Genet Resour Crop Evol (2022) 69:737–754
https://doi.org/10.1007/s10722-021-01260-y

RESEARCH ARTICLE

Organelle DNA sequence data provide new insights into the maternal and paternal lineages of *Musa* species germplasms

Huimin Feng · You Chen · Changming Li · Xiaoxiong Xu · Hongli Luo · Chaozu He

Received: 12 June 2020 / Accepted: 9 January 2021 / Published online: 14 November 2021
© The Author(s), under exclusive licence to Springer Nature B.V. 2021

Abstract Single-copy organelle DNA loci are widely used to infer plants' phylogenetic relationships. In bananas (*Musa* spp.), the maternal transmission of chloroplast DNA and paternal transmission of mitochondrial DNA provides an excellent possibility to follow both maternal and paternal lineages by using chloroplast and mitochondrial markers, respectively. In the present study, 25 chloroplast and 12 mitochondrial DNA sequences of *Musa* spp. were used to elucidate the genetic diversity and phylogenetic relationships of wild and cultivated bananas. Sequence alignment and phylogenetic analyses grouped the 34 wild and 26 cultivars into different clades and subclades. The use of cytoplasmic genes to analyze the origin and evolution of cultivated bananas revealed two main maternal and paternal origins: via *Musa acuminata* Colla and via *Musa balbisiana* Colla. Relationships among wild accessions and intra- and interspecific hybrids, as well as between the latter, evidenced eight chloroplast and six mitochondrial gene pools, and 18 cytotypes were identified. The maternal origin of most *M. acuminata* hybrids was the Ca2 gene pool, while the A-B hybrids were derived from the Ca3 or Ca5 gene pools. The most common paternal origin was the Ma3 gene pool, except for ABB genotypes with Mb1 origin. Furthermore, we found a role for *Musa itinerans* Cheesman in the paternal origin of banana cultivars. The present findings will help refine *Musa* spp. phylogeny, and enrich the available cytoplasmic data for *Musa* spp. germplasms that will be useful for improving the breeding of banana cultivars.

Keywords Maternal origin · *Musa* species · Organelle DNA loci · Paternal origin

H. Feng · H. Luo · C. He (✉)
Hainan Key Laboratory for Sustainable Utilization of Tropical Bioresources, Hainan University, No.58 Renmin Avenue, Haikou 5702228, Hainan, China
e-mail: czhe@hainanu.edu.cn

H. Feng · Y. Chen · C. Li · X. Xu
Sanya Key Laboratory for Breeding of Tropical Crops, Hainan Tropical Ocean University, Sanya 572022, Hainan, China

Introduction

Bananas (*Musa* spp., family Musaceae) originated in southeast Asia and western Pacific, and spread widely throughout the tropics and subtropics to become one of the most important sources of tropical food, next to rice, wheat, and maize. The genus *Musa* comprises 70 wild species (Häkkinen 2013) and 500 cultivars (Simmonds 1966).

The phylogenetics of bananas were first studied in 1753 (Linnaeus 1753), and the genus was named *Musa* by Carl Linnaeus. Later, Sagot (1887) divided the genus into giant bananas, fleshy edible bananas, and ornamental bananas. Based on Sagot's work,

Baker (1893) further divided *Musa* into three subgenus: *physocaulis*, *eumusa*, and *rhodochlamys*. Classification at the chromosomal level was first performed by Cheesman (1947), who divided *Musa* into four sections: *eumusa* and *rhodochlamys* (2n = 22), and *callimusa and australimusa* (2n = 20). Subsequently, Simmonds (1960) added a group, *ingentimusa*, containing only two species: *Musa ingens* Simmonds (2n = 14) and *Musa beccarii* Simmonds (2n = 18). Thereafter, Argent (1976) created a new section *ingentimusa*, into which *M. ingens* was placed, and Häkkinen et al. (2005) suggested the inclusion of *M. beccarii* into the *callimusa* section. For half a century there has been little change in the genus classification system for *Musa* as proposed by Cheesman (1947). However, studies examining the taxonomic relationships within *Musa* using molecular approaches have questioned the validity and practicability of this system and have generally classified the genus into two groups, namely, *Musa* (2n = 22) and *Callimusa* (2n = 20/18) (Gawel et al. 1992; Wong et al. 2002; Nwakanma et al. 2003; Li et al. 2010; Liu et al. 2010; Bekele and Shigeta 2011; Christelová et al. 2011; Häkkinen 2013). Most recently, Feng et al. (2016) confirmed this classification using simple sequence repeat (SSR) markers to determine the molecular phylogeny of the genus *Musa*.

To date, more than 70 wild *Musa* species have been identified; of these, *Musa acuminata* Colla (A genome) and *Musa balbisiana* Colla (B genome) are the most prominent. The genomic constitutions of AA, AB, AAA, AAB, and ABB (Stover and Simmonds 1987) observed in present-day banana cultivars evolved through intra- and inter-specific crosses (Cheesman 1948; Simmonds and Shepherd 1955) of these two species. *M. acuminata* is genetically rich comprising 10 subspecies (*banksii, burmannica, burmannicodes, errans, malaccensis, microcarpa, siamea, truncata,* and *zebrina*) and the variety (var.) chinensis (Feng et al. 2009). Although higher genetic diversity has been observed in *M. balbisiana* (Sotto and Rabara 2000), intraspecific classification has not been reported to date.

Banana cultivars are mostly diploid, triploid, or tetraploid, with characteristics of sterility, parthenocarpy, polyploidy, or unknown origin, which has led to the slow progress of banana genetic improvements. Examining the genetic diversity and phylogenetic relationships of banana germplasms would therefore help clarifying the origin and evolution of banana cultivars and accelerate their breeding.

Three different genomes exist within plant cells: a nuclear genome, a chloroplast genome, and a mitochondrial genome. Several studies have shown the potential of using complete organellar genomes to analyze phylogenetic relationships in plants. The nuclear genome size of *M. acuminata* is 523 Mbps (D'Hont et al. 2012) and the *M. balbisiana* genome size is 79% that of *M. acuminata* (Davey et al. 2013). In contrast, the organellar genomes are much smaller than the nuclear genome; the chloroplast genome is 0.17 Mbps (Barrett et al. 2014; Shetty et al. 2016; Li et al. 2017), while the exact size of the mitochondrial genome is currently unknown. Moreover, the complete chloroplast genomes of *M. acuminata* (Martin et al., 2013) and *Musa itinerans* Cheesman (Li et al., 2017) have been published. In *Musa* spp., inheritance of the chloroplast genome is strongly biased toward the maternal lineage, while the mitochondrial genome is paternally inherited (Fauré et al. 1994). Thus, organellar genomes enable maternal and paternal lineages to be followed using chloroplast and mitochondrial markers, respectively.

Gawel and Jarret (1991a) used chloroplast DNA restriction fragment length polymorphisms (RFLPs) to analyze the phylogenetics of *Musa* species and subspecies, and reported cytoplasmic diversity among *Musa* cultivars (Gawel and Jarret 1991b). Later, Carreel et al. (2002) combined RFLPs with the hybridization of heterologous mitochondrial and chloroplastic probes to characterize 71 wild accessions, and 131 diploid and 103 triploid cultivars of *Musa,* and identified 10 chloroplastic patterns and more than 100 mitochondrial DNA patterns. Umali and Nakamura (2003) reported a single nucleotide polymorphism (SNP) marker from the trnL-F intergenic spacer region of chloroplast DNA, which could be used to discriminate *M. acuminata* from *M. balbisiana*. Nwakanma et al. (2003) constructed a molecular phylogeny of *Musa* species using restriction-site polymorphisms of organelles, and suggested that the evolutionary status of *M. balbisiana* was primitive. More recently, Swangpol et al. (2007) analyzed SNPs from selected non-coding chloroplast DNA sequences of *Musa* interspecific hybrids and found that the *M. acuminata* and *M. balbisiana* genomes could be clearly distinguished. Boonruangrod et al. (2008) analyzed the relationship between

chloroplast and mitochondrial haplotypes of 54 accessions and identified six chloroplastic and seven mitochondrial gene pools. A combination of chloroplast and mitochondria gene pools identified 14 cytotypes; Cytotype VIII, resulting from the crossing of maternal Cytotypes I and II and paternal Cytotype III ancestors, was identified in the majority of the analyzed cultivars.

In the present study, sequence data from 25 chloroplast and 12 mitochondria DNA genomes were used to assess the phylogenetic relationships of 60 *Musa* species, including a wide range of wild and cultivated species. We aimed to elucidate the genetic diversity and phylogenetic relationships of wild and cultivated *Musa* species and subspecies, especially Chinese species. Furthermore, we generated additional cytoplasmic data on *Musa* germplasms to refine phylogenetic research, and reconstructed the paternal and maternal lineages of diploid wild species as well as that of banana cultivars.

Materials and methods

Plant materials and DNA extraction

Sixty *Musa* spp. samples were used in the present study (Table 1). Of these, 49 accessions were collected from the Biodiversity International *Musa* Germplasm Transit Centre (ITC, Leuven University, Belgium) and 11 were collected by the authors during trips throughout south China. Samples were identified and morphological characters were described using the *Musa* descriptors (INIBAP/CIRAD, 1996). Total genomic DNA was extracted from young leaves using a modified cetyltrimethylammonium bromide (CTAB) method (Paterson et al. 1993). The quality of extracted DNA was determined by visualization on 1% agarose gel and on NanoDrop 2000 (Thermo Fisher Scientific, MA, USA). Total DNA samples were diluted to 50 ng/μL with sterile water.

PCR amplification and sequencing

Twenty-five pairs of chloroplast DNA primers and 12 pairs of mitochondrial DNA primers were selected for use in PCR amplifications (Table 2). These were carried out in 50-μL reaction mixtures containing 2 μL of 50 ng/μL DNA, 2 μL of each primer (10 μM), and 25 μL of 2 × Taq Master Mix (Vazyme Biotech Co., Ltd., China), with the final volume adjusted using double distilled water. PCR amplifications were performed in a FlexCycler (Analytikjena, Germany) under the following reaction conditions: 94 °C for 5 min, 30 cycles of 94 °C for 30 s, 48–58.5 °C for 30 s, 72 °C for 60 s, and a final extension at 72 °C for 7 min. PCR products were visualized on 1% agarose gels and subsequently purified and sequenced at BGI Technology Co., Ltd. (China) using the Sanger method. The primers used for sequencing were the same as those used for the PCR. Sequencher v.4.2 software (Gene Codes Corp., MI, USA) was used to assemble the sequences.

Sequence alignment and phylogenetic analyses

Nucleotide sequences were aligned by MAFFT v.7 (Katoh and Standley. 2013). Characteristics of genetic diversity, including conserved sites, variable sites, and parsimony-informative (Pi) sites were computed in DnaSP v.6.12.03 (Rozas et al. 2017). The incongruence length difference (ILD) test was performed in PAUP v.4.0b (Swofford 2002) to estimate the level of potential incongruence in the data. The aligned chloroplast and mitochondrial DNA sequences were then concatenated in SequenceMatrix-Windows v.1.7.8 (Vaidya et al. 2011), and used for further phylogenetic analyses.

Phylogenetic relationships were inferred using maximum likelihood (ML) and Bayesian inference (BI) methods. For ML analysis, IQ-TREE v.1.6 software was first used for best-fit model estimation of the sequence matrix (Kalyaanamoorthy et al. 2017), and then to calculate and select suitable nucleotide substitution models and corresponding parameters. The ML tree was then constructed with the ultrafast bootstrap (BS) of IQ-TREE v.1.6 (Hoang et al. 2018; Nguyen et al. 2015) following repeated searches for 1000 repeats, using a tree with the largest likelihood value, 1000 repeated self-expansion detections to tests the confidence of each branch, and the Figtree option to view and optimize the generated ML tree.

For the BI analysis, the data sets were first tested for the best-fit model of evolution with IQ-TREE v.1.6 using the Akaike information criterion (AIC) and then analyzed using MrBayes v.3.2 software (Ronquist et al. 2012). The BI analysis utilized the Markov Chain Monte Carlo algorithm, starting with a random tree,

Table 1 Analyzed accessions and their cytotype affiliations

Number	Common name	Nuclear genotype	Species	Sub-species	Source	Chloroplast gene-pool	Mitochondrial gene-pool	Cytotype
M01	*Itinerans*	wild	*itinerans*	–	22°06.772′N, 111°11.239′E	–	–	–
M02	*Itinerans*	wild	*itinerans*	–	18°40.398′N, 109°52.772′E	–	–	–
M03	*Itinerans*	wild	*itinerans*	–	22°28.358′N, 109°29.498′E	–	–	–
M04	*Itinerans*	wild	*itinerans*	–	23°58.428′N, 97°33.672′E	–	–	–
M05	Xiaoguoyejiao	AAwt	*acuminata*	var. *chinensis*	22°39.170′N, 103°03.763′E	Ca1	Ma1	I
M06	*Balbisiana*	BBwt	*balbisiana*	Ssp501	ITC0080	Cb1	Mb1	XIV
M07	*Musa laterita*	wild	*laterita*	–	ITC0627	–	–	–
M08	*Musa beccarii*	wild	*beccarii*	–	ITC1070	–	–	–
M09	Pisang lilin	AAcv	AA	Ssp/sgr503	ITC1121	Ca2	Ma1	II
M10	Prata	AAB	AAB	Pome	ITC0207	Ca3	Ma3	VII
M11	Safet velchi	AB	AB	Ney Poovan	ITC0245	Ca3	Ma1	V
M12	Fhia-01	AAAB	AAAB	Ssp/sgr502	ITC0504	Ca3	Ma3	VII
M13	Fenjiao	ABB	ABB		18°36.888′N, 109°30.164′E	Cb1	Mb2	XV
M14	Type 3 ×	AAwt	*acuminata*	Ssp503	ITC0060	Ca4	Ma1	VIII
M15	Type 2 ×	AAwt	*acuminata*	Ssp505	ITC0069	Ca4	Ma1	VIII
M16	Pahang IRFA	AAwt	*acuminata*	Ssp517	ITC0070	Ca2	Ma2	III
M17	Tavoy	AAwt	*acuminata*	*burmannica*	ITC0072	Ca2	Ma2	III
M18	*malaccensis*	AAwt	*acuminata*	*malaccensis*	ITC0074	Ca2	Ma1	II
M19	Calcutta 4	AAwt	*acuminata*	*burmannicoides*	ITC0249	Ca2	Ma1	II
M20	Borneo	AAwt	*acuminata*	*microcarpa*	ITC0253	Ca5	Ma3	XII
M21	Higa	AAwt	*acuminata*	Ssp528	ITC0378	Ca2	Ma3	IV
M22	Hybrid	AAwt	*acuminata*	Ssp507	ITC0382	Ca5	Ma1	XI
M23	Pa (Musore) no.3	AAwt	*acuminata*	Ssp521	ITC0406	Ca2	Ma1	II
M24	Pisang Cici Alas	AAwt	*acuminata*	Ssp529	ITC0415	Ca1	Ma1	I
M25	Hybrid	AAwt	*acuminata*	Ssp513	ITC0608	Ca2	Ma1	II
M26	Selangor 2	AAwt	*acuminata*	Ssp523	ITC0629	Ca2	Ma1	II
M27	Khae (Phrae)	AAwt	*acuminata*	*siamea*	ITC0660	Ca1	Ma1	I

Table 1 continued

Number	Common name	Nuclear genotype	Species	Sub-species	Source	Chloroplast gene-pool	Mitochondrial gene-pool	Cytotype
M28	Pisang Cici	AAwt	*acuminata*	Ssp516	ITC0681	Ca3	Ma2	VI
M29	THA018	AAwt	*acuminata*	Ssp527	ITC1067	Ca2	Ma1	II
M30	*zebrina*	AAwt	*acuminata*	*zebrina*	ITC1177	Ca4	Ma2	IX
M31	*Musa chunii*	wild	*chunii*	–	24°37.049′N, 97°34.954′E	–	–	–
M32	*Musa balbisiana*	BBwt	*balbisiana*	–	24°38.973′ N, 97°36.210′ E	Cb1	Mb2	XV
M33	*Musa yunnanensis*	wild	*yunnanensis*	–	24°43.553′ N, 97°34.343′ E	–	–	–
M34	*Musa nagensium*	wild	*nagensium*	–	24°38.230′ N, 97°34.906′ E	–	–	–
M35	Cameroun	BBwt	*balbisiana*	Ssp510	ITC0246	Cb1	Mb3	XVI
M36	Eti kehel	BBwt	*balbisiana*	Ssp504	ITC0271	Cb2	Mb2	XVII
M37	*balbisiana*	BBwt	*balbisiana*	Ssp513	ITC0545	Cb3	Mb2	XVIII
M38	*Butuhan*	wild	*balbisiana* × *textilis*	Ssp/sgr501	ITC1074	–	–	–
M39	Tjau lagada	AAcv	AA	Ssp/sgr553	ITC0090	?	Ma3	?
M40	Guyod	AAcv	AA	Ssp/sgr555	ITC0299	Ca4	Ma3	X
M41	Gu nin chio	AAcv	AA	Ssp/sgr561	ITC0442	Ca2	Ma3	IV
M42	Khai (kampengpeth)	AAcv	AA	Ssp/sgr678	ITC0532	Ca2	Ma3	IV
M43	Kluai lep mu nang	AAcv	AA	Ssp/sgr671	ITC0533	Ca2	?	?
M44	Amas (south johnstone)	AAcv	AA	Ssp/sgr551	ITC0567	Ca2	Ma3	IV
M45	Pisang berlin	AAcv	AA	Ssp/sgr521	ITC0611	Ca2	Ma1	II
M46	Maduranga	ABB	ABB	Ssp/sgr503	ITC0035	Ca3	Mb2	XIII
M47	Vudi wai wai	AAB	AAB	Ssp/sgr506	ITC0357	Ca5	Ma3	XII
M48	Datil	AB	AB	Ssp/sgr502	ITC0392	Ca2	Ma3	IV
M49	Kluai roi wi	AAB	AAB	Pisang Kelat	ITC0535	Ca3	Ma1	V
M50	Namwa khom	ABB	ABB	Pisang Awak	ITC0659	Cb1	Mb2	XV
M51	Pisang raja bulu	AAB	AAB	Pisang Raja	ITC0843	Ca5	Ma3	XII
M52	Tetraploide embrapa 401	AAAB	AAAB	Ssp/sgr503	ITC1194	Ca3	Ma1	V
M53	Pc12-05	AAAB	AAAB	Ssp/sgr510	ITC1260	Ca3	Ma1	V
M54	Chuoi man	AAB	AAB	AB Ssp/sgr 507	ITC1379	Ca3	Ma1	V

Table 1 continued

Number	Common name	Nuclear genotype	Species	Sub-species	Source	Chloroplast gene-pool	Mitochondrial gene-pool	Cytotype
M55	Chuoi mat	AAB	AAB	Pome	ITC1381	Ca5	*M.itinerans*	XIX
M56	Fhia-18	AAAB	AAAB	Ssp/sgr 520	ITC1412	Ca3	Ma3	VII
M57	Mdzodji	AAB	AAB	Plantain	ITC1469	Ca3	Ma3	VII
M58	Highgate	AAA	AAA	Gros Michel	ITC0263	Ca2	Ma3	IV
M59	*Musa peekelii* ssp. *peekeii*	wild	*peekelii*	*peekelii*	ITC0917	–	–	–
M60	Williams	AAA	AAA	Williams	18°44.200′N, 09°28.702′E	Ca2	Ma3	IV

four chains (one cold chain and three hot chains) running for 10,000,000 generations, and sampling once every 1000 generations. After reaching equilibrium, 25% of the burn-in samples were discarded, and the remaining samples were used to build a consistent tree. Bayesian trees were evaluated using posterior probability (PP).

Results

Phylogenetic analysis of *Musa* spp. based on 25 chloroplast gene sequences

The results of the ILD test supported the combination of the 25 chloroplast genes dataset ($\pi = 0.01$). The combined dataset covered 22,306 bp, which generated 617 variable sites with 265 Pi sites. The IQ-TREE analysis found that the best-fit models for the ML and BI trees were K3Pu + F + I and GTR + G + I, respectively. The phylogenetic trees constructed by both methods revealed the same topological structure (Fig. 1).

Three major clades were identified by the phylogenetic trees. Clade A was formed by the *M. acuminata* complex, *Musa laterita* Cheesman, *Musa yunnanensis* Häkkinen & H. Wang, *Musa chunii* Häkkinen, *M. balbisiana* from the ITC, *M. balbisiana* × *Musa textiles* Née, and most banana cultivars (BS = 100, PP = 1). Clade B (BS = 81, PP = 0.94) consisted of *M. itinerans*, *Musa nagensium* Prain, most *M. balbisiana,* and two cultivars with a B genome. Clade C included *M. beccarii* as the outgroup, which belongs to section *callimusa*, with a basic chromosome number of 2n = 18. The number of chromosomes in other test materials was 2n = 22 or 2n = 20.

Clade A contained eight subclades. Subclade A1 (BS = 100, PP = 1) consisted of *M. acuminata* var. *chinensis*, *M. acuminata* subsp. *siamea*, *M. acuminata ssp529*, and *M. laterita*. Among these, *M. acuminata* var. *chinensis* is a unique variant found in China. Subclade A2 (BS = 95, PP = 1) comprised cultivars with the A genotype, which are widely spread, and most of *M. acuminata* wild accessions, including subspecies *malaccensis*, *burmannica*, and *burmannicoides;* these wild accessions are mainly distributed in Malaysia. Subclade A3 (BS = 89, PP = 1) comprised most A-B genotype cultivars and a single *M. acuminata* wild accession, *M. acuminata ssp516*. Subclade A4 (BS = 54, PP = 1) comprised three wild

Table 2 Primers used to amplify chloroplast and mitochondrial genes

No	Gene	Size (bp)	Tm (°C)	Forward (5′-3′)	Reverse (5′-3′)	Reference
Chloroplast genes						
1	matK	683	48	CGATCTATTCATTCAATATTTC	TCTAGCACACGAAAGTCGAAGT	Cuenoud et al. 2002
2	psaA	831	50	AAATCGTGAGCATCAGCATG	CCGAGGAGAACAGGCCATTC	Swangpol et al. 2007
3	psbA	831	50	ATGACTGCTACTTTAGAAAGACG	TCATGCATWACTTCCATACCTA	Yoon et al. 2002
4	rpL16	940	55	GCTATGCTTAGTGTGTGACT	CATTCTTCCTCTATGTTGTTT	Swangpol et al. 2007
5	rpoC1	498	50	GTGGATACACTTCTTGATAATGG	TGAGAAAACATAAGTAAACGGGC	Kress and Erickson 2007
6	trnL	578	55	CGAAATCGGTAGACGCTACG	GGGGATAGAGGGACTTGAAC	Taberlet et al.1991
7	trnL-trnF	354	50	GGTTCAAGTCCCTCTATCCC	ATTTGAACTGGTGACACGAG	Taberlet et al.1991
8	psbC-trnS	1349	55	GGTCGTGACCAAGAAACCAC	GGTTCGAATCCCTCTCTCTC	Nicolosi et al. 2000
9	trnS-trnT	1108	55	CGAGGGTTCGAATCCCTCTC	AGAGCATCGCATTTGTAATG	Nicolosi et al. 2000
10	trnC-rpoB	1098	50	TGCCTTACCACTCGGCCAT	GTAGATATTCCCTCATTTCC	Takanori and Ohmi 2000
11	psbM-trnD^{GUC}	970	50	AGCAATAAATGCRAGAATATTTACTTCCAT	GGGATTGTAGYTCAATTGGT	Small et al. 2005
12	rps16	722	60	GTGGTAGAAAGCAACGTGCGACTT	TCGGGATCGAACATCAATTGCAAC	Oxelman et al.1997
13	ycf3	1733	50	GCWTTTACYTATTAYAGAGATG	TNGAATGGCCTGTTCTCC	Small et al. 2005
14	trnS^{GGA}-rpS4	832	50	TTACCGAGGGTTCGAATCCCTC	ATGTCSCGTTAYCGAGGACCT	Small et al. 2005
15	trnV^{UAC}-trnM^{CAU}	902	50	GGCTATACGGRYTYGAACCGTA	CCTACTATTGGATTYGAACCAATGACTC	Small et al. 2005
16	trnP^{UGG}-petG	407	50	TGTAGCGCAGCYYGGTAGCG	CAATAYCGACGKGGYGATCAATT	Small et al. 2005
17	rpL20-rpS12	726	50	ATTAGAAANRCAAGACAGCCAAT	CGYYAYCGAGCTATATATCC	Small et al. 2005
18	psbB-psbH	563	50	TCCAAAAANKKGGAGATCCAAC	TCAAYRGTYTGTGTAGCCAT	Small et al. 2005
19	rpoB	1167	55	CTAAGGGGTTGTTGTGTAAC	AATATGCAACGTCAAGCAGT	Lin et al. 1999
20	petA	758	55	AYGCATATCCCATTTTTGC	TCATTTCGWACAAYTGAACC	Scarcelli et al. 2011
21	psbC-trnS	1369	57	GGTCGTGACCAAGAAACCAC	GGTTCGAATCCCTCTCTCTC	Demesure et al., 1995
22	rpl36-rps8	534	55	CACAAATTTTACGAACGAAG	TAATGACAGAYCGAGARGCTCGAC	Kress et al. 2005
23	ycf6-psbM	1272	55	GGATATAGTAAGTCTTGCTTGGG	TTCTTGCATTTATTGCTACTGC	Kress et al. 2005
24	trnV-atpE	1209	55	GTGTAAACGAGTTGCTCTACCA	CGACATTTGCACATTTAGATGCTAC	Kress et al. 2005
25	trnK-rps16	872	55	TACTCTACCRTTGAGTTAGCAAC	AAAGGKGCTCAACCTACARGAAC	Kress et al. 2005
Mitochondrial genes						
1	nad4L/orf25	512	50	CTGTYTTTTCGCACTTAGGC	GTCCGRGGTACTATTGCTGT	Duminil et al. 2002
2	nad5/4–5-F	1243	55	CCAATTTTTGGGCCAATTCC	CATTGCAAAGGCATAATGAT	Dumolin-Lapegue et al. 1997
3	orf25	433	46	AAGACCRCCAAGCYYTCTCG	TTGCTGCTATTCTATCTATT	Duminil et al. 2002

Table 2 continued

No	Gene	Size (bp)	Tm (°C)	Forward (5′-3′)	Reverse (5′-3′)	Reference
4	atpA	475	55	AGATGGGATTGCACGTGTTT	CAGTTGAGCGTTTCTGTCCA	Cipriani et al. 2003
5	ccb206	505	50	TCAATCTTGTRAACTAATCG	CYYCTCCACACCAATCACGA	Duminil et al. 2002
6	Cox3	709	50	ATGATTGAATCTCAGAGGCAC	ACCACCCCACCAATAGATAGA	Zhang 2011
7	Cob	727	50	ATTCCTCTTCCAACTCGTC	ATGCGAGTTATAGCAGTCC	Yan 2009
8	ccmC	456	50	GTTGTTCTTAACAGCGATGG	GAAGTTAGCAAAGTTAGACA	Zhang 2011
9	ccmFCb	297	50	TCACTCAACGATTGCCTTTG	TCGAAGCATATAGATCCGTA	Zhang 2011
10	ccmFN2	417	55	TTGGACCGCCGGCGCGAACA	AGATGCACAAGAGTACTTCG	Zhang 2011
11	rps12-1/nad3-2	311	55	TTTCTTCTCTACCATGACGA	TGATCCYACTCGGTSTTCCT	Duminil et al. 2002
12	rrn18	717	55	GCTCAGAAGGAACGCTAGCT	TTACGGCATGGACTACCAGG	Zhang 2011

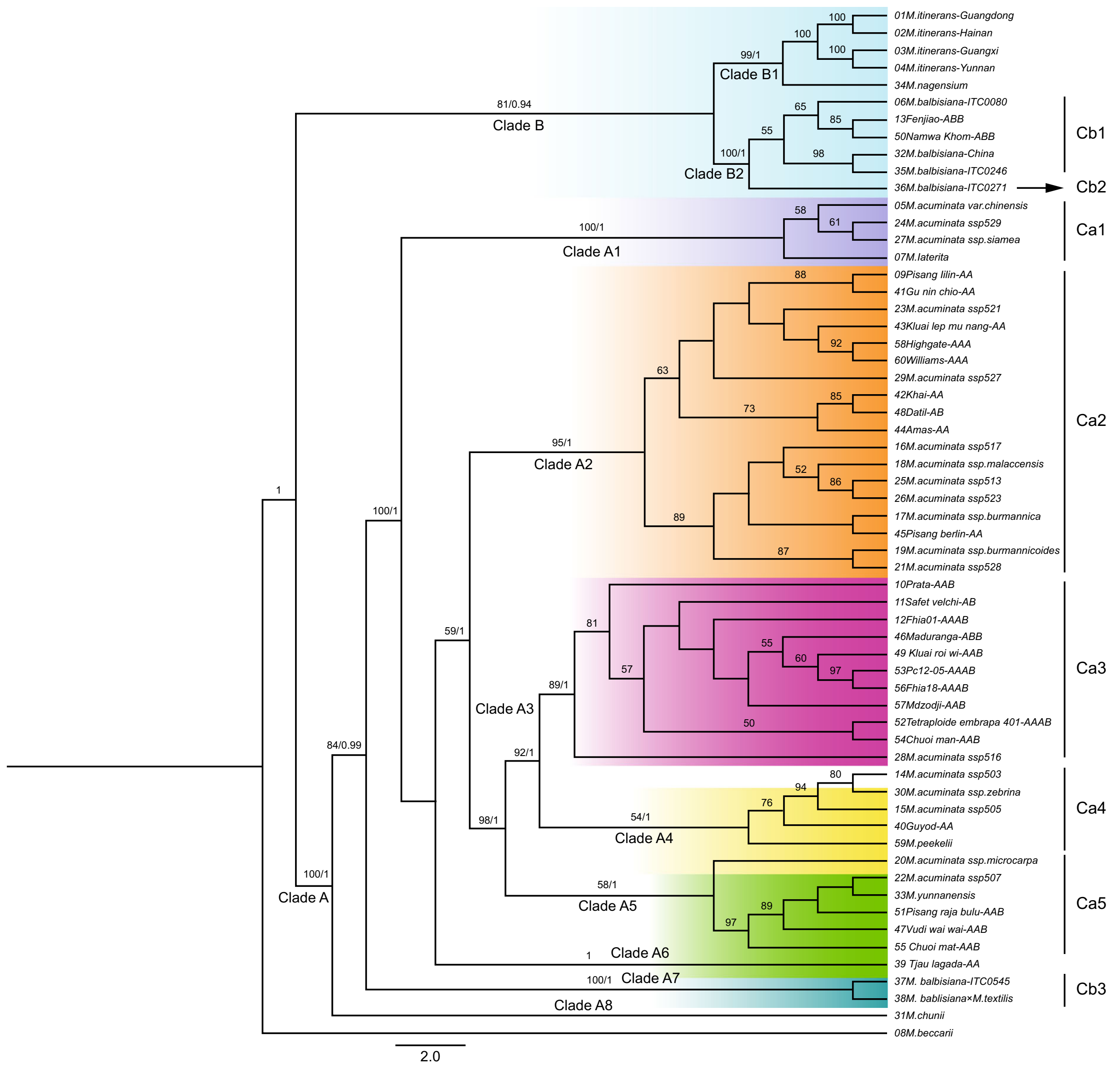

Fig. 1 Maximum Likelihood (ML) tree of 60 *Musa* spp. accessions based on 25 chloroplast gene sequences (including chloroplast genotyping)

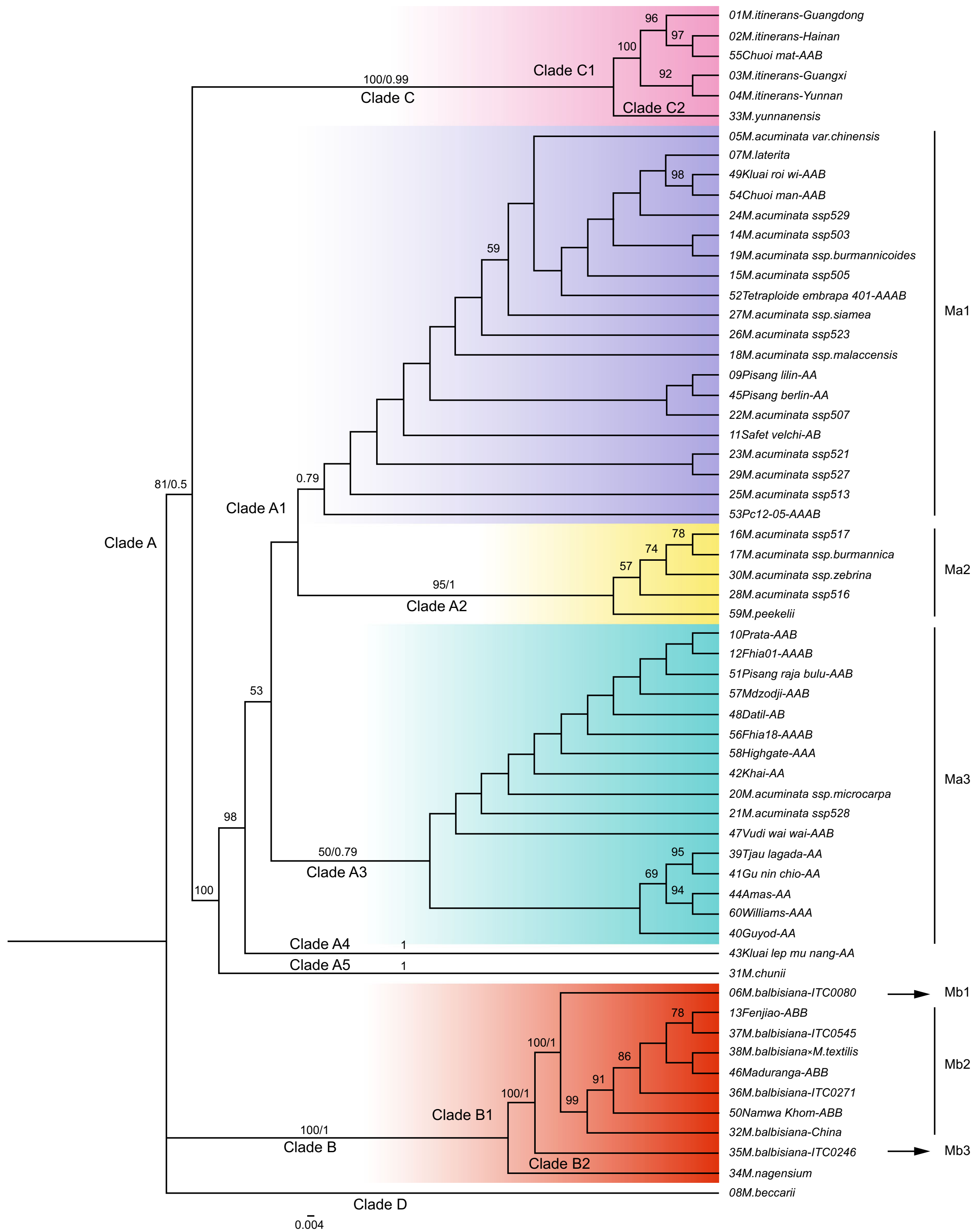

Fig. 2 Maximum Likelihood (ML) tree of 60 *Musa* spp. accessions based on 12 mitochondrial gene sequences (including mitochondrial genotyping)

accessions of *M. acuminata*, including subspecies *zebrina*, *ssp503*, and *ssp505*, *M. peekelii* Lauterb, and one AA genotype cultivar. Subclade A5 (BS = 58, PP = 1) consisted of *M. acuminata* subsp. *microcarpa*, *M. yunnanensis*, and three AAB genotype cultivars. Subclade A6 (PP = 1) included one AA genotype cultivar but no wild types, being an independent branch. Subclade A7 (BS = 100, PP = 1) comprised two wild accessions, *M. balbisiana* and *M. balbisiana* × *M. textilis*. Subclade A8 contained only wild accessions of *M. chunii*, which are found exclusively in Yunnan, China.

Clade B was divided into two subclades. Subclade B1 (BS = 99, PP = 1) contained *M. nagensium* and four *M. itinerans* accessions, which were collected from different populations in China. *M. itinerans* accessions collected from the Guangdong and Hainan populations were grouped together, while those from the Guangxi and Yunnan populations formed another group. Subclade B2 (BS = 100, PP = 1) contained four *M. balbisiana* accessions, ITC0080, ITC0246, and ITC0271, which were collected from China, and two ABB genotype cultivars.

Phylogenetic analysis of Musa spp. based on 12 mitochondrial gene sequences

The dataset based on 12 mitochondrial genes was supported by the ILD test (π = 0.01); these fragments were therefore combined to form a single dataset for subsequent phylogenetic analyses. The combined dataset covered 6802 bp, generating 553 variable sites with 279 Pi sites. The IQ-TREE analyses found that the best-fit model was TVM + F + I for the ML tree and GTR + G + I for the BI tree. The phylogenetic trees constructed using these two methods presented the same topological structure (Fig. 2).

Four major clades were identified from the phylogenetic trees. Clade A (BS = 100, PP = 1) contained the *M. acuminata* complex, *M. laterita*, *M. chunii,* and most of the banana cultivars in our sample set. Clade B (BS = 100, PP = 1) included all tested *M. balbisiana*, *M. nagensium*, the natural wild hybrid *M. balbisiana* × *M. textilis,* and three A-B genotype cultivated bananas. Clade C (BS = 100, PP = 0.99) contained samples from different populations of *M. itinerans* and *M. yunnanensis,* and the AAB genotype cultivar 'Cluoi mat'. Clade D included *M. beccarii* as the outgroup.

Clade A consisted of five subclades. Subclade A1 contained *M. acuminata* var. *chinensis*, *M. acuminata* subspecies *burmannicoides* and *malaccensis*, eight unclassified wild *M. acuminata* accessions, *M. laterita,* two AA genotype cultivars ('Pisang lilin' and 'Pisang berlin'), one AB genotype cultivar ('Safet velchi'), two AAB genotype cultivars ('Kluai roi wi' and 'Chuoi mat'), and two AAAB genotype cultivars ('Tetraploide EMBRAPA 401' and 'Pc12-05'). Subclade A2 included *M. acuminata* subspecies *burmannica* and *zebrina*, two unclassified wild *M. acuminata* accessions, and *M. peekelii*, with no related cultivars in our sample set. Subclade A3 contained *M. acuminata* subsp. *microcarpa*, *M. acuminata ssp528,* and 14 cultivars, including five AA genotypes ('Tjau lagada', 'Gunn chio', 'Guyod', 'Amas', and 'Khai'), one AB genotype ('Datil'), two AAA genotypes ('Highgate' and 'Williams'), four AAB genotypes ('Prata', 'Pisang raja bulu', 'Mdzodji', and 'Vudi wai wai'), and two AAAB genotypes ('FHIA01' and 'FHIA18'). Subclades A4 and A5 each contained one sample; the cultivar 'Kluai lep mu nang' (AA genotype) and *M. chunii,* respectively.

Clade B consisted of two subclades, with *M. nagensium* forming an independent subclade (BS = 100, PP = 1). *M. balbisiana*, the natural wild hybrids *M. balbisiana* × *M. textilis,* and three A-B genotype cultivated bananas were grouped in another subclade (BS = 100, PP = 1), which was further divided into three branches (BS = 100, PP = 1). *Musa balbisiana*-ITC0246 and *M. balbisiana*-ITC0080 formed independent branches, while other *M. balbisiana* samples, *M. balbisiana* × *M. textilis,* and cultivars 'Fenjiao' (AB), 'Maduranga' (ABB), and 'Namwa khom' (ABB) clustered together.

Clade C consisted of two subclades: subclade C1 contained four samples of *M. itinerans* obtained from different geographical sources in China, and one AAB genotype cultivar, 'Chuoi mat'. *M. itinerans* from the Guangdong and Hainan populations grouped together, and those from the Guangxi and Yunnan populations also grouped together. In contrast, subclade C2 contained only *M. yunnanensis*.

Origin and evolution of cultivated bananas based on cytoplasmic genes

To elucidate the gene pools of cultivated bananas ancestors, we analyzed 49 wild and cultivated types of

M. acuminata and *M. balbisiana* using chloroplast and mitochondrial genes inherited from single parents. Based on 25 chloroplast genes, eight chloroplast gene pools were identified; five from *M. acuminata* and three from *M. balbisiana* (Table 1). The first gene pool (Ca1) contained three *M. acuminata* wild types (var. *chinensis* and subspecies *siamea* and *ssp529*), with no related cultivars in our sample set. The presence of diploid wild type *M. acuminata* subspecies *burmannica*, *malaccensis*, *burmannicoides, ssp517, ssp513, ssp528, ssp521, ssp523,* and *ssp527* characterized the largest gene pool (Ca2), along with six AA cultivars ('Pisang lilin', 'Gu nin chio', 'Khai', 'Kluai lep mu nang', 'Amas', and 'Pisang berlin'), one AB diploid hybrid cultivar ('Datil'), and two triploid AAA cultivars ('Highgate' and 'Willams'). This gene pool contained the most abundant wild type *M. acuminata* accessions, most of the AA cultivars, and all AAA cultivars. One unclassified *M. acuminata* wild type (*ssp516*) and 10 cultivars formed the third gene pool (Ca3). The cultivars in Ca3 included one AB diploid hybrid ('Safet velchi'), five triploids (AAB: 'Prata', 'Kluai roi wi', 'Chuoi mat', and 'Mdzodji'; AAB: 'Maduranga'), and all four tetraploid AAAB cultivars ('Fhia-01', 'Fhia-18', 'Tetraploide EMBRAPA 401', and 'Pc12-05'). All cultivars in this gene pool contained genome B. Three *M. acuminata* wild type subspecies (*zebrina*, *ssp503*, and *ssp505*) and one AA diploid cultivar ('Guyod') formed the fourth gene pool (Ca4).The Ca5 gene pool comprised two *M. acuminata* wild type subspecies (*microcarpa* and *ssp507),* one AA cultivar ('Tiau lagada'), and three AAB triploids ('Vudi wai', 'Pisang raja bulu', and 'Chuoi mat').

Based on 25 chloroplast genes, three subgroups of *M. balbisiana* were identified. Three diploid *M. balbisiana* wild types (one collected from China and subspecies *ssp501* and *ssp510*) formed the larger gene pool (Cb1) along with two ABB triploid cultivars ('Fenjiao' and 'Namwa Khom'), while subspecies *ssp504* and *ssp513* were found in the Cb1 and Cb2 gene pools, respectively.

Similarly, based on 12 mitochondrial genes, six mitochondrial gene pools were identified among the 49 analyzed accessions (Table 1). *M. acuminata* var. *chinensis, M. acuminata* subspecies *siamea, malaccensis,* and *burmannicoides,* and eight unclassified *M. acuminata* wild accessions formed the largest gene pool (Ma1) along with two diploid AA cultivars ('Pisang lilin' and 'Pisang berlin'), one diploid AB cultivar ('Safet velchi'), two triploid AAB cultivars ('Kluai roi wi' and 'Chuoi mat'), and two tetraploid AAAB cultivars ('Tetraploide EMBRAPA 401' and 'Pc12-05'). Three diploid wild type *M. acuminata* subspecies (*burmannica*, *zebrina,* and *ssp516*) formed the second gene pool (Ma2) with no related cultivars. Gene pool Ma3 consisted of two diploid *M. acuminata* wild subspecies (*microcarpa* and *ssp528*), six diploid cultivars (AA: 'Tjau lagada', 'Gu nin chio', 'K hai', 'Amas', 'Guyod'; and AB: 'Datil'), six triploid cultivars (AAA: 'Highgate' and 'Williams'; AAB: 'Prata', 'Mdzodji', 'Vudi wai', and 'Pisang raja bulu'), and two tetraploid AAAB cultivars ('Fhia-01' and 'Fhia-18'). In contrast, *M. balbisiana* comprised three gene pools. The Mb1 and Mb3 gene pools were represented by one subspecies each (*ssp501*, Mb1; *ssp510*, Mb3), while Mb2 consisted of six samples, including three diploid *M. balbisiana* wild types (one collected from China, *ssp504,* and *ssp513*), and three ABB triploid cultivars ('Fenjiao', 'Namwa Khom', and 'Maduranga').

Eighteen cytotypes (a combination of chloroplast and mitochondrial gene pools) were identified among the analyzed samples (Table 1). The analyzed wild type *M. acuminata* accessions yielded nine cytotypes (I, II, III, IV, VI, VIII, IX, XI, and XII), while five cytotypes (XIV, XV, XVI, XVII, and XVIII) were found among the *M. balbisiana* wild types. Seven different cytotypes were found among 23 cultivars; three of these cytotypes resembled those found in the wild types (cytotypes II, IV, and XV), while the remaining four represented new combinations.

Three different cytotypes (II, IV, and X) were identified among the six diploid AA genotypes. The chloroplast genomes of 'Pisang lilin' and 'Pisang berlin' belonged to the Ca2 gene pool, which contained the wild type *M. acuminata* subspecies *malaccensis*, *burmannica,* and *burmannicoides*. The mitochondrial genome of those subspecies originated from the Ma1 gene pool, which contained the wild type subspecies *malaccensis*, *siamea*, and *burmannicoides,* and the var. *chinensis*, all of cytotype II. A similar chloroplast type was identified in the diploid AA cultivars 'Gu nin chio', 'Khai', and 'Amas'; however, its mitochondrial genome represented the Ma3 gene pool, which contained *M. acuminata* subsp. *microcarpa*. The remaining AA hybrid cultivar 'Guyod' had Ca4 chloroplast and Ma3 mitochondrial

genomes, representing the gene pools containing *M. acuminata* subsp. *zebrina* (chloroplast) and *M. acuminata* subsp. *microcarpa* (mitochondrial). Two diploid AA cultivars were identified, 'Kluai lep mu nang' (AA) and 'Tjau lagada' (AA). The chloroplast gene pool of 'Kluai lep mu nang' was Ca2, which contained *M. acuminata* subspecies *malaccensis*, *burmannica,* and *burmannicoides*. However, the identity of the mitochondrial genome remains unknown. Conversely, the mitochondrial gene pool of 'Tjau lagada' was Ma3, which contained *M. acuminata* subsp. *microcarpa*, while the chloroplast genome remains unknown. Therefore, we were unable to classify these two cultivars into the abovementioned cytotypes.

Similar to the diploid AA cultivars, two triploid AAA genotypes were of cytotype IV ('Highgate' and 'Williams'). Therefore, cytotype IV was found most frequently among the intraspecific *M. acuminata* hybrids. Among the interspecific hybrids analyzed in the present study, the two diploid AB cultivars were of two cytotypes, IV and V.

The six triploid AAB cultivars were of three cytotypes (V, VII, and XI). However, the cytotype of 'Chuoi mat' with an AAB genome was identified as cytotype XIX, as it combined the Ca5 chloroplast genome and the *M. itinerans* mitochondrial genome. The three ABB cooking bananas were of cytotypes XIII and XV.

Similarly, two of the four tetraploid AAAB cultivars ('Tetraploid EMBRAPA 401' and 'Pc12-05') were of cytotype V, while the remaining two tetraploid cultivars ('Fhia-01' and 'Fhia-18') were of cytotype VII.

Discussion

In *Musa* spp., inheritance of the chloroplast genome is strongly biased toward the maternal lineage, while the mitochondrial genome is paternally inherited (Fauré et al. 1994). Consequently, the organellar genomes enable the maternal as well as the paternal lineages to be followed through the use of chloroplast and mitochondrial markers, respectively. In the present study, ML and BI trees of 60 *Musa* spp. accessions were constructed based on 25 chloroplast and 12 mitochondrial gene sequences. The topologies identified using both approaches were consistent for the two organellar genomes.

Maternal phylogenetic analysis of *Musa* spp. based on chloroplast genes

Based on 25 chloroplast gene sequences, the *M. acuminata* wild types and cultivars grouped together and were distinguished from *M. balbisiana* wild types and cultivars. Gawel and Jarret (1991a) reported similar findings using different chloroplast probes and Southern blot analysis. However, in the present study, differences in the maternal origin of *M. balbisiana* were identified. Most *M. balbisiana* samples grouped with *M. itinerans* and *M. nagensium,* while *M. balbisiana* (ITC0545) and *M. balbisiana* × *M. textilis* clustered with *M. acuminata*. Although relatively independent from *M. acuminata*, both *M. balbisiana* (ITC0545) and *M. balbisiana* × *M. textilis* were expected to have a common maternal origin. Therefore, we hypothesized that some of the *M. balbisiana* germplasms might have contacted *M. acuminata* germplasms during their evolution. In addition, *M. balbisiana* (ITC0545) and *M. balbisiana* × *M. textilis* clustered together, indicating that *M. balbisiana* and *M. textilis* have a closer genetic relationship in terms of their maternal origin (Gawel and Jarret 1991b).

Most *M. balbisiana* samples were closely related to *M. itinerans*, indicating that they may have shared a common maternal ancestor. In China, the wild germplasm of *M. balbisiana* is only distributed in Yunnan and cannot form a large population. In contrast, *M. itinerans* is distributed throughout Hainan, Guangdong, Guangxi, and Yunnan (Häkkinen, 2008), and can form large populations in all of these provinces. Here, we showed that *M. itinerans* from Guangdong and Hainan populations clustered together, while those from Guangxi and Yunnan populations formed a different group, indicating that in southern China *M. itinerans* has distinct chloroplast genomes. Our findings are in contrast to those of Ge et al. (2005) who analyzed the populations of *M. balbisiana* distributed in China using chloroplast PCR–RFLP markers and identified two major clades corresponding to two geographical regions; thus, the wild *Musa* spp. germplasms might have been incorrectly identified by Ge et al. (2005).

Except for *M. chunii*, the wild germplasms of other *Eumusa* groups in the tested materials were interspersed between *M. acuminata* complexes. Among them, *M. laterita, M. acuminata* subsp. *siamea*, and *M.*

acuminata var. *chinensis* have the same maternal origin. *M. yunnanensis* was first identified by Häkkinen and Hong (2007). Based on its maternal evolution, this species has the same origin as *M. acuminata* subsp. *microcarpa*. However, Feng et al. (2016) analyzed the nuclear genome of this species using SSR markers and revealed a closer relationship with *M. balbisiana*. *M. chunii* was first identified by Häkkinen (2009) in Yunnan, China. The maternal origin of this species is unique; it did not cluster with any of the tested materials used in the present study.

Based on the 25 chloroplast gene sequences, we conclude that the maternal evolution of *Musa* spp. followed two main routes: via *M. acuminata* and via *M. balbisiana*.

Paternal phylogenetic analysis of *Musa* spp. based on mitochondrial genes

Based on the 12 mitochondrial gene sequences, the test materials could be divided into two independent branches of *M. acuminata* and *M. balbisiana*. Therefore, we believe that the patrilineal evolution of *Musa* spp. also followed two evolutionary routes: via *M. acuminata* and via *M. balbisiana*. However, *M. itinerans* and *M. balbisiana* presented the same matrilineal evolutionary path, while *M. itinerans* and *M. acuminata* were closer when considering paternal evolution. *M. laterita* and *M. acuminata* var. *chinensis* grouped together in both patrilineal and maternal evolution, indicating the same parental origin. Feng et al. (2016) used SSR markers for phylogenetic research on *Musa* spp., and also found that *M. laterita* was most closely related to *M. acuminata* var. *chinensis*. *M. peekelii* clustered with *M. acuminata* subsp. *zebrina* in both patrilineal and maternal evolution, having the same parental origin. *M. chunii* represents an independent branch in mitochondrial evolution. This species is unique in terms of both its maternal and paternal origin.

Few studies on the mitochondrial genome have investigated the phylogenetic evolution of *Musa* spp. Most previous studies have used some molecular marker technologies, such as PCR–RFLP (Nwakanma et al. 2003; Boonruangrod et al. 2008) and RFLP (Carreel et al. 2002). Additionally, the wild germplasm resources of *Musa* spp. are very limited, and there are limited references to the wild germplasms distributed in China. Therefore, we believe that our study is the first to report the use of multiple mitochondrial gene sequences for phylogenetic analysis, providing insight into the patrilineal evolution of the *Musa* genus.

Cytoplasm gene pools of *M. acuminata* ancestors

In the present study, 25 chloroplast gene sequences resulted in the identification of five chloroplast gene pools in the *M. acuminata* complex, namely, *chinensis/siamea*, *burmannica/malaccensis/burmannicoides*, and *M. acuminata* subspecies *ssp516*, *zebrina*, and *microcarpa*. Carreel et al. (2002) previously reported five chloroplast patterns for *M. acuminata*: *zebrina*, *malaccensis*, *siamea*, *banksii*, and *errans/burmannica/burmannnicoides/siamea/malaccensis/microcarpa/truncata*. Conversely, three chloroplast gene pools were identified by Boonruangrod et al. (2008): *errans/banksii*, *microcarpa,* and *burmannicoides/siamea/burmannica/zebrina/malaccensis*. The results of the present study support the findings of Carreel et al. (2002), which are, in part, consistent with those of Boonruangrod et al. (2008), and consider *burmannica/malaccensis/burmannicoides* as representing the same gene pool, while *zebrina* and *siamea* have independent chloroplast gene pools. Moreover, consistent with Boonruangrod et al. (2008), we also found that *microcarpa* belongs to an independent chloroplast gene pool. The limited disparity between our findings and those of Carreel et al. (2002) and Boonruangrod et al. (2008) could be due to the large number of *M. acuminata* subspecies used by those earlier studies, which might have influenced their separation. However, we believe that this ambiguity was addressed here through the use of chloroplast gene loci up to 22,306 bp, generating 352 Pi loci and 1663 insertion/deletion loci, which is sufficiently large to make the distinction.

Based on 12 mitochondrial gene sequences, three mitochondrial gene pools were identified: *chinensis/siamea/malaccensis/burmannicoides*, *burmannica/zebrina,* and *microcarpa.* The results are, in part, supported by the findings of previous studies (Boonruangrod et al. 2008; Carreel et al. 2002).

Cytoplasm gene pools of *M. balbisiana* ancestors

Based on 25 chloroplast gene sequences, three chloroplast gene pools were identified in *M.*

balbisiana: *balbisiana* (ITC0080)/Cameroun (ITC0246)/China, Eti Kehel (ITC0271), and *balbisiana* (ITC0545). As reported by Carreel et al. (2002) and Boonruangrod et al. (2008), *M. balbisiana* formed two maternal gene pools. However, comparisons are not possible because only one sample was common between the present and the previous studies (Cameroun [ITC0246]). However, Cameroun formed an independent gene pool (Carreel et al. 2002) and clustered with Singapurii, and Butuhan (Boonruangrod et al. 2008).

Based on the 12 mitochondrial gene sequences, *M. balbisiana* could also be divided into three mitochondrial gene pools: *balbisiana* (ITC0080), Eti Kehel/*balbisiana* (ITC0545)/China, and Cameroun. Boonruangrod et al. (2008) also divided *M. balbisiana* into three mitochondrial gene pools. The material shared by the two studies (Cameroun) formed an independent gene pool. However, Carreel et al. (2002) reported no differences in the mitochondrial genome of *M. balbisiana*.

The current research results support that *M. balbisiana* originated via different evolutionary routes; however, the intraspecific classification of *M. balbisiana* warrants further discussion.

Origin and evolution of banana cultivars based on organelle DNA sequences

Banana cultivars originated from intraspecific crosses of *M. acuminata* or interspecific crosses of *M. balbisiana*. *M. acuminata* provided the A genome and *M. balbisiana* provided the B genome to form a series of banana cultivars of different genotypes: AA, AB, AAA, AAB, ABB, and AAAB. Because most banana cultivars are parthenocarpic and sterile, and parthenogenetic genomes are susceptible to mutations, it is possible that the ancient organelle genome was trapped in existing cultivars remaining more or less unchanged since the initial cultivar formation. Considering that present-day wild types are the offspring of ancient species, comparing these to cultivars may reveal gene pools of common origin. In the present study, 25 chloroplast genes and 12 mitochondrial genes were used to study 18 *M. acuminata* wild types, 5 *M. balbisiana* wild types, and 26 cultivars with different genotypes.

The maternal origin of 26 cultivars with different genotypes was found to follow three patterns. First, the maternal origin of most AA/AAA genotype cultivars was derived from the Ca2 gene pool, which was represented by the wild type *M. acuminata* subspecies *malaccensis*, *burmannica,* and *burmannicoides*. This finding was similar to those of Carreel et al. (2002) and Boonruangrod et al. (2008). Second, the maternal origin of the A-B genotype cultivars (for example, AB, AAB, AAAB) was mostly of Ca3 (*M. acuminata ssp516*) and Ca5 (*M. acuminata* subsp. *microcarpa*) gene pools. Third, when the genome of cultivated bananas contained only one B genome (for example, AB or AAB genotypes), the female parent tended to be *M. acuminata.* Conversely, when the genome of cultivated banana contained two B genomes (for example, the ABB genotype), most of the maternal sources were derived from *M. balbisiana*, consistent with previous studies (Carreel et al., 2002; Boonruangrod et al., 2008).

Three paternal origins were identified for the 26 cultivated bananas tested in the present study: Ma1 (*chinensis*/*siamea*/*malaccensis*/*burmannicoides*), Ma3 (*microcarpa*), and Mb2 (Eti Kehel/*balbisiana* [ITC0545]/China), with Ma3 being the most common paternal gene pool. However, Boonruangrod et al. (2008) and Carreel et al. (2002) reported that the most common paternal gene pool of the A genome was *errans*/*banksii*. Additionally, all tested ABB cultivated bananas belonged to the Mb1 gene pool ('Pisang klutuk wulung'/'Pisang batu'/'Honduras'/'Lal vechi'/'Tani'). Although results cannot be directly compared because they used different materials, the common material, *M. balbisiana* (Cameroun), used in the present and previous studies was not involved in the development of banana cultivars. In addition, in the present study, the AAB genotype 'Cluoi mat' grouped with *M. itinerans* but not with *M. balbisiana* and *M. acuminata*. Therefore, further exploration is needed to determine whether *M. itinerans* also acted as a parent during the hybridization of banana cultivars.

Acknowledgements We would like to thank Prof. Yaoting Wu from Hainan University for providing the plant materials. We would also like to thank Editage (www.editage.com) for English language editing. This study was funded by the Hainan Provincial Natural Science Foundation of China (grant number 318QN249).

Author contributions All authors contributed to the study conception and design. Material preparation, data collection, and analysis were performed by Huimin Feng, You Chen, and Changming Li. The first draft of the manuscript was written by

Huimin Feng. All authors commented on the previous versions of the manuscript. All authors have read and approved the final manuscript.

Funding This study was funded by the Hainan Provincial Natural Science Foundation of China (grant number 318QN249).

Availability of data and material The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

Argent GCG (1976) The wild bananas of Papua new Guinea. Notes Roy Bot Gard Edinb 35:77–114

Baker JG (1893) A synopsis of the genera and species of Musaceae. Ann Bot 7:189–229. https://doi.org/10.1093/aob/os-7.2.189

Barrett CF, Specht CD, Leebens-Mack J, Stevenson DW, Zomlefer WB, Davis JI (2014) Resolving ancient radiations: can complete plastid gene sets elucidate deep relationships among the tropical gingers (Zingiberales)? Ann Bot 113:119–133. https://doi.org/10.1093/aob/mct264

Bekele E, Shigeta M (2011) Phylogenetic relationships between *Ensete* and *Musa* species as revealed by the trnT trnF region of cpDNA. Genet Resour Crop Ev 58:259–269. https://doi.org/10.1007/s10722-010-9568-2

Boonruangrod R, Desai D, Fluch S, Berenyi M, Burg K (2008) Identification of cytoplasmic ancestor gene-pools of *Musa acuminata* Colla and *Musa balbisiana* Colla and their hybrids by chloroplast and mitochondrial haplotyping. Theor Appl Genet 118:43–55. https://doi.org/10.1007/s00122-008-0875-3

Carreel F, Gonzalez de Leon D, Lagoda P, Lanaud C, Jenny C, Horry JP, du Montcel T (2002) Ascertaining maternal and paternal lineage within *Musa* by chloroplast and mitochondrial DNA RFLP analyses. Genome 45:679–692. https://doi.org/10.1139/g02-033

Cheesman EE (1947) Classification of the bananas. II. The Genus Musa 1 Kew Bull 2:106–117. https://doi.org/10.2307/4109207

Cheesman EE (1948) Classification of the bananas. Kew Bull 3:17–28. https://doi.org/10.2307/4119750

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. https://doi.org/10.1186/1471-2148-11-103

Cipriani G, Fiori A, Moroldo M, De Pauli P, Messina R, Testolin R (2003). Screening chloroplast, mitochondrial, and nuclear DNA sequences suitable for taxonomic studies in *Actinidiaceae*. Acta Hortic 610:337–342. https://doi.org/10.17660/ActaHortic.2003.610.44

Cuenoud P, Savolainen V, Chatrou LW, Powell M, Grayer RJ, Chase MW (2002) Molecular phylogenetics of *Caryophyllales* based on nuclear 18S rDNA and plastid *rbcL*, *atpB*, and *matK* DNA sequences. Am J Bot 89:132–144. https://doi.org/10.3732/ajb.89.1.132

D'Hont A, Denoeud F, Aury J et al (2012) The banana (*Musa acuminata*) genome and the evolution of monocotyledonous plants. Nature 488:213–217. https://doi.org/10.1038/nature11241

Davey MW, Gudimella R, Harikrishna JA, Sin LW, Khalid N, Keulemans J (2013) A draft *Musa balbisiana* genome sequence for molecular genetics in polyploid, inter- and intra-specific Musa hybrids. BMC Genomics 14:683. https://doi.org/10.1186/1471-2164-14-683

Demesure B, Sodzi N, Petit RJ (1995) A set of universal primers for amplification of polymorphic noncoding regions of mitochondrial and chloroplast DNA in plants. Mol Ecol 4:129–131. https://doi.org/10.1111/j.1365-294x.1995.tb00201.x

Duminil J, Pemonge MH, Petita RJ (2002) A set of 35 consensus primer pairs amplifying genes and introns of plant mitochondrial DNA. Mol Ecol Notes 2:428–430. https://doi.org/10.1046/j.1471-8286.2002.00263.x

Dumolin-Lapegue S, Pemonge MH, Petit RJ (1997) An enlarged set of consensus primers for the study of organelle DNA in plants. Mol Ecol 6:393–397. https://doi.org/10.1046/j.1365-294X.1997.00193.x

Fauré S, Noyer JL, Carreel F, Horry JP, Bakry F, Lanaud C (1994) Maternal inheritance of chloroplast genome and paternal inheritance of mitochondrial genome in bananas (*Musa acuminata*). Curr Gen 25:265–269. https://doi.org/10.1007/BF00357172

Feng HM, Chen Y, Deng CJ, Wu YT (2009) Distribution of wild *Musa* species. J Fruit Sci 26:361–368 ((**in Chinese**))

Feng HM, Chen Y, Li B, Wu YT (2016) Molecular phylogeny of genus *Musa* determined by simple sequence repeat markers. Plant Genet Resour 14:192–199. https://doi.org/10.1017/S1479262115000222

Gawel NJ, Jarret RL (1991a) Cytoplasmic genetic diversity in bananas and plantains. Euphytica 52:19–23. https://doi.org/10.1007/BF00037852

Gawel NJ, Jarret RL (1991b) Chloroplast DNA restriction fragment length polymorphisms (RFLPs) in *Musa* species. Theor Appl Genet 81:783–786. https://doi.org/10.1007/BF00224990

Gawel NJ, Jarret RL, Whittemore AP (1992) Restriction fragment length polymorphism (RFLP)-based phylogenetic analysis of *Musa*. Theor Appl Genet 84:286–290. https://doi.org/10.1007/BF00229484

Ge XJ, Liu MH, Wang WK, Schaal BA, Chiang TY (2005) Population structure of wild bananas, *Musa balbisiana*, in China determined by SSR fingerprinting and cpDNA PCR-RFLP. Mol Ecol 14:933–944. https://doi.org/10.1111/j.1365-294X.2005.02467.x

Häkkinen M (2008) *Musa itinerans* (Musaceae) and its intraspecific taxa in China. Novon 18:50–60. https://doi.org/10.3417/1055-3177(2007)17[440:NSAVOM]2.0.CO;210.3417/2006162

Häkkinen M (2009) *Musa chunii* Häkkinen, a new species (Musaceae) from Yunnan, China and taxonomic identity of

Musa rubra. J Syst Evol 47:87–91. https://doi.org/10.1111/j.1759-6831.2009.00005.x

Häkkinen M, Hong W (2007) New species and variety of *Musa* (Musaceae) from Yunnan, China. Novon 17:440–446. https://doi.org/10.3417/1055-3177(2007)17[440:NSAVOM]2.0.CO;2

Häkkinen M, Suleiman M, Gisil J (2005). *Musa beccarii* (Musaceae) varieties in Sabah northern Borneo. Acta Phytotax Geobot 56:135–140. https://doi.org/10.18942/apg.KJ00004623239

Häkkinen M (2013). Reappraisal of sectional taxonomy in *Musa* Musaceae. Taxon 62:809–813. https://doi.org/10.12705/624.3

Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS (2018) UFBoot2: Improving the ultrafast bootstrap approximation. Mol Biol Evol 35:518–522. https://doi.org/10.1093/molbev/msx281

INIBAP/CIRAD (1996) Descriptors for banana (Musa spp.). International plant genetic resources institute

Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS (2017) ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods 14:587–589. https://doi.org/10.1038/nmeth.4285

Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol 30:772. https://doi.org/10.1093/molbev/mst010

Kress WJ, Erickson DL (2007) A two-locus global DNA barcode for land plants: the coding *rbcL* gene complements the non-coding *trnH–psbA* spacer region. PLoS ONE 2:e508. https://doi.org/10.1371/journal.pone.0000508

Kress WJ, Wurdack KJ, Zimmer EA, Weigt LA, Janzen DH (2005) Use of DNA barcodes to identify flowering plants. Proc Natl Acad Sci USA 102:8369–8374. https://doi.org/10.1073/pnas.0503123102

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*. Mol Phyl Evol 57:1–10. https://doi.org/10.1016/j.ympev.2010.06.021

Li W, Liu Y, Gao LZ (2017) The complete chloroplast genome of the endangered wild *Musa itineras* (Zingiberales: Musacaea). Conserv Genet Resour 9:667–669. https://doi.org/10.1007/s12686-017-0737-x

Lin YR, Zhu L, Ren S, Yang J, Schertz KF, Paterson AH (1999) A *Sorghum propinquum* BAC library, suitable for cloning genes associated with loss-of-function mutations during crop domestication. Mol Breeding 5:511–520. https://doi.org/10.1023/A:1009642201139

Linnaeus C (1753) *Musa* L. Species Plantarum, vol. 2. Impensis Laurentii Salvii, Stockholm, p. 1043.

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. https://doi.org/10.2307/27757047

Martin G, Baurens FC, Cardi C, Aury JM, D'Hont A (2013) The complete chloroplast genome of banana (*Musa acuminata*, Zingiberales): insight into plastid monocotyledon evolution. PLoS ONE 8:e67350. https://doi.org/10.1371/journal.pone.0067350

Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ (2015) IQ-TREE: A fast and effective stochastic algorithm for estimating Maximum-Likelihood phylogenies. Mol Biol Evol 32:268–274. https://doi.org/10.1093/molbev/msu300

Nicolosi E, Deng ZN, Gentile A, La Malfa S, Continella G, Tribulato E (2000) Citrus phylogeny and genetic origin of important species as investigated by molecular markers. Theor Appl Genet 100:1155–1166. https://doi.org/10.1007/s001220051419

Nwakanma DC, Pillay M, Okoli BE, Tenkouano A (2003) Sectional relationships in the genus *Musa* L. inferred from the PCR-RFLP of organelle DNA sequences. Theor Appl Genet 107:850–856. https://doi.org/10.1007/s00122-003-1340-y

Oxelman B, Liden M, Berglund D (1997) Chloroplast rps 16 intron phylogeny of the tribe *Sileneae* (Caryophyllaceae). Plant Syst Evol 206:393–410. https://doi.org/10.1007/BF00987959

Paterson AH, Brubaker C, Wendel JFA (1993) A rapid method for extraction of cotton (*Gossypium* spp.) genomic DNA suitable for RFLP or PCR analysis. Plant Mol Biol Rep 11:122–127. https://doi.org/10.1007/BF02670470

Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Hohna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP (2012) MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst Biol 61:539–542. https://doi.org/10.1093/sysbio/sys029

Rozas J, Ferrer-Mata A, Sánchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE, Sánchez-Gracia A (2017) DnaSP 6:DNA Sequence polymorphism analysis of large datasets. Mol Biol Evol 34:3299–3302. https://doi.org/10.1093/molbev/msx248

Sagot P (1887) Sur le Genre Bananier. Bull Soc Bot 34:328–330. https://doi.org/10.1080/00378941.1887.10830263

Scarcelli N, Barnaud A, Eiserhardt W, Treier UA, Seveno M, d'Anfray A, Vigouroux Y, Pintaud JC (2011) A set of 100 chloroplast DNA primer pairs to study population genetics and phylogeny in Monocotyledons. PLoS ONE 6:e19954. https://doi.org/10.1371/journal.pone.0019954

Shetty SM, Md Shah MU, Makale K, Mohd-Yusuf Y, Khalid N, Othman RY (2016) Complete chloroplast genome sequence of corroborates structural heterogeneity of inverted repeats in wild progenitors of cultivated bananas and plantains. Plant Genome 9:1–14. https://doi.org/10.3835/plantgenome2015.09.0089

Simmonds NW (1960) Notes on banana taxonomy. Kew Bull 14:198–212. https://doi.org/10.2307/4114778

Simmonds NW, Shepherd K (1955) The taxonomy and origins of the cultivated bananas. Bot J Linn Soc 55:302–312. https://doi.org/10.1111/j.1095-8339.1955.tb00015.x

Simmonds NW (1966) Bananas, 2nd edn. Tropica Agri Series, London, UK, Longmans

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. Mol Phylogenet Evol 36:509–522. https://doi.org/10.1016/j.ympev.2005.04.018

Sotto RC, Rabara RC (2000) Morphological diversity of *Musa balbisiana* Colla in the Philippines. Infomusa 9:28–30

Stover RH, Simmonds NW (1987) Bananas, 3rd edn. UK, Longman, London

Swangpol S, Volkaert H, Sotto RC, Seelanan T (2007) Utility of selected non-coding chloroplast DNA sequences for lineage assessment of *Musa* interspecific hybrids. J Biochem Mol Biol 40:577–587. https://doi.org/10.5483/bmbrep.2007.40.4.577

Swofford DL (2002) PAUP* 4.0: Phylogenetic analysis using parsimony and other methods. Beta version 4.0 b10[M]. Sunderland: Sinauer Associates

Taberlet P, Gielly L, Pautou G, Bouvet J (1991) Universal primers for amplification of three non-coding regions of chloroplast DNA. Plant Mol Biol 17:1105–1109. https://doi.org/10.1007/BF00037152

Takanori O, Ohmi O (2000) Intra- and inter-specific phylogeny of wild *Fagopyrum* (Polygonaceae) species based on nucleotide sequences of noncoding regions in chloroplast DNA. Am J Bot 87:573–582. https://doi.org/10.2307/2656601

Umali R, Nakamura I (2003) Identification of dCAPS markers that discriminate A and B cytoplasms in banana (*Musa* Spp.). Plant Biotechnol 20:159–164. https://doi.org/10.5511/plantbiotechnology.20.159

Vaidya G, Lohman DJ, Meier R (2011) SequenceMatrix: concatenation software for the fast assembly of multi-gene datasets with character set and codon information. Cladistics 27:171–180. https://doi.org/10.1111/j.1096-0031.2010.00329.x

Wong C, Kiew R, Argent G, Set O, Lee SK, Gan YY (2002) Assessment of the validity of the sections in *Musa* (Musaceae) using AFLP. Ann Bot 90:231–238. https://doi.org/10.1093/aob/mcf170

Yan ZX (2009) Study on mitochondrial respiratory chain of Zidao type cytoplasmic male sterility rice (Oryza sativa L). Wuhan University, Wuhan

Yoon HS, Hackett JD, Bhattacharya D (2002) A single origin of the peridinin-and fucoxanthin-containing plastids in dinoflagellates through tertiary endosymbiosis. Proc Natl Acad Sci USA 99:11724–11729. https://doi.org/10.1073/pnas.172234799

Zhang X (2011) Study on differences of mitochondrial genome between cytoplasmic male sterile line and maintainer line in cotton (Gossypium hirsutum L). Chinese Academy of Agricultural Sciences, Beijing

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.