specific crosses between two wild diploid (2n = 22) species *M. acuminata* and *M. balbisiana*
- Despite the socio-economic importance of this crop and the need of improved varieties, there is a lack of knowledge on genome structure and shortage of reliable genetic maps

AIMS OF THE STUDY

- Isolation of single nuclei and chromosomes from banana *M. acuminata* 'DH Pahang'
- Whole genome amplification (WGA) of isolated nuclei and chromosomes
- Quality evaluation of amplified DNA
- Verification of the 'Pahang' genetic map by anchoring SSR markers on single chromosomes

MATERIAL AND METHODS

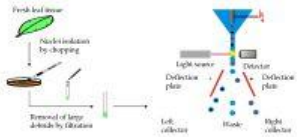
PLANT MATERIAL

- Double haploid *M. acuminata* 'DH Pahang' (ITC 1511) was used in this study
- Nuclear genome size of *M. acuminata* 'DH Pahang' is relatively small (1C ~ 600 Mb)
- Diploid plant contains 22 chromosomes (1 chromosome represents ~ 54,5 Mb)

M. acuminata 'DH Pahang'

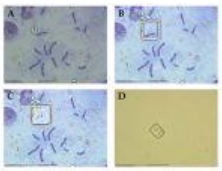
NUCLEI SORTING USING FLOW CYTOMETRY

- Suspension of intact nuclei was prepared according to Bartoš *et. al* (2005)
- Flow-sorted nuclei were used for optimization of WGA conditions
- DNA of isolated nuclei was amplified using WGA method based on linker adapter PCR (GenomPlex Single Cell Whole Genome Amplification Kit, Sigma-Aldrich)



LASER MICRODISSECTION OF CHROMOSOMES

- Metaphase spreads were prepared according to Doleželová *et al.* (1998); protoplast suspension was dropped on a polyethylene napthalene membrane, which was attached to a glass slide
- Single chromosomes of *M. acuminata* 'DH Pahang' were isolated using laser microbeam microdissection and catapulting
- DNA of isolated chromosomes was amplified using WGA method based on linker adapter PCR (GenomPlex Single Cell Whole Genome Amplification Kit, Sigma-Aldrich)



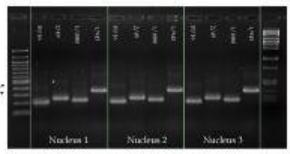
Laser microdissection of chromosomes (illustrative pictures): A suitable metaphase plate is localized under inverted microscope (A), the membrane is cut around the selected chromosome using a laser microbeam (B) and the chromosome is catapulted by a single laser pulse (C) into the cap of a PCR tube (D)

PRELIMINARY RESULTS

DNA AMPLIFIED FROM SINGLE NUCLEI

- The size of DNA amplified from nuclei ranges from 100 bp to 1000 bp
- All banana specific markers for reverse transkriptase (RT) domains of Tnt1, Angela and SIRE retroelements gave positive products after PCR on DNA amplified from single nuclei

WGA products from single nuclei of *M. acuminata* 'DH Pahang'

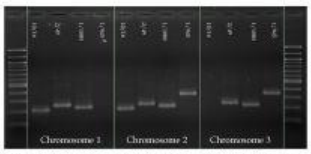


PCR with amplified nuclei and specific primers for reverse transkriptase domain of Tnt1 (10/16), Angela (2/49) and SIRE (1/1080; 1/945)

DNA AMPLIFIED FROM CHROMOSOMES

- The size of DNA amplified from chromosomes ranges from 100 bp to 1000 bp
- Most of banana specific markers for RT domains of Tnt1, Angela and SIRE retroelements gave positive PCR products on DNA amplified from single chromosomes

WGA products from single chromosomes of *M. acuminata* 'DHPahang'



PCR with amplified chromosomes and specific primers for reverse transkriptase domain of Tnt1 (10/16), Angela (2/49) and SIRE (1/1080; 1/945)

CONCLUSIONS AND FUTURE WORK

- WGA method based on linker adapter PCR was used to amplify microdissected chromosomes of *M. acuminata* 'DH Pahang'
- Preliminary results indicate representative amplification even if only one copy of chromosome is used as a template
- The work is in progress to verify linkage groups of the 'Pahang' genetic map using a set of SSR markers and to study chromosomal translocations in different subspecies and clones of *Musa*

This work has been supported by the Grant Agency of the Academy of Sciences of the Czech Republic; grant award no. IAA 600380703

4.3.3 Nuclear rDNA ITS Sequence Region Diversity and Phylogenetic Inference in *Musaceae*

Hřibová E, Čížková J, Christelová P, Taudien S, Roux N, de Langhe E., Doležel J

Abstracts of the International Conference "Plant and Animal Genome XIX". W068. Sherago International, Inc., San Diego, 2011.

[Oral presentation]

Nuclear rDNA ITS Sequence Region Diversity and Phylogenetic Inference in *Musaceae*

E. Hřibová[1], J. Čížková[1], P. Christelová[1], S. Taudien[2], N. Roux[3], E. de Langhe[4], J. Doležel[1]

[1] Centre of the Region Haná for Biotechnological and Agricultural Research, Institute of Experimental Botany, Sokolovska 6, Olomouc, Czech Republic
[2] Genome Analysis, Leibniz Institute for Age Research, Fritz Lipmann Institute, Beutenbergstr. 11, Jena, Germany
[3] Bioversity International, Parc ScientifiQue Agropolis II, Montpellier Cedex 5, France
[4] Laboratory of Crop Improvement, Katholieke Universiteit Leuven, Leuven, Belgium

Nuclear 45S rDNA loci contain many copies of 18S, 5.8S and 26S rRNA genes arranged in tandem and transcribed as single units. The three rRNA genes are separated by internal transcribed spacers ITS1 and ITS2. While the rRNA units are evolutionary conserved, the ITS show high divergence and are popular targets in genetic diversity and phylogenetic studies. Interestingly, in the family *Musaceae*, ITS have not been studied in detail and their use in genetic diversity and phylogenetic studies remains scarce. Our study provided the first detailed analysis of the ITS sequence diversity in *Musaceae* and revealed frequent occurance of more than one type of ITS sequence within a species. The ITS1-5.8S-ITS2 region was studied in 89 representatives of *Musaceae* and ITS-based phylogenetic reconstruction supported the division of genus *Musa* into two clades: Callimusa and Australimusa, and Eumusa and Rhodochlamys. We show that it is necessary to identify putative pseudogenic ITS sequences, which may have negative effect on phylogenetic reconstruction. Almost all banana hybrid clones contained conserved parental ITS sequences, which enabled determination of genomic constitution of hybrids using ITS. However, the presence of only one type of ITS in clones assumed to be of hybrid origin calls for furthed study to confirm their origin and unravel processes leading to evolution of their genomes. This work has been supported by Czech Academy of Sciences (awards IAA600380703 and KJB 500380901) and was performed in collaboration with the International Atomic Energy Agency (Research Agreement no. 13192).

4.3.4 Molecular Analysis and Genomic Organization of Major Satellite DNA in *Musa* spp.

Čížková J, Hřibová E, Christelová P, Humplíková L, Suchánková P, Doležel J

Abstracts of the 18th International Chromosome Conference. P. 21. Manchester, 2011.

[Poster presentation]

Molecular Analysis and Genomic Organization of Major Satellite DNA in *Musa* spp.

J. Čížková[1], E. Hřibová[1], P. Christelová[1], L. Humplíková[1], P. Suchánková[1], J. Doležel[1]

[1] Centre of the Region Haná for Biotechnological and Agricultural Research, Institute of Experimental Botany, Olomouc, Czech Republic

Bananas are giant perennial tropical plants within the genus *Musa*. Most of the currently cultivated bananas are seed sterile diploid, triploid and tetraploid clones derived from intra- and inter-specific crosses between two wild diploid species, *M. acuminata* Colla and *M. balbisiana* Colla. The nuclear genome of banana is relatively small (1C = 500 – 650 Mbp) and comprises small and morphologically very similar chromosomes. In order to enable identification of individual chromosomes, a set of molecular cytogenetic markers is necessary. In this work we employed fluorescence *in situ* hybridization in a set of *Musa* species and cultivated clones to study genomic distribution of banana satellites CL18 and CL33, which we have identified in our previous study, 5S rDNA, 45S rDNA and a single-copy BAC clone 2G17. Furthermore, the diversity and evolution of the major tandem repetitive DNA sequences were characterized at molecular level. This work has been supported by the Grant Agency of the Academy of Sciences of the Czech Republic (grant award no. KJB 500380901) and by the Ministry of Education, Youth and Sports of the Czech Republic and the European Regional Development Fund (Operational Programme Research and Development for Innovations No. CZ.1.05/2.1.00/01.0007).

MOLECULAR ANALYSIS AND GENOMIC ORGANIZATION OF MAJOR SATELLITE DNA IN *MUSA* SPP.

JANA ČÍŽKOVÁ, EVA HŘIBOVÁ, PAVLA CHRISTELOVÁ, LENKA HUMPLÍKOVÁ, PAVLA SUCHÁNKOVÁ AND JAROSLAV DOLEŽEL

CENTRE OF THE REGION HANÁ FOR BIOTECHNOLOGICAL AND AGRICULTURAL RESEARCH, INSTITUTE OF EXPERIMENTAL BOTANY, OLOMOUC, CZECH REPUBLIC

BACKGROUND INFO

- Bananas are giant perennial plants within the genus *Musa*
- Genus *Musa* is divided into four sections: Eumusa, Australimusa, Callimusa and Rhodochlamys
- Most of cultivated bananas are seed sterile diploid, triploid or tetraploid clones derived from intra- and inter-specific crosses between two wild diploid species belonging to the section Eumusa: *M. acuminata* (A genome) and *M. balbisiana* (B genome)
- Evolution of some clones also involves hybridization with *M. schizocarpa* (S genome) and representatives of the Australimusa section (T genome)
- The nuclear genome of banana is relatively small (1C = 500 – 650 Mbp) and comprises small and morphologically very similar chromosomes – a set of molecular cytogenetic markers is necessary to enable identification of individual chromosomes
- Two satellite sequences (CL18 and CL33) were previously identified after 454 sequencing of *M. acuminata* 'Calcutta 4'

AIMS OF THE STUDY

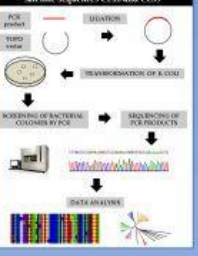
- Determination of the genomic distribution of satellite sequences CL18 and CL33 in a set of *Musa* species and cultivated clones
- Co-localization of satellite sequences CL18 and CL33 with 5S rDNA, 45S rDNA and a single-copy BAC clone 2G17
- Analysis of sequence diversity of satellite sequences CL18 and CL33 at molecular level

MATERIAL AND METHODS

Studied accessions

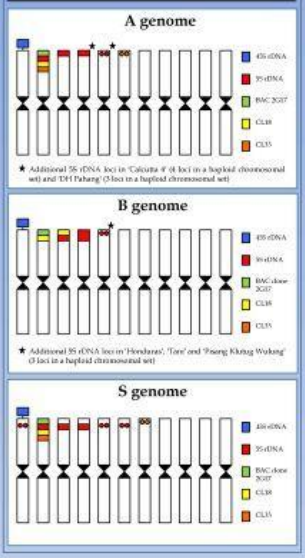
Accession name	Genome constitution	Chromosome number (2n)
Calcutta 4	AA	22
Long Tieray	AA	22
TunGia	AA	22
Mala Ga	AA	22
DH Pahang	AA	22
Cameroon	BB	22
Honduras	BB	22
Tani	BB	22
Pisang Klutag Wulung	BB	22
M. schizocarpa	SS	22
M. schizocarpa	SS	22
Ako	AS	22
Tanlon Kepa	AS	22
Pelipita	AFB	33
Rebenkatter	AFB	33
Obino l'Ewai	AAB	33
Maribi	AAB	33
3 Hands Plantly	AAB	33
Umbrella	AT	21
M. ornata		22

Cloning, sequencing and data analysis of satellite sequences CL18 and CL33

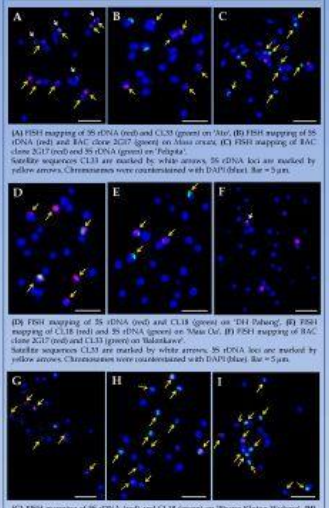


RESULTS

Idiograms



Examples of genomic distribution of satellite DNA as determined on mitotic metaphase chromosomes of *Musa* after FISH with labeled probes



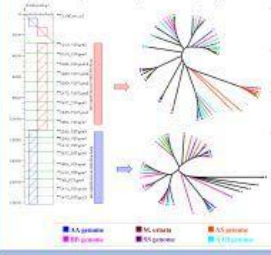
(A) FISH mapping of 5S rDNA (red) and CL33 (green) on 'Ako'. (B) FISH mapping of 5S rDNA (red) and BAC clone 2G17 (green) on *Musa ornata*. (C) FISH mapping of BAC clone 2G17 (red) and 5S rDNA (green) on 'Pelipita'. Satellite sequences CL33 are marked by white arrows. 5S rDNA loci are marked by yellow arrows. Chromosomes were counterstained with DAPI (blue). Bar = 5 μm.

(D) FISH mapping of 5S rDNA (red) and CL18 (green) on 'DH Pahang'. (E) FISH mapping of CL18 (red) and 5S rDNA (green) on 'Mini Ou'. (F) FISH mapping of BAC clone 2G17 (red) and CL33 (green) on 'Malaikahaw'. Satellite sequences CL33 are marked by white arrows. 5S rDNA loci are marked by yellow arrows. Chromosomes were counterstained with DAPI (blue). Bar = 5 μm.

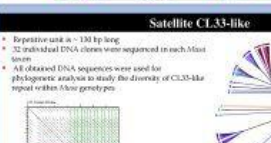
(G) FISH mapping of 5S rDNA (red) and CL18 (green) on 'Pisang Klutag Wulung'. (H) FISH mapping of CL18 (red) and 5S rDNA (green) on *Musa schizocarpa* . (I) FISH mapping of 45S rDNA (red) and 5S rDNA (green) on *Musa schizocarpa*. 5S rDNA loci are marked by yellow arrows. Chromosomes were counterstained with DAPI (blue). Bar = 5 μm.

Satellite CL18-like

- Repetitive unit is ~ 2 kb long
- Two individual parts of CL18 repetitive unit were sequenced from the genomes of evaluated species
- DNA sequences obtained from each *Musa* accession were identified and used for phylogenetic analyses to study the diversity of CL18-like repeat within *Musa* genotypes



- The sequences representing both parts of CL18-like repetitive units were obtained from most of evaluated taxa
- Sequencing of *M. acuminata* 'TunGia' and *M. schizocarpa* did not result in the sequences similar to part 1
- Sequences of AS hybrid banana clones homologous to part 1 of CL18-like repeat contained a long deletion
- Sequences homologous to part 2 of CL18-like repetitive units were not identified in hybrid banana clone '3 Hands Plantly'

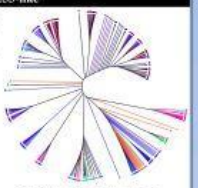


NJ cladograms based on uncorrected p-distance were created using SplitsTree

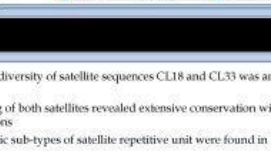
Satellite CL33-like

- Repetitive unit is ~ 130 bp long
- 32 individual DNA clones were sequenced in each *Musa* taxon
- All obtained DNA sequences were used for phylogenetic analysis to study the diversity of CL33-like repeat within *Musa* genotypes

- CL33-like satellite DNA units were found highly conserved within and between the evaluated banana accessions



NJ cladograms based on uncorrected p-distance were created using SplitsTree



CONCLUSIONS

- Genomic distribution of satellite sequences CL18 and CL33, 5S rDNA, 45S rDNA and BAC clone 2G17 was established in a set of *Musa* species and cultivated clones
- The use of CL18, CL33, 5S rDNA, 45S rDNA and BAC clone 2G17 as probes for FISH increased the number of chromosomes which could be identified in *Musa*
- Different patterns of probe co-localization in *M. acuminata* sp. and *M. balbisiana* sp. enabled determination of genomic constitution in their interspecific hybrids
- For the first time, diversity of satellite sequences CL18 and CL33 was analyzed at molecular level
- Sanger sequencing of both satellites revealed extensive conservation within and between the evaluated accessions
- No genome-specific sub-types of satellite repetitive unit were found in A, B or S *Musa* genomes

This work was supported by the Grant Agency of the Academy of Sciences of the Czech Republic (grant award no. KJB 500380901) and by the Ministry of Education, Youth and Sports of the Czech Republic and the European Regional Development Fund (Operational Programme Research and Development for Innovations no. CZ.1.05/2.1.00/01.0007).

4.3.5 Physical Mapping of the Banana (*Musa* spp.) Genome Using Microdissected Chromosomes

Čížková J, Černohorská H, D'Hont A, Hřibová E, Rubeš J, Doležel J

Book of Abstracts "Olomouc Biotech 2011 - Plant Biotechnology: Green for Good". P.42. Olomouc, 2011.

[Poster presentation]

Physical Mapping of the Banana (*Musa* spp.) Genome Using Microdissected Chromosomes

J. Čížková[1], H. Černohornská[2], A. D'Hont[3], E. Hřibová[1], J. Rubeš[2], J. Doležel[1]

[1] Centre of the Region Haná for Biotechnological and Agricultural Research, Institute of Experimental Botany, Olomouc, Czech Republic
[2] Laboratory of Genetics and Reproduction, Veterinary Research Institute, Hudcova 70, CZ-62132 Brno, Czech Republic
[3] CIRAD, UMR 1098 DAP, Avenue Agropolis, 34398 Montpellier cedex 5, France

Bananas and plantains are giant perennial tropical plants within the genus *Musa*. Most of the currently cultivated bananas are seed sterile diploid, triploid or tetraploid clones derived from intra- and inter-specific or intersub-specific crosses between seed bearing diploid progenitors *M. acuminata* Colla and *M. balbisiana* Colla. The breeding of new, improved and resistant varieties is hampered by clonal propagation and sterility of cultivated bananas. Despite their socio-economic importance and the need of improved varieties, there is a lack of knowledge on genome structure and shortage of reliable genetic maps. Observation of chromosome pairing in meiosis suggests that banana sub-species and thus banana cultivars differ by translocations and inversions. The chromosomal rearrangements in parents of mapping populations cause segregation distortion, result in pseudo-linkages and hamper ordering of markers into linkage groups. Together with difficulties in establishing mapping population, these features hamper genetic mapping in *Musa*. One possibility to verify the position of markers in the genome is to locate them using PCR on isolated chromosomes. As it is not possible to sort individual *Musa* chromosomes by flow cytometry, we have used laser microbeam microdissection and catapulting to isolate single chromosomes from double haploid *M. acuminata* 'Pahang' and used them as template for whole genome amplification (WGA). DNA of isolated chromosomes was amplified using different methods and the quality of DNA amplified from a single banana chromosome was assessed. The work is in progress to verify linkage groups of the 'Pahang' genetic map using a set of SSR and DArT markers and to study chromosomal translocations among different subspecies and clones of *Musa*.
This work has been supported by the Ministry of Education, Youth and Sports of the Czech Republic and the European Regional Development Fund (Operational Programme Research and Development for Innovations No. CZ.1.05/2.1.00/01.0007).

PHYSICAL MAPPING OF THE BANANA (*MUSA* SPP.) GENOME USING MICRODISSECTED CHROMOSOMES

JANA ČÍŽKOVÁ[1], HALINA ČERNOHORSKÁ[2], ANGELIQUE D'HONT[3], EVA HŘIBOVÁ[1], JIŘÍ RUBEŠ[2], JAROSLAV DOLEŽEL[1]

[1] Centre of the Region Haná for Biotechnological and Agricultural Research, Institute of Experimental Botany, Olomouc, Czech Republic
[2] Laboratory of Genetics and Reproduction, Veterinary Research Institute, Hudcova 70, CZ-62132 Brno, Czech Republic
[3] CIRAD, UMR 1098 DAP, Avenue Agropolis, 34398 Montpellier cedex 5, France

INTRODUCTION

- Bananas are giant perennial tropical plants within the genus *Musa*
- Most of cultivated bananas are seed sterile diploid, triploid or tetraploid clones derived from interspecific and intersub-specific crosses between two wild diploid (2n = 22) species *M. acuminata* and *M. balbisiana*
- Despite the socio-economic importance of this crop and the need of improved varieties, there is a lack of knowledge on genome structure and shortage of reliable genetic maps

AIMS OF THE STUDY

- Isolation of single nuclei and chromosomes from banana *M. acuminata* 'DH Pahang'
- Whole genome amplification (WGA) of isolated nuclei and chromosomes
- Quality evaluation of amplified DNA
- Verification of the 'Pahang' genetic map by anchoring SSR markers on single chromosomes

MATERIAL AND METHODS

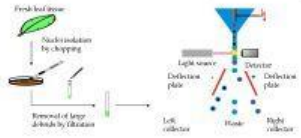
PLANT MATERIAL

- Double haploid *M. acuminata* 'DH Pahang' (ITC 1511) was used in this study
- Nuclear genome size of *M. acuminata* 'DH Pahang' is relatively small (1C ~ 600 Mb)
- Diploid plant contains 22 chromosomes (1 chromosome represents ~ 54,5 Mb)

M. acuminata 'DH Pahang'

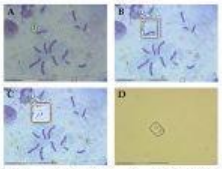
NUCLEI SORTING USING FLOW CYTOMETRY

- Suspension of intact nuclei was prepared according to Bartoš *et. al* (2005)
- Flow-sorted nuclei were used for optimization of WGA conditions
- DNA of isolated nuclei was amplified using WGA method based on linker adapter PCR (GenomPlex Single Cell Whole Genome Amplification Kit, Sigma-Aldrich)



LASER MICRODISSECTION OF CHROMOSOMES

- Metaphase spreads were prepared according to Doleželová *et. al.* (1998); protoplast suspension was dropped on a polyethylene napthalene membrane, which was attached to a glass slide
- Single chromosomes of *M. acuminata* 'DH Pahang' were isolated using laser microbeam microdissection and catapulting
- DNA of isolated chromosomes was amplified using WGA method based on linker adapter PCR (GenomPlex Single Cell Whole Genome Amplification Kit, Sigma-Aldrich)



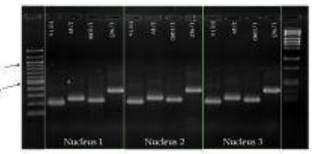
Laser microdissection of chromosomes (illustrative pictures): A suitable metaphase plate is localized under inverted microscope (A), the membrane is cut around the selected chromosome using a laser microbeam (B) and the chromosome is catapulted by a single laser pulse (C) into the cap of a PCR tube (D)

PRELIMINARY RESULTS

DNA AMPLIFIED FROM SINGLE NUCLEI

- The size of DNA amplified from nuclei ranges from 100 bp to 1000 bp
- All banana specific markers for reverse transcriptase (RT) domains of Tnt1, Angela and SIRE retroelements gave positive products after PCR on DNA amplified from single nuclei

WGA products from single nuclei of *M. acuminata* 'DH Pahang'

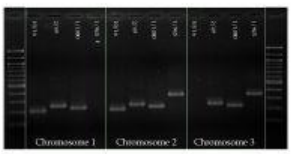


PCR with amplified nuclei and specific primers for reverse transcriptase domain of Tnt1 (10/16), Angela (2/49) and SIRE (1/1080; 1/945)

DNA AMPLIFIED FROM CHROMOSOMES

- The size of DNA amplified from chromosomes ranges from 100 bp to 1000 bp
- Most of banana specific markers for RT domains of Tnt1, Angela and SIRE retroelements gave positive PCR products on DNA amplified from single chromosomes

WGA products from single chromosomes of *M. acuminata* 'DH Pahang'



PCR with amplified chromosomes and specific primers for reverse transcriptase domain of Tnt1 (10/16), Angela (2/49) and SIRE (1/1080; 1/945)

CONCLUSIONS AND FUTURE WORK

- WGA method based on linker adapter PCR was used to amplify flow-sorted nuclei and microdissected chromosomes of *M. acuminata* 'DH Pahang'
- Preliminary results indicate representative amplification even if only one copy of a chromosome is used as a template
- The work is in progress to verify linkage groups of the 'Pahang' genetic map using a set of SSR markers and to study chromosomal translocations in different subspecies and clones of *Musa*

This work has been supported by the Ministry of Education, Youth and Sports of the Czech Republic and the European Regional Development Fund (Operational Programme Research and Development for Innovations No. CZ.1.05/2.1.00/01.0007).

4.3.6 Molecular and Cytogenetic Characterization of Wild *Musa* Species Newly Introduced to ITC Collection

Čížková J, Hřibová E, Christelová P, Van den Houwe I, Roux N, Doležel J

Book of Abstracts "Olomouc Biotech 2013 - Plant Biotechnology: Green for Good II". P.57. Olomouc, 2013.

[Poster presentation]

Molecular and Cytogenetic Characterization of Wild *Musa* Species Newly Introduced to ITC Collection

J. Čížková[1], E. Hřibová[1], P. Christelová[1], I. Van den Houwe[2], N. Roux[3], J. Doležel[1]

[1] Centre of the Region Haná for Biotechnological and Agricultural Research, Institute of Experimental Botany, Šlechtitelů 31, Olomouc CZ-78371, Czech Republic
[2] Laboratory of Tropical Crop Improvement, Katholieke Universiteit Leuven, Kasteelpark Arenberg 13, B-3001 Leuven, Belgium
[3] Bioversity International, Parc Scientifique Agropolis II, Montpellier Cedex 5, 34397, France

Bananas and plantains (*Musa* spp.) are a major staple food and an important export commodity in many countries of humid tropics. The production of bananas is, however, threatened by rapid spreading of various diseases and adverse environmental conditions. The preservation and characterization of banana diversity is essential for the purposes of crop improvement. The world's largest banana and plantain collection is maintained at the Bioversity International Transit Centre (ITC) in Belgium and contains more than 1000 accessions. This reference collection is continuously extended by various edible cultivars and wild species. Detailed morphological and molecular characterization of the deposited as well as newly introduced accessions is of a great significance for efficient collection and protection of banana diversity.

The aim of this work was to characterize a set of wild *Musa* species, which were newly introduced into the ITC collection. Altogether, 23 wild *Musa* accessions were described in terms of cytogenetic as well as molecular characters. Nuclear genome size, chromosome number and genomic distribution of rRNA genes were determined in individual accessions to shed light on their genome structure and evolution. The molecular characterization was performed using a set of 19 microsatellite markers. The SSR genotyping platform enabled us to identify species most closely related to the unknown accessions and thus clarify their phylogeny. In several cases, SSR genotyping was coupled with the sequence analysis of ITS DNA spacer region, to elucidate the putative hybrid character of some accessions. The ITS1-5.8S-ITS2 region was amplified using specific primers and PCR products were sequenced after cloning. Sanger sequencing of ITS region was also used to support incongruent results of SSR analysis.

MOLECULAR AND CYTOGENETIC CHARACTERIZATION OF WILD *MUSA* SPECIES NEWLY INTRODUCED TO ITC COLLECTION

Čížková J[1], Hřibová E[1], Christelová P[1], Van den Houwe I[2], Roux N[3], Doležel J[1]

1 Centre of the Region Haná for Biotechnological and Agricultural Research, Institute of Experimental Botany, Olomouc, Czech Republic
2 Laboratory of Tropical Crop Improvement, Katholieke Universiteit Leuven, Belgium
3 Bioversity International, Montpellier, France

INTRODUCTION

- Bananas (*Musa* spp.) are a major food crop and an important export commodity in many countries of humid tropics
- The production of banana is threatened by various diseases and adverse environmental conditions
- The preservation and characterization of banana diversity is essential for the purposes of crop improvement
- World's largest banana collection is maintained at the Bioversity International Transit Centre (ITC) in Belgium
- This reference collection is continuously extended by edible cultivars and wild species
- Efficient collection and conservation of *Musa* genetic diversity depends on unambiguous sample identification

TAXONOMY OF THE FAMILY MUSACEAE

Family: Musaceae
- **Genus:** *Musa*
  - **Section:** Eumusa (2n = 2x = 22)
  - **Section:** Rhodochlamys (2n = 2x = 22)
  - **Section:** Australimusa (2n = 2x = 20)
  - **Section:** Callimusa (2n = 2x = 18, 20)
- **Genus:** *Ensete*
- **Genus:** *Musella*

Taxonomy of the family Musaceae is based on morphological characters and basic chromosome number

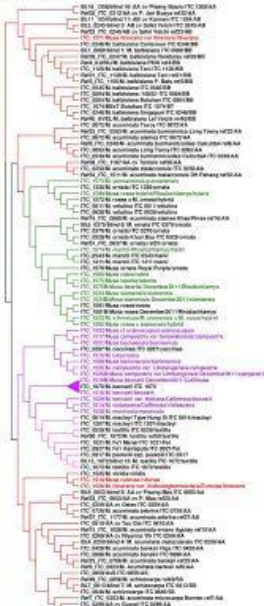
AIMS OF THE STUDY

- Contribution to characterization of the genetic diversity of wild *Musa* species
- Determination of chromosome number, nuclear DNA content and genomic distribution of rRNA genes to shed light on the genome structure of evaluated accessions
- Verification of the classification of evaluated accessions based on SSR analysis
- Verification of the putative hybrid origin of evaluated *Musa* accessions

RESULTS

SSR ANALYSIS

- Standardized SSR genotyping platform (Christelová *et al.*, 2011, AoB PLANTS, doi: 10.1093/aobpla/plr024) was used to analyze genetic diversity of studied accessions
- Newly characterized *Musa* accessions are labeled in colors in the tree bellow



UPGMA cluster analysis of diploid accessions based on Nei (1973) genetic distance

NUCLEAR DNA CONTENT

Accession name	ITC code	Section	Chromosome number	2C nuclear DNA content [pg]
M. sheareri var. *siikuangkuanensis*	1526	Eumusa	22	1.311
M. sheareri var. *sheareri*	1571	Eumusa	22	1.217
M. yunnanensis	1575	Eumusa	22	1.306
M. rubikae	1518	Rhodochlamys	22	1.310
M. x flemingiae (*M. siamensis* x *M. rosea*)	1522	Rhodochlamys	22	1.261
M. siamensis	1534	Rhodochlamys	22	1.280
M. wasai	1574	Rhodochlamys	22	1.282
M. laterita	1575	Rhodochlamys	22	1.315
M. rubra	1940	Rhodochlamys	22	1.306
M. rosea x *siamensis*	1992	Rhodochlamys	22	1.279
M. rosea (hybrid)	1998	Rhodochlamys	22	1.285
M. cf. urashoreyon	1532	Australimusa	20	1.442
M. violascens	1514	Callimusa	20	1.428
M. lolae	1515	Callimusa	20	1.432
M. beccarii var. *beccarii*	1516	Callimusa	18	1.537
M. campestris var. *sarawakensis*	1517	Callimusa	20	1.417
M. acuminata	1528	Callimusa	20	1.390
M. beccarii var. *borneo*	1529	Callimusa	18	1.673
M. borneensis	1531	Callimusa	20	1.772
M. campestris var. *berkangensis*	1535	Callimusa	20	1.454
M. barioensis	1568	Callimusa	20	1.480

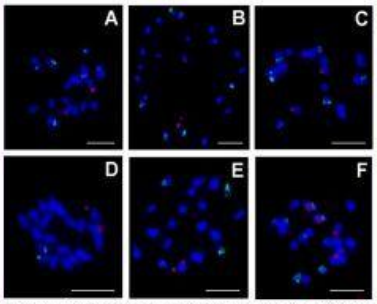
ANALYSIS OF ITS REGION

- Clone based sequencing strategy was used to analyze the ITS1-5.8S-ITS2 sequence region of 45S rDNA locus
- Analysis of ITS region was used with the aim to verify hybrid origin of evaluated *Musa* accessions

Multiple alignment of a part of ITS1 sequence region created using mafft

- Except for 10 accessions (ITC 1514; 1515; 1517; 1526; 1528; 1534; 1529; 1571 and 1598) all others contained three or more ITS types which support their hybrid origin
- Hybrid origin was observed also for *M. beccarii* var. *buffum* (ITC 1529) which did not cluster together with other *beccarii* species analysed in this as well as in previous study using SSR genotyping

CYTOGENETIC MAPPING OF rDNA



Localization of 45S rDNA (red) and 5S rDNA (green) on mitotic metaphase chromosomes. (A) *M. yunnanensis*, (B) *M. rubra*, (C) *M. rosea* x *siamensis*, (D) *M. campestris* var. *sarawakensis*, (E) *M. acuminata*, (F) *M. borneensis*. Bar = 5 μm.

CONCLUSIONS

- Nuclear DNA content, chromosome number and genomic distribution of 45S and 5S rDNA were estimated in studied accessions showing higher variability in these characters in comparison with previous studies
- For *M. borneensis* (ITC 1531), the highest nuclear DNA content known for diploid *Musa* species was estimated
- SSR analysis did not confirm genetic origin of several accessions (ITC 1518, ITC 1532 and ITC 1573)
- Classification of these accessions should be re-evaluated
- Analysis of ITS region suggested hybrid origin of most of studied accessions

This work was supported by the Ministry of Education, Youth and Sports of the Czech Republic (award MSMT INGOII), by the Bioversity International (LoA CII. 2011, LoA C# 2012) and the European Regional Development Fund (Operational Programme Research and Development for Innovations No. ED0007/01/01).

5 CONCLUSIONS

This Ph.D. thesis focused on the structure and organization of banana genome and the genetic diversity and phylogenetic relationships within *Musa* species.

In the first part of the study, analysis of nucleotide sequence of the ITS region was used to characterize the genetic diversity of banana and to clarify phylogenetic relationships within the family *Musaceae*. 87 accessions of wild diploid species and triploid hybrid clones were chosen to cover all main taxonomic groups of the family *Musaceae*. The presence of more than one type of the ITS1-5.8S-ITS2 region was revealed within some *Musa* species and hybrid clones. The importance of using clone-based sequencing strategy and of identification of putative pseudogenic ITS sequences were shown. DNA sequences obtained by classical Sanger sequencing and by massively parallel 454 sequencing were analysed with the aim to thoroughly describe the nucleotide structure and diversity of the ITS region. Neigbor-Joining and Bayesian analysis lead to similar phylogenetic inferences. The monophyletic origin of the family *Musaceae* was confirmed and within the family, three main clades were revealed. The monophyly of *Ensete* and *Musella* was not supported as they shared the same clade. The genus *Musa* was divided into two distinct clades. One of them contained Eumusa and Rhodochlamys accessions, the other included representatives of sections Australimusa and Callimusa. Positions of *Ensete*/*Musella* clade and Australimusa/Callimusa clade were reversed in the phylogenetic tree. This suggests that the phylogenetic reconstruction based merely on ITS sequence analysis may lead to inconsistent inference on the generic level. On the other hand, the locus has proven powerful in phylogenetic studies on lower taxonomic level. Incomplete concerted evolution was inferred in most of intraspecific banana hybrids by the presence of conserved parental ITS sequences. Independent evolution of parental rDNA in hybrid clones enables determination of their genomic constitution. The identification of only one type of ITS sequence in some of the presumed interspecific hybrids warrants further study to confirm their origin and to unravel processes leading to evolution of their genomes. The study provides a plausible picture of the evolution of *Musa* species which can be further utilized in improving the classification within the genus.

In the second part of the study, cytogenetic mapping was used to characterize genomic organization of two main banana DNA satellites newly identified from 454 sequencing data of *M. acuminata* 'Calcutta 4'. DNA satellites CL18 and CL33 together with other DNA sequences (a LINE element, rRNA genes and a single-copy BAC clone 2G17) were localized on mitotic metaphase chromosomes of nineteen *Musa* accession. Representatives of A, B and S genomes and their inter-specific hybrids were included in the study. Multicolor FISH with probes for satellite DNA sequences, rRNA genes and the BAC clone 2G17 resulted in characteristic chromosome banding patterns specific for A, B and S genome. It has been shown that the differences in probe localizations may be used for the determination of genomic constitution in interspecific hybrids. LINE element was preferentially localized to centromeric regions of all studied accessions. Probe for the LINE element thus appears as a convenient way of labeling primary constrictions, which are not always easily visible on small and condensed *Musa* chromosomes. The study increased the number of cytogenetic markers and also the number of individual chromosomes which can be identified in *Musa*. Molecular analysis of the two satellites revealed a high level of sequence conservation within, and a high homology between *Musa* species. Cluster analysis based on Neighbor-Joining method showed no A or B genome-specific satellite units. These findings suggest that both satellites originated before the divergence of the section Eumusa and that their sequence remained conserved during speciation.

The third part was focused on molecular and cytogenetic characterization of wild *Musa* species, which have been recently introduced into the banana and plantain germplasm collection maintainded at the Bioversity International Transit Centre. Detailed characterization of newly introduced accessions is necessary for effective collection and use of banana genetic diversity. In this work, nuclear DNA content and genomic distribution of 45S and 5S rDNA were examined in 21 diploid accessions recently added to ITC collection, representing all four sections of the genus *Musa*. 2C DNA content is sections Eumusa and Rhodochlamys ranged from 1.217 to 1.315 pg. Species belonging to sections Australimusa and Callimusa had 2C DNA contents ranging from 1.390 to 1.772 pg. While the number of 45S rDNA loci was conserved in Eumusa and Rhodochlamys, it was highly variable in Callimusa species. The variation on the number of 5S rDNA loci was significantly higher. To assess the genetic diversity and relationships of studied ITC accessions, a set of SSR markers

was used, following a standardized platform for genotyping in banana. The grouping of 21 studied accessions within the diploid reference set was mostly consistent with the characterization based on traditional morphotaxonomic classification. Detailed analysis of ITS region revealed polymorphic sequences in most accessions and indicated the presence of putative pseudogenic ITS types. The results also suggested that some of analyzed accessions might be interspecific hybrids and/or their backcross progeny.

6 LIST OF ABBREVIATIONS

AFLP	amplified fragment length polymorphism
BAC	bacterial artificial chromosome
cM	centimorgan
DArT	Diversity Array Technology
DNA	deoxyribonucleic acid
eccDNA	extrachromosomal DNA
ETS	external transcribed spacer
FISH	fluorescence *in situ* hybridization
GISH	genomic *in situ* hybridization
IGS	intergenic spacer
ISH	*in situ* hybridization
ITC	International Transit Centre
ITS	internal transcribed spacer
LINE	long interspersed nuclear element
LTR	long terminal repeat
NOR	nucleolar organizing region
NTS	non-transcribed spacer
PCR	polymerase chain reaction
RAPD	random amplified polymorphic DNA
*rbc*L	ribulose-1,5-bisphosphate carboxylase
rDNA	ribosomal DNA
RFLP	restriction fragment length polymorphism
RNA	ribonucleic acid
rRNA	ribosomal RNA
SINE	short interspersed nuclear element
SNP	single nucleotide polymorphism
SSR	simple sequence repeat
TIR	terminal inverted repeat

WGD	whole genome duplication
YAC	yeast artificial chromosome