*Journal of Biochemistry and Molecular Biology*, Vol. 40, No. 4, July 2007, pp. 577-587

JOURNAL OF
Biochemistry and Molecular Biology

# Utility of Selected Non-coding Chloroplast DNA Sequences for Lineage Assessment of *Musa* Interspecific Hybrids

**Sasivimon Swangpol[1,2,†], Hugo Volkaert[3,4], Rachel C. Sotto[5] and Tosak Seelanan[1,2,*]**

[1]Biological Sciences Program, Faculty of Science, Chulalongkorn University, Phayathai Road, Pathumwan, Bangkok 10330, Thailand
[2]Department of Botany, Faculty of Science, Chulalongkorn University, Phayathai Road, Pathumwan, Bangkok 10330, Thailand
[3]Center for Agricultural Biotechnology, Kasetsart University, Kamphaeng Saen Campus, Kamphaeng Saen, Nakhorn Pathom 73140, Thailand
[4]NSTDA-BIOTEC, Thailand Science Park, 113 Phahonyothin Road, Klong 1, Klong Luang, Pathum Thani 12120, Thailand
[5]Plant Biology Division, Institute of Biological Sciences, College of Arts and Sciences, University of the Philippines, Los Baños, College, Laguna 4031, Philippines

*Received 2 October 2006, Accepted 14 March 2007*

**Single-copy chloroplast loci are used widely to infer phylogenetic relationship at different taxonomic levels among various groups of plants. To test the utility of chloroplast loci and to provide additional data applicable to hybrid evolution in *Musa*, we sequenced two introns, *rpl16* and *ndhA*, and two intergenic spacers, *psaA-ycf3* and *petA-psbJ-psbL-psbF* and combined these data. Using these four regions, *Musa acuminata* Colla (A)- and *M. balbisiana* Colla (B)-containing genomes were clearly distinguished. Some triploid interspecific hybrids contain A-type chloroplasts (the AAB/ABB) while others contain B-type chloroplasts (the BBA/BBB). The chloroplasts of all cultivars in 'Namwa' (BBA) group came from the same wild maternal origin, but the specific parents are still unrevealed. Though, average sequence divergences in each region were little (less than 2%), we propose that *petA-psbJ* intergenic spacer could be developed for diversity assessment within each genome. This segment contains three single nucleotide polymorphisms (SNPs) and two indels which could distinguish diversity within A genome whereas this same region also contains one SNP and an indel which could categorize B genome. However, an inverted repeat region which could form hairpin structure was detected in this spacer and thus was omitted from the analyses due to their incongruence to other regions. Until thoroughly identified in other members of Musaceae and Zingiberales clade, utility of this inverted repeat as phylogenetic marker in these taxa are cautioned.**

**Keywords:** Chloroplast DNA, Interspecific hybrids, *Musa acuminata*, *Musa balbisiana*, Sequence polymorphisms

## Introduction

A long-held hypothesis is that most cultivated bananas are derived from or are hybrids of two *Musa* species, *M. acuminata* Colla (AA) and *M. balbisiana* Colla (BB) (Cheesman, 1947; Simmonds, 1962, 1995). Based on chromosome numbers and morphological determination, *Musa* cultivars and hybrids arisen from these wild species are traditionally divided into seven genomic groups-namely, AA, BB, AAA, AAB, ABB, BBB, ABBB (Simmonds and Shepherd, 1955; Chomchalow and Silayoi, 1984). However, the exact origins and relationships among these bananas are still uncertain due to very broad genetic diversity within *M. acuminata*, and yet lesser known about diversity in *M. balbisiana*. Several subspecies have been described within *M. acuminata*, i.e. subspp. *banksii*, *burmanica*, *burmanicoides*, *malaccensis*, *microcarpa*, *siamea*, *truncata*, *errans* and *zebrina*. On the other hand, no infraspecific entity has ever been designated within *M. balbisiana* (INIBAP 2003). Nevertheless, various investigations revealed that *M. balbisiana* is a polymorphic species, e.g. isozyme analysis by Lebot *et al.* (1993), sequence-tagged microsatellite site analyses by Kaemmer *et al.* (1997), and simple sequence repeat (SSR) and cpDNA PCR-RFLP analyses by Ge *el al.* (2005).

Geographic origins of cultivated hybrids between *M. acuminata* and *M. balbisiana* is subject of debates. Simmonds

†Present address: Department of Plant Science, Faculty of Science, Mahidol University, Rama VI Road, Ratchathewee, Bangkok 10400, Thailand

*To whom correspondence should be addressed.
Tel: 66-2218-5508; Fax: 66-2218-5502
E-mail: tosak.s@chula.ac.th

(1962) suggested that India were centers of hybrid origins where AA cultivars, possibly originated in Malaya, were carried into *M. balbisiana* distribution areas along Indian east coast. However, he did not believed that wild *M. balbisiana* was native to Thailand, Malaysia, or Indonesia because - although is commonly cultivated for male bud consumption and its leaves for wrappings- wild *M. balbisiana* lacked any specific vernacular names.

Alternatively, De Langhe and De Maret (1999) proposed that the location of hybridization could be in eastern Indonesia and the Bismarck Archipelago where *M. acuminata* subsp. *banksii* was hybridized to anthropogenic *M. balbisiana* brought in by Neolithic Austronesia settlers between 4,500 and 3,500 years ago (Simmonds 1962; Argent 1976; De Langhe and De Maret 1999; Birds *et al.*, 2003). They believed that subsequent introductions of the hybrids into other tropical areas over the world, also back to mainland Asia, should be attributed to ancient sea travelers.

Though none of *balbisiana* genome-containing bananas are significant in the world trade, they are invaluable for rural people in Southeast Asia and other tropical countries. Production of bananas in Southeast Asia, especially for local consumption, is mostly based on triploid hybrid cultivars such as 'Saba' in the Philippines, 'Hin' in Thailand and 'Pisang Kepok' in Indonesia; 'Namwa' in Thailand and its synonym 'Pisang Awak' in Malaysia and Indonesia; and 'Chuoi Tay' in Vietnam (see Valmayor *et al.* 2000 for names). Knowledge of the relationships within and among wild accessions, diploid and triploid cultivars, is required for the successful selections of parental resistance to pests and diseases in the breeding programs (INIBAP 2001).

The study of non-coding chloroplast (cp) sequences, has proved to be useful in plant phylogenetic analyses at lower taxonomic levels (Gielly and Taberlet, 1994). Several investigations have reported on using cpDNA sequence polymorphisms to determine inter- and infra-specific variation in different plant groups e.g. in *Cicer* in Fabaceae (Javadi, 2004), *Panax* (Kim and Lee, 2004) and *Jasminum* (Kim and Lee, 2005). Lee and Wen (2003) succesfully used the *trnC-trnD* to access interspecific diversity within a wide range of plant families including *Aralia* (Araliaceae), *Nelumbo* (Nelumbonaceae), *Nolana* (Solanaceae), *Prunus* (Rosaceae), and others. Recently, Kress *et al.* (2005) proposed the use of short (~450 bp) DNA sequences of plastid *trnH-psbA* intergenic spacer, in conjunction with nuclear internal transcribed spacer, as DNA barcoding for interspecific discrimination of flowering plants. Single nucleotide substitutions, indels, and inverted repeats are reported being very useful for assessment of species diversity (Kim and Lee, 2005; Kress *et al.*, 2005; Shaw *et al.*, 2005). In *Musa*, Careel *et al.* (2002) found that RFLP (Restriction Fragment Length Polymorphism) analyses of cpDNA could identify maternally related cultivars.

We choose to sequence non-coding regions of the chloroplast genome that may provide data for potential resolution of the origin of the *Musa* complex, especially *M. balbisiana*. Sequences from multiple banana accessions showed that there are more than one maternal haplotypes, not only among, but within each genome groups. Relationships of these haplotypes were inferred from network reconstruction by statistical parsimony algorithm and from most parsimonious trees generated by maximum parsimony method.

## Materials and Methods

**Plant materials.** Forty-six accessions of *M. acuminata*, *M. balbisiana* (both from the section *Musa*, basic chromosome number x = 11), and cultivars and hybrids from these two species were chosen to represent eight genomic constitutions and to reflect their geographical distributions (Table 1). Four accessions from the section *Australimusa* (basic chromosome number, x = 10), i.e. *M. jackeyi*, *M. textilis*, and *Musa* [Fehi group] cultivars, 'Aata' and 'Tongkat Langit', were included as outgroup taxa. Young curling leaves in pseudostems were collected for DNA extraction using DNeasy Plant Mini Kit (Qiagen) according to the manufacturer's instructions. In addition, 20 DNA samples (Table 1, labeled NEU, CIRAD, or ITC) were generously provided by Dr. Françoise Carreel, CIRAD-FLHOR, Guadeloupe, and by the International Transit Center (ITC) of the International Network for the Improvement of Banana and Plantain (INIBAP), Leuven, Belgium. Two more samples from the Philippines (Table 1, labeled IPB) were supplied by the Institute of Plant Breeding, Los Baños, College, Laguna, Philippines.

**DNA amplification and sequencing.** To screen for appropriate levels of sequence divergence in the *Musa* plastid genome, we designed several primers from published sequences across a wide range of plant groups and compared polymorphisms in amplifying fragments of *M. acuminata*, *M. balbisiana*, and a few *Australimusa* species. Finally, four primer sets (Table 2), i.e. the *ndhA* intron, *rpl16* intron, *psaA-ycf3* intergenic spacer and *petA-psbJ-psbL-psbF* intergenic spacer (N-IN, R-IN, PY-S and AF-S, respectively), were selected. PCR amplification was carried out in the 50 mL reaction mixture containing 20 ng DNA template, 200 mM dNTP, 2.5 mM $MgCl_2$, 0.2 mM each primer, 1× *Taq* Buffer, and 1 unit of *Taq* DNA polymerase (Qiagen). Double-stranded PCR amplification was done in the P-100 thermocycler (MJR Inc.) using the condition as following: 30 cycles of 5 min denaturing at 95°C, 2 min annealing at 50°C, 3 min extension at 72°C, and additional 10 min at 72°C afterward.

Detection of cp-haplotypes was done using PCR-RF-SSCP (Polymerase Chain Reaction-Restriction Fragment-Single Strand Conformation Polymorphism or PRS) analyses. The procedure was as following. Except for the PCR products of PY-S, the R-IN and N-IN fragments were digested with *Hinf*I restriction enzyme, and the AF-S fragment was digested with *Taq*I (New England BioLabs, Inc, USA), using conditions recommended by the supplier. Two volumes of denaturing solution (95% (v/v) formamide, 0.025% bromphenol blue, 0.025% (w/v) xylene cyanol and 10 mM NaOH) were mixed with PCR or digested products, then denatured for 10 min at 95°C, and immediately placed on ice to stabilize single strands. Then 3.5 mL aliquots were loaded on a non-denaturing

**Table 1.** List of *Musa* accessions, their genome composition, origins and sources of accessions used in this study

| # | Species/Cultivar | Genome[a] | Origin[b] | Source[c] |
|---|---|---|---|---|
| 1 | *M. jackeyi* | TTw | Northeast Queensland, Australia | NEU0389 |
| 2 | *M. textilis* | TTw | Philippines | NEU0001 |
| 3 | *Musa* [Fehi Group] 'Aata' | TT | Papua New Guinea | NEU0385 |
| 4 | *Musa* [Fehi Group] 'Tongkat Langit' | TT | Indonesia | NEU0388 |
| 5 | *M. acuminata* subsp. *banksii* 'Madang' | AAw | Papua New Guinea | ITC0254 via CIRAD |
| 6 | *M. acuminata* subsp. *malaccensis* 1 | AAw | Malaysia | ITC0250 via CIRAD |
| 7 | *M. acuminata* subsp. *malaccensis* 2 | AAw | Ranong, Thailand | TM068 |
| 8 | *M. acuminata* subsp. *siamea* 1 | AAw | Phrae, Thailand | KU s.n. |
| 9 | *M. acuminata* subsp. *siamea* 2 | AAw | Mae Hong Son, Thailand | TM051 |
| 10 | *Musa* 'Grande Naine' | AAA | ? | NEU0172 |
| 11 | *Musa* 'Hom Thong' | AAA | Thailand | KU s.n. via QSG |
| 12 | *Musa* 'Kofi' | AAB | ? | ITC0912 |
| 13 | *Musa* 'Nga Chang' | AAB | Nonthaburi, Thailand | TM139 |
| 14 | *Musa* 'Popoulou' | AAB | Papua New Guinea | ITC0335 |
| 15 | *Musa* 'Auko' | AB | ? | ITC0983 |
| 16 | *Musa* 'Hak Muk Khao' | ABB | Thailand | KU s.n. |
| 17 | *Musa* 'Hak Muk Khiao' | ABB | Thailand | KU s.n. |
| 18 | *Musa* 'Hin' (Yala) | ABB | Yala, Thailand | TM128 |
| 19 | *Musa* 'Hin' (Songkhla) | ABB | Songkhla, Thailand | TM129 |
| 20 | *Musa* 'Pisang Abuperak' | ABB | Indonesia | ITC0056 |
| 21 | *Musa* 'Pisang Gajih Merah' | ABB | Indonesia | ITC0677 |
| 22 | *Musa* 'Saba' (ITC) | ABB | Philippines | ITC0116 |
| 23 | *Musa* 'Hin' (Nakhorn Sri Thammarat) | BBA | Nakhorn Sri Thammarat, Thailand | KU s.n. |
| 24 | *Musa* 'Moe Nang' | BBA | Thailand | ITC1066 |
| 25 | *Musa* 'Namwa Dum' | BBA | Thailand | KU s.n. |
| 26 | *Musa* 'Namwa Khao' | BBA | Phrae, Thailand | KU s.n. |
| 27 | *Musa* 'Namwa Khom' (dwarft) | BBA | Nakhorn Ratchasima, Thailand | KU s.n. |
| 28 | *Musa* 'Namwa Mali-Ong' | BBA | Chantaburi, Thailand | KU s.n. |
| 29 | *Musa* 'Namwa Nuan' | BBA | Ang Thong, Thailand | KU s.n. |
| 30 | *Musa* 'Namwa Sai Daeng' | BBA | Thailand | KU s.n. |
| 31 | *Musa* 'Namwa Sai Dum' | BBA | Thailand | KU s.n. |
| 32 | *Musa* 'Namwa Sai Lueang' | BBA | Thailand | KU s.n. |
| 33 | *Musa* 'Saba' (IPB) | BBA | Philippines | IPB-P93 |
| 34 | *Musa* 'Pli Hai' | BBBA | Chiang Rai, Thailand | KU s.n. |
| 35 | *Musa* 'Thipparot' | BBBA | Sisaket, Thailand | KU s.n. |
| 36 | *M. balbisiana* (wild) 1 | BBw | Nan, Thailand | TM123 |
| 37 | *M. balbisiana* (wild) 2 | BBw | Cagayan, Philippines | IPB-P28 |
| 38 | *M. balbisiana* 'ITC626' | BB | ? | ITC0626 |
| 39 | *M. balbisiana* 'Cameroon' | BB | ? | ITC0246 via CIRAD |
| 40 | *M. balbisiana* 'Honduras' | BB | ? | ITC0247 via CIRAD |
| 41 | *M. balbisiana* 'Lal Velchi' | BB | ? | NEU0051 |
| 42 | *M. balbisiana* 'Montpellier' | BB | ? | ITC0212 |
| 43 | *M. balbisiana* 'Pang La Nuan' | BB | Trang, Thailand | TM121 |
| 44 | *M. balbisiana* 'Pisang Klutuk Wulung' | BB | Indonesia | NEU0054 |
| 45 | *M. balbisiana* 'Tani' (Chachaeng Sao) | BB | Chachaeng Sao, Thailand | TM083 |
| 46 | *M. balbisiana* 'Phalo' | BB | Chaiyaphum, Thailand | KU s.n. |
| 47 | *M. balbisiana* 'Phong La' | BB | Nakhorn Si Thammarat, Thailand | KU s.n. |
| 48 | *M. balbisiana* 'Pa' | BB | Phrae, Thailand | KU s.n. |
| 49 | *M. balbisiana* 'Tani' (Sukhothai) | BB | Sukhothai, Thailand | TM127 |
| 50 | *Musa* 'Lep Chang Ku' | BBB | Phatthalung, Thailand | KU s.n. |

[a]Genome compositions of species/cultivar names used in this study are based on Chomchalow and Silayoi (1984); INIBAP (2003); Silayoi and Babprasert (1983); Valmayor (2000). Letter 'w' = 'wild'.

[b]Except for TM accessions which were recently collected, origins were from the oldest known collecting records and '?' indicates unknown origins

[c]NEU and CIRAD = Centre de Coopération Internationale en Recherche Agronomique pour le Dévelopement, Guadeloupe; ITC = International Transit Center, International Network for the Improvement of Banana and Plantain (INIBAP), Leuven, Belgium; TM = Germplasm collected by the authors in Thailand; KU = *ex situ* collection at Kasetsart University, Pak Chong Research Station, Nakhorn Ratchasima, Thailand, consult Silayoi and Babprasert (1983) for accession details; QSG = Queen Sirikit Garden, Bangkok, Thailand; IPB = Institute of Plant Breeding, University of the Philippines, Los Baños, College, Laguna, Philippines

**Table 2** Primers for amplifications

| Name | Nucleotide sequence | Tm (°C) | Amplified region | Sequence code | Location[a] | Protein encoded |
|---|---|---|---|---|---|---|
| *rpl16*-F | 5'GCTATGCTTAGTGTGTGACT3' | 55.3 | *rpl16* intron | R-IN | LSC | ribosomal protein L16 |
| *rpl16*-R | 5'CATTCTTCCTCTATGTTGTTT3' | 52.0 | | | | |
| *psaA*-F | 5'AAATCGTGAGCATCAGCATG3' | 54.7 | *psaA-ycf3* exon3 intergenic spacer | PY-S | LSC | photosystem I P700 apoprotein A1 and photosystem I assembly protein Ycf3 |
| *psaA*-R | 5'CCGAGGAGAACAGGCCATTC3' | 58.6 | | | | |
| *ndhA*-F | 5'GCTGCTCAATCTATTAGTTATGA3' | 55.3 | *ndhA* intron | N-IN | SSC | NADH dehydrogenase subunit 1 |
| *ndhA*-R | 5'TGTGCTTCAACTATATCAACTGT3' | 53.7 | | | | |
| *petA*-F | 5'TATGAAAATCCACGAGAAGC3' | 51.3 | *petA-psbJ-psbL-psbF* intergenic spacer | AF-S | LSC | cytochrome f, photosystem II protein J and L, and cytochrome b559 beta chain, respectively |
| *petA*-R | 5'TATCAGCAATGCAGTTCATC3' | 51.7 | | | | |

[a]LSC = large single copy chloroplast region and SSC = small single copy chloroplast region

polyacrylamide gel (Sequagel MD, National Diagnostics). Electrophoresis was performed in 1× TBE buffer using Hoefer SQ3 Sequencer (Amersham Pharmacia Biotech) and run in a 4°C refrigerator at constant 8 watt for 12-14 h. The gel was then silver stained using standard procedure (Guillemette and Lewis, 1983). From PRS results, accessions showed similar pattern were classified as having the same cp-haplotypes. One to six accessions depending on geographical distributions from each cp-haplotype were chosen for DNA sequencing by automated DNA sequencer (Macrogen, Inc.).

**Sequence alignments and hairpin structural analyses.** The DNA sequences were verified and correction were made when necessary using Chromas version 1.45 (McCarthy, 1996). DNA sequence alignment was manually done using GeneDoc version 2.6.002 (Nicholas and Nicholas, 1997). Nucleotide diversity was determined by analyses of the aligned matrix (excluding gaps and missing data) using the DnaSP version 4.00.2 (Rozas *et al.*, 2003). Predictions of RNA folding structure and minimum free energy were calculated using Mfold web server (Zuker, 2003). Variable regions with mononucleotide repeats and a few sites where alignments cannot be unambiguously done were omitted from further phylogenetic analyses (Zhang, 2000).

**Treatment of Indels.** Reports have shown that indels are widespread in land plants and often phylogenetically informative (Gielly and Taberlet, 1994; Kim and Lee, 2005) though may be homoplasious at low taxonomic ranks (Golenberg, 1993, Shaw *et al.*, 2005). To test the phylogenetic utilities of indels, we had analysed nucleotide substitutions and indels separately and found that both were equally useful (data not shown). In our analyses, each indel, regardless of its size, was accounted as one mutational event and treated as fifth state. Nucleotide substitutions within indels were treated as additional events. For maximum parsimony analyses, three different gap-coding were employed: (i) retaining all gaps and treating as missing data, (ii) complete deletion of indel regions, and (iii) removing all indels from the data matrix, identifying every potential mutation event and applying as unordered multistate characters (0/1/2) for each event, then adding the characters back to the matrix (Baum *et al.*, 1994; Peralta and Spooner, 2001).

**Analyses of genetic relationships.** Incongruence length difference (ILD) tests (Farris *et al.*, 1995) - as implemented under partition homogeneity test in the PAUP* version 4.0b10 (Swofford 2002) - were conducted to determine whether the four regions were statistically incongruent. Gene genealogies were estimated using a network reconstruction by statistical parsimony algorithm (Templeton *et al.*, 1992) generated by the TCS version 1.21 (Clement *et al.*, 2000). Furthermore, phylogenetic trees were reconstructed by the maximum parsimony (MP) method with all characters equally weighted in the PAUP* version 4.0b10 (Swofford 2002). Heuristic search with 100 random sequence addition, tree bisection-reconnection (TBR) branch-swapping (Steepest descent option in effect), and MULPARS options were used. To evaluate branch supports, bootstrap analyses with 500 replicates were conducted.

## Results

**cpDNA sequence and hairpin structural analyses.** The combined cpDNA sequences of the accessions consisted of mostly non-coding regions (72%) (Table 3). Total length of four cpDNA regions ranged from 5,460-5,506 bp in the A accessions and 5,583-5,631 bp in the B accessions. After alignment (GenBank alignment number ALIGN_000853), the sequences resulted in a final data matrix of 5,766 bp by introducing 29 gaps, among which 10 were due to outgroup indels. The total gap length was 408 bp, which is about 7.1% of the aligned sequence length. Thirteen gaps were small indels (2-10 bp) and the rest were large, ranging from 11-55 bp.

Notwithstanding, an 86-bp inverted repeats was located in AF-S fragment, situated in *petA-psbJ* intergenic spacer and 29-bp downstream of *psbJ* coding region. This repeat could form a hairpin secondary structure with a 24-bp loop and 31-bp stem. Three types of loops were found in *Musa* accessions (Fig. 1), namely, (i) INVT found in the *Australimusa* bananas, (ii) INV1, differing from INVT by a single nucleotide substitution, found in *M. acuminata* subsp. *malaccensis* 2 from Ranong Province, Thailand, and ABB 'Hak Muk

**Table 3.** Quantitative data of *Musa* combined chloroplast DNA analysis. Each region contains data regarding: GenBank accession numbers, position on data matrix, length comparison, the number of PICs or Potentially Informative Characters (total nucleotide substitution and indels) and nucleotide diversity (percentage of total PICs per aligned length)

| Sequence Code[a]/ GenBank Accession# | Region | Position on Data Matrix[b] | Aligned Length (bp) | Total PICs | Nucl. Diversity (%) |
|---|---|---|---|---|---|
| R-IN/ AY855875- AY855904 | *rpl16* exons | 1-43<br>1111-1122 | 55 | 0 | 0.00 |
| | *rpl16* intron | 44-1110 | 1067 | 18 | 1.69 |
| PY-S/ AY870111-AY870140 | *psaA* | 1123-1289 | 167 | 0 | 0.00 |
| | *psaA-ycf3* IGS | 1290-1970 | 681 | 10 | 1.47 |
| | *ycf3* | 1971-2124 | 154 | 0 | 0.00 |
| N-IN/ AY870767-AY870796 | *ndhA* exons | 2125-2164<br>3325-3385 | 101 | 0 | 0.00 |
| | *ndhA* intron | 2165-3324 | 1160 | 38 | 3.28 |
| AF-S/ AY874832-AY874861 | *petA* | 3386-4219 | 834 | 10 | 1.20 |
| | *petA-psbJ* IGS | 4220-5350 | 1131 | 29 | 2.56 |
| | *psbJ* | 5351-5473 | 123 | 0 | 0.00 |
| | *psbJ-psbL* IGS | 5474-5598 | 125 | 3 | 2.40 |
| | *psbL* | 5599-5715 | 117 | 2 | 1.71 |
| | *psbL-psbF* IGS | 5716-5737 | 22 | 1 | 4.55 |
| | *psbF* | 5738-5766 | 29 | 0 | 0.00 |
| Total | | | 5766 | 111 | 1.93 |

[a]Sequence code appears according to Table 2.
[b]Complete data matrix is presented in GenBank alignment number ALIGN_000853.

Khiao', and (iii) INV2 found commonly in the rest of the A and B accessions. Inverted repeats were also found within this *petA-psbJ* intergenic spacer, though at different length, of *Boesenbergia curtisii* and *Hedychium biflorum* (BOC and HEB, respectively, in Fig. 1; J. Techaprasan, pers. comm.).

Fifteen haplotypes, differing by single nucleotide polymorphisms (SNP) and indels, were detected by cpDNA sequence comparison among the studied accessions (Table 4). All wild accessions represented unique haplotypes. The edible cultivars were classified into eigth different haplotypes, five of which (H8-H12, Table 4) represent A genome-like accessions. The tetraploid 'Pli Hai' showed a unique haplotype while the rest of the cultivars showed B genome-like haplotypes (H13-H15, Table 4).

**Network analysis.** Thirty selected accessions from fifteen cp-haplotypes were connected into a network as shown in Fig. 2. The cp-haplotype network was arranged into three distinct clades representing *M. acuminata* (A), *M. balbisiana* (B), and the section *Australimusa* (T). The A clade consisted of ten cp-haplotypes, arranged into one lineage (*M. acuminata* subsp. *siamea* 2 from Mae Hong Son Province, Thailand) and two subclades. *Musa* (AAA) 'Hom Thong' was placed within the subclade of *M. acuminata* subsp. *malaccensis* and *M. acuminata* subsp. *banksii*, whereas all hybrid cultivars, *i.e.*, AB, AAB, and ABB, were placed in the subclade of *M. acuminata* subsp. *siamea* 1 from Prae Province. Within the B clade, there were arranged from innermost *Musa* 'Pli Hai', to wild *M. balbisiana* from the Philippines, and wild *M. balbisiana* from Nan Province at the network tip. The cp-haplotype of *M. balbisiana* from the Philippines was also found in 'Tani' (Sukhothai), 'Lal Velchi', 'Honduras', 'ITC626', and three ABB hybrids, 'Saba' (IPB), 'Namwa Nuan', and 'Hin' (Nakhorn Sri Thammarat). Likewise, the cp-haplotype of *M. balbisiana* from Nan Province was also found in 'Pisang Klutuk Wulung', 'Cameroon', a BBB 'Lep Chang Ku', and an ABB hybrid, 'Moe Nang'. It should be noted that two 'Saba' and 'Hin' accessions were placed in two different clades: 'Saba' (ITC) and 'Hin' (Yala) in the A clade while 'Saba' (IPB) and 'Hin' (Nakhorn Sri Thammarat) in the B clade.

**Phylogenetic analyses.** The four chloroplast regions were not statistically incongruent when the hairpin region was excluded from the *petA* region. The MP analyses of the combined sequence using three different methods of gap coding yielded comparable topologies. In general, the *Musa* section formed two subclades, A and B, with bootstrap value of 100%. More parsimony-informative (Pi) characters were identified when gaps were recoded. When indels were coded as missing, there were 138 variable sites with 114 Pi sites. The analysis yielded 3,288 equally most-parsimonious trees with tree length of 144, the consistency index (CI), excluding uninformative characters, of 0.9500 and the retention index (RI) of 0.9861.

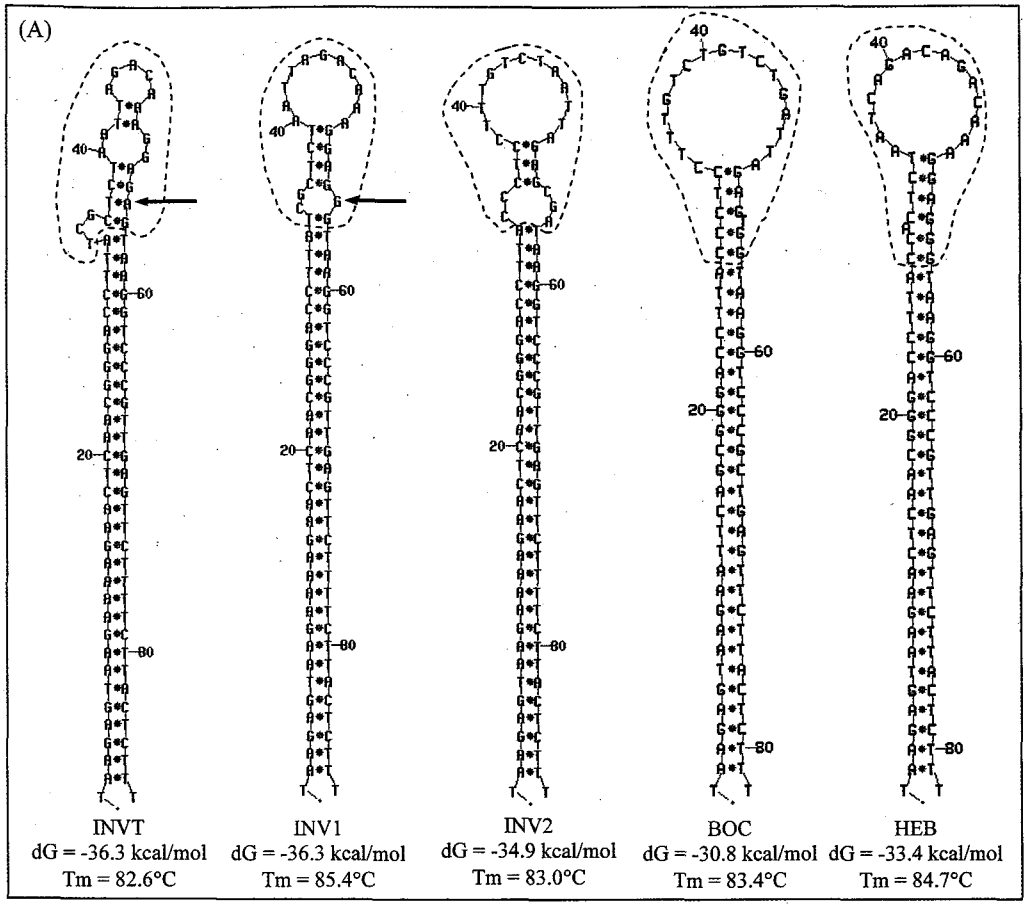


**Fig. 1.** Hairpin structure of *pet*A (3')-(3') *psb*J intergenic spacer region in *Musa* (INVT, INV1, and INV2), *Boesenbergia curtisii* (BOC) and *Hedychium biflorum* (HEB). INVT was found in the Australimusa bananas, INV1 in *M. acuminata* subsp. *malaccensis* 2 from Ranong Province, Thailand, and *Musa* 'Hak Muk Khieo' and INV2 occurred in the rest of the banana accessions. (A) Dot lines surround comparable loop regions of the different structures. Notice the single nucleotide difference between INVT and INV1 (arrows) causes structural change of hairpins. (B) Short 5-bp inverted repeats within the 24-bp loop are underlined. A/G single substitution difference between INVT and INV1 are shaded. Arrows highlighted nucleotide changes.

When indels were completely excluded, there were 122 variable sites of which 100 were Pi. There were 776 equally most-parsimonious trees with tree length = 127, CI (excluding uninformative characters) = 0.9524, and RI = 0.9873. The A and B clades were arranged as in previous analysis (data not shown). When indels were removed, recoded as unordered multistate characters, and added back to the data matrix (Table 4), there were 152 variable sites, of which 112 were Pi. The analysis resulted in 18 most-parsimonious trees - one of which is shown in Fig. 3 - with tree length = 167, CI (excluding uninformative characters) = 0.9241 and RI = 0.9800. In the A clade, wild *M. acuminata* subsp. *siamea* 2 was separated from the other A accessions based on four unique changes. The remaining A accessions consisted of *M. acuminata* subsp.

**Table 4.** Fifty-seven parsimony informative sites excluding outgroup found within four chloroplast non-coding regions of *Musa*. Unordered multistate characters indicated for indels are 0=deletion, 1=insertion, and 2=insertion with extra nucleotide difference. *M a.* = *Musa acuminata*, *M.b.* = *Musa balbisiana*. Sites 1-11 are in R-IN, sites 12-19 are in PY-S, sites 20-37 are in N-IN, and sites 38-57 are in AF-S. Asterisk (*) indicate *Musa* accessions 'Saba' and 'Hin', which are separated in two genome groups

| Haplotypes: Genome | Accession names | Pi Sites | | | |
|---|---|---|---|---|---|
| | | R-IN | PY-S | N-IN | AF-S |
| | | 1 11111111 2 2222222222 2222223 334 4444444444 4555555 | | | |
| | | 12234455660 3455667 82 2233355555 7888890 560 2234467888 9022567 | | | |
| | | 54910426097 8247344 16 7803602258 1036704 186 3573518369 2412042 | | | |
| | | 98649592267 8435190 29 0912820170 8698679 451 6576292165 8578948 | | | |
| H1: TT1 | *M. jackeyii*/ *Musa* 'Aata' | 0C01TG1CA0 CC1TAT0CC C G1GA1AAG0C CG100GT CTG A0TG0GT001 GCA0AGC | | | |
| H2: TT2 | *M. textilis*/ *M.* 'Tongkat Langit' | . . . . . | | | |
| H3: AA1 | *M.a.* subsp. *malaccensis* 1 | ...0GT..G1 .A0...1..T A2AG..G..A .T0.1.C .C. .....10 T.T.G.. | | | |
| H4: AA2 | *M.a.* subsp. *malaccensis* 2 | ...0.T..G1 .A0...1..T A2AG..G..A .T0.1.C .C. ..C....10 T.T.G.. | | | |
| H5: AA3 | *M.a.* subsp. *banksii* | ...0.T..G1 .A0.....T A2AG..G..A .T0.1.C .CA .....110 T.T.G.. | | | |
| H6: AA4 | *M.a.* subsp. *siamea* 1 | ...0.T.TG1 .A0.T.1A.T A2AG..G..A .T0.1.C .C. ....1..110 T.T.G.. | | | |
| H7: AA5 | *M.a.* subsp. *siamea* 2 | ...0GT..G1 .0.....T A2AG..G..A .T0.1.C .C. .....110 T.T.G.. | | | |
| H8: AAA | *Musa* 'Hom Thong' | ...0.T..G1 .A0...1..T A2AG..G..A .T0.1.C .CA ..C....10 T.T.G.. | | | |
| H9: AAB/ABB1 | *Musa* 'Nga Chang'/ *Musa* 'Hin' (Yala)* | ...0.T..G1 .A0.....T A2AG..G..A .T0.1.C .C. .....110 TTT.G.. | | | |
| H10: AB/ABB2 | *Musa* 'Auko'/ *Musa* 'P. Gajih Merah' | ...0.T..G1 .A0.....T A2AG..G..A .T0.1.C .C. ...A...110 TTT.G.. | | | |
| H11: ABB3 | *Musa* 'Saba' (ITC)* | ...0.T.TG1 .A0...1A.T A2AG..G..A .T0.1.C .C. ...A...110 TTT.G.. | | | |
| H12: ABB4 | *Musa* 'Hak Muk Khiao' | ...0.T.TG1 .A0.T.1A.T A2AG..G..A .T0.1.C .C. ....1..110 T.T.G.. | | | |
| H13: BBBA | *Musa* 'Pli Hai' | 1T1...0... T..C.C..T. .0..0..A1. A..12T.A .. .....TA..2 ...1.TA | | | |
| H14: BB1/BBB/BBA1 | *M.b.* (wild) 1 and others[a] | 1T1...0... T..C.C..T. .0..0G.A1. A..12T.A .. G1...TA..2 ...1.TA | | | |
| H15: BB2/BBA2 | *M b.* (wild) 2 and others[b] | 1T1...0... T..C.C..T. .0..0G.A1. A..12T.A .. .....TA..2 ...1.TA | | | |

[a]The other accessions showing H14 haplotype included *M.b.* 'P. Klutuk Wulung'/*M.b.* 'Cameroon'/*Musa* 'Lep Chang Ku'/*Musa* 'Moe Nang'

[b]The other accessions showing H15 haplotype included *M.b.* 'Tani' (Sukhothai)/*M.b.* 'Lal Velchi'/*M.b.* 'Honduras'/*M.b.* 'ITC626'/ *Musa* 'Saba' (IPB)*/*Musa* 'Namwa Nuan'/*Musa* 'Hin' (NST)*

*banksii* lineage and two clades, A1 and A2. The A1 clade was divided into two subclades: (i) A1-1 included two *M. acuminata* subsp. *malaccensis* accessions and *Musa* 'Hom Thong' and (ii) A1-2 included *M. acuminata* subsp. *siamea* and *Musa* 'Hak Muk Khiao'. Three lineages of the A type hybrids - three genomic constitutions, AB, AAB, and ABB - formed the A2 clade without any wild *M. acuminata*. In the B clade, *Musa* 'Pli Hai' appeared as the basal lineage to the rest with moderate bootstrap support (63%). Within the remaining B-genome containing accessions which formed as polytomy, the B1 subclade was emerged with high bootstrap value (84%) including wild and cultivated *M. balbisiana*, BBA and BBB accessions.

## Discussion

**Phylogenetic utility of four cpDNA sequences.** Our investigation suggested the use of the *petA-psbJ* intergenic spacer to access diversity among *Musa* cultivars and hybrids. The results from the analyses of the *petA-psbJ* intergenic spacer with 29 Pi sites elucidated relationships among the A accessions by four substitutions and two indels, and those among the B accessions by one substitution and one indel (Table 4). The *rpl16* intron and *psaA-ycf* intergenic spacer gave less number of Pi sites (18 and 10, respectively), while *psbL-psbF* intergenic spacer was too short (22 bp) to give meaningful result. Although *ndhA* intron yielded more Pi sites (38) than did other regions, no inference about relationships among members within either A or B genomes was obtained. It is noteworthy that only the *petA-psbJ* provided relationship inferences within the B-genome containing accessions.

**Phylogenetic utility of hairpin structures.** Hairpin secondary structures formed by small inversions (4 bp to 16 bp), associated with inverted repeats ranged from 11 to 24 bp, are widespread in angiosperms (Kim and Lee, 2004). The hairpin structures in *petA-psbJ* spacer were not only found in *Musa* (reported here), *Boesenbergia* and *Hedychium* (J. Techaprasan, pers. comm.) of the Zingiberales, but also in Poaceae, Fabaceae, Solanaceae, and Araliacea (Kim and Lee, 2004). Their occurrences among several distantly related plants suggested that they are a conserved structural feature and could have played an important role in transcription termination of protein-encoding transcription unit as proposed by Wilson and von Hippel (1995) and Lesnik *et al.* (2001).

The three different orientations of the stem-loop structures found in *Musa* confirm Kim and Lee (2004) suggestion that base substitutions and flip-flop orientations within the loops can be generated within a short period of time. The new structure may have a selective advantage over the original one. However, small inversion may have been generated by parallel or back mutation events during chloroplast genome evolution. As we found in our analyses that the inverted repeat in *petA-psbJ* intergenic spacer was not phylogenetically congruence to other segment, the repeats may not be useful for phylogenetic reconstruction. The small inversion may also induce a large number of sequence differences during phylogenetic reconstruction, thus it may be misleading (Kelchner and Wendel, 1996) and should be omitted prior to

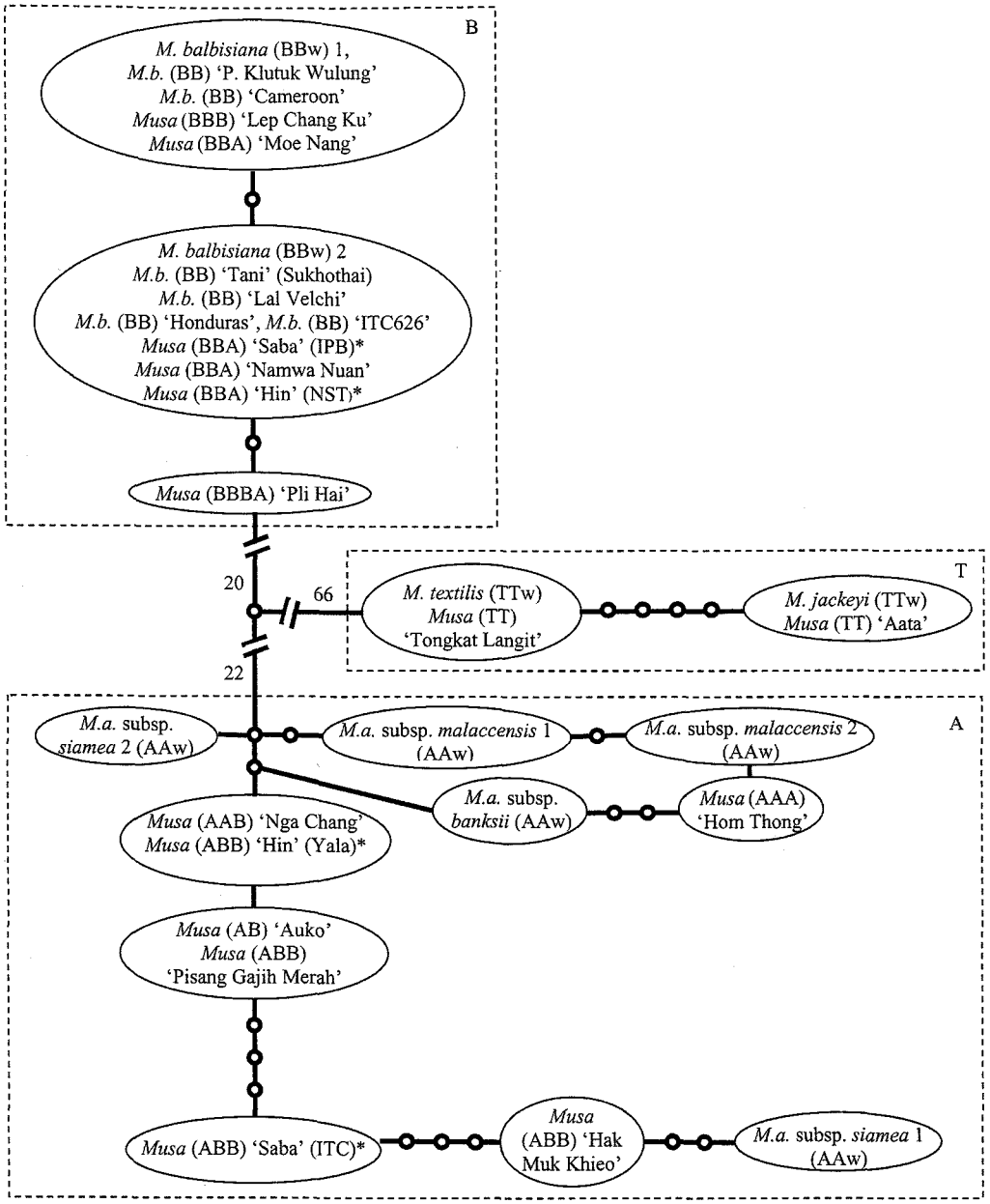


**Fig. 2.** Statistical parsimony haplotype networks of cpDNA sequences. Three clades of *M. acuminata* (A), *M. balbisiana* (B), and *Australimusa* (T) genomes were squared. Accessions in the same haplotypes appear in ovals. Small open circles indicate missing intermediates. The number of mutational changes between the different genomes is indicated next to the branches. Asterisk (*) indicate *Musa* accessions 'Saba' and 'Hin', which are separated in two genome groups. Accession names appear according to those in Table 3.

the analyses. In conclusion, caution is needed when using chloroplast non-coding sequences for phylogenetic analysis and the inverted repeat events should be thoroughly examined.

**Phylogenetic analyses of *Musa*.** The statistical parsimony haplotype network constructed from combined cpDNA data solidly supported the theory proposed by Cheesman (1947) and Simmonds and Shepherd (1955) that hybrid and polyploid cultivars came from two wild species, *M. balbisiana* and *M. acuminata*.

The presence of genetically distinct *M. balbisiana* in a remote valley in northern Thailand indicated strongly that it is wild and native to Thailand. To our knowledge, this is the first report of its natural distribution in Thailand. This population connects wild *M. balbisiana* distribution range from Assam, India (Simmond, 1962), Hainan Island, China (Ge *et al.*, 2005), and the Phillippines (Sotto and Rabara, 2000).

Our cpDNA sequence data support the observed lower level of genetic variability within *M. balbisiana* compared to that within *M. acuminata* (Simmonds, 1962; De Langhe and De Maret, 1999). However, it can be inferred from the network analysis (Fig. 2) that hybridization could happen by

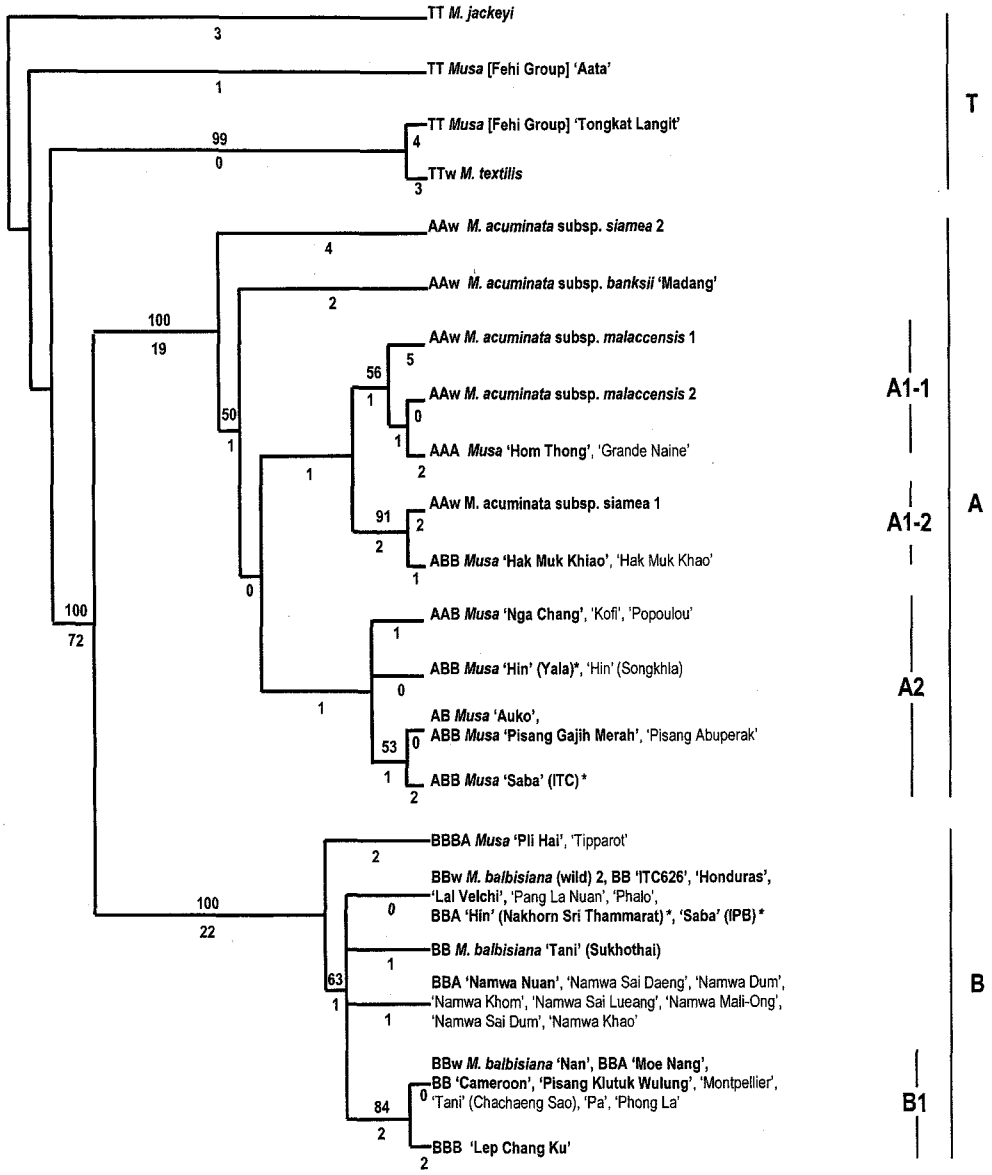


**Fig. 3.** One of 18 most parsimonious trees from the analysis of *Musa* combined cpDNA sequence data with all gaps excluded and recoded (tree length = 167, CI = 0.9241, RI = 0.9800). Bootstrap percentage (>50) and branch lenghts are indicated above and below the branches, respectively. Letters T = *Australimusa* banana, A = *M. acuminata*, B = *M. balbisiana* lineages. Sequences of 30 accessions, appeared in bold-type face, were deposited in GenBank. (Table 3). Asterisk (*) indicate *Musa* accessions 'Saba' and 'Hin', which are separated in two genome groups.

at least two different B-genome donors: One occurred with wild *M. balbisiana* from the North of Thailand while the other occurred via anthopogenic *M. balbisiana* in the Pacific Islands.

Polymorphic regions within selected non-coding cpDNA sequences are useful for diversity assessments and relationship inferences of *Musa*. The A clade can be segregated into three subclades. 'Hom Thong', an AAA triploid cultivar, is likely an intersubspecific hybrid between wild *M. acuminata* subsp. *malaccensis* and subsp. *banksii*. Our report of the close relationship between the Cavendish group and the Malay Peninsular subspecies agreed largely with previous reports, though not completely due to different samples used (Gawel and Jarret, 1991; Lebot *et al.*, 1993; Howell *et al.*, 1994; Bhat *et al.*, 1994; Ude *et al.* 2002).

The interspecific hybrid cultivars, ABB, AB, and AAB, were closely related, though the location of their origin is still uncertain (Fig. 2). These haplotypes were nested within all *M. acuminata* subspecies but their relationships were not robust as bootstrap supports for some clades were lower than 50% (Fig. 3). Nevertheless, triploid hybrids, 'Saba' and its synonym 'Hin' required special review. As inferred from the network and phylogenetic analyses, each clone originated from two different maternal lineages. To reflect their maternal origins,

we therefore support and encourage the use of two genome designations, ABB and BBA triploids as proposed by Carreel *et al.* (2002). As such, 'B Saba' and 'B Hin' should be assigned as BBA whereas 'A Saba' and 'A Hin' should be ABB. This instance is also applicable to the 'Namwa' cultivar group (BBA) and 'Moe Nang' (BBA). Our cpDNA sequence data provided strong evidence for a better classification of these triploids and cautioned the selection of triploid hybrid materials for future genetic analyses.

From our results, cpDNA sequences are useful to infer gene genealogies and origins of hybrids. The sequence analysis provides high resolution of genetic variability and is rather promising for further polyploid and genome evolutionary studies. As more *Musa* DNA sequences accumulate and become available, analyses of these information should ease explanations of hitherto unsolved inquiries in different research fields, such as classification, evolution, hybridization, genetics, and breeding.

**Acknowledgments** For leaf and DNA samples, the authors wish to thank Dr. F. Carreel (CIRAD, Guadeloupe); ITC (Leuven, Belgium); IPB (Univ. of the Philippines, Philippines); Queen Sirikit Garden, Bangkok; Ms. Kalayanee Suwitthawat, Pak Chong Research Station, Kasetsart University; and Mr. Samadchai Chattrakhom. The authors would like to acknowledge Ms. Jiranan Techaprasan (BIOTEC, Thailand) for her data on inverted repeat sequences of *Boesenbergia* and *Hedychium*. We express our appreciation to Mr. Nathan D. Austin and Dr. Narong Chomchalow for their editorial advice. This work was financially supported by the TRF/BIOTEC Special Program for Biodiversity Research and Training grant R_245003.

## References

Argent, G. C. G. (1976) The wild bananas of Papua New Guinea. *Notes Roy. Bot. Gard. Edinb.* **35**, 77-114.

Baum, D. A., Systsma, K. J. and Hoch, P. C. (1994) A phylogenetic analysis of *Epilobium* (Onagraceae) based on nuclear ribosomal DNA sequences. *Syst. Bot.* **19**, 363-388.

Bhat, K., Jarret, R. L. and Liu, Z.-W. (1994) RFLP characterization of Indian *Musa* germplasm for clonal identification and classification. *Euphytica* **80**, 95-103.

Bird, M. I., Hope, G. and Taylor, D. (2003) Population PEPII: the dispersal of humans and agriculture through Austral-Asia and Oceania. *Quaternary Intl.* **118-119**, 145-163.

Carreel, F., Gonzales de Leon, D., Lagoda, P., Lanaud, C., Jenny, C., Horry, J. P. and Tezenas du Montcel, H. (2002) Ascertaining maternal and paternal lineage within *Musa* by chloroplast and mitochondrial DNA RFLP analyses. *Genome* **45**, 679-692.

Cheesman, E. E. (1947) Classification of the bananas. *Kew Bull.* **2**, 97-117.

Chomchalow, N. and Silayoi, B. (1984) Banana Germplasm in Thailand. *IBPGR/SEAP Newsletter* **8**, 23-28.

Clement, M., Posada, D. and Crandall, D. (2000) TCS: a computer program to estimate gene genealogies. *Mol. Ecol.* **9**, 1657-1660.

De Langhe, E. and De Maret, P. (1999) Tracking the banana: its significance in early agriculture; in *The Prehistory of Food: Appitites for Change*, Gosden, C. and Hather, J. (eds.), pp. 377-396, Routledge, London, UK.

Farris, J. S., Kallersjo, M., Kluge, A. G. and Bult, C. (1995) Testing significance of incongruence. *Cladistics* **10**, 315-319.

Gawel, N. J. and Jarret, R. L. (1991) Cytoplasmic genetic diversity in bananas and plantains. *Euphytica* **52**, 15-23.

Ge, X. J., Liu, M. H., Wang, W. K., Schaal, B. A. and Chiang, T. Y. (2005) Population Structure of Wild Bananas, *Musa balbisiana*, in China Determined by SSR Fingerprinting and cpDNA PCR-RFLP. *Mol. Ecol.* **14**, 933-944.

Gielly, L. and Taberlet, P. (1994) The use of chloroplast DNA to resolve plant phylogenies: noncoding versus *rbcL* sequences. *Mol. Biol. Evol.* **11**, 769-777.

Golenberg, E. M. (1993) Evolution of a noncoding region of the chloroplast genome. *Mol. Phylogenet. Evol.* **2**, 52-64.

Guillemette, J. G. and Lewis, P. N. (1983) Detection of subnanogram quantities of DNA and RNA on native and denaturing polyacrylamide and agarose gels by silver staining. *Electrophoresis* **4**, 92-94.

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

INIBAP (2001) Strategy for the Global Musa Genomics Consortium, Arlington, USA

INIBAP (2003) *Musa* Germplasm Information System: Search by *Musa acuminata* subspecies. http://singertk.cgiar.org/MGIS2/SearchbyAcuminataSubSpecies.htm

Javadi, F. and Yamaguchi, H. (2004) Interspecific relationships of the genus *Cicer* L. (Fabaceae) based on *trnT-F* sequences. *Theor. Appl. Genet.* **109**, 317-322.

Kaemmer, D., Fischer, D., Jarret, R. L., Baurens, F. C., Grapin, A., Dambier, D., Noyer, J. L., Lanaud, C. and Kahl, G. (1997) Molecular breeding in the genus *Musa*: a strong case for STMS marker technology. *Euphytica* **96**, 49-63.

Kelchner, S. A. and Wendel, J. F. (1996) Hairpins create minute inversions in non-coding regions of chloroplast DNA. *Curr. Genet.* **30**, 259-262.

Kim, K.-J. and Lee, H.-L. (2004) Complete chloroplast genome sequences from Korean ginseng (*Panax schinseng* Nees) and comparative analysis of sequence evolution among 17 vascular plants. *DNA Res.* **11**, 247-261.

Kim, K.-J. and Lee, H.-L. (2005) Widespread occurrence of small inversions in the chloroplast genomes of land plants. *Mol. Cells* **19**, 104-113.

Kress, W. J., Wurdack, K. J., Zimmer, E. A., Weigt, L. A. and Janzen, D. H. (2005) Use of DNA barcodes to identify flowering plants. *Proc. Natl. Acad. Sci. USA* **102**, 8369-8374.

Lebot, V., Aradhya, K. M., Manshardt, R. and Meilleur, B. (1993) Genetic relationships among cultivated bananas and plantains from Asia and the Pacific. *Euphytica* **67**, 163-175.

Lee, C. and Wen, J. (2004) Phylogeny of *Panax* using chloroplast *trnC–trnD* intergenic region and the utility of *trnC–trnD* in interspecific studies of plants. *Mol. Phylogen. Evol.* **31**, 894-903.

Lesnik, E. A., Sampath, R., Levene, H. B., Henderson, T. J., McNeil, J. A. and Ecker, D. J. (2001) Prediction of rho-

independent transcriptional terminators in *Escherichia coli*. *Nucleic Acids Res.* **29**, 3583-3594.

McCarthy, C. (1996) Chromas. www.technelysium.com.au/chromas.html, Queensland, Australia

Nicholas, K. B. and Nicholas, H. B. J. (1997) GeneDoc: a tool for editing and annotating multiple sequence alignment, distributed by the authors. www.psc.edu/biomed/genedoc.

Peralta, I. E. and Spooner, D. M. (2001) Granule-bound starch synthase (*GBSSI*) gene phylogeny of wild tomatoes (*Solanum* L. section *Lycopersicon* (Mill.) Wettst. subsection *Lycopersicon*). *Am. J. Bot.* **88**, 1888-1902.

Rozas, J., Sanchez-DelBarrio, J., Messeguer, X. and Rozas, R. (2003) DnaSP: DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* **19**, 2496-2497.

Shaw, J., Lickey, E. B., Beck, J. T., Farmer, S. B., Liu, W., Miller, J., Siripun, K. C., Winder, C. T., Schilling, E. E. and Small, R. L. (2005) The tortoise and the Hare II: relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis. *Am. J. Bot.* **92**, 142-166.

Shepherd, K. (1990) Observations on *Musa* taxonomy: a note on *Musa* germplasm of the Philippines; in *Identification of genetic diversity in the genus Musa: proceeding of an international workshop held at Los Baños, Philippines, 5-10 September 1988*, Jarret, R. L. (eds.), pp. 158-165, INIBAP, Los Baños, Philippines

Silayoi, B. and Babprasert, C. (1983) *Banana Genetic Resource Exploration in Thailand.* Kasetsart University, Bangkok, Thailand.

Simmonds, N. W. (1962) *The Evolution of the Bananas*, Longmans, London, Great Britain.

Simmonds, N. W. (1995) Bananas; in *Evolution of Crop Plants*, Smartt, J. and Simmonds, N. W. (eds.), 2nd ed., pp. 370-379, Longman Scientific & Technical, Essex, England

Simmonds, N. W. and Shepherd, K. (1955) The taxonomy and origins of the cultivated bananas. *J. Linn. Soc. Bot.* **55**, 302-312.

Sotto, R. C. and Rabara, R. C. (2000) Morphological diversity of *Musa balbisiana* Colla in the Philippines. *InfoMusa* **9**, 28-30.

Swofford, D. L. (2002) *PAUP*: Phylogenetic analysis using parsimony (* and other methods)*. version 4.0b.10 for Macintosh, Sinauer Associates, Sunderland, MA, USA.

Templeton, A. R., Crandall, K. A. and Sing, C. F. (1992) A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data III. Cladogram estimation. *Genetics* **132**, 619-633.

Ude, G., Pillay, M., Nwakanma, D. and Tenkouano, A. (2002) Genetic diversity in *Musa acuminata* Colla and *Musa balbisiana* Colla and some of their natural hybrids using AFLP markers. *Theor. Appl. Genet.* **104**, 1246-1252.

Valmayor, R. V., Jamaluddin, S. H., Silayoi, B., Kusumo, S., Danh, L. D., Pascua, O. C. and Espino, R. R. C. (2000) *Banana Cultivar Names and Synonyms in Southeast Asia*, International Network for the Improvement of Banana and Plantain - Asia and the Pacific Office, Los Baños, Laguna, the Philippines

Wilson, K. S. and von Hippel, P. H. (1995) Transcription termination at intrinsic terminators: The role of the RNA hairpin. *Proc. Natl. Acad. Sci. USA* **92**, 8793-8797.

Wong, C., Kiew, R., Loh, J. P., Gan, L. H., Lee, S. K., Ohn, S., Lum, S. and Gan, Y. Y. (2001) Genetic Diversity of the Wild Banana *Musa acuminata* Colla in Malaysia as evidenced by AFLP. *Ann. Bot.-London* **88**, 1017-1025.

Zhang, W. (2000) Phylogeny of the grass family (Poaceae) from *rpl*16 intron sequence data. *Mol. Phylogen. Evol.* **15**, 135-146.

Zuker, M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* **31**, 3406-3415.