

# Development of SSR markers from *Musa balbisiana* for genetic diversity analysis among Thai bananas

Tiwa Rotchanapreeda<sup>1,3</sup> · Sirapope Wongniam<sup>2</sup> · Sasivimon C. Swangpol<sup>3</sup> · Piyarat P. Chareonsap<sup>4</sup> · Narongsak Sukkaewmanee<sup>5</sup> · Jamorn Somana<sup>1</sup>

Received: 10 February 2015 / Accepted: 17 December 2015 / Published online: 21 April 2016  
© Springer-Verlag Wien 2016

**Abstract** Bananas in Thailand have been surveyed by our team to be at least 140 cultivars in the plantations, 10 wild species and, 4 introduced species. To characterize the genetic relationship of species and cultivars, a set of novel SSR markers was developed. Totalling 53 clones containing SSR motifs were isolated from SSR-enriched library of wild *Musa balbisiana* Colla ‘Tani’ (BB). Selected positive clones were used to design 28 primer pairs for amplification of 12 wild and 82 cultivar accessions with genome designations AA, AB, AAA, AAB, ABB, and BBB. These SSR markers loci were homology searched to the banana genomes to map their locations. The seven-sets multiplex PCR approach using four fluorescent-labeled universal primers were utilized for cost effectiveness. Capillary fragment analysis yielded the accurate size of amplicons

for evaluation of particular patterns for each cultivar. Phylogram and Structure analysis presented the specific genotype of genome groups (A and B genotypes, polyploid hybrid genomes) and cultivar groups. By A:B specific alleles ratio, accurate genome designations of hybrids can be determined. Additionally, a marker, characterized to be partial plastid *ycf2* gene, indicated the maternal identification of hybrid cultivars. One SSR marker was also preliminary tested with some wild species and advised to be the candidate fingerprinting marker for species identification. In conclusion, SSR marker sets developed here proved their exploitation in detailed identity and relationship of cultivated bananas, which would be useful for genetic conservation and ongoing breeding programs in Thailand and other areas.

Handling editor: Eric Schranz.

**Electronic supplementary material** The online version of this article (doi:10.1007/s00606-015-1274-2) contains supplementary material, which is available to authorized users.

✉ Jamorn Somana  
jamorn.som@mahidol.ac.th

<sup>1</sup> Department of Biochemistry, Faculty of Science, Mahidol University, Bangkok 10400, Thailand

<sup>2</sup> Central Instrument Facility, Faculty of Science, Mahidol University, Bangkok 10400, Thailand

<sup>3</sup> Department of Plant Science, Faculty of Science, Mahidol University, Bangkok 10400, Thailand

<sup>4</sup> Plant Genetic Conservation Project under the Royal Initiative of Her Royal Highness Princess Maha Chakri Sirindhorn, Bangkok 10303, Thailand

<sup>5</sup> Art Education Section, Faculty of Anthropology and Social Science, Muban Chombueng Rajabhat University, Ratchaburi 70150, Thailand

**Keywords** Capillary fragment analysis · DNA fingerprinting · Genome specific alleles · Maternal inheritance · *Musa balbisiana* · SSR markers

## Introduction

Banana is one of the most popular fruits worldwide, and also the economical and nutritional important food crop in several areas of the world. Most edible banana cultivars originated from intra- and/or inter-specific hybridizations of two wild species in the genus *Musa*, *M. acuminata* (A genome) and *M. balbisiana* (B genome) (Simmonds and Shepherd 1955). These cultivars include diploids, triploids and a few tetraploids in various genome combinations, e.g., AA, AB, AAB, ABB, BBB, and AABB (Valmayor et al. 2000). Edible bananas are widely geographical distributed, accomplished solely by transportation of vegetative planting materials by human agencies, diversified by somatic

mutation, and multiplied by vegetative propagation and artificial selection (Simmonds 1962). These features made the edible banana genetic background rather complex.

SSR, as the DNA marker, consists of tandem repeats of short nucleotides motif which allocating throughout the genome of eukaryotes (Weising et al. 2005). The polymorphisms obtained from SSR are highly informative with codominant inheritance. Genotyping with SSR markers using fluorescent end-labelling of PCR fragments provides the accuracy and reproducibility of alleles call used for genetic studies. However, the cost for construction of many fluorescent-labeled primers is very expensive. To overcome the labelling of entire primers set, the three-primer protocol to labelling fluorophore to an universal primer was introduced, for either single fluorescent universal primer against each locus-specific primer (Schuelke 2000), or single universal primer labeled with multiple fluorophores for each locus-specific primer (James et al. 2011). By this manner, small additional cost would be spent for tailing universal primer sequence to one of each pair, and only few fluorescent universal primers are paid. Even though the cost for labelling is greatly reduced, reactions containing similar-sized alleles from different loci needs to be performed in singleplex PCR and analyzed individually, or pool-plexing after many singleplex PCR reactions with multiple fluorescent universal primers (Missiaggia and Grattapaglia 2006). The more cost-efficient technique is using multiplex PCR, in which, multiple fluorescent universal primers and multiple locus-specific primers are co-amplified in single run (Blacket et al. 2012; Culley et al. 2013; Ge et al. 2014). The samples are feasible for simultaneous capillary fragment analysis. The work-loading time and reagents consumption are dramatically reduced. The lower project cost has been allowed for higher number of samples and SSR loci, and became an alternative choice to access for high throughput SSR profiling.

In banana, hundreds of SSR markers have been isolated and developed for genetic studies. Despite this profusion of markers, it was considerable that they still be far behind available genetic analyses in other important crop plants. Most of the SSR markers sets in bananas were mainly developed in *M. acuminata* (Crouch et al. 1997; Kaemmer et al. 1997; Lagoda et al. 1998; Creste et al. 2006; Miller et al. 2010; Wang et al. 2010), a few were done in *M. balbisiana* (Buhariwalla et al. 2005; Ravishankar et al. 2013), possibly due to the abundance of germplasms in the former species, whereas limited in the later.

In the genome sequencing era, several important organisms have been consecutively reported for their

complete genome sequences, both fully annotated and drafted. For plants, *Arabidopsis thaliana*, was the first reported for its genome sequence (The Arabidopsis Genome Initiative 2000). Then, several important crop plants, e.g., rice (International Rice Genome Sequencing Project 2005), maize (Schnable et al. 2009), potato (The Potato Genome Sequencing Consortium 2011), and rubber tree (Rahman et al. 2013), were also subsequently reported. For banana, two important wild species were recently completed for their genome sequences. The doubled-haploid genome of *M. acuminata* subsp. *malaccensis* accession 'DH-Pahang' (D'Hont et al. 2012) was the first monocotyledon reported for whole-genome sequence apart from Poales. With the sufficient reads and coverage, the genome was assembled, fully annotated, and reported to be 523 megabase (Mb). Followed by a year, the drafted genome resource of *M. balbisiana* accession 'Pisang Klutuk Wulung' ('PKW') (Davey et al. 2013) was also reported to be a consensus genome size of 402.5 Mb. Both A and B genomes contain 11 *Musa* chromosomes ( $2n = 2x = 22$ ). The density of SSR motifs (di- to deca-nucleotide repeats) found in A genome of 'Pahang' (0.074 SSRs/Kbp genome; total of 34,871) was analyzed to become slightly lower than in the B genome of 'PKW' (0.090 SSRs/Kbp genome; total of 30,599) (Davey et al. 2013). Beside the chromosomal attention, the plastid genome of *M. acuminata* subsp. *malaccensis* has also become in the focus and reported to be at approximately 169 kb with the total of 112 SSRs throughout its genome (Martin et al. 2013). With the completed genome databases of the two *Musa* species, the advantages of genomic era are more obvious, providing precision of SSR markers rather than conventional SSR isolating strategies.

With the mentioned information, this study was successful to isolate the total of 28 novel SSR markers from *M. balbisiana* (B genome) with their exact loci in both banana genomes. After the multiplex PCR with capillary fragment analysis using these markers, the genetic relationship among Thai cultivated bananas was analyzed and revealed. By the number of A or B specific alleles, many hybrid cultivars can be clearly grouped, revised and estimated to proper genome designation. Moreover, one plastid SSR marker was approved as a tool to indicate the maternal side of hybrid cultivars. In addition, another chromosomal marker was proposed as the candidate fingerprinting marker to identify the wild banana species. Finally, the methodologies used in this study could be a guideline for further microsatellite researches in crop improvement and cultivar characterization, not just mainly emphasis on phylogenetic analysis.

## Materials and methods

### Plant materials, flow cytometry, total genomic DNA isolation

Banana samples used in the present study consisted of 12 wild accessions (native and introduced species) and 82 cultivar accessions (polyploids and hybrids) found in Thailand (Table 1). The passport data to source locations is presented in Online Resource 1. Native species were collected from wild habitats with specimens conservation methods as described in Wongniam et al. (2010) including recently reported species, *Musa yunnanensis* (SS&JS 235, Häkkinen and Hong 2007) and *M. serpentina* (SS&JS 246, Swangpol and Soman 2011) and introduced species, *M. ornata* (SS&JS 433), *M. coccinea* (SS&JS 223), and *Musella lasiocarpa* (SS&JS 231), were ornamental plants and maintained in institute or private collections. Cultivar samples and the given names called by the local growers were obtained from several banana distributors. Some accessions were checked for possible closest similarity to collection in the Musa Germplasm Information System; MGIS. Designated genomes of these cultivars were followed the description by Chomchalow and Silayoi (1984) relevant to six genomic types: AA, AB, AAA, AAB, ABB, and BBB with some corrections based on our results. The ploidy level of cultivars was determined based on flow cytometry technique as previously described (Doležel et al. 2007) with some modification, and the total genomic DNA samples were isolated using the modified CTAB method as detailed in Online Resource 2.

### Construction of SSR-enriched library

A SSR-enriched library was constructed from a diploid genome of wild *M. balbisiana* (BBw, SS&JS 12) according to the FIASCO (Fast Isolation by AFLP of Sequences COntaining repeats) isolation protocol (Zane et al. 2002). Approximately 250 ng of genomic DNA was simultaneously digested with *MseI* (NEB) and ligated with *MseI* AFLP adaptors (5'-TACTCAGGACTCAT-3'/5'-GAC-GATGAGTCCTGAG-3') using T4 DNA ligase (NEB) for 3 h at 37 °C. The ligated fragments were PCR amplified using *MseI* adaptor-specific primer (5'-GATGAGTCCT-GAGTAA-3'; abbreviated to MS primer) with no 'selective' bases. Probes used for hybridization to enrich the SSR motifs consisted of the 5'-biotinylated (GT)<sub>15</sub>, (CT)<sub>15</sub>, (AAC)<sub>8</sub>, and (GAA)<sub>8</sub> oligonucleotides. The SSR-enriched fragments bound to the probes were captured by streptavidin coated beads (Dynabeads Magnetic Beads, Invitrogen) and separated using a magnetic particle concentrator (Dyna MPC-S, Invitrogen). After the washing and elution

steps, the eluted fragments were purified using QIAquick PCR Purification Kit (Qiagen), re-amplified by PCR using the MS primer, and cloned into pJET1.2/blunt cloning vector (Thermo Scientific). The recombinant plasmids were transformed into the competent *Escherichia coli* DH5 $\alpha$  cells (Invitrogen). The recombinant clones containing the SSR-enriched library were screened by colony PCR with MS primer to investigate the single band product with fragment size of between 100 and 500 bp. Screened colonies were propagated, purified plasmids and subsequently sequenced on ABI 3730XL Genetic Analyzer (1st BASE Company, Malaysia).

### SSR clones identification and primers design

The nucleotide sequences were manipulated using Prophet 5.0 package (BBN Systems and Technologies) to trim the regions of vector and MS primer. The positive clones were verified by containing of any SSR motifs within their sequences using the analysis program Simple Sequence Repeat Identification Tool (SSRIT, Temnykh et al. 2001). Di- and tri-nucleotide motifs with at least 4 repeats were selected as the criteria. To avoid the construction of redundant markers set, the Sequence Assembly Program CAP3 (Huang and Madan 1999) was used to find out the contiguous sequences among the positive clones. The singleton and consensus sequences which contained the unique SSR motifs were used to design the primers for each SSR marker using the web-based program Primer3 (Rozen and Skaletsky 2000). The forward and reverse primers were designed to anneal to the flanking regions of each SSR motif with a few conditions. Restrictive parameters for auto-annealing were considered. The primer length was between 18 and 24 nucleotides with an optimal length of 20 nucleotides. Then, the optimal GC content and melting temperature was narrowed to approximately 50 % and 60 °C, respectively. Each SSR marker was named with the prefix 'BB' genome followed by the resident repeat motif and clone number. Sequences of SSR loci were submitted to the GenBank database and the nucleotide BLAST program was used in attempt to identify their probable functions. Additionally, all sequences were searched for homology against two complete genome sequences of *M. acuminata* subsp. *malaccensis* 'DH-Pahang' and *M. balbisiana* 'PKW' to map the exact loci position on designated 11 *Musa* chromosomes and plastid genome. The searching was accessed via The Banana Genome Hub database (Droc et al. 2013) which is available at the following URL: <http://banana-genome.cirad.fr>. The ClustalW2 program was used to perform the sequence alignment between SSR loci and the corresponding loci on DH-Pahang and PKW genomes.

**Table 1** Wild and cultivar banana accessions from over Thailand, with information on DNA content and ploidy level by flow cytometry, and cultivars genotypic affinity and maternity identification by SSR

No.	Accession <sup>a</sup> SS&JS	Species/cultivar names	Genome <sup>b</sup>	Flow cytometry		SSR		
				2C (pg) <sup>c</sup>	Ploidy	N <sub>a</sub> A-specific <sup>d</sup>	N <sub>a</sub> B-specific <sup>d</sup>	Maternity <sup>e</sup>
1	12	<i>Musa balbisiana</i>	BBw	1.16 ± 0.01	2x	0	12	B genome
2	377	<i>M. acuminata</i> subsp. <i>malaccensis</i>	AAw	1.22 ± 0.01	2x	15	0	A genome
3	223	<i>M. coccinea</i>	CCw	1.38 ± 0.01	2x	–	–	–
4	235	<i>M. yunnanensis</i>	YYw	1.26 ± 0.01	2x	–	–	–
5	246	<i>M. serpentina</i> <sup>f</sup>	SSw	1.36 ± 0.02	2x	–	–	–
6	388	<i>M. laterita</i>	LLw	1.35 ± 0.01	2x	–	–	–
7	433	<i>M. ornata</i>	OOw	1.30 ± 0.01	2x	–	–	–
8	443	<i>M. itinerans</i>	IIw	1.34 ± 0.02	2x	–	–	–
9	486	<i>M. gracilis</i>	GGw	1.44 ± 0.01	2x	–	–	–
10	231	<i>Musella lasiocarpa</i>	MIMlw	1.24 ± 0.01	2x	–	–	–
11	386	<i>Ensete superbum</i>	EsEsw	1.44 ± 0.01	2x	–	–	–
12	397	<i>E. glaucum</i>	EgEgw	1.45 ± 0.01	2x	–	–	–
13	24	Thong Khi Maeo	AA	1.27 ± 0.01	2x	24	0	A genome
14	37	Khai Kamphaeng Phet	AA	1.27 ± 0.01	2x	21	0	A genome
15	38	Khai Suan Phueng	AA	1.28 ± 0.01	2x	21	0	A genome
16	39	Pa Rayong	AA	1.27 ± 0.01	2x	23	0	A genome
17	49	Sa	AA	1.28 ± 0.01	2x	30	0	A genome
18	52	Nam Mak	AA	1.24 ± 0.01	2x	23	0	A genome
19	59	Hom Chan	AA	1.26 ± 0.01	2x	28	0	A genome
20	62	Lep Mue Nang	AA	1.25 ± 0.01	2x	29	0	A genome
21	68	Nua Thong	AA	1.23 ± 0.01	2x	26	0	A genome
22	71	Thong Dok Mak	AA	1.26 ± 0.01	2x	31	0	A genome
23	83	Sae Ma	AA	1.25 ± 0.02	2x	26	0	A genome
24	84	Hom Champa	AA	1.27 ± 0.01	2x	23	0	A genome
25	120	Saero	AA	1.27 ± 0.02	2x	31	0	A genome
26	69	Khai Chumphae	AB <sup>g</sup>	1.26 ± 0.01	2x	21	13	A genome
27	29	Teen Tao	AAA	1.81 ± 0.02	3x	29	0	A genome
28	31	Khai Thong Ngoei	AAA <sup>g</sup>	1.84 ± 0.02	3x	27	0	A genome
29	33	Nom Sao	AAA <sup>g</sup>	1.93 ± 0.04	3x	30	0	A genome
30	34	Hom Kariang	AAA	1.89 ± 0.03	3x	28	0	A genome
31	36	Nak	AAA	1.86 ± 0.01	3x	27	0	A genome
32	44	Tam Nuan	AAA	1.90 ± 0.01	3x	27	0	A genome
33	47	Man	AAA	1.88 ± 0.04	3x	29	0	A genome
34	48	Kung Khieo	AAA	1.89 ± 0.02	3x	26	0	A genome
35	50	Hom Khieo Ton Sung	AAA	1.90 ± 0.05	3x	30	0	A genome
36	54	Thong Khak	AAA	1.80 ± 0.01	3x	25	0	A genome
37	55	Nam Nom	AAA	1.96 ± 0.01	3x	25	0	A genome
38	61	Hom Thong Pa	AAA	1.82 ± 0.02	3x	28	0	A genome
39	74	Pisang Ampiang	AAA <sup>g</sup>	2.01 ± 0.01	3x	29	0	A genome
40	77	Hom Thong Phon San	AAA	1.86 ± 0.04	3x	31	0	A genome
41	78	Hom Thip Nakhon Sawan	AAA	1.89 ± 0.03	3x	30	0	A genome
42	86	Hom Khom	AAA	1.85 ± 0.04	3x	30	0	A genome
43	119	Niu Chorake Amphawa	AAA	1.85 ± 0.05	3x	30	0	A genome
44	134	Hom Si Sa Ket	AAA <sup>g</sup>	1.87 ± 0.05	3x	30	0	A genome
45	148	Hom Thong Taiwan	AAA	1.90 ± 0.03	3x	26	0	A genome
46	149	Hom Thong Phetchaburi	AAA	1.90 ± 0.01	3x	29	0	A genome
47	158	Khai Phra Tabong	AAA	1.89 ± 0.02	3x	28	0	A genome
48	161	Khieo	AAA	1.85 ± 0.04	3x	30	0	A genome
49	171	Wan	AAA <sup>g</sup>	1.84 ± 0.03	3x	28	0	A genome
50	189	Nak Yak Thong Pha Phum	AAA	1.86 ± 0.03	3x	25	0	A genome
51	194	Khai Thong Ruang	AAA	1.84 ± 0.02	3x	30	0	A genome
52	198	Chai	AAA	1.90 ± 0.04	3x	32	0	A genome
53	16	Tip Mukdahan	AAB <sup>g</sup>	1.79 ± 0.04	3x	27	10	A genome

**Table 1** continued

No.	Accession <sup>a</sup> SS&JS	Species/cultivar names	Genome <sup>b</sup>	Flow cytometry		SSR		
				2C (pg) <sup>c</sup>	Ploidy	N <sub>a</sub> A-specific <sup>d</sup>	N <sub>a</sub> B-specific <sup>d</sup>	Maternity <sup>e</sup>
54	42	Langka	AAB	1.89 ± 0.01	3x	26	8	A genome
55	46	Nang Klai Surin	AAB <sup>g</sup>	1.84 ± 0.02	3x	27	10	A genome
56	53	Roi Wi	AAB	1.81 ± 0.01	3x	22	14	B genome
57	56	Thong Som	AAB	1.79 ± 0.03	3x	25	8	A genome
58	58	Khai Boran	AAB <sup>g</sup>	1.83 ± 0.04	3x	25	10	A genome
59	67	Langka Nakhon Sawan	AAB	1.87 ± 0.01	3x	22	9	A genome
60	72	Khai Si Sa Ket	AAB	1.82 ± 0.02	3x	29	10	A genome
61	73	Nam Chiang Rai	AAB <sup>g</sup>	1.85 ± 0.02	3x	27	10	A genome
62	76	Sam Duan Phichit	AAB <sup>g</sup>	1.89 ± 0.02	3x	28	10	A genome
63	85	Nang Phaya	AAB <sup>g</sup>	1.86 ± 0.02	3x	25	13	B genome
64	133	Wan Tap Maeo	AAB <sup>g</sup>	1.85 ± 0.04	3x	27	10	A genome
65	179	Klai	AAB	1.80 ± 0.06	3x	25	8	A genome
66	186	Nga Chang	AAB	1.82 ± 0.05	3x	14	6	A genome
67	187	Kani	AAB	1.84 ± 0.02	3x	15	7	A genome
68	188	Chi	AAB	1.82 ± 0.04	3x	26	4	A genome
69	193	Nam Thai	AAB	1.82 ± 0.04	3x	24	15	B genome
70	195	Nam Fat	AAB	1.83 ± 0.05	3x	23	14	B genome
71	26	Hin	ABB	1.79 ± 0.02	3x	11	17	B genome
72	28	Nom Mi	ABB	1.78 ± 0.04	3x	13	13	A genome
73	35	Hak Muk Nuan	ABB	1.80 ± 0.01	3x	14	14	A genome
74	41	Nam Wa Ngoen	ABB	1.89 ± 0.01	3x	14	20	B genome
75	51	Nam Wa Mali-Ong	ABB	1.80 ± 0.04	3x	19	20	B genome
76	63	Nam Wa Maharat	ABB	1.77 ± 0.01	3x	19	20	B genome
77	64	Nam Wa Nuan	ABB	1.88 ± 0.01	3x	20	20	B genome
78	65	Hak Muk Thong	ABB	1.76 ± 0.01	3x	11	16	A genome
79	66	Nam Wa Luk Sai Dam	ABB	1.81 ± 0.04	3x	18	18	B genome
80	75	Nam Wa Khom	ABB	1.79 ± 0.02	3x	18	18	B genome
81	79	Hak Muk Khieo	ABB	1.84 ± 0.04	3x	11	17	A genome
82	81	Phama Hak Kuk	ABB	1.79 ± 0.02	3x	10	16	A genome
83	82	Nam Wa Sai Dang	ABB	1.82 ± 0.03	3x	14	19	B genome
84	152	Niu Mue Nang	ABB	1.86 ± 0.04	3x	11	13	B genome
85	157	Nam Wa Sai Lueang Ubon	ABB	1.81 ± 0.02	3x	17	19	B genome
86	160	Nam Wa Tanao Si	ABB	1.90 ± 0.02	3x	19	18	B genome
87	163	Thep Phanom	ABB	1.81 ± 0.02	3x	11	13	B genome
88	166	Nam Wa Thong Ma-Eng	ABB	1.80 ± 0.04	3x	21	18	B genome
89	170	Nam Wa Khieo	ABB	1.88 ± 0.02	3x	20	19	B genome
90	190	Nam Lawo	ABB	1.88 ± 0.01	3x	11	16	B genome
91	191	Nam Wa Nuan Chan	ABB	1.82 ± 0.00	3x	21	19	B genome
92	196	Chang	ABB	1.77 ± 0.03	3x	5	18	B genome
93	13	Tani Kiu Chan	BBB	1.78 ± 0.03	3x	0	16	B genome
94	25	Lep Chang Kud	BBB	1.76 ± 0.02	3x	0	17	B genome

<sup>w</sup> wild accessions

<sup>a</sup> Accession numbers were as identified on S. Swangpol and J. Somana (SS&JS) collection

<sup>b</sup> Genome designations followed the description by Chomchalow and Silayoi (1984)

<sup>c</sup> The 2C nuclear DNA contents at G<sub>0</sub>/G<sub>1</sub> state are shown in pg of mean ± standard deviation ( $n = 3$ ); minimal CV = 1.6 %; maximal CV = 5.0 %; average CV = 3.6 %

<sup>d</sup> Numbers of alleles identified to be specific to A (N<sub>a</sub>A) and B (N<sub>a</sub>B) genome, total allele numbers are 33 and 20, respectively, details in Online Resource 6

<sup>e</sup> Maternity of cultivars was considered by the presence of two distinguished alleles on the maternal inheritance marker (BB\_AAC-3)

<sup>f</sup> Swangpol and Somana (2011)

<sup>g</sup> Genome designations which were revised and suggested from experimental evidences in this study, previous genome designations and description of revised genome designations are shown in Online Resource 3



### Multiplex PCR with fluorescent-labeled universal primers

Genomic DNA samples of the 82 cultivar accessions were evaluated for the polymorphic patterns. Two wild accessions, *M. balbisiana* (SS&JS 12) and *M. acuminata* subsp. *malaccensis* (SS&JS 377) were also included. From 53 positive clones containing the SSR motifs, 28 loci were unique and used for designing of 28 primer pairs. The multiplex PCR system was established to perform in a total of seven PCR sets (Table 2). In each set, one of forward or reverse primer was designed to tail with one of the following universal primers; M13(-29) [5'-CGACGTTGTAACGAC-3'], M13(-20) [5'-AAAA CGACGGCCAGT-3'], M13(-40) [5'-GTTTTCCCAGTC ACGAC-3'], or BGHr [5'-GGCACAGTCGAGGC-3'], in which, adjacent bases that shared similarity were designed to become overlapping. Four fluorescent-labeled universal primers were constructed to contain the complementary sequence with individual universal primer and 5'-labeled with different fluorophores: 5'-FAM\_M13(-29) (blue dye), 5'-HEX\_M13(-20) (green dye), 5'-NED\_M13(-40) (yellow dye), and 5'-PET\_BGHr (red dye). The multiplex PCR reaction was performed using QIAGEN Multiplex PCR Kit (Qiagen). The reaction was done in a 96-well PCR plate with a total volume of 10  $\mu$ l and contained 25 ng of genomic DNA, 1  $\times$  QIAGEN Multiplex PCR Master Mix (containing HotStarTaq DNA polymerase, multiplex PCR buffer, dNTP mix), and 0.2  $\mu$ M each of total 12 oligonucleotides (4 forward primers, 4 reverse primers, 4 fluorescent-labeled universal primers). The cycling program was performed on a thermal cycler and consisted of hot-started activation step at 95 °C for 15 min, then 35 cycles including denaturation at 94 °C for 30 s, annealing at 45 °C for 3 min, extension at 72 °C for 1 min, and final extension step at 60 °C for 30 min. The annealing temperature was set to the lowest annealing temperature of fluorescent-labeled universal primers at 45 °C. The annealing time of 3 min was used to maximize the sensitivity of the reaction. The amplicons generated by the early cycles contained the sequence of universal primer, which in later, served as template for fluorescent-labeled universal primers annealing and then amplification to generate final PCR products tagged with corresponding fluorescent dye. Capillary fragment analysis was performed on ABI 3730XL Genetic Analyzer (1st BASE Company, Malaysia) with GeneScan 500 LIZ (orange, Applied Biosystems) as size standard dye. The dilution factor of 100 was become account to pool each sample in the loading mixture prior to electrophoresed.

### Polymorphisms and statistical analysis

SSR alleles from combined dyes were separately analyzed by individual marker using Peak Scanner software version 1.0 (Applied Biosystems). Each raw fragment size was manipulated with the offset value from internal control, subtracted by one nucleotide from plus A artifact, and neutralized the universal primer extension for actual allele call. Polymorphic pattern of all SSR markers were evaluated as the binary data due to the presence (1) or absence (0) of the alleles. The Polymorphism Information Content (PIC) value (Anderson et al. 1993) was calculated to estimate the polymorphisms of each SSR marker. For the phylogenetic analysis, a dissimilarity matrix was computed using DICE coefficient (Nei and Li 1979) applied in NTSYS-pc software package version 2.11X (Exeter Software). The DARwin program version 5.0.158 (Perrier and Jacquemoud-Collet 2006) was used to construct a clustering phylogram from the matrix following the Weighted Neighbor-Joining method (Saitou and Nei 1987). The binary scores were treated as dominant marker and evaluated with model-based clustering approach by Structure software version 2.3.4 (Pritchard et al. 2000; Falush et al. 2003, 2007). The admixture and correlated allele frequencies models were conducted with 100,000 periods of burn-in prior to 100,000 replicates of Markov chain Monte Carlo (MCMC) run for 10 iterations. The numbers of inferred populations ( $K$ ) were tested from 1 to 30. The optimal number of population was estimated through the Structure Harvester program (Earl and von Holdt 2012) by the calculation of delta  $K$  values (Evanno et al. 2005). The results obtained from Structure software were permuted with CLUMPP program (Jakobsson and Rosenberg 2007). All accessions were grouped into clusters by sorting of permuted association coefficient ( $Q$ ) proportion of each allelic pool using Microsoft Excel program then figures were generated with stacked column chart.

For the genome-specific alleles, one was identified to be specific to A or B genome with following criteria. The A genome specific alleles must not present in any BBw and BBB accessions but only present in AAw, AA, AAA, AB, AAB, or ABB accessions with at least half of the population (considered to be 41 accessions). While the B genome specific alleles must not present in any AAw, AA, and AAA accessions but only present in BBw, BBB, AB, AAB, or ABB accessions with at least half of the population (considered to be 22 accessions). The significant differences between data were statistically tested with single factor (one-way) ANOVA analysis module in Microsoft Excel program.

**Table 2** The 28 SSR markers from *M. balbisiana* genome, with the primers details, polymorphic identification results, predicted homological functions, and estimated locations on *Musa* genome (DH-Pahang and Pisung Klutuk Wulung) of each SSR loci

Markers / probes	Repeat motifs	Sequences (5' to 3')	$N_a$	$N_r$ A/B specific	Allele sizes (bp)	PIC	GenBank homology	<i>Musa</i> genome location	GenBank Acc. no.
<i>Multiplex set 1</i>									
BB_GT-34.1/ FAM_M13(-29)	(AG) <sub>5</sub>	F: CGACGTTGTA <del>AAAA</del> CGACTGCCCTTCAATTCATGTCAA R: TGGTATGCAGCAAGCAAAAAG	15	A genome: 5 B genome: 1	108–287	0.8767	Unidentified repeat motif	chrUn_random	KM190147
BB_GAA-7/ HEX_M13(-20)	(CTT) <sub>4</sub>	F: GAATGTGAGCCCTCTTCTCG R: AAAACGACGGCCAGTACTGGACCCGTTTTCAGTGAC	20	A genome: 2 B genome: –	82–308	0.8784	Unidentified repeat motif	Chromosome 10	KM190148
BB_CT-7/ NED_M13(-40)	(AG) <sub>5</sub>	F: GTTTTCCCAGTCAAGCAAGCAAGACACACAAAC R: GAACGAGAAACTGCCTTTGC	12	A genome: – B genome: –	109–231	0.8260	Unidentified repeat motif	Chromosome 11	KM190149
BB_CT-18/ PET_BGHR	(GA) <sub>10</sub> (AGA) <sub>5</sub>	F: GGCACAGTCGAGGCACACCACTGTCTGCATC R: CTGAGCAGTTC <del>CCCC</del> ACTAGC	No results	No results			Unidentified repeat motif	Chromosome 10	KM190150
<i>Multiplex set 2</i>									
BB_CT-40/ FAM_M13(-29)	(GT) <sub>10</sub> (GA) <sub>17</sub>	F: CGACGTTGTA <del>AAAA</del> ACGACGATAGAGTGCCCACTCCAA R: ATGCCCTTCTGGCTCGFCTT	7	A genome: – B genome: –	147–238	0.6206	Unidentified repeat motif	Chromosome 3	KM190151
BB_CT-19/ HEX_M13(-20)	(AG) <sub>12</sub> (AGA) <sub>6</sub>	F: AGGAAATTGCTAGTGGGCAGA R: AAAACGACGGCCAGTACCAATTTTGACCACCGTCTC	11	A genome: – B genome: –	142–306	0.7914	Unidentified repeat motif	Chromosome 3	KM190152
BB_AAC-9/ NED_M13(-40)	(AG) <sub>9</sub>	F: GTTTTCCCAGTCAAGCAATTTTCCGCTGCCACTG R: CACGACTGGAAGACGAGTCA	12	A genome: – B genome: –	148–397	0.7709	Unidentified repeat motif	Chromosome 5	KM190153
BB_GAA-31/ PET_BGHR	(AG) <sub>6</sub>	F: ACCGAAAGAAAACGAAAGCAGA R: GGCACAGTCGAGGCACACCTCGGATGTGTGTACC	12	A genome: 1 B genome: 2	155–432	0.7851	Unidentified repeat motif	Chromosome 6	KM190154
<i>Multiplex set 3</i>									
BB_CT-38.1/ FAM_M13(-29)	(GA) <sub>17</sub>	F: TACTCCCCACCATAAAGAGC R: CGACGTTGTA <del>AAAA</del> ACGACGTTCTTTCGGTGTCTG	No results	No results			Unidentified repeat motif	Chromosome 9	KM190155
BB_CT-11/ HEX_M13(-20)	(TC) <sub>8</sub>	F: GGCTGTACTCTGTGGTGGT R: AAAACGACGGCCAGTGTCAACCAATCCATGACCAG	34	A genome: 4 B genome: –	95–459	0.9320	Unidentified repeat motif	Chromosome 6	KM190156
BB_CT-37/ NED_M13(-40)	(TG) <sub>9</sub> (GA) <sub>21</sub>	F: GTTTTCCCAGTCAAGCACTGTTGGATTGCCATGGTATG R: TCCGAGTCTACTCCCAAGGA	41	A genome: 1 B genome: 1	88–456	0.9340	Unidentified repeat motif	Chromosome 11	KM190157
BB_CT-39/ PET_BGHR	(GA) <sub>29</sub>	F: GGCACAGTCGAGGCTTGAGCCTATGATGGGGATG R: TTGGAGTTTGCCTCTCTCAC	No results	No results			Unidentified repeat motif	Chromosome 3	KM190158
<i>Multiplex set 4</i>									
BB_CT-2/ FAM_M13(-29)	(AG) <sub>10</sub>	F: CGACGTTGTA <del>AAAA</del> ACGACTGTTTGTGCTGATGCTGACC R: GCGATAACATTC <del>CCCG</del> CAAT	16	A genome: 1 B genome: 1	123–271	0.8429	Unidentified repeat motif	Chromosome 4	KM190159
BB_CT-15/ HEX_M13(-20)	(TC) <sub>20</sub>	F: AAAACGACGGCCAGTCAACCAATTTGTGATGCCACTC R: TAGGCCACATACCCAGCTTC	29	A genome: – B genome: –	85–244	0.8991	<i>Hdr</i> gene [JQ922096]	Chromosome 8	KM190160
BB_CT-35.1/ NED_M13(-40)	(GA) <sub>16</sub> (GA) <sub>6</sub>	F: GTTTTCCCAGTCAAGCACTGCTTCAGTGAGACCCACA R: ACTAGCAAGACCGAGGAAA	27	A genome: 3 B genome: 3	124–409	0.9262	Unidentified repeat motif	Chromosome 4	KM190161
BB_AAC-3/ PET_BGHR	(TC) <sub>4</sub> (TC) <sub>4</sub>	F: GGCACAGTCGAGGCAAAATTCGGGGTCAAAAAGT R: GAGGGATTTATGGGACGACA	2	A genome: 1 B genome: 1	157–172	0.4111	<i>ycf2</i> gene [EU016998]	Plastid	KM190162

Table 2 continued

Markers / probes	Repeat motifs	Sequences (5' to 3')	$N_a$	$N_a$ /B specific	Allele sizes (bp)	PIC	GenBank homology	<i>Mitsa</i> genome location	GenBank Acc. no.
<i>Multiplex set 5</i>									
BB_CT-8/ FAM_M13(-29)	(CT) <sub>18</sub>	F: GTTCAAGCATCTCTCAGCACA R: CGACGTTGTAAAACGACCGAAAAGGAGAAAACCAAGTTG	27	A genome: - B genome: -	102-274	0.9328	Unidentified repeat motif	Chromosome 4	KM190163
BB_CT-33/ HEX_M13(-20)	(TG) <sub>8</sub> (GA) <sub>10</sub>	F: GGCAATGTCTCATAAAGAAAGAGAG R: AAAACGACGGCCAGTTTTCACCTTTGCAGAGAA	32	A genome: 3 B genome: 1	99-239	0.9392	Unidentified repeat motif	Chromosome 4	KM190164
BB_GAA-32/ NED_M13(-40)	(TTC) <sub>8</sub>	F: GTTTTCCCAGTCAAGACTGCCACCAATATATACCA R: CCAGTTGGTTCCTTGCTCT	32	A genome: 3 B genome: 1	105-236	0.9276	Unidentified repeat motif	Several hits with low confidence	KM190165
BB_GT-10/ PET_BGHR	(GA) <sub>5</sub>	F: GAGTGATCCCACTTGAGGA R: GGCACAGTCGAGGGCAACCATCATTTGGAGAC	13	A genome: - B genome: 1	109-337	0.8050	Unidentified repeat motif	Chromosome 11	KM190166
<i>Multiplex set 6</i>									
BB_CT-20/ FAM_M13(-29)	(AG) <sub>18</sub>	F: CGACGTTGTAAAACGACGGGAGATAAAGGGCAACTT R: GATACAACAGGATCCGAAGCA	24	A genome: 3 B genome: -	93-327	0.8844	Unidentified repeat motif	Chromosome 4	KM190167
BB_GT-47/ HEX_M13(-20)	(TG) <sub>5</sub>	F: TGTGGTTCTTAGCCTTGCACCT R: AAAACGACGGCCAGTCCAAATACTAGGGACTCCCA	6	A genome: 1 B genome: 2	194-238	0.7234	Unidentified repeat motif	Chromosome 6	KM190168
BB_GAA-4/ NED_M13(-40)	(AG) <sub>22</sub> (CA) <sub>4</sub>	F: GATTGCTTGCAGATAATGAACCTT R: GTTTTCCCAGTCAAGACTGCCACCAATATATACCA	32	A genome: 2 B genome: 2	92-236	0.9154	Unidentified repeat motif	Chromosome 9	KM190169
BB_GT-37/ PET_BGHR	(TC) <sub>17</sub> (CA) <sub>15</sub>	F: GGCACAGTCGAGGCAACATCACGTTGCCCATAA R: ACATATGTTGGGGCATTGCT	No results				Unidentified repeat motif	Chromosome 1	KM190170
<i>Multiplex set 7</i>									
BB_CT-38.2/ FAM_M13(-29)	(GA) <sub>17</sub> (CTC) <sub>4</sub>	F: TACTCCCCCACCATAAAGAGC R: CGACGTTGTAAAACGACCAAATCCCTTGGTTGTTC	8	A genome: - B genome: -	202-287	0.8086	Unidentified repeat motif	Chromosome 9	KM190171
BB_GAA-35/ HEX_M13(-20)	(AGA) <sub>4</sub>	F: AAAACGACGGCCAGTGGGTTTTACCCACCTCTTT R: CGAAACCAATTCCTCTGAT	14	A genome: 1 B genome: 2	125-278	0.8627	Unidentified repeat motif	Chromosome 2	KM190172
BB_CT-36/ NED_M13(-40)	(GA) <sub>13</sub>	F: GGAGCTTGTGATGTGGGAAT R: GTTTTCCCAGTCAAGACTGCCACCAATATATACCA	22	A genome: 1 B genome: 1	115-376	0.8781	<i>Laccase</i> gene [U73106]	Chromosome 9	KM190173
BB_CT-6/ PET_BGHR	(TC) <sub>19</sub> (AC) <sub>5</sub>	F: GGCTTGGTCAATCAGAGGAAG R: GGCACAGTCGAGGCTGAAGCCAAACCTTTATTGCT	23	A genome: 1 B genome: 1	115-288	0.8985	Unidentified repeat motif	Chromosome 10	KM190174



## Identification of maternal inheritance

With the analysis of evaluated polymorphic pattern of all SSR markers, a marker named BB\_AAC-3, characterized to be a locus of *ycf2* gene in the banana plastid genome, showed two unique alleles with distinguished presence beneath the A and B genome groups of cultivars. To further identification, these heterologous alleles were PCR amplified for sequencing using DNA samples of *M. balbisiana* (SS&JS 12) and *M. acuminata* subsp. *malaccensis* (SS&JS 377). Each reaction was done in a total volume of 25  $\mu$ l, containing 25 ng of genomic DNA, 1 $\times$  Standard *Taq* Reaction Buffer (NEB), 0.2 mM of dNTP mix, 0.2  $\mu$ M each of forward and reverse primers of BB\_AAC-3 marker, and 1 unit of *Taq* DNA polymerase (NEB). The cycling program was performed on a thermal cycler and consisted of initial denaturation step at 95  $^{\circ}$ C for 5 min, then 35 cycles including denaturation at 94  $^{\circ}$ C for 30 s, annealing at 50  $^{\circ}$ C for 30 s, extension at 72  $^{\circ}$ C for 30 s, and final extension step at 72  $^{\circ}$ C for 10 min. PCR products were cloned into the pSC-A-amp/kan vector using the StrataClone PCR Cloning Kit (Stratagene) following the manufacturer's protocol. Recombinant plasmids containing required inserts were sequenced, and the nucleotide sequences of two alleles were searched for homology to the GenBank database. Finally, the multiple sequence alignment of sequences with the same identity was performed and analyzed.

## Candidate markers set for species fingerprinting

By an attempt to further investigate the novel developed SSR markers here, some wild accessions were selected to perform the preliminary examination using multiplex PCR with all SSR markers, to evaluate the polymorphic pattern among population of wild species (data not shown). After analysis of the patterns, a set of fingerprinting candidate were selected based on following criteria; each marker is differently fluorescent end-labeled, presence of unique alleles among different species, and no homology sharing with cultivar accessions. Using those criteria, the fingerprinting candidate set was demanded to consist of markers: BB\_CT-8 (FAM-labeled), BB\_GAA-35 (HEX-labeled), BB\_GAA-4 (NED-labeled), and BB\_GAA-31 (PET-labeled). To intensify the advantages of the set, DNA samples of all 12 wild accessions (Table 1) were used to perform the multiplex PCR with fluorescent-labeled universal primers. The procedures ran similarly with respect to prior mention, excepting that the PCR reaction was done with 100 ng of genomic DNA samples in the 0.2-ml PCR tube with 20- $\mu$ l total volume.

## Results

### Determination of genome size and ploidy level

Genome size and standard deviation of all tested accessions were shown in Table 1. Two wild accessions, SS&JS 377 (AAw) and SS&JS 12 (BBw), contained genome sizes of  $1.22 \pm 0.01$  and  $1.16 \pm 0.01$  pg, respectively, which were significantly different ( $P < 0.001$ ). Comparing between the two sections of *Musa* genus, *Musa* section (SS&JS 377, 12, 235, 246, 388, 433, 443) and Callimusa section (SS&JS 223, 486), the genome size of Callimusa (mean = 1.41 pg) was significantly higher than that of *Musa* (mean = 1.27 pg) ( $P < 0.01$ ). Within the genus *Ensete*, the genome size of *E. superbum* (SS&JS 386) ( $1.44 \pm 0.01$  pg) and *E. glaucum* (SS&JS 397) ( $1.45 \pm 0.01$  pg) appeared to be pronouncedly related. However, they were significantly larger than the genome size of *Musella lasiocarpa* (SS&JS 231) ( $1.24 \pm 0.01$  pg) and wild *Musa* spp. ( $P < 0.001$ ).

The ploidy levels diploid ( $2C = 2x$ ) or triploid ( $2C = 3x$ ) were assigned by comparing their genome sizes to those of AAw and BBw accessions. A total of 14 diploid cultivar accessions (AA, AB) contained a mean of 1.26 pg in genome size, ranging from 1.23 to 1.28 pg, meanwhile 68 triploid accessions (AAA, AAB, ABB, BBB) contained a larger genome size with the mean of 1.84 pg, ranging from 1.76 to 2.01 pg. Some previously described ploidy level accessions were revealed by the flow cytometry results to be invalid (Online Resource 3). Therefore, they were corrected for actual ploidy level as shown in Table 1. The average genome sizes with standard deviations for each genome designation were as followed; AAA =  $1.88 \pm 0.05$  pg, AAB =  $1.83 \pm 0.03$  pg, and ABB =  $1.82 \pm 0.04$  pg. There was no significantly different genome size between AAB and ABB cultivars ( $P > 0.05$ ) but both genome groups were significantly smaller than AAA cultivars ( $P < 0.001$ , both).

### Characterization of markers from SSR-enriched library

In this study, the SSR-enriched library was constructed from *M. balbisiana* (BBw) which yielded a total of 672 recombinant clones. While this library was enriched with hybridization using (GT)<sub>15</sub>, (CT)<sub>15</sub>, (AAC)<sub>8</sub>, and (GAA)<sub>8</sub> oligonucleotide probes, 257, 270, 122, and 23 clones were fished out with those probes, respectively. Of these, 107 clones were verified by colony PCR to contain the single band with fragments in range of 100 to 500 bp. Nucleotide sequencing fished out 53 positive clones to contain SSR motifs, of which were tandem repeats of di- and tri-

nucleotides: (AC)<sub>n</sub>, (AG)<sub>n</sub>, (CA)<sub>n</sub>, (CT)<sub>n</sub>, (GA)<sub>n</sub>, (GT)<sub>n</sub>, (TC)<sub>n</sub>, (TG)<sub>n</sub>, (AGA)<sub>n</sub>, (CTC)<sub>n</sub>, (CTT)<sub>n</sub>, or (TTC)<sub>n</sub>, ranging from 4 to 29 in number of repeats. From these clones, a number of unique fragments were used to design for a total of 28 primer pairs for 28 non-redundancy SSR markers here.

The sequences of all 28 SSR loci were searched for probably homological functions against the GenBank database to reveal their function (Table 2). Most markers, 25 loci, showed similarity to subject with low expected values and the alignments covered only the short tandem repeat motifs with no significant identity which were presented as ‘unidentified repeat motifs’ of unknown inter- or intra-genic regions. The rest few, three loci, showed similarity with high expected values against the characterized function genes. A marker, BB\_CT-15, displayed the most similarity to a 4-hydroxy-3-methylbut-2-enyl diphosphate reductase (*Hdr*) gene in *Hordeum chilense* subject (GenBank accession JQ922096). Another marker, BB\_CT-36, showed the highest similarity to a laccase gene *LAC2-4* in *Liriodendron tulipifera* (GenBank accession U73106). Finally, the marker BB\_AAC-3 was identified to be a partial sequence of a hypothetical plastid protein RF2 (*ycf2*) gene in *M. acuminata* (GenBank accession EU016998).

Two banana species (*M. acuminata* and *M. balbisiana*) have been reported for their complete genome sequences. The *Musa* chromosomes are reported to be 11 designates and a chrUn\_random set, which contains unlocalized and unplaced sequences. The BLAST tool of The Banana Genome Hub database was used to allocate the location of each SSR locus on these chromosomes. Beside the *Musa* nuclear genome, a complete plastid genome sequence of *M. acuminata* subsp. *malaccensis* has been recently published (Martin et al. 2013) (GenBank accession HF677508), and its genome was revealed to possess one of the loci. By looking beneath, only one SSR locus, BB\_GAA-32, was unable to be allocated due to its too short sequence. All chromosomes except chromosome 7, were shown to possess the remaining loci residences (Table 2 and Fig. 1), in which, chromosome 4 contains most loci number; five in total. A SSR marker, BB\_GT-34.1, was shown to locate on the chrUn\_random set, suggesting that this loci may present within a highly repetitive region or an interspersed element which is hardly to locate and assemble. All chromosome-allocable SSR markers were shown to be only one copy per locus. Contrary to nuclear SSR markers, the marker BB\_AAC-3, which is identified to be the plastidic *ycf2* gene, was found in two loci of the plastid genome due to that *M. acuminata* plastid genome contains two copies of *ycf2* gene.

In an attempt to investigate the polymorphisms in differentiation of tandem repeat numbers between A and B

**Fig. 1** Physical map showing the SSR loci mapping on the *Musa* genome. Eleven *Musa* chromosomes (Chr1 to Chr11) of *M. acuminata* subsp. *malaccensis* ‘DH-Pahang’ (A genome, left chromatid) and *M. balbisiana* ‘PKW’ (B genome, right chromatid), and the circular plastid genome of *M. acuminata* subsp. *malaccensis* ‘CIRAD:930’, are illustrated as the physical map. Scale bars on the right approximately indicate the size of chromosomes in Mb and the location of SSR loci, but not to scale for the plastid genome. The exact location of SSR loci in range of nucleotide base pairs aligned on chromosomes sequences of A and B genome are shown in Online Resource 4 and 5, respectively

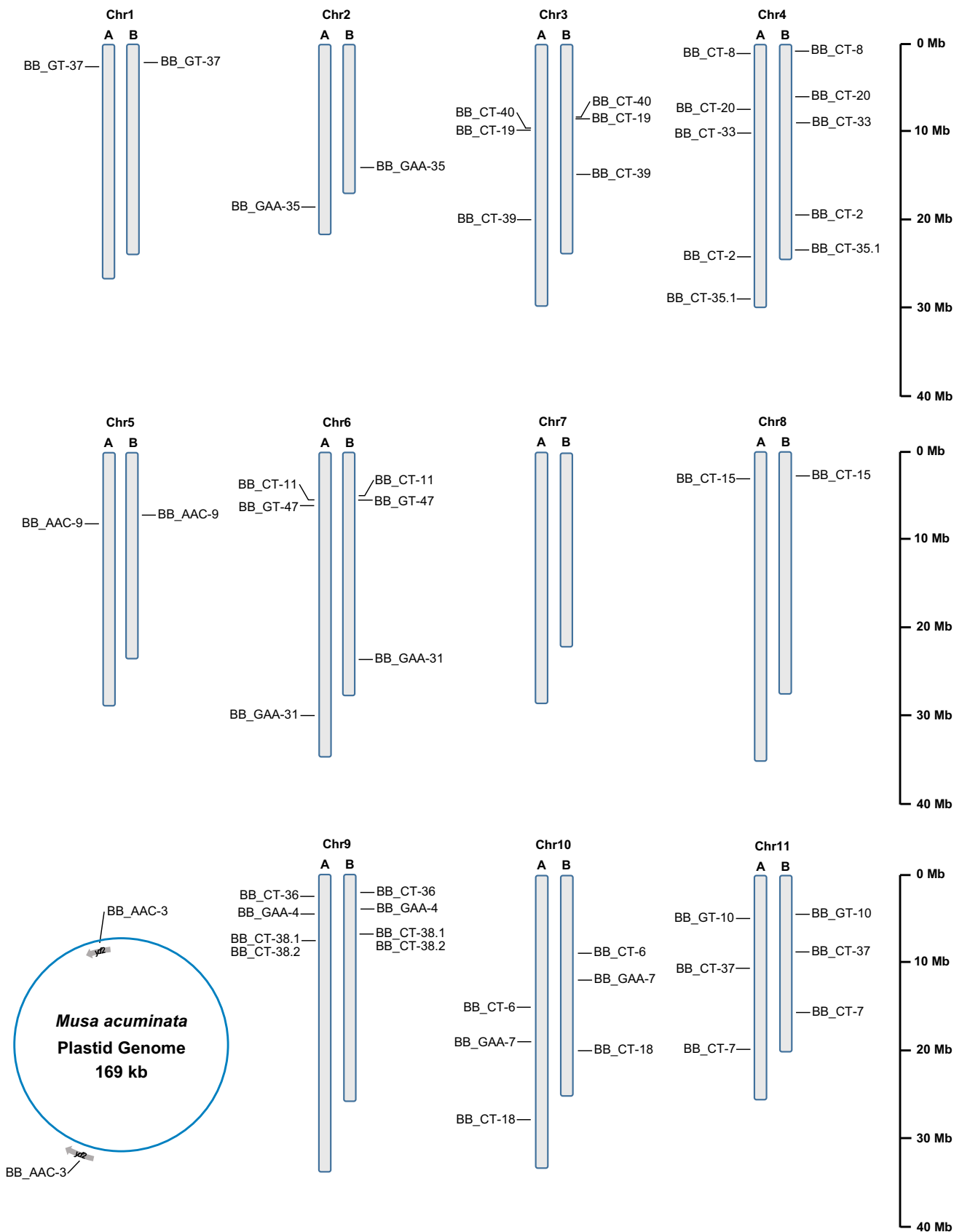
genome, each SSR clone sequence was searched against both banana genome databases for retrieving the roughly residences, followed by pairwise alignment trial between the query and that resident *Musa* genome sequence to finally obtain the actual location. The exact locations of all SSR loci were revealed in range of nucleotide base pair which covering the entire of each clone. The alignment data between the queries and DH-Pahang (A genome) or PKW (B genome) are available in Online Resource 4 and 5, respectively.

### Polymorphism of SSR markers set

Multiplex PCR system with fluorescent-labeled universal primers was exploited in the study. The reactions were expanded to become seven sets, each set to be labeled with four fluorophores: FAM, HEX, NED, and PET. With the accurate alleles call obtained from capillary fragment analysis, the SSR markers polymorphism contents were detailed in Table 2. The multiplex reactions showed that 4 SSR markers, BB\_CT-18, BB\_CT-38.1, BB\_CT-39, and BB\_GT-37, produced no detectable allele signals. Despite of these unfortunate results, the rest 24 markers succeeded to produce the signal peaks presenting a total of 471 polymorphic alleles. The SSR fragments size ranged from 82 to 459 bp with the average of 205 bp. The number of alleles ( $N_a$ ) per locus ranged from 2 alleles to 41 alleles with the average of 19.6, and the PIC values ranged from 0.4111 to 0.9392 with the average of 0.8363. Except for the PIC value obtained from the marker BB\_AAC-3 (0.4111) which containing only two alleles, all other markers displayed the PIC values higher than 0.5000, which indicating the highly polymorphic effectiveness of these markers.

### Clarity of cultivar classification to A and B genome groups

With the analysis of polymorphic pattern obtained from allele calls of all cultivar accessions, these cultivars were found to contain different numbers of A and B genome specific alleles using criteria as mentioned in the methods. By the nature of these criteria, *M. acuminata* and cultivars



**Table 3** Data of genome specific alleles in cultivars with four genome groups

Genome	Average $N_aA^a$	$P$ value <sup>b</sup>	Average $N_aB^a$	$P$ value <sup>b</sup>	Ratio $N_aA/N_aB^c$
AA	25.8 ± 3.65	0.0063*	–	–	–
AAA	28.4 ± 1.94		–		–
AAB	24.3 ± 4.04	0.0000*	9.8 ± 2.86	0.0000*	2.70 (1.57 – 6.50)
ABB	14.9 ± 4.47		17.3 ± 2.34		0.86 (0.28 – 1.17)

<sup>a</sup> Average number of alleles which specific to A genome ( $N_aA$ ) and B genome ( $N_aB$ ), data represented with mean ± SD

<sup>b</sup>  $P$  value from single factor ANOVA ( $\alpha = 0.01$ ), \* statistically significant values

<sup>c</sup> Calculated from  $N_aA$  divided by  $N_aB$  in each accession then average all, data represented with mean (minimum – maximum)

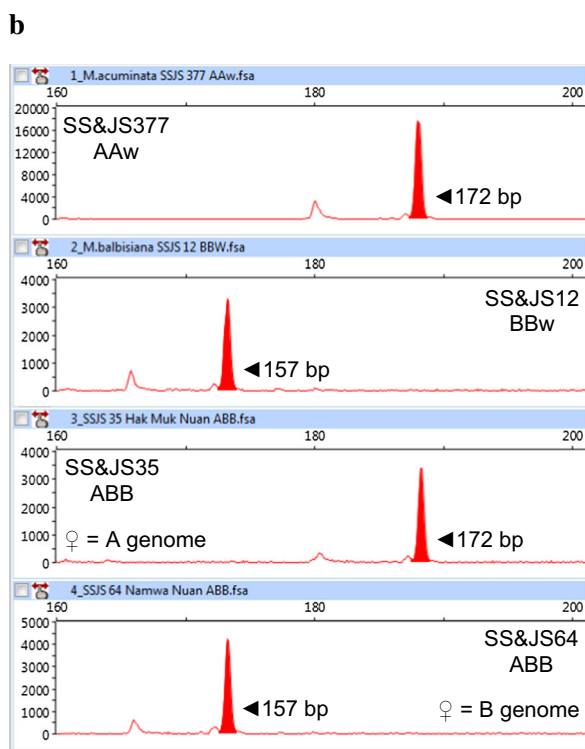
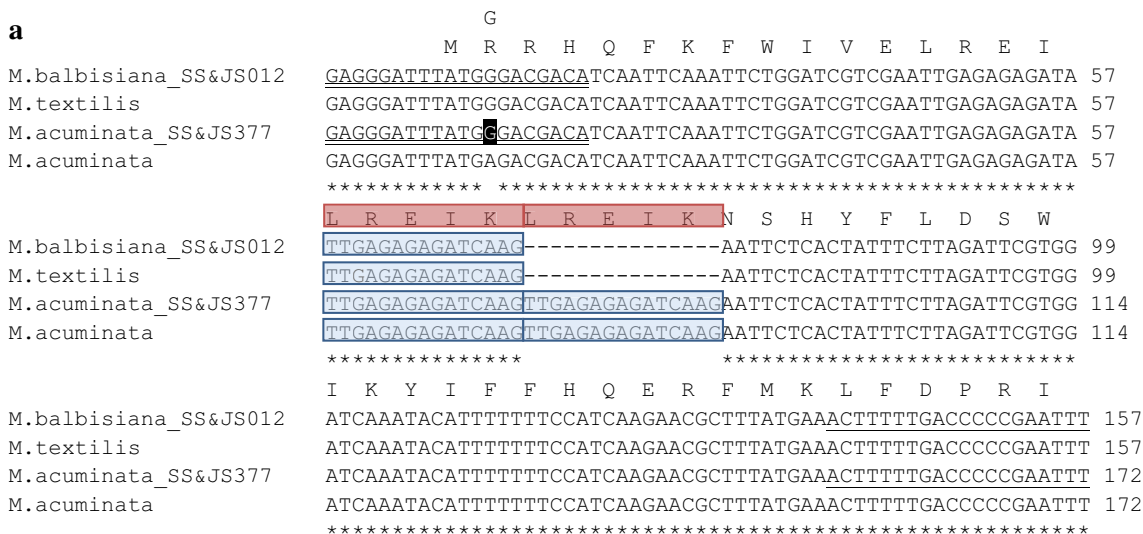
with AA and AAA genome contained only A genome specific alleles, *M. balbisiana* and cultivars with BBB genome contained only B genome specific alleles, and the hybrids AB, AAB, and ABB contained both A and B genome specific alleles (Table 1). The maximum numbers of A and B genome specific alleles counted per accession were 32 and 20 alleles, respectively. The selected A and B genome specific alleles from each SSR marker used for the classification are in Online Resource 6.

The numbers of genome specific alleles revealed their usefulness after ANOVA analysis for comparison between genome designations (Table 3). Comparing between AA and AAA accessions, they contained the average number of A genome specific alleles of 25.8 and 28.4 alleles, respectively. The AAA accessions usually contained slightly greater number than AA accessions and the difference was significant ( $P < 0.01$ ). When compare the number of both A and B genome specific alleles between AAB and ABB accessions, there were also significantly different. The AAB accessions contained the average A genome specific allele numbers of 24.3 alleles and 14.9, respectively. Meanwhile, the AAB and ABB accessions contained B genome specific alleles in average number of 9.8 and 17.3, respectively. In both cases, these differences were significant ( $P < 0.001$ , both). Moreover, the relationship between A and B genome specific alleles was additionally revealed with their ratio, which calculated from A genome specific alleles number divided by B genome specific alleles number of each AAB and ABB accession. In AAB cultivars, the averaged ratio was pronouncedly higher (2.70, ranging from 1.57 to 6.50) than in ABB cultivars (0.86, ranging from 0.28 to 1.17). In addition, there was a gap between the ratio of 1.17 and 1.57 which not presented in any AAB nor ABB accessions.

Among the numbers of SSR markers generating A and B genome specific alleles, the results of allele call from capillary fragment analysis showed that the BB\_AAC-3 marker generated two alleles of 157 and 172 bp, each discriminately presented among population here. The 157-bp allele existed in wild *M. balbisiana* (SS&JS 12,

BBw), BBB cultivars and some cultivars containing the B genome (AAB, ABB), but undetectable in accessions with only A genome (AAw, AA, and AAA). On the other hand, the 172-bp allele found in wild *M. acuminata* subsp. *malaccensis* (SS&JS 377, AAw), AAA and the remaining cultivar accessions containing A genome (AB, AAB, and ABB), but undetectable in accessions with only B genome (BBw and BBB). Such phenomenon suggested that the 157- and 172-bp alleles were B and A extra-nuclear genome specific alleles, respectively. Among 41 A and B genome hybrid cultivars, 20 accessions were shown to contain the 172-bp allele while the 157-bp allele presented in 21 accessions.

The homology search and sequence alignment indicated that the BB\_AAC-3 locus was *ycf2* gene compartmentalized in plastid genome. With the PCR amplification, the nucleotide sequences of the partial *ycf2* gene from two wild accessions, *M. balbisiana* (SS&JS 12) and *M. acuminata* subsp. *malaccensis* (SS&JS 377), were obtained and annotated to be 157 and 172 bp, respectively, which supporting the results from capillary fragment analysis. These sequences were multiple aligned to the *ycf2* gene sequences retrieved from the complete plastid genome sequences of two wild banana species, *M. textilis* (Barrett et al. 2014) (GenBank accession KF601567) and *M. acuminata* subsp. *malaccensis* (CIRAD:930, GenBank accession HF677508). The multiple sequence alignment in Fig. 2a showed that the BB\_AAC-3 locus was on 5'-end coding region of *ycf2* gene. The sequences among three *Musa* species were highly conserved except a transition between guanine and adenine bases in the second amino acid residue codon which resulting in differential translation into glycine (GGA codon) and arginine (AGA codon), respectively. The alignment showed that *M. acuminata* 'CIRAD:930' also possessed the 172-bp allele as in SS&JS 377 accession. Interestingly, the locus in *M. textilis* appeared to be 157-bp allele as in *M. balbisiana*. The 15-bp difference between the A and B genome was clearly identified from the gap in the alignment in Fig. 2a, and also expanded by the allele peaks read by the capillary fragment analysis as shown in

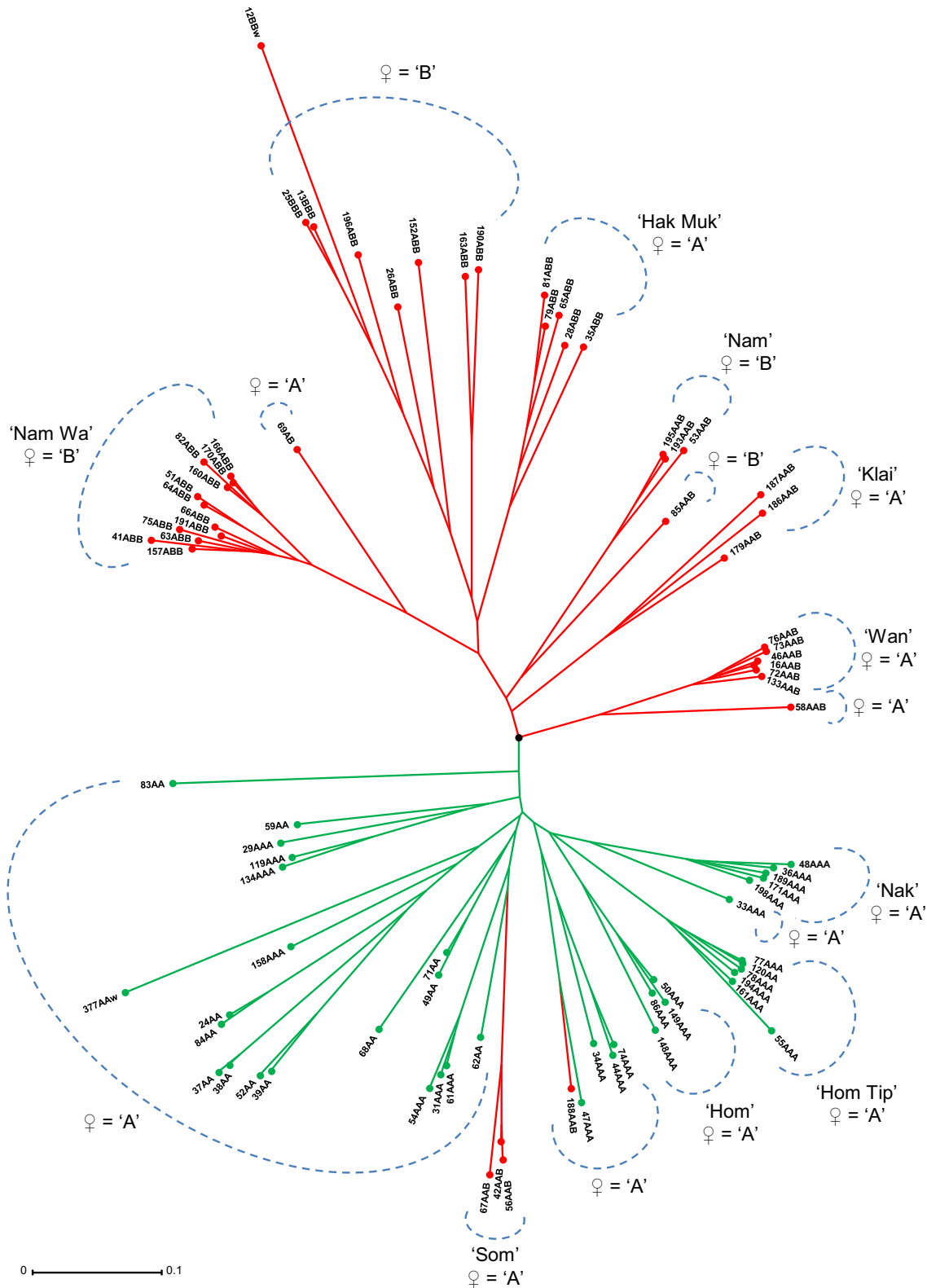


**Fig. 2** The two maternal inheritance identification alleles from SSR marker BB\_AAC-3. (a) The multiple alignment of partial chloroplast *ycf2* gene sequences of two cloned fragments from *Musa balbisiana* (SS&JS 012) and *M. acuminata* subsp. *malaccensis* (SS&JS 377) against the same locus on completed plastid genome of *M. textilis* (GenBank accession KF601567) and *M. acuminata* subsp. *malaccensis* (CIRAD:930, GenBank accession HF677508). The blue boxes mark the 15-bp repetitive nucleotide sequence and the red boxes mark the 5-residues repetitive amino acid sequence. The region of forward and reverse primers of SSR marker ‘BB\_AAC-3’ used in cloning are represented with single and double underline, respectively. The sequence of two clones was reverse complemented

prior to alignment to matching the sense strand. The highlighted guanine base in SS&JS 377 is generated from the PCR using primer designed from *M. balbisiana* SSR-enriched library, hence the actual base might be adenine as in CIRAD:930 genome which would code for arginine instead of glycine. (b) Allele peaks generated from capillary fragment analysis of two wild AAw and BBw accessions as mentioned above, and two ABB hybrid cultivars (SS&JS 35 ‘Hak Muk Nuan’ and SS&JS 64 ‘Nam Wa Nuan’). Peak positions on the chromatogram are related to the raw fragments size, prior to manipulation to obtain the actual allele calls as shown in labels. ♀, maternal side

Fig. 2b. Such difference was generated from the 15-bp nucleotide repeat which occurred only in *M. acuminata* accessions (SS&JS 377 and CIRAD:930), and made the

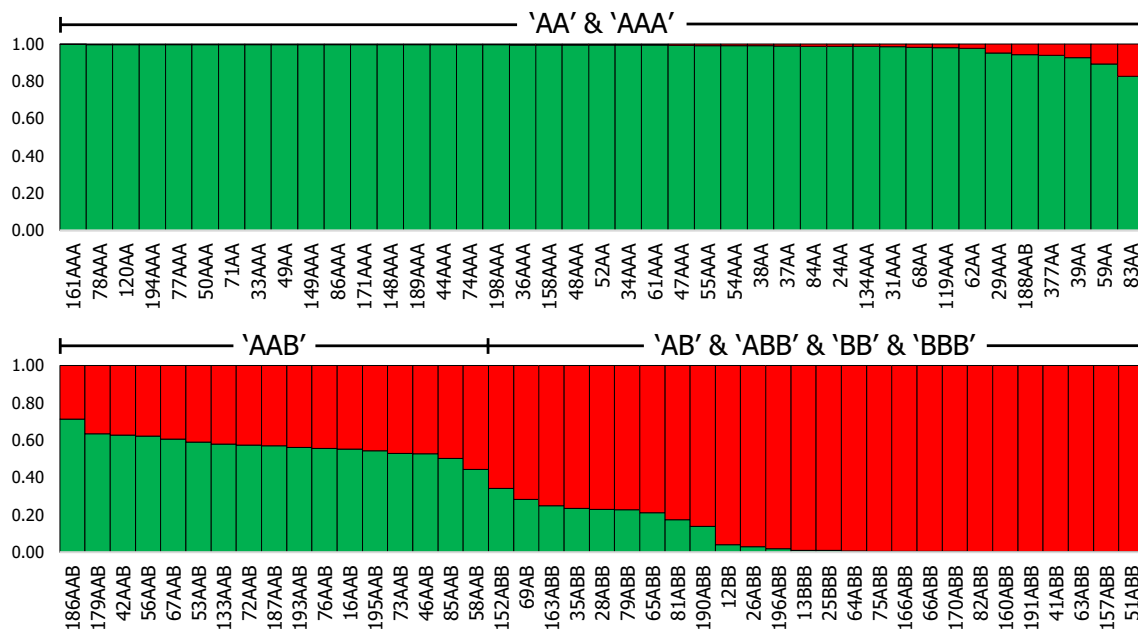
172-bp allele. The repeat sequence of 15 bp, 5'-TTGAGAGAGATCAAG-3', also generated an addition of five amino acid residues (LREIK) to the *ycf2* protein in *M.*



**Fig. 3** Hierarchical cluster showing the genetic relationship of banana cultivars in Thailand. Accessions with only A genome (AAw, AA, and AAA) are displayed as green lines, while accessions containing B genome (AB, AAB, ABB, BBw, and BBB) are red. The

maternity identification, obtained from the SSR marker BB\_AAC-3, of each cluster is represented with '♀' sign following with 'A' and 'B' which indicating the maternal side to A and B genome, respectively





**Fig. 4** Clustering of individual cultivars by Structure software. Each of the 84 accessions is represented with a single column. The *scaleleft* indicates for  $Q$ . With optimal  $K = 2$ , two colored segments with lengths proportional to permuted  $Q$  are represented for each  $K$ : *green*

(A genome) and *red* (B genome). Individuals were grouped into clusters according to sorting of  $Q$ . Three clusters were identified according to the genotypes of cultivars

*acuminata*. Therefore, the long nucleotide repeat in BB\_AAC-3 locus makes this SSR marker to become maternal marker for identifying the parental side of hybrid cultivars. For the example, from Fig. 2b, two cultivar accessions, SS&JS 35 ‘Hak Muk Nuan’ and SS&JS 64 ‘Nam Wa Nuan’, despite they possess the same genome designation of ABB, the former presented the 172-bp allele indicating that its maternity appears to be from A genome while the later presented the 157-bp allele indicating B genome maternity. This data supported the previous chloroplast sequence analysis from these cultivar groups ‘Hak Muk’ and ‘Nam Wa’ (Swangpol et al. 2007).

#### Genetic relationship among different genotypes of banana cultivars

Totally 471 polymorphic alleles were obtained from amplification of 24 SSR markers, while another 4 SSR markers were unable to generate the results. The allele calls were scored as binary data to construct a hierarchical cluster to reveal the genetic relationship among cultivar accessions (Fig. 3). The total of 82 cultivar accessions and 2 wild accessions (AAw and BBw) were clearly divided into two main clusters of bananas: A genome cluster (containing AAw, AA, and AAA accessions) and B genome cluster (containing AB, AAB, ABB, BBw, and BBB accessions). However, four AAB accessions (SS&JS 188 and three Mysore type or ‘Som’, SS&JS 42, 56, 67) were

also included in A genome cluster. By looking at each cluster, in B genome cluster, three major monophyletic cultivar groups were able to be distinguished among B-genome containing accessions; Pisang Awak type or ‘Nam Wa’ (ABB), Bluggoe type or ‘Hak Muk’ (ABB), and Plaintain type or ‘Klai’ (AAB). Other accessions were assembled of heterologous genotypes and defined to be polyphyletic clusters. All 12 accessions of ‘Nam Wa’ with ABB genome were tightly assembled within the cluster indicating the low variation of this cultivar group. The one accession of AB group, SS&JS 69, was located close to ‘Nam Wa’ group but not included into the group. Meanwhile in A genome cluster, most accessions were heterologous assembled as well as in B genome cluster, however, three cultivar groups still be retrieved; Red type or ‘Nak’ (AAA), Gros Michel and Cavendish or ‘Hom’ (AAA) and ‘Hom Tip’ (AAA).

By combining the results from BB\_AAC-3 SSR marker to the phylogenetic clustering, the maternity parental side of hybrid cultivars was informative to cultivar groups in both genome clusters. All clusters in A genome cluster were maternally inherited from A genome species, including Mysore or ‘Som’ (AAB) group even they also possessed the B genotype. In contrast, cultivar groups in B genome clusters were maternally inherited either from A or B genome. Up to four from seven clusters contained maternity from A genome, including ‘Hak Muk’ (ABB), ‘Klai’ (AAB), and ‘AB’ groups. The rest of the clusters



contained maternity from B genome, including ‘Nam Wa’ (ABB) and BBB groups.

Besides the phylogenetic clustering, the model-based clustering with Bayesian analysis was also constructed using the Structure software (Fig. 4). By testing with  $K$  from 1 to 30, the optimal number of  $K$  was estimated to be  $K = 2$  (Online Resource 7). This  $K$  was correlated with phylogram, in which, each  $K$  represented for the A and B genome clusters. By sorting of permuted  $Q$ , three clusters were identified based on the genotypes of cultivars. The first cluster contained cultivars with AA and AAA genotypes, excepting for one accession SS&JS 188 (AAB) which also presented in the A genome cluster of phylogram. The high  $Q$  values proportional to A genome ( $Q_A$ , 0.9979–0.8263) supported these pure-A accessions into this cluster, while low  $Q$  values proportional to B genome ( $Q_B$ , 0.0021–0.1737) were also presented. The second cluster contained cultivars with only AAB genotype, including of three accessions (SS&JS 42, 56, 67) which previously grouped in A genome cluster of phylogram. The roughly equivalent between A ( $Q_A$ , 0.7135–0.4432) and B ( $Q_B$ , 0.2865–0.5568) genomes proportion was able to be assumed, however, most accessions still had  $Q_A$  more than  $Q_B$ . And the third cluster contained the rest of cultivars with mostly genotypes of ABB and BBB, with low  $Q_A$  values (0.3413–0.0020) and high  $Q_B$  values (0.6587–0.9980). The position of each cluster certainly correlated to the phylogram, in which, the cluster of AAB cultivars was in the middle between AA, AAA cultivars and ABB, BBB cultivars. However, each individual accession position displayed no correlation to the cultivar groups as in phylogram. The position of SS&JS 83 was presented as the frontier between A and B genome cluster. And two BBB accessions (SS&JS 13, 25) and all ‘Nam Wa’ accessions were strictly grouped together as in phylogram.

### Wild banana species fingerprinting

Using the given criteria, a set of four SSR markers, consisting of BB\_CT-8, BB\_GAA-35, BB\_GAA-4, and BB\_GAA-31, was chosen and tested with 12 wild species. However, the results from capillary fragment analysis showed unexpected consequence. Three out of four, BB\_GAA-35, BB\_GAA-4, and BB\_GAA-31, did not yield good results after analysis, which at least one sample failed to generate any detectable peaks for allele call (data not shown). With the failure to generate allele pattern of all accessions in selected species, these markers were ambiguous for interpretation and excluded from the candidates. The FAM-labeled BB\_CT-8 marker, however, was able to generate 1–3 allele peaks each in all species, ranging from 206 to 253 bp (Fig. 5). Four species were

**Fig. 5** The candidate banana species fingerprinting. Allele peaks generated from capillary fragment analysis of 12 banana wild species found in Thailand using one SSR marker BB\_CT-8 (labeled with FAM). Peak positions on the chromatogram are related to the raw fragments size, prior to manipulation to obtain the actual allele calls as shown in the labels

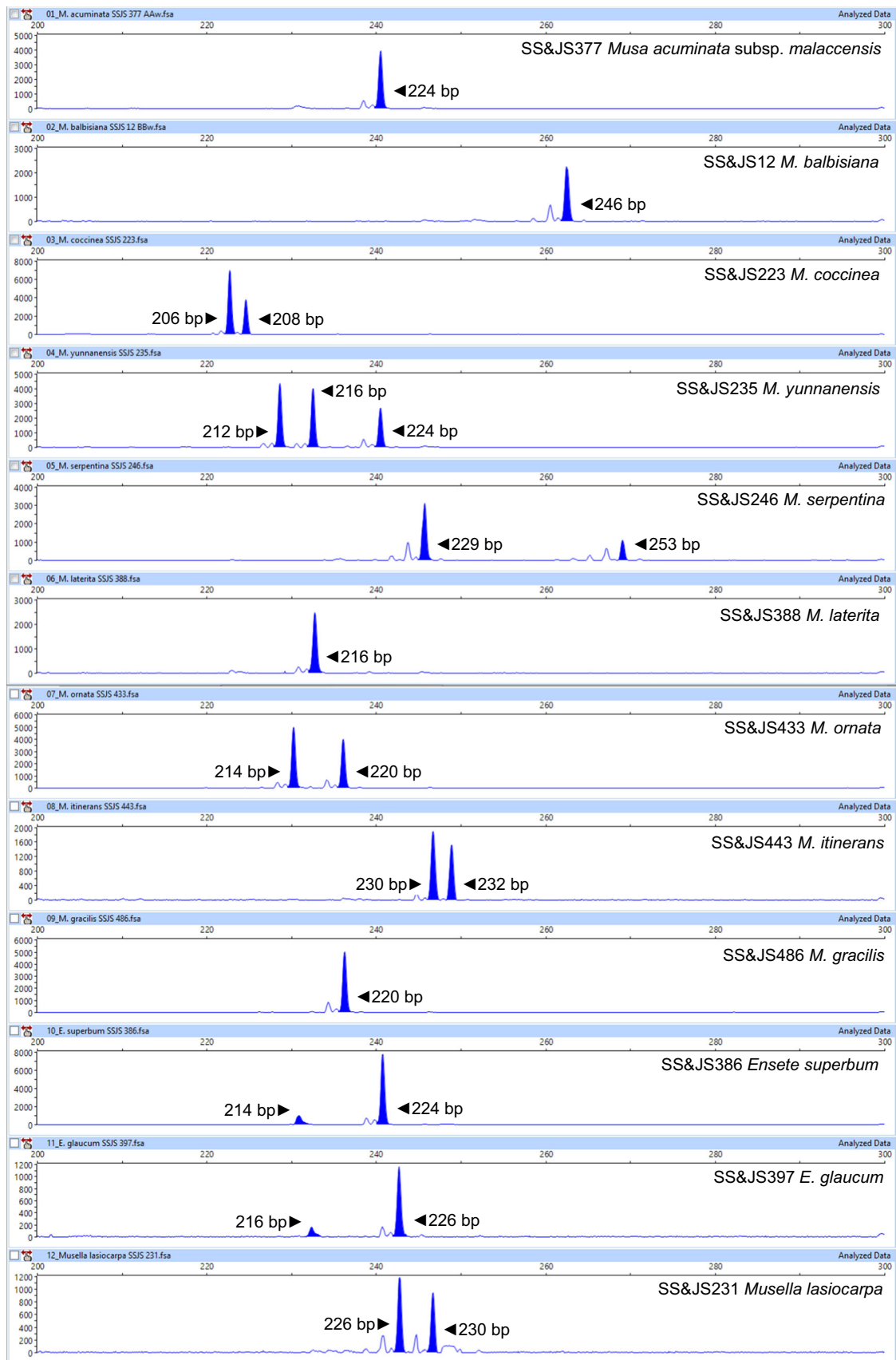
revealed to contain single allele of this locus, while seven species contained two alleles, and only one species contained up to three alleles. The allele peaks pattern was unique in each species, on the other hand, distinguishable for individual species.

## Discussion

### Nuclear genome size and ploidy level

Nuclear DNA content estimation by flow cytometry was conducted in this study with 12 wild banana species and also 82 cultivars. The results were statistically supported to clarify the genome size differences between some taxa of wild bananas. The smaller genome size of BB genome than AA genome has been suggested in some studies (Doležal et al. 1994; Lysák et al. 1999), however, discrepant result was also reported (de Jesus et al. 2013). By the way, more solid evidence to dissolve this conflict recently came from the genome sequencing approaches and supported the smaller genome size of BB genome (D’Hont et al. 2012; Davey et al. 2013). Again, the BB genome was reported here to be smaller. Several molecular analyses revised the old five sections of genus *Musa* to become acceptant with only two sections, *Musa* and *Callimusa* (Häkkinen 2013). By comparing these two sections, the genome size of bananas in section *Callimusa* was bigger than *Musa*, as supported by previous report (Bartoš et al. 2005). This study is the first to report the genome size of *Musella* genus, which was found here to be smaller than genus *Ensete*. However, genome size of two *Ensete* taxa, *E. glaucum* and *E. superbum*, reported here is pronouncedly different from *E. gillettii* (Bartoš et al. 2005), hence, estimation with more taxa is inevitable for this conclusion. Nevertheless, groupings by different genome sizes may not be applicable to all taxa since variation in genome size within- and between-species can occur from many molecular events (Biémont 2008).

Estimation of genome size is more useful to clarify ploidy level of banana cultivars (Pillay et al. 2001), since estimation of ploidy by morphological characterization alone may lead to misinterpret (Nsabimana and van Staden 2006). From the total 82 cultivars presented here, only 3 accessions (3.7 %) were revised to their ploidy amendment



(Online Resource 3). Besides the ploidy, different proportion of A and B genome in hybrid triploids is also suggested here. Although the genome size difference between AAB and ABB cultivar groups is found to be insignificant, both groups were significantly smaller genome size than AAA group. The similar result was existed with lower number of accessions (Lysák et al. 1999). This incident should be from the smaller B genome size than A genome. Furthermore, the actual proportion of each genome in hybrid cultivars might require inference from several molecular approaches, e.g., in situ hybridization (D'Hont et al. 2000; Jeridi et al. 2011), or dosage of genome-specific alleles as described in this study.

### Isolation and characterization of microsatellite markers

In spite of the little resources of microsatellite markers constructed from B genome banana, a marker set representing for 28 SSR loci was novel developed from *M. balbisiana* 'Tani' here to be used with genetic study among both wild and cultivar bananas in Thailand. With SSR enrichment using repeat-motif oligonucleotide probes, this study also reflexes di-nucleotide microsatellites in higher frequency than tri-nucleotide microsatellites as compared in banana genomes (Davey et al. 2013) and many other plants (Buhariwalla et al. 2005; Creste et al. 2006; Ray et al. 2006; Shokeen et al. 2007; Eustice et al. 2008). Therefore, we presented SSR specific primers mostly for di-nucleotide repeats of microsatellite markers (Table 2). The triploid bananas trend to have higher allele numbers per loci than diploid bananas. This might be due to replication slippage (Schlötterer and Harr 2004) and unequal homologous recombination (Devos et al. 2002) of each specific locus collectively during long time cultivation, or from tri-parental origin event as similarly occurred in roses (Iwata et al. 2000) and onions (Fredotović et al. 2014). This caused higher variation and higher frequency of alleles.

The common fashion of SSR markers developing strategies, which based on SSR enrichment, cloning, and screening for candidate loci, have been considered to be work-loaded and none cost-effective. The next-generation genome sequencing technologies could provide higher advantageous and feasible by direct selection of repeat-sequences containing reads for primers design (Zalapa et al. 2012). This has been proved for their discriminative power to yield higher rate of polymorphism detection in recent studies (Addisalem et al. 2015; Vukosavljev et al. 2015). However, the SSR loci in this study were isolated before the publication of both *Musa* genomes (D'Hont et al. 2012; Davey et al. 2013). Thus, the markers were inevitably developed by using such common procedures. By the way, by the time of achievable genome sequencing

resources, with progressiveness, all SSR loci were located on both A and B genomes to present in the physical map of 28 SSR loci developed here (Fig. 1). Considering that the markers were valid at the location of all loci on both genomes. The nucleotide ranges between the two genomes were different. The nucleotide range location on B genome was always lower than on A genome due to the smaller B genome size. These differences were between 100 kb and up to 8 Mb as shown in the map. Although the physical map presented here is not dense as the saturated SSR/DaRT linkage map (Hippolyte et al. 2010) or maternal linkage map (Mbanjo et al. 2012) in *M. acuminata*, or incompatible to genetic maps directly retrieved from complete genome sequences, the attempt here to locate SSR loci to the designated *Musa* chromosomes may provide more accurate description of the actual length of DNA that separates loci from each other in basepairs rather than linkage distances in centiMorgans. Yet each loci location is also comprehensively sustained by the clone sequence data. Nevertheless, comparing to the construction of highly informative genetic maps based on combination of genotyping by sequencing (GBS) and single nucleotide polymorphism (SNP) chip array techniques in other crops, e.g., maize (Ganal et al. 2011), rice (Thomson et al. 2012), sunflower (Bowers et al. 2012), oilseed rape (Delourme et al. 2013), and cassava (Soto et al. 2015), studies of genetic mapping in bananas which mostly based on microsatellite markers will have to follow trends into those techniques as well.

Based on advantageous budget, the use of multiplex PCR with fluorescent universal primers here dramatically reduced the cost for both fluorescent labelling and automated fragment analysis service charge. Also, when compared to preliminary PCR runs with commercial standard *Taq* DNA polymerase (data not shown), the use of QIAGEN Multiplex PCR Kit as described in the method yielded higher intensity and cleaner fragment peaks, as also previously reported (Culley et al. 2013). This study demonstrated the first exploitation of multiplex PCR with multiple fluorescent-labeled universal primers in banana, as the previous studies were only based on the use of single M13 primers tailed with different fluorophores (Christelová et al. 2011; Mbanjo et al. 2012). Although the valuable outcome of accurate allele calling was retrieved in this study, some implications may still need to be improved. The low annealing temperature were unavoidable due to high melting temperature at higher than 45 °C generated very low intensity peaks in some loci. Peak intensity ratios of each locus also cannot be used due to primers competition and amplification saturation of PCR cycles. Up to date, the high GC content (Blacket et al. 2012) and modified M13-based universal primers with suggestion of high cycle number (Culley et al. 2013) were proposed to

enhance the primers versatility and allowing higher annealing temperature of multiplex PCR. Unfortunately, those useful advices could not be exploited here since the setup for multiplex PCR approach was done prior to the publications. Another implication came from the failure of 4 SSR markers which generated no detectable fluorescent signals even though the amplified fragments were successfully verified with gel electrophoresis. The similar results were also previously reported with inapplicably 3 out of totally 22 markers (Christelová et al. 2011). In spite of the annealing temperature was yet minimized the resources competition between co-amplifying primers may also be taking into account. Because of three failures from seven primer sets were dealing with BGHr universal sequence, suggesting for its possibly low versatility and should be avoided. Considering on these implications, better advantageous choices to raise multiplex PCR performance would be recommended in future projects.

Genotyping of alleles for SSR loci conveniently conducted with binary scoring, however, the retrieved information was descended as dominant marker. To achieve the codominant nature of SSR marker, several methods were proposed to quantitatively estimate allele dosage or frequencies among samples including of Microsatellite DNA Allele Counting-Peak Ratios (MAC-PR) (Esselink et al. 2004), an R package 'POLYSAT' (Clark and Jasieniuk 2011), and Microsatellite Allele Dose and Configuration Establishment (MADCE) (van Dijk et al. 2012). Although the methods were designed to use with polyploid plants, strategies to analyses may require well knowledge in genetic background of plant subjects to achieve applicable results. However, the genetic background of cultivated bananas is rather complicated than those successfully analyzed subjects, due to many ploidy levels and genome combinations, high genetic diversity, multiple origins from different wild hybrids between two principle ancestral species and, difficulty to produce high number of viable off springs from cultivars. These cause low informative genetic mapping resources compared to other important crops (Heslop-Harrison and Schwarzhacher 2007). Additionally, due to date as our knowledge, no banana genetic studies have accessed to feasible of any codominant scoring methods, and also this study mainly focused on proposition of SSR markers from B genome.

### Genetic relationship analysis among cultivars

From the 24 SSR loci, total 471 alleles were revealed in AA<sub>w</sub>, BB<sub>w</sub> and cultivars with the average of 19.6 alleles per loci. When combined to high diversity background of samples, all loci were capable to present rather high PIC values, with the average of 0.8363. With SSR profiling, the relationship of Thai cultivated bananas was revealed

through two clustering approaches, the phylogenetic (Fig. 3) and Structure analysis (Fig. 4). Two main clusters of A and B genomes were presented in the phylogram, which correlated to the result from Structure analysis that inferring for the optimal number of population at  $K = 2$ . However, inferred  $K$  here was very different from the previously reported at  $K = 21$  in higher number of banana subjects (de Jesus et al. 2013). The reasons to explain this may come from the fact that this study excluded other non-A and non-B wild species from the analysis. Only one accession of each AA<sub>w</sub> and BB<sub>w</sub> genome was included in profiling here, whereas, variation of *Musa* species, including many subspecies of AA<sub>w</sub> genome, was included in previous study. These furnished high informative of allelic pool which subsequently be reflected as high inferred  $K$  of those wilds and cultivars population.

The A genome cluster of the phylogram mostly comprises of AA<sub>w</sub>, AA, and AAA bananas but some AAB cultivars are also present, which pushing the remaining AAB and all other BB<sub>w</sub>, BBB, and ABB accessions into the B genome cluster. While the Structure analysis gave more precise classification by grouping all AA<sub>w</sub>, AA, and AAA accessions into the 'AA' and 'AAA' cluster, all 'AAB' accessions into the 'AAB' cluster, and all BB<sub>w</sub>, BBB, AB, and ABB accessions into the last cluster. These AAB cultivars have the A maternal origin allele and low number of B specific alleles (Table 1). The calculated scores included them into the A segment but still in their own cluster. Formerly, some cultivars in some clusters had different genome designations from the groups. This led to a careful revision of data by checked at the plants and DNA samples, counting number of A and B specific alleles, calculating of allele ratio, and comparing to cultivar and genome designation groups. Eventually, it came to the conclusion that some accessions had been misclassified since the beginning and, therefore, was edited in this study (Online Resource 3).

These information again affirms the discrimination power of the established marker sets here. With limited numbers of the markers, they were still very effective to clarify cultivar groups which are very diverse in Thailand. Not only well known cultivar groups such as Thai native cultivars of Gross Michel and Cavendish or 'Hom' (AAA), Plantain or 'Klai' (AAB), and Pisang Awak or 'Nam Wa' (ABB) were clearly distinguish but several other cultivars also clearly grouped according to their similarities. These groups have no specific subgroup term internationally for example, 'Nam' (AAB) and 'Wan' (AAB). Moreover, unique individual cultivars were decisively separated which correspondent to their unique phenotypes such as SS&JS 33 'Nom Sao' (AAA), SS&JS 58 'Khai Boran' (AAB), and SS&JS 152 'Niu Mue Nang' (ABB) (Swangpol et al. 2009). Finally, these markers together with other



developing markers from genomic sequences will help to solve the complicated evolutionary history of both wild and cultivated bananas. They might also help to confirm new concepts of banana domestication such as actual genome dosage ratio which might not be integrally precise, for example, between triploid A and B genome of 3:0, 2:1 or 1:2; multiple interspecific hybrids (not only once and not only from A or B); and somatic triploidization (asexually changes from diploid to triploid).

### Wild banana species fingerprinting

Only few reports used SSR markers for study in population of wild bananas such as *M. balbisiana* (Ge et al. 2005) and *Ensete ventricosum* (Getachew et al. 2014). None of them has validated markers to be used across the banana species diversity. The study here is the first preliminary study to evaluate SSR marker sets with several wild species together with cultivars and their wild parental species. Most of the makers in this study could generate some specific allele sizes for most wild species tested here. However, altogether in the marker set, it would create an informative unique pattern for each individual. After data analysis, we selected a number of good candidate markers, BB\_CT-8, BB\_GAA-35, BB\_GAA-4, and BB\_GAA-31, which could generate clear results with minimal absent alleles. Among these markers, only BB\_CT-8 could generate specific allele products for all species tested. Whether or not these allele products from BB\_CT-8 and other markers are really specific for each species is needed to be validated more with individuals in wild population of each species in addition to investigation more on genomic details of each species.

### Conclusion

In summary, this study achieved the isolation and characterization of total 28 microsatellite markers from the B genome of wild banana species, *M. balbisiana*, and been tested among wild and cultivated bananas population in Thailand. The physical map, both A and B genome, of all SSR loci was presented here as anticipation for further genetic researches such as QTL mapping. As literatures were reviewed, this study was the first to report the utilization of A and B genome specific alleles and their ratio dosage, which verifying for the usefulness of SSR-based tools to investigate the genotypic characters of polyploidy and hybrid cultivars, rather than in situ hybridization approaches. An SSR locus on plastid genome was proved for its feasibility to identify the maternal side of hybrid cultivars, which is also the considerable first attempt by SSR in banana studies. With the use of maternal inheritance marker in large populations of

wilds and cultivars, the gene flows of banana taxa should be revealed in this banana cradle area of Southeast Asia or even worldwide. This study also proposed the first candidate fingerprinting SSR marker for identification of wild banana species in all three genera of Musaceae. Finally, the efforts in this study expect to exploit the usefulness of microsatellite markers for genetic study rather than solely construction of dendrogram and may express the better polymorphic efficiency of markers in genetic characterization power. These markers benefit for banana genotypic and species identifications, crop improvement, germplasm analysis and assessment of genetic diversity and relationship.

**Acknowledgments** This research was financially supported by the Thailand Research Fund in collaboration with the Commission on Higher Education of Thailand to JS (MRG5380133) and to SCS (MRG5280100) and Plant Genetic Conservation Project, under the Royal Initiative of Her Royal Highness Princess Maha Chakri Sirindhorn to PPC. Research Assistant Supporting Grant was provided by Faculty of Science, Mahidol University to TR. The authors are grateful for banana samples and knowledge from all institutes and collectors including Pra Sobhon Khanaporn and Mr. Samadchai Chattrakhom.

### Compliance with ethical standards

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

### Information on Electronic Supplementary Material

**Online Resource 1.** Banana samples, 12 wild and 82 cultivar accessions from the collections in Thailand, with their source locations.

**Online Resource 2.** Supplementary materials and methods.

**Online Resource 3.** List of banana cultivars in Thailand with previous and suggested genome designations.

**Online Resource 4.** SSR loci information and position on designated chromosomes of AA genome: *Musa acuminata* accession 'Pahang'.

**Online Resource 5.** SSR loci information and position on designated chromosomes of BB genome: *Musa balbisiana* accession 'Pisung Klutuk Wulung'.

**Online Resource 6.** Microsatellite profiling of A- and B-genome specific alleles in 17 SSR markers.

**Online Resource 7.** Estimation of optimal *K* evaluated by Structure Harvester program.

### References

- Addisalem AB, Esselink GD, Bongers F, Smulders MJM (2015) Genomic sequencing and microsatellite marker development for *Boswellia papyrifera*, an economically important but threatened tree native to dry tropical forests. AoB PLANTS 7:plu086. doi:10.1093/aobpla/plu086
- Anderson JA, Churchill GA, Sutriquet JE, Tanksley SD, Sorrells ME (1993) Optimizing parental selection for genetic linkage maps. Genome 36:181–186. doi:10.1139/g93-024

- Barrett CF, Specht CD, Leebens-Mack J, Stevenson DW, Zomlefer WB, Davis JI (2014) Resolving ancient radiation: can complete plastid gene sets elucidate deep relationships among the tropical ginger (Zingiberales)? *Ann Bot (Oxford)* 113:119–133. doi:10.1093/aob/mct264
- Bartoš J, Alkhimova O, Doleželová M, De Langhe E, Doležel J (2005) Nuclear genome size and genomic distribution of ribosomal DNA in *Musa* and *Ensete* (Musaceae): taxonomic implications. *Cytogenet Genome Res* 109:50–57. doi:10.1159/000082381
- Biémont C (2008) Genome size evolution: within-species variation in genome size. *Heredity* 101:297–298. doi:10.1038/hdy.2008.80
- Blacket MJ, Robin C, Good RT, Lee SF, Miller AD (2012) Universal primers for fluorescent labelling of PCR fragments—an efficient and cost-effective approach to genotyping by fluorescence. *Molec Ecol Resources* 12:456–463. doi:10.1111/j.1755-0998.2011.03104.x
- Bowers JE, Bachlava E, Brunick RL, Rieseberg LH, Knapp SJ, Burke JM (2012) Development of a 10,000 locus genetic map of the sunflower genome based on multiple crosses. *G3 (Bethesda)* 2:721–729. doi:10.1534/g3.112.002659
- Buhariwalla HK, Jarret RL, Jayashree B, Crouch JH, Oritz R (2005) Isolation and characterization of microsatellite markers from *Musa balbisiana*. *Molec Ecol Notes* 5:327–330. doi:10.1111/j.1471-8286.2005.00916.x
- Chomchalow N, Silayoi B (1984) Banana germplasm in Thailand. IBPGR/SEAP Newslett 8:23–28. [http://pikul.lib.ku.ac.th/Full\\_text\\_kukr/KU0222046c.pdf](http://pikul.lib.ku.ac.th/Full_text_kukr/KU0222046c.pdf)
- Christelová P, Valárik M, Hřibová E, Van den Houwe I, Channelière S, Roux N, Doležel J (2011) A platform for efficient genotyping in *Musa* using microsatellite markers. *AoB Plants* 2011:plr024. doi:10.1093/aobpla/plr024
- Clark LV, Jasieniuk M (2011) POLYSAT: an R package for polyploid microsatellite analysis. *Molec Ecol Resources* 11:562–566. doi:10.1111/j.1755-0998.2011.02985.x
- Creste S, Benatti TR, Orsi MR, Risterucci A-M, Figueira A (2006) Isolation and characterization of microsatellite loci from a commercial cultivar of *Musa acuminata*. *Molec Ecol Notes* 6:303–306. doi:10.1111/j.1471-8286.2005.01209.x
- Crouch JH, Crouch HK, Ortiz R, Jarret RL (1997) Microsatellite markers for molecular breeding of *Musa*. *InfoMusa* 6:5–6. doi:10.1300/J144v01n01\_06
- Culley TM, Stamper TI, Stokes RL, Brzyski JR, Hardiman NA, Klooster MR, Merritt BJ (2013) An efficient technique for primer development and application that integrates fluorescent labeling and multiplex PCR. *Appl Plant Sci* 1:1300027. doi:10.3732/apps.1300027
- 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 Genom* 14:683. doi:10.1186/1471-2164-14-683
- de Jesus ON, Silva SDOE, Amorim EP, Ferreira CF, de Campos JMS, Silva GDG, Figueira A (2013) Genetic diversity and population structure of *Musa* accessions in ex situ conservation. *BMC Plant Biol* 13:41. doi:10.1186/1471-2229-13-41
- Delourme R, Falentin C, Fomeju BF et al (2013) High-density SNP-based genetic map development and linkage disequilibrium assessment in *Brassica napus* L. *BMC Genom* 14:120. doi:10.1186/1471-2164-14-120
- Devos KM, Brown JKM, Bennetzen JL (2002) Genome size reduction through illegitimate recombination counteracts genome expansion in *Arabidopsis*. *Genome Res* 12:1075–1079. doi:10.1101/gr.132102
- D'Hont A, Paget-Goy A, Escoute J, Carreel F (2000) The interspecific genome structure of cultivated banana, *Musa* spp. revealed by genomic DNA in situ hybridization. *Theor Appl Genet* 100:177–183. doi:10.1007/s001220050024
- D'Hont A, Denoeud F, Aury J-M et al (2012) The banana (*Musa acuminata*) genome and the evolution of monocotyledonous plants. *Nature* 488:213–217. doi:10.1038/nature11241
- Doležel J, Doleželová M, Novák FJ (1994) Flow cytometric estimation of nuclear DNA amount in diploid bananas (*Musa acuminata* and *M. balbisiana*). *Biol Plant* 36:351–357
- Doležel J, Greilhuber J, Suda J (2007) Estimation of nuclear DNA content in plants using flow cytometry. *Nat Protocols* 2:2233–2244. doi:10.1038/nprot.2007.310
- Droc G, Larivière D, Guignon V et al (2013) The banana genome hub. *Database (Oxford)* 2013:035. doi:10.1093/database/bat035
- Earl DA, von Holdt BM (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genet Resources* 4:359–361. doi:10.1007/s12686-011-9548-7
- Esselink GD, Nybom H, Vosman B (2004) Assignment of allelic configuration in polyploids using the MAC-PR (microsatellite DNA allele counting—peak ratios) method. *Theor Appl Genet* 109:402–408. doi:10.1007/s00122-004-1645-5
- Eustice M, Yu Q, Lai CW, Hou S, Thimmapuram J, Liu L, Alam M, Moore PH, Presting GG, Ming R (2008) Development and application of microsatellite markers for genomic analysis of papaya. *Tree Genet Genomes* 4:333–341. doi:10.1007/s11295-007-0112-2
- Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molec Ecol* 14:2611–2620. doi:10.1111/j.1365-294X.2005.02553.x
- Falush D, Stephens M, Pritchard JK (2003) Inference of population structure: extensions to linked loci and correlated allele frequencies. *Genetics* 164:1567–1587
- Falush D, Stephens M, Pritchard JK (2007) Inference of population structure using multilocus genotype data: dominant markers and null alleles. *Molec Ecol Notes* 7:574–578. doi:10.1111/j.1471-8286.2007.01758.x
- Fredotović Z, Šamanić I, Weiss-Schneeweiss H, Kamenjarin J, Jang T, Pužina J (2014) Triparental origin of triploid onion, *Allium × cornutum* (Clementi ex Visiani, 1842), as evidenced by molecular, phylogenetic and cytogenetic analyses. *BMC Plant Biol* 14:24. doi:10.1186/1471-2229-14-24
- Ganal MW, Durstewitz G, Polley A et al (2011) A large maize (*Zea mays* L.) SNP genotyping array: development and germplasm genotyping, and genetic mapping to compare with the B73 reference genome. *PLOS One* 6:e28334. doi:10.1371/journal.pone.0028334
- 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. *Molec Ecol* 14:933–944. doi:10.1111/j.1365-294x.2005.02467.x
- Ge C, Cui YN, Jing PY, Hong XY (2014) An alternative suite of universal primers for genotyping in multiplex PCR. *PLOS One* 9:e92826. doi:10.1371/journal.pone.0092826
- Getachew S, Mekbib F, Admassu B, Kelemu S, Kidane S, Negisho K, Djikeng A, Nzuki I (2014) A look into genetic diversity of Enset (*Ensete ventricosum* (Welw.) Cheesman) using transferable microsatellite sequences of banana in Ethiopia. *J Crop Improv* 28:159–183. doi:10.1080/15427528.2013.861889
- Häkkinen M (2013) Reappraisal of sectional taxonomy in *Musa* (Musaceae). *Taxon* 62:809–813. doi:10.12705/624.3
- Häkkinen M, Hong W (2007) New species and variety of *Musa* (Musaceae) from Yunnan, China. *Novon* 17:440–446. doi:10.3417/1055-3177(2007)17[440:NSAVOM]2.0.CO;2

- Heslop-Harrison JS, Schwarzacher T (2007) Domestication, genomics and the future for banana. *Ann Bot (Oxford)* 100:1073–1084. doi:[10.1093/aob/mcm191](https://doi.org/10.1093/aob/mcm191)
- Hippolyte I, Bakry F, Seguin M et al (2010) A saturated SSR/DArT linkage map of *Musa acuminata* addressing genome rearrangements among bananas. *BMC Plant Biol* 10:65. doi:[10.1186/1471-2229-10-65](https://doi.org/10.1186/1471-2229-10-65)
- Huang X, Madan A (1999) CAP3: a DNA sequence assembly program. *Genome Res* 9:868–877. doi:[10.1101/gr.9.9.868](https://doi.org/10.1101/gr.9.9.868)
- International Rice Genome Sequencing Project (2005) The map-based sequence of the rice genome. *Nature* 436:793–800. doi:[10.1038/nature03895](https://doi.org/10.1038/nature03895)
- Iwata H, Kato T, Ohno S (2000) Triparental origin of Damask roses. *Gene* 259:53–59. doi:[10.1016/S0378-1119\(00\)00487-x](https://doi.org/10.1016/S0378-1119(00)00487-x)
- Jakobsson M, Rosenberg NA (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* 23:1801–1806. doi:[10.1093/bioinformatics/btm233](https://doi.org/10.1093/bioinformatics/btm233)
- James EA, Brown GK, Citroen R, Rossetto M, Porter C (2011) Development of microsatellite loci in *Triglochin procera* (Juncaginaceae), a polyploid wetland plant. *Conservation Genet Resources* 3:103–105. doi:[10.1007/s12686-010-9301-7](https://doi.org/10.1007/s12686-010-9301-7)
- Jeridi M, Bakry F, Escoute J, Fondi E, Carreel F, Ferchichi A, D'Hont A, Rodier-Goud M (2011) Homoeologous chromosome pairing between the A and B genomes of *Musa* spp. revealed by genomic in situ hybridization. *Ann Bot (Oxford)* 108:975–981. doi:[10.1093/aob/mcr207](https://doi.org/10.1093/aob/mcr207)
- Kaemmer D, Fischer D, Jarret RL, Baurens F-C, Grapin A, Dambier D, Noyer J-L, Lanaud C, Kahl G, Lagoda PJL (1997) Molecular breeding in the genus *Musa*: a strong case for STMS marker technology. *Euphytica* 96:49–63. doi:[10.1023/A:1002922016294](https://doi.org/10.1023/A:1002922016294)
- Lagoda PJ, Noyer JL, Dambier D, Baurens FC, Grapin A, Lanaud C (1998) Sequence tagged microsatellite site (STMS) markers in the Musaceae. *Molec Ecol* 7:659–663
- Lysák MA, Doleželová M, Horry JP, Swennen R, Doležel J (1999) Flow cytometric analysis of nuclear DNA content in *Musa*. *Theor Appl Genet* 98:1344–1350
- Martin G, Baurens FC, Cardi C, Aury J, D'Hont A (2013) The complete chloroplast genome of banana (*Musa acuminata*, Zingiberales): insight into plastid monocotyledon evolution. *PLOS One* 8:e67350. doi:[10.1371/journal.pone.0067350](https://doi.org/10.1371/journal.pone.0067350)
- Mbanjo EGN, Tchoumboungang F, Mouelle AS, Oben JE, Nyine M, Dochez C, Ferguson ME, Lorenzen J (2012) Molecular marker-based genetic linkage map of a diploid banana population (*Musa acuminata* Colla). *Euphytica* 188:369–386. doi:[10.1007/s10681-012-0693-1](https://doi.org/10.1007/s10681-012-0693-1)
- Miller RNG, Passos MAN, Menezes NNP, Souza MT, Costa MMC, Azevedo VCR, Amorim EP, Pappas GJ, Ciampi AY (2010) Characterization of novel microsatellite markers in *Musa acuminata* subsp. *burmannicoides*, var. Calcutta 4. *BMC Res Notes* 3:148. doi:[10.1186/1756-0500-3-148](https://doi.org/10.1186/1756-0500-3-148)
- Missiaggia A, Grattapaglia D (2006) Plant microsatellite genotyping with 4-color fluorescent detection using multiple-tailed primers. *Genet Molec Res* 5:72–78
- Nei M, Li WH (1979) Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc Natl Acad Sci USA* 76:5269–5273
- Nsabimana A, van Staden J (2006) Ploidy investigation of bananas (*Musa* spp.) from the National Banana Germplasm Collection at Rubona-Rwanda by flow cytometry. *S African J Bot* 72:302–305. doi:[10.1016/j.sajb.2005.10.004](https://doi.org/10.1016/j.sajb.2005.10.004)
- Perrier X, Jacquemoud-Collet JP (2006) DARwin software. CIRAD Agricultural Research for Development. <http://darwin.cirad.fr/>
- Pillay M, Ogundiwin E, Nwakanma DC, Ude G, Tenkouano A (2001) Analysis of genetic diversity and relationships in East African banana germplasm. *Theor Appl Genet* 102:965–970. doi:[10.1007/s001220000500](https://doi.org/10.1007/s001220000500)
- Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155:945–959
- Rahman AY, Usharraj AO, Misra BB et al (2013) Draft genome sequence of the rubber tree *Hevea brasiliensis*. *BMC Genom* 14:75. doi:[10.1186/1471-2164-14-75](https://doi.org/10.1186/1471-2164-14-75)
- Ravishankar KV, Raghavendra KP, Athani V, Rekha A, Sudeepa K, Bhavya D, Srinivar V, Ananad L (2013) Development and characterisation of microsatellite markers for wild banana (*Musa balbisiana*). *J Hort Sci Biotechnol* 88:605–609
- Ray T, Dutta I, Saha P, Das S, Roy SC (2006) Genetic stability of three economically important micropropagated banana (*Musa* spp.) cultivars of lower Indo-Gangetic plains, as assessed by RAPD and ISSR markers. *Plant Cell Tissue Organ Cult* 85:11–21. doi:[10.1007/s11240-005-9044-4](https://doi.org/10.1007/s11240-005-9044-4)
- Rozen S, Skaletsky H (2000) Primer3 on the WWW for general users and for biologist programmers. In: Krawetz S, Misener S (eds) *Bioinformatics methods and protocols: methods in molecular biology*. Humana Press, Totowa, pp 365–386
- Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molec Biol Evol* 4:406–425
- Schlötterer C, Harr B (2004) *Microsatellite instability*. Wiley, Hoboken, NJ, USA. doi:[10.1038/npg.els.0000840](https://doi.org/10.1038/npg.els.0000840)
- Schnable PS, Ware D, Fulton RS et al (2009) The B73 maize genome: complexity, diversity, and dynamics. *Science* 326:1112–1115. doi:[10.1126/science.1178534](https://doi.org/10.1126/science.1178534)
- Schuelke M (2000) An economic method for the fluorescent labeling of PCR fragments. *Nat Biotechnol* 18:233–234. doi:[10.1038/72708](https://doi.org/10.1038/72708)
- Shokeen B, Sethy NK, Kumar S, Bhatia S (2007) Isolation and characterization of microsatellite markers for analysis of molecular variation in the medicinal plant Madagascar periwinkle (*Catharanthus roseus* (L.) G. Don). *Plant Sci (Elsevier)* 172:441–451. doi:[10.1016/j.plantsci.2006.10.010](https://doi.org/10.1016/j.plantsci.2006.10.010)
- Simmonds NW (1962) *The evolution of the bananas*. Longmans, London
- Simmonds NW, Shepherd K (1955) The taxonomy and origins of the cultivated bananas. *J Linn Soc Bot* 55:302–312. doi:[10.1111/j.1095-8339.1955.tb00015.x](https://doi.org/10.1111/j.1095-8339.1955.tb00015.x)
- Soto JC, Ortiz JF, Perlaza-Jiménez L, Vásquez AX, Lopez-Lavalle LA, Mathew B, León J, Bernal AJ, Ballvora A, López CE (2015) A genetic map of cassava (*Manihot esculenta* Crantz) with integrated physical mapping of immunity-related genes. *BMC Genom* 16:190. doi:[10.1186/s12864-015-1397-4](https://doi.org/10.1186/s12864-015-1397-4)
- Swangpol S, Somana J (2011) *Musa serpentina* (Musaceae): a new banana species from western border of Thailand. *Thai Forest Bull Bot* 39:31–36
- Swangpol S, Volkaert HA, Sotto RC, Seelanan T (2007) Utility of selected non-coding chloroplast DNA sequences for lineage assessment of *Musa* interspecific hybrids. *J Biochem Molec Biol* 40:577–587
- Swangpol S, Somana J, Chattrakom S, Sukkaewmanee N, Kiatprapai P (2009) 108 Thai Banana Cultivars [in Thai]. Bangkok Publishing, Bangkok
- Temnykh S, DeClerck G, Lukashova A, Lipovich L, Cartinhour S, McCouch S (2001) Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): frequency, length variation, transposon associations, and genetic marker potential. *Genome Res* 11:1441–1452. doi:[10.1101/gr.184001](https://doi.org/10.1101/gr.184001)
- The Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408:796–815. doi:[10.1038/35048692](https://doi.org/10.1038/35048692)
- The Potato Genome Sequencing Consortium (2011) Genome sequence and analysis of the tuber crop potato. *Nature* 475:189–195. doi:[10.1038/nature10158](https://doi.org/10.1038/nature10158)
- Thomson MJ, Zhao K, Wright M et al (2012) High-throughput single nucleotide polymorphism genotyping for breeding applications



- in rice using the BeadXpress platform. *Molec Breed* 29:875–886. doi:[10.1007/s11032-011-9663-x](https://doi.org/10.1007/s11032-011-9663-x)
- Valmayor RV, Jamaluddin SH, Silayoi B, Kusumo S, Danh LD, Pascua OC, Espino RRC (2000) Banana cultivar names and synonyms in Southeast Asia. International Network for Improvement of Banana and Plantain—Asia and the Pacific Office, Los Baños
- van Dijk T, Noordijk Y, Dubos T, Bink MC, Meulenbroek BJ, Visser RG, van de Weg E (2012) Microsatellite allele dose and configuration establishment (MADCE): an integrated approach for genetic studies in allopolyploids. *BMC Plant Biol* 12:25. doi:[10.1186/1471-2229-12-25](https://doi.org/10.1186/1471-2229-12-25)
- Vukosavljev M, Esselink GD, Van't Westende WPC, Cox P, Visser RGF, Arens P, Smulders MJM (2015) Efficient development of highly polymorphic microsatellite markers based on polymorphic repeats in transcriptome sequences of multiple individuals. *Molec Ecol Resources* 15:17–27. doi:[10.1111/1755-0998.12289](https://doi.org/10.1111/1755-0998.12289)
- Wang JY, Zheng LS, Huang BZ, Liu WL, Wu YT (2010) Development, characterization, and variability analysis of microsatellites from a commercial cultivar of *Musa acuminata*. *Genet Resources Crop Evol* 57:553–563. doi:[10.1007/s10722-009-9493-4](https://doi.org/10.1007/s10722-009-9493-4)
- Weising K, Nybom H, Wolff K, Kahl G (2005) DNA fingerprinting in plants: principles, methods, and applications, 2nd edn. CRC Press, Boca Raton
- Wongniam S, Somana J, Swangpol S, Seelanan T, Chareonsap P, Chadchawan S, Jenjittikul T (2010) Genetic diversity and species-specific PCR-based markers from AFLP analyses of Thai bananas. *Biochem Syst Ecol* 38:416–427. doi:[10.1016/j.bse.2010.03.015](https://doi.org/10.1016/j.bse.2010.03.015)
- Zalapa JE, Cuevas H, Zhu H, Steffan S, Senalik D, Zeldin E, McCown B, Harbut R, Simon P (2012) Using next-generation sequencing approaches to isolate simple sequence repeat (SSR) loci in the plant sciences. *Amer J Bot* 99:193–208. doi:[10.3732/ajb.1100394](https://doi.org/10.3732/ajb.1100394)
- Zane L, Bargelloni L, Patarnello T (2002) Strategies for microsatellite isolation: a review. *Molec Ecol* 11:1–16. doi:[10.1046/j.0962-1083.2001.01418.x](https://doi.org/10.1046/j.0962-1083.2001.01418.x)