CHAPTER 4

PHYLOGENETIC UTILITY OF
THE GRANULE-BOUND STARCH SYNTHASE (*GBSS*) GENE
IN SOUTHEAST ASIAN BANANAS

ABSTRACT

For more than 50 years the leading hypothesis for the origin of edible bananas has been that they are derived from two ancestral species, *Musa acuminata* Colla (A genome) and *M. balbisiana* Colla (B genome). This hypothesis has been based mainly on morphological characterization, chromosome counting, and more recently, molecular makers. While nuclear sequences have been used to infer phylogenetic relationships in several plant species, our analysis is the first application to the bananas. This investigation aims to detect sources of genome constitution in banana hybrids, with special emphasis on the Southeast Asian triploid BBA/ABB complex; 'Namwa'/ 'Hin' and an AAB plantain 'Nga Chang'. Phylogenetic analyses were conducted on the granule-bound starch synthase (*GBSS*) of *Musa*, using an amplified fragment of approximately 1.0-kb containing six introns. Variation within this *GBSS* locus is marked by several nucleotide substitutions, simple sequence repeats in intron 3, and insertions/deletions (indels) in intron 4, 5, 7, and 8. Maximum parsimony analyses identified three alleles of which one can be designated as B allele. The results suggested multiple origins of nuclear genome of the triploid bananas, which are comparatively congruent to those from our previous analyses on chloroplast DNA (*rpl*16 and *ndh*A introns, *psa*A-*ycf*3 and *pet*A-*psb*J-*psb*L spacers). Our data proved the usefulness of the *GBSS* gene in intraspecific phylogenetic analyses of the *Musa*.

4.1 INTRODUCTION

It is believed that banana is one of the earliest cultivated crops and has been domesticated and intensively selected over several millennia by farmers in the Southeast Asian region (Simmonds 1962; De Langhe and De Maret 1999). Hypothesis on the origin of edible bananas that has been put forward more than 50 years ago by Cheesman (1947) and Simmonds and Shepherd (1955). Two wild diploid species in the Eumusa section, namely *Musa acuminata* Colla (A genome) and *M. balbisiana* Colla (B genome) are the putative progenitors of almost all edible seedless banana cultivars (Cheesman 1947; Simmonds and Shepherd 1955; Shepherd 1990). Hybridization between the two parental species might have occurred in eastern Indonesia and Bismarck archipelago. Since the area is out of natural distribution range of *M. balbisiana*, artificial introduction of this species from the mainland Southeast Asia by Neolithic Austronesian people to the hybridization zone is proposed (Simmonds 1962; Argent 1976; De Langhe and De Maret 1999). Subsequent introductions of the hybrids into tropical areas over the world, also back to mainland Asia, were attributed to ancient sea travelers.

Though triploid hybrid bananas have low economic value in the world trade, they are important for rural people in Southeast Asia and other tropical countries. In some places, bananas are staple food crop in the shortage of cereal commodities (INIBAP 2001). Several triploid hybrid cultivars have been grown widely in traditional farming systems for local consumption in Southeast Asia. These cultivars included 'Saba' (BBA/BBB) in the Philippines and its synonym 'Hin' in Thailand and 'Pisang Kepok' in Indonesia; 'Namwa' (BBA) in Thailand and its

synonym 'Pisang Awak' in Malaysia and Indonesia; and 'Chuoi Tay' (BBA) in Vietnam (Valmayor et al. 2000). Although it is probable that *M. acuminata* and *M. balbisiana* could have played important roles in the evolution of these edible clones, their diversity, origins, and genomic evolution have not clearly been understood (Jarret 1990). While several subspecies of *M. acuminata* have been found (Simmonds 1956, 1962), no subspecies has been described in *M. balbisiana*. It was found through isozyme analysis, though, that *M. balbisiana* is also a polymorphic species (Lebot et al. 1993), which was in agreement with Kaemmer et al. (1997) who found that B genome presented in plantains (AAB) and the B genome detected in cooking bananas (ABB/BBA) came from different *M. balbisiana* donors. Shepherd (1990) suggested that this difference occurred because AAB and ABB/BBA cultivars evolved in different areas and/or at different times.

Little genetic differentiation between several crops and their wild progenitors has been found despite the sometimes large morphological differences because of human selective bias toward a small set of genes controlling specific morphological traits (Doebley 1992). It has been demonstrated, however, that non-coding sequences of cpDNA are useful to distinguish maternal lineages of the bananas (Swangpol et al., in preparation) as well as within lower taxonomic levels in other plants (Golenberg 1993; Gielly and Taberlet 1994; Ohsako and Ohnishi 2000; Zhang 2000). Compared to the highly conserved plastidic DNA, single or low copy nuclear-encoded genes could provide more phylogenetically informative characters between or within species. However, interlocus gene conversion and recombination have also been observed in nuclear gene sequences (Small et al. 1998). In addition, nuclear genes preserve the evolutionary history of both maternal and paternal lineages. Nuclear

genes and especially low-copy-number genes have been used widely to study domestication, e.g. in corn (Gaut and Clegg 1993) and in cotton (Seelanan et al. 1997).

Granule-bound starch synthase (*GBSS*, EC 2.4.1.11), also called *waxy* gene, is a nuclear gene, which encodes an important enzyme involved in the amylose synthesis in plant plastids (Vos-Scheperkeuter et al. 1986). The *GBSS* gene has been used successfully for phylogenetic analyses in several plant species; e.g. grasses (Mason-Gamer and Kellogg 1996; Ingram and Doyle 2003), *Ipomoea* (Convolvulaceae) (Miller et al. 1999), Rosaceae (Evans et al 2000), and *Solanum* (Peralta and Spooner 2001). The *GBSS* introns provided more phylogenetically informative characters and higher levels of variation, than the ITS region in grasses (Mason-Gamer et al. 1998) and in 651 bp sequence consisting three exons (9, 10, 11) and two introns in *Ipomoea* (Miller et al 1999). However, ITS sequences showed more diversity than *GBSS* sequences of 700 bp within the genus *Physalis* (Whitson and Manos 1999). The *GBSS* gene is a single copy gene in most monocot and dicot species and which has been examined (van der Leij et al. 1991; Mason-Gamer et al. 1998), and two copies have been found in all Rosaceae (Evans et al. 2000; Evans and Campbell 2002).

The main goal of our research was to evaluate the use of *GBSS* sequences to infer the phylogenetic relationships of *Musa* triploid hybrids with an emphasis on the Southeast Asian cultivars. Secondary goal was to compare these results to previous banana classification based on morphological characters and chromosome constitution and other molecular data, i.e. RFLPs (Gawel and Jarret 1991a; Bhat et al. 1994),

PCR-RFLP (Nwakanma et al. 2003), molecular cytogenetics (Osuji et al. 1997), and AFLPs (Loh et al. 2000; Ude et al. 2002b).

4.2 MATERIALS AND METHODS

4.2.1 Plant Materials

Ten accessions of triploid hybrid cultivars from the Eumusa section ($x = 11$) were included in the analyses along with three *M. balbisiana* accessions (Table 4.1). Young curling leaves in pseudostems were collected. DNA was extracted using DNeasy Plant Mini Kit (Qiagen, USA) according to the manufacturer's instruction.

4.2.2 Primers

The oligonucleotide primer set was developed based on the *GBSS* sequences of several monocot and dicot species obtained from GenBank. The PCR amplification primers were designed for short sequences highly conserved across several plant species in the 3' end of the third exon region (5'-CAG TAC AAG GAT GCG TGG GAC AC-3') and in the 5' end of the ninth exon region (5'-CC AGC CTT CAT CCA TTT TAT CT-3'). The amplified fragment contains six introns and five complete exons.

Amplification reactions were performed in a total reaction mixture of 30 μL, containing 200 μM dNTPs (Promega), 1.5 mM $MgCl_2$, 0.5 μM each primers, 1x *Taq* DNA polymerase buffer, 1 unit of *Taq* DNA polymerase (Qiagen) and 20-50 ng of total genomic DNA as template. Amplification was carried out at 94°C for 3 min, followed by 32 cycles of 45 sec at 94°C, 45 sec at 50°C, 90 sec at 72°C, and final extension at 72°C

Table 4.1 List of *Musa* accessions, their genome composition, origins and sources of accessions used in this study.

No.	Species/Cultivar	Code	Origin[a]	Source[b]
1	*M. balbisiana* 'Cameroon'	BB-CMR	?	ITC0246 / CIRAD
2	*M. balbisiana* (wild)	BB-NAN	Nan, Thailand	TM123
3	*M. balbisiana* 'Pisang Klutuk Wulung'	BB-PKW	Indonesia	ITC1063 / CIRAD
4	*Musa* 'Hin'	BBA-HIN	Nakorn Sri Thammarat, Thailand	KU
5	*Musa* 'Namwa Mali-Ong'	BBA-NWM	Thailand	KU
6	*Musa* 'Namwa Nuan'	BBA-NWN	Ang Thong, Thailand	KU
7	*Musa* 'Namwa Nuan Chan'	BBA-NWS	Thailand	KU
8	*Musa* 'Hak Muk Khiaw'	ABB-HMI	Thailand	KU
9	*Musa* 'Hak Muk Khaow'	ABB-HMO	Thailand	KU
10	*Musa* 'Hin'	ABB-HIN	Yala, Thailand	TM128
11	*Musa* 'Kupulik'	AAB-KPL	?	ITC0909
12	*Musa* 'Nga Chang'	AAB-NGA	Thailand	TM139
13	*Musa* 'Teeb Kum'	AAB-TBK	Thailand	ITC0667

[a] ? indicate unknown origins

[b] ITC= International Transit Center, International Network for the Improvement of Banana and Plantain (INIBAP), Leuven, Belgium;
CIRAD= Centre de Coopération Internationale en Recherche Agronomique pour le Dévelopement, Guadeloupe;
TM= Germplasms collected by the authors in Thailand;
KU= Kasetsart University Student Training Farm, Nakorn Ratchasima, Thailand

for 5 min using a T1 Thermocycler 96 (Whatman, Biometra, Germany).
PCR products were checked on 1% agarose gel using 1xTAE buffer (40
mM Tris, 0.114% glacial acetic acid, 1 mM EDTA, pH 8). A single PCR
amplification product was used for cloning.

4.2.3 Cloning and Sequencing of Fragments

The PCR fragments were purified using MinElute PCR
Purification Kit (Qiagen) and cloned into *E. coli* 'DH10B' using either
TOPO TA Cloning Kit for sequencing (Invitrogen) or pGEM-T Vector
Systems (Promega). Several clones containing an insert of the expected
size were selected. Plasmids were extracted using Wizard® *Plus* SV
Minipreps DNA Purification System (Promega) and sent for sequencing
(Macrogen, Inc, Seoul, South Korea).

4.2.4 Sequence alignments and analyses

Nucleotide sequences of the *GBSS* were first aligned with the
GeneDoc program version 2.6.002 (Nicholas and Nicholas 1997) and
manually edited. Exon/intron boundaries are based upon *Oryza sativa* cv.
group *japonica* (GenBank accession number AF031162, Frances et al.
1998). Nucleotide diversity (π) (Nei 1987) were calculated using DnaSP
4.0 (Rozas et al. 2003).

4.2.5 Phylogenetic analysis

Phylogenetic relationships among *GBSS* gene of *Musa* were
inferred by the maximum parsimony (MP) method. Three analyses were
done treating gaps as missing i.e. analysis of (i) exons only (ii) introns
only (iii) exons and introns. Another analysis (iv) was performed using
the data matrix with all variable sites including exons and introns, but
indels were removed, identifying every potential mutation events and

applying only one unordered multistate character (0/1/2) for each event, then adding the characters back to the matrix (Baum et al. 1994; Peralta and Spooner 2001). Variable sites within missing regions, e.g. about 500 bp from 5' and 3' ends of NAN and HIN, except parsimony informative sites within indels, were not included in the analysis.

The most parsimonious trees were constructed by PAUP* version 4.0b.10 (Swofford 2002), all characters were equally weighted, and heuristic search with 100 random addition analyses and tree bisection-reconnection (TBR) branch-swapping (Steepest descent option in effect), stepwise addition, and MULPARS options was used in the search for most-parsimonious trees. The amount of homoplasy was evaluated with the consistency index (CI), excluding uninformative sites, and the retention index (RI). A strict consensus tree was generated from the equally most-parsimonious trees. To evaluate the branch supports, bootstrap and majority-rule consensus from 500 replicates were conducted.

4.3 RESULTS

4.3.1 *GBSS* sequence analyses

A total of 25 *GBSS* clones were obtained from thirteen *Musa* accessions. These clones included four clones from three *M. balbisiana* accessions, ten clones from three BBA 'Namwa' accessions, six clones from AAB 'Nga Chang', one clone each from the rest. The 1.0-kb *GBSS* sequence of *Musa* investigated include 39 bp from the 3' end of the third translated exon, five complete exons ranging from 64 to 110 bp, and 48 bp from the 5' end of the ninth exon (Fig 4.1). Sequence alignment is shown in Appendix C.
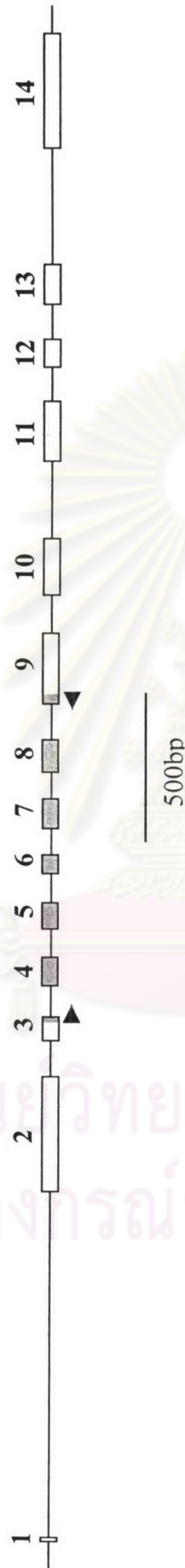
49

Fig. 4.1 Structure of *GBSSI* sequence in *Oryza sativa* cv group *japonica* (AF031162) and the portion of *GBSS* in *Musa* investigated in this study (shaded). The structure showing 14 translated exons (boxes) and their intervening introns (lines). Arrowheads indicate position and direction of sequencing primers used.

49

Average nucleotide diversity was 0.02902. The sequence length ranged from 1,032-1,093 bp. After alignment, the sequences resulted in a final data matrix of 1,118 positions by introducing 17 gaps. Total gap length was 109 bp, which is about 10% of the aligned sequence length. Four gaps were single nucleotide indels, nine were small (2-10 bp), and four were large (15-19 bp). Average GC content at coding positions is 0.483 and at non-coding positions is 0.409.

4.3.2 Phylogenetic analyses

The MP analyses of *GBSS* exon region only had 118 variable positions, of which 44 were parsimony informative. This analysis yielded 98,001 most parsimonious trees with tree length of 166, CI = 0.5679, RI = 0.8148. The topology distinguishes two subclades, of which one clade has more resolution (Fig 4.2). This clade consisting of all diploid *M. balbisiana* sequences as well as hybrids. The analysis of the intron region provided 140 variable characters, of which 61 are parsimony informative. It resulted in 34,775 most parsimonious trees with tree length of 211, CI = 0.5691, RI = 0.7905. The 50% majority-consensus tree (Fig. 4.3) also contained two subclades with little or no resolution. Nonetheless, diploid *M. balbisiana* are still clustered in the same subclade similar to previous analysis.

The analysis of both exons and introns (Fig. 4.4) contained 258 variable sites, of these 105 were parsimony informative. The analysis resulted in 240 most parsimonious trees with tree length of 385, CI = 0.5472 and RI = 0.7828. The 50% majority-consensus tree contained two subclades, which was mostly congruent with consensus trees from earlier analyses.
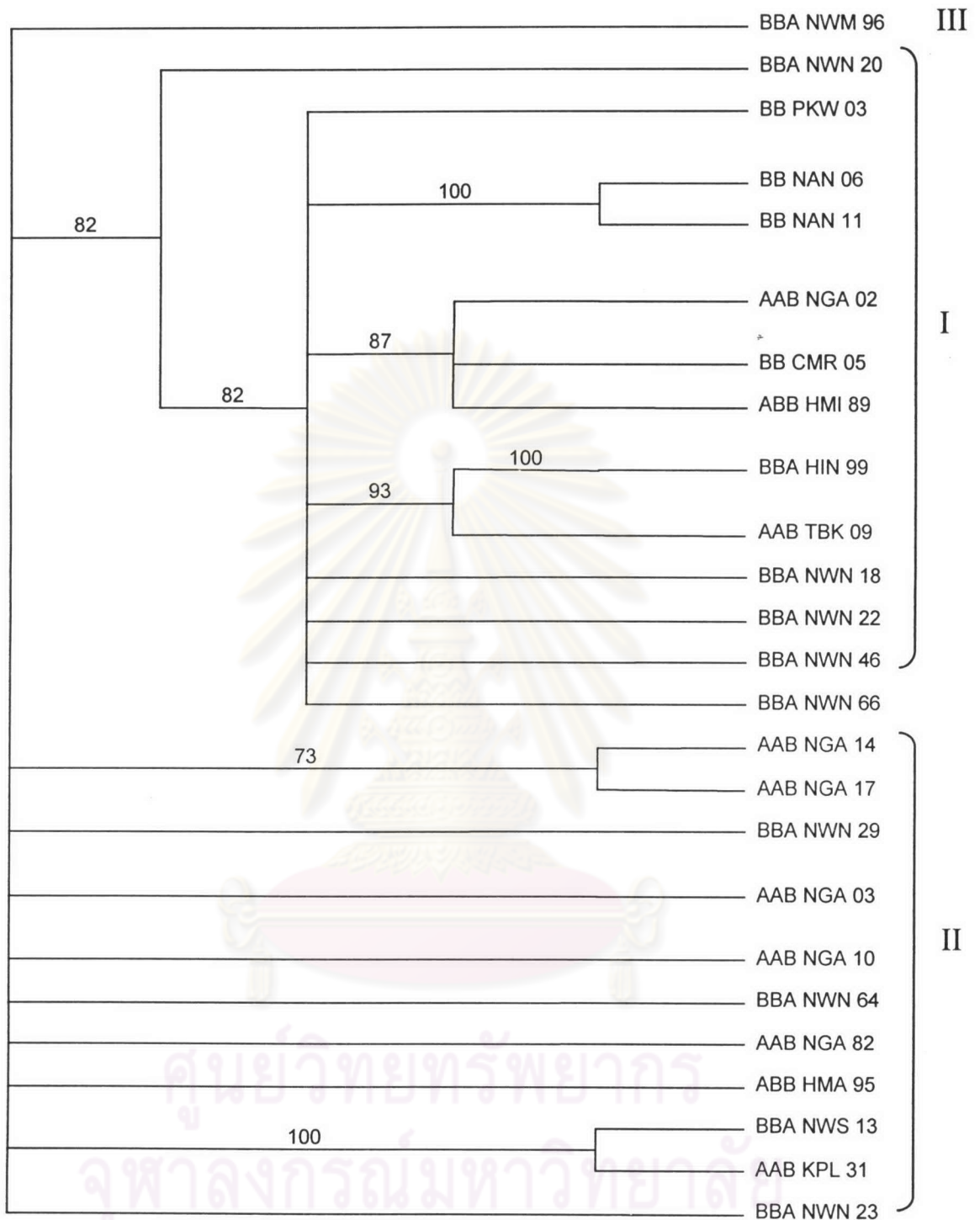
Fig. 4.2 A 50% majority-rule consensus tree of 98,001 most parsimonious trees from the analysis of *GBSS* coding regions of *Musa* with gaps coded as missing. Number above branch represents bootstrap values. Name of taxa appear according to Table 4.1, number behind name indicated clone number.
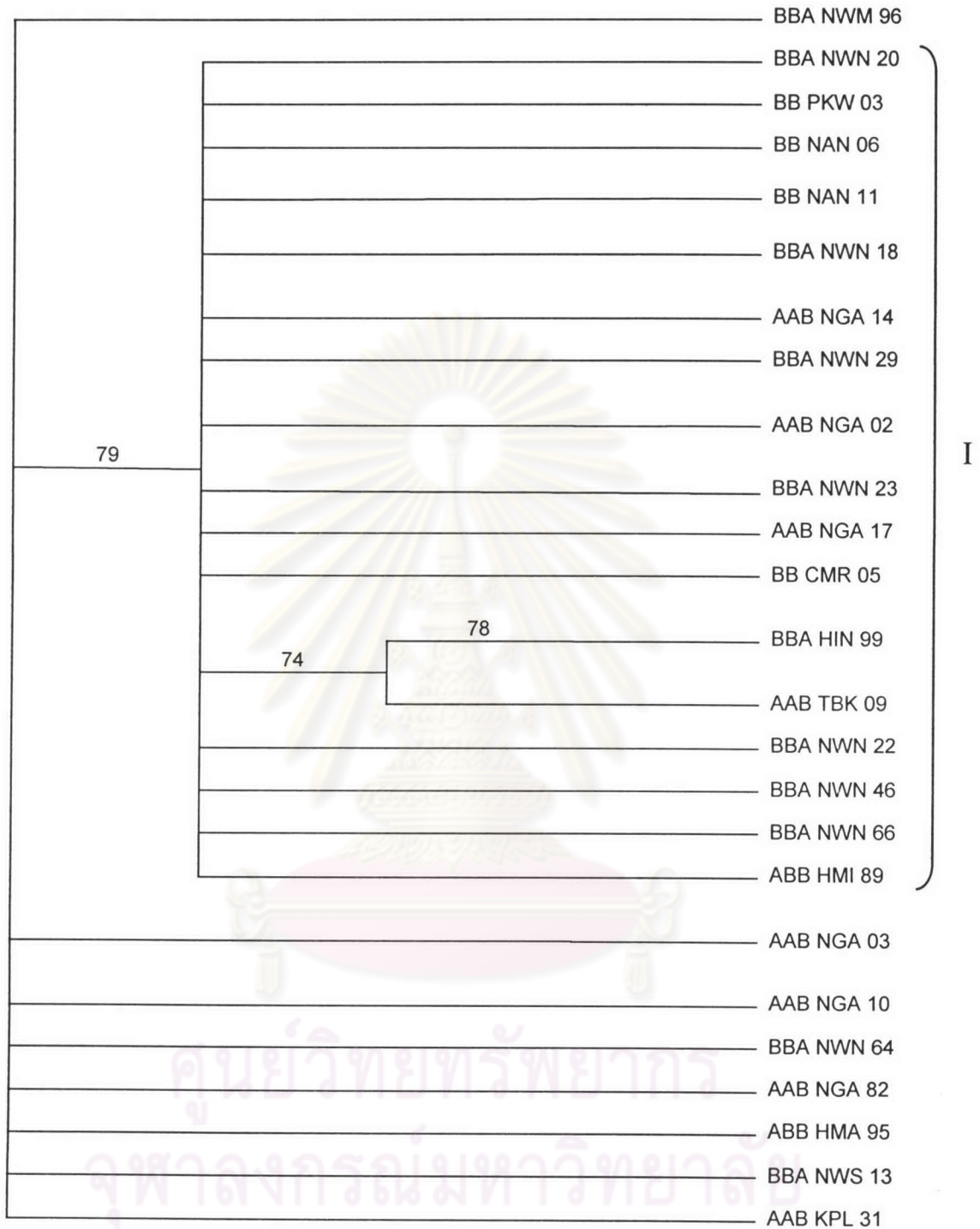
Fig. 4.3 A 50% majority-rule consensus tree of 34,775 most parsimonious trees from the analysis of *GBSS* non-coding regions of *Musa* with gaps coded as missing. Number above branch represents bootstrap values. Name of taxa appear according to Table 4.1, number behind name indicated clone number.

Fig. 4.4 A 50% majority-rule consensus tree of 240 most parsimonious trees from the analysis on all variable sites of *GBSS* in *Musa* with gaps coded as missing. Number above branch represents bootstrap values. Name of taxa appear according to Table 4.1, number behind name indicated clone number.
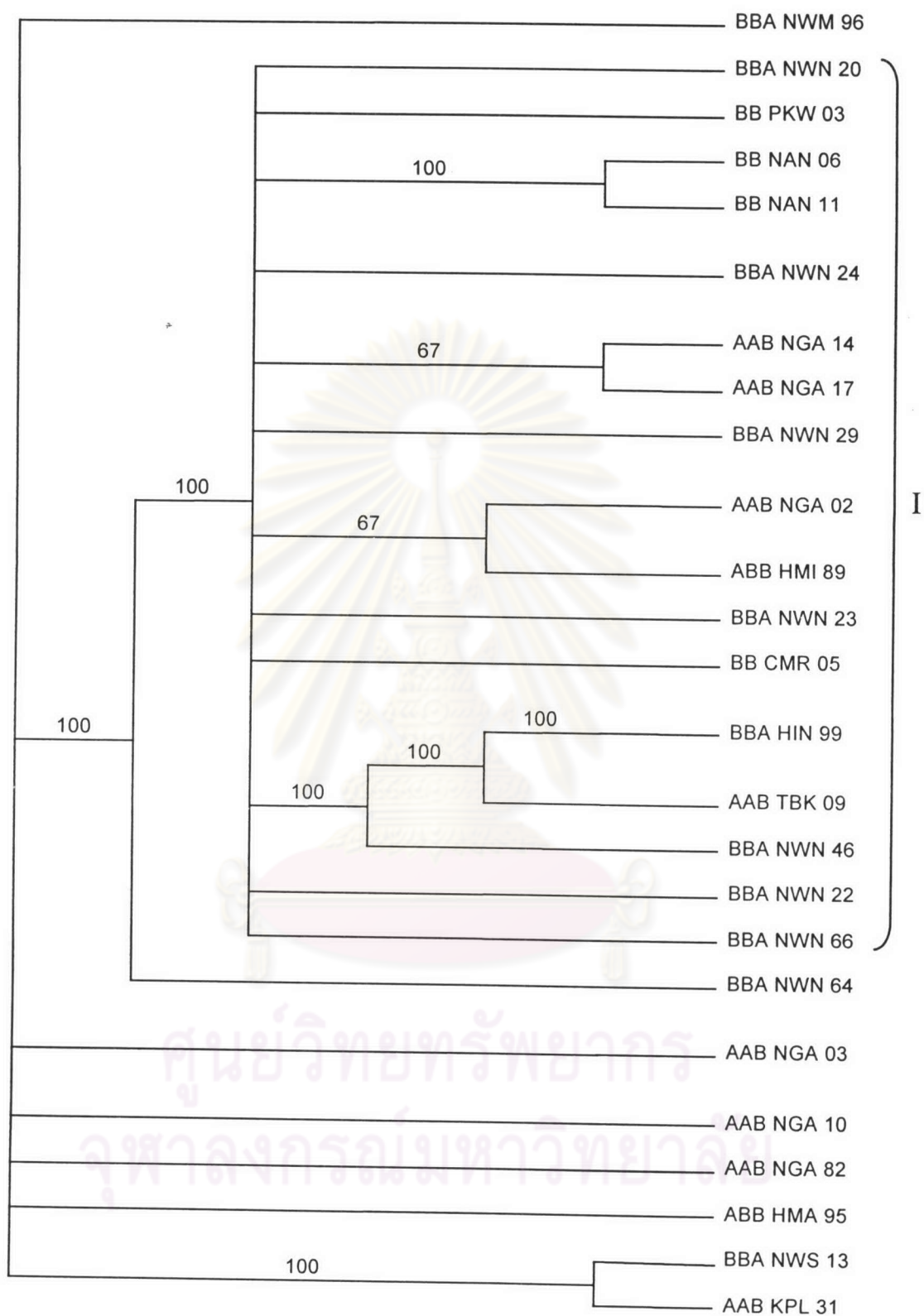
The MP analysis of the complete sequence with all gaps removed and replaced by the coding matrix appears in Fig. 4.5 provided 159 variable characters, of which 54 were parsimony informative. The lower number of variable sites than previous three analysis was due to missing data e.g. in NAN (see aligned matrix in Appendix C). The analysis produced 25,690 most parsimonious trees with tree length of 208, CI = 0.6703, and RI = 0.8718.

The consensus tree (Fig. 4.6) yielded much more resolution than previous three analyses. There were three lineages denoted as I, II, and III. There are two aspects worth considering in subclade I, (1) all three diploid *M. balbisiana* belonged to this subclade. However, (2) diploid *M. balbisiana* 'NAN' from Thailand and 'CMR' formed a clade with 'NGA' and 'HMI' accessions while *M. balbisiana* from Indonesia formed another clade with 'NWN' and 'TBK'. Subclade II consisted of the clones from the same accessions as obtained in subclade I, i.e. 'NWN' and 'NGA'. This incident probably indicated the presence of the second allele of *GBSS* in *Musa*. Lastly, the third lineage consisting of 'NWN 20' hinted that the third allele present in the banana genome.

When only the 17 recoded gaps are analyzed separatedly, it resulted in 99,000 parsimonious trees which 50% majority rule consensus tree was mostly similar to that from the fourth analysis (data not shown). Few differences were (i) the placement of the clones 'NWN 29' with 'NWN 20' instead of with the subclade II in Fig. 4.6, and (ii) there was no resolution in subclade I and II.

```
                              5           10          15
                              .           .           .
BBA_NWN_20    1 1 1 1 0 1 1 0 0 0 1 1 1 0 0 0 0
BBA_NWN_29    1 1 1 1 0 1 1 0 0 0 1 1 1 0 0 1 1
AAB_NGA_03    2 1 1 0 1 1 1 1 1 0 1 0 1 0 0 1 1
AAB_NGA_10    2 1 1 0 1 1 1 1 1 0 1 0 1 0 0 1 1
BBA_NWN_64    2 1 1 0 1 1 1 0 1 0 1 1 1 0 0 1 1
AAB_NGA_82    2 1 1 0 1 1 1 2 1 0 1 0 1 0 0 1 0
ABB_HMO_95    ? ? ? ? ? ? 1 0 1 0 1 0 1 0 0 1 1
BBA_NWS_13    ? ? ? ? ? ? ? 0 1 0 1 0 1 0 0 1 1
AAB_KPL_31    ? ? ? ? ? ? ? ? 1 0 1 0 1 0 0 1 1
BBA_NWM_96    2 1 1 1 1 1 1 1 1 0 1 0 1 0 0 1 1
BBA_NWN_23    1 1 1 1 0 1 1 0 0 1 0 0 0 0 0 1 1
AAB_NGA_17    1 0 0 0 0 0 2 0 0 2 0 0 0 0 0 1 1
AAB_NGA_14    2 1 1 0 1 1 1 1 0 2 0 0 0 1 1 0 1
AAB_NGA_02    1 0 0 0 0 0 2 0 0 2 0 0 0 1 1 0 0
ABB_HMI_89    ? ? ? ? ? ? ? ? 0 2 0 0 0 1 1 0 0
BB__CMR_05    ? ? ? ? ? ? ? ? 0 1 0 0 0 1 1 0 0
BBA_NWN_18    1 1 0 1 0 0 0 0 0 1 0 0 0 1 0 0 0
BBA_HIN_99    ? ? ? ? ? ? ? ? 0 1 0 0 0 1 0 0 ?
AAB_TBK_09    ? ? ? ? ? ? ? ? 0 1 0 0 0 1 0 0 ?
BBA_NWN_22    1 1 1 1 0 1 1 0 0 1 0 0 0 1 0 0 0
BBA_NWN_46    1 1 1 1 0 1 1 0 0 1 0 0 0 1 0 0 0
BBA_NWN_66    1 1 1 1 0 1 1 0 0 1 0 0 0 1 0 0 0
BB__PKW_03    1 1 1 1 0 1 1 0 0 1 0 0 0 1 0 0 0
BB__NAN_06    ? ? ? ? ? ? ? ? 0 2 0 0 0 ? ? ? ?
BB__NAN_11    ? ? ? ? ? ? ? ? 0 2 0 0 0 ? ? ? ?
```

Fig. 4.5 Indel matrix of *GBSS* sequences in *Musa*. The positions of the 17 indels with reference to the aligned data matrix are as follows: 40-56, 57-60, 61-64, 65, 69-71, 72, 73, 96, 309-313, 475-483, 498-516, 517-518, 519-534, 788-790, 791-794, 833-836, 1033-1047. Unordered multistate codes were used in the analysis (see text); 0 = indel, 1 and 2 = nucleotide substitution, ? = missing data.
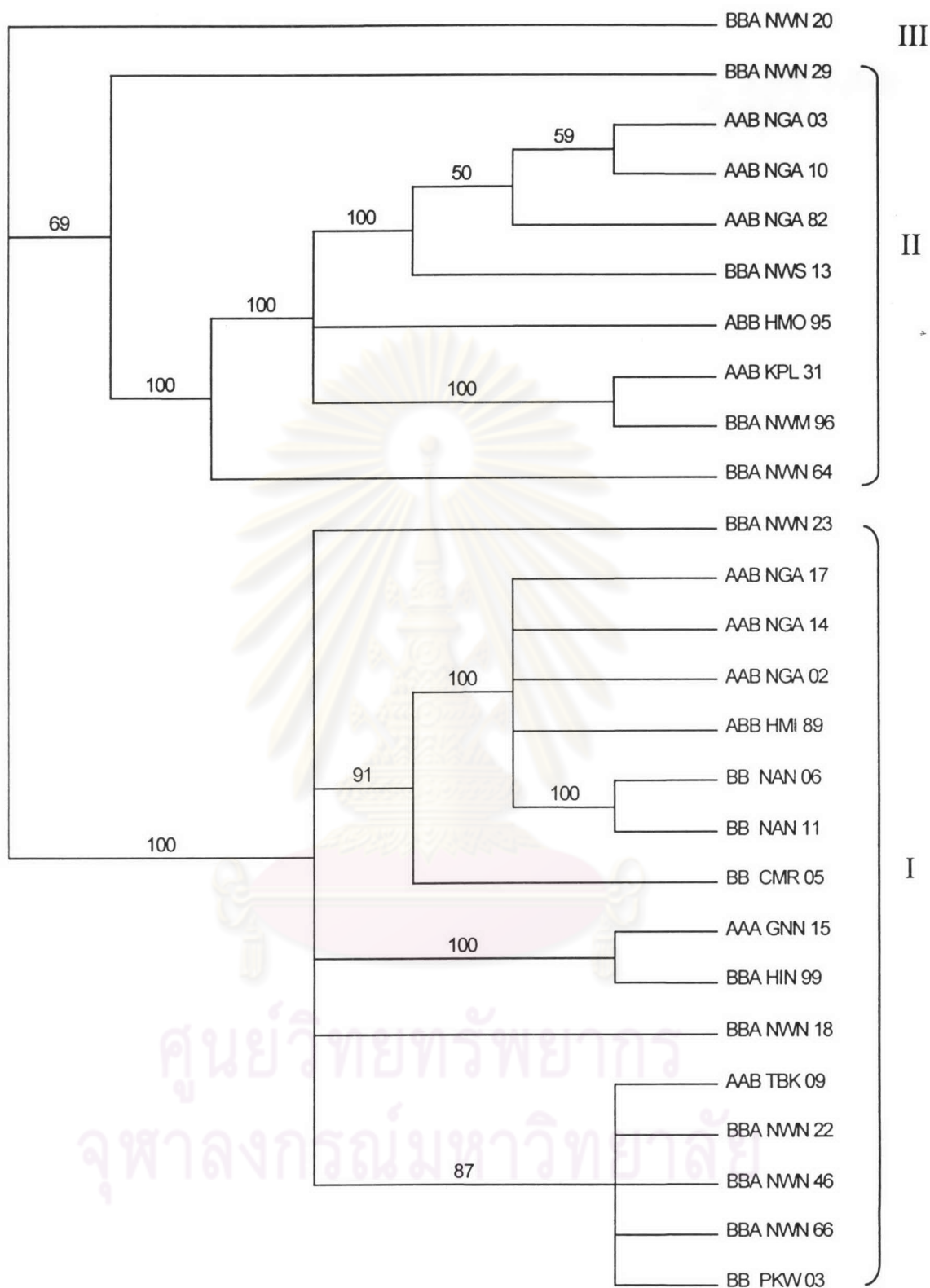
Fig. 4.6 A 50% majority-rule consensus tree of 25,690 most parsimonious trees from the analysis of 1,118 bp of *GBSS* in *Musa* with all gaps excluded and recoded. Number above branch represents bootstrap values. Name of taxa appear according to Table 4.1, number behind name indicated clone number.

4.5 Discussion

4.5.1 Analysis of *GBSS* Structural Gene in *Musa*

Length of *GBSS* gene from 3' end of exon 3 to 5' end of exon 9 in *Musa* is relatively shorter than rice *waxy* gene (*Oryza sativa* cv. group *japonica*-- AF031162, Frances et al. 1998) by ~40 nucleotides due to shorter introns. In *Musa*, all investigated introns were shorter than those in rice by 2 to 30 bp except intron 5, which is longer by ~42 bp. This fifth intron provided two large sites of length variation in *Musa*, at 475-483 and 498-534. Dinucleotide sequences found at the 5'- and 3'-ends of all six introns followed the universal GT-AG rule (Breathnach and Chambon 1981). High homology of banana *GBSS* sequences was indicated by low value of nucleotide diversity (2.9%).

Of the *GBSS* gene in *Musa*, the GC contents in the coding regions (48.3%) and in the non-coding positions (40.9%) were lower than those found in wheat (70% and 47.8%) and other grass (Murai et al. 1999). It has been found that nuclear genes of monocot plants could be classified into two types: GC-biased genes with higher GC content at the third codon position and GC-nonbiased genes with well balanced GC occupation rates (Campbell and Gowri 1990). Unlike in grass (ranged from 80.0-97.4%, Murai et al. 1999), but comparable to GC content at the third codon positions in dicot (41.6% in potatoes, van der Leij et al. 1991), the banana *GBSS* exhibited nonbiased towards GC: the GC distribution patterns at third-codon positions were 47.9%. The difference GC content could be resulted from different gene characteristics (Campbell and Gowri 1990).

4.5.2 Phylogenetic treatment of gaps

The indels in our *GBSS* of the bananas were particularly useful in resolving terminal clades in the topology and strengthening support for certain branches. Inclusion of 17 gaps in *GBSS* sequence data added the parsimony-informative characters to the gene, exon, and intron matrixes, therefore resulted in the higher resolution of phylogenetic analysis (Fig. 4.3-4.6). In clade II, the analyses by nucleotide substitutions failed to distinguish its low diversity (Fig. 4.3-4.5). However, inclusion of gaps as unordered multistate characters in the analysis resulted in splitting of this clade into one subclade and one lineage (Fig. 4.6) with a strong bootstrap value support (100%). Gaps provided some phylogenetic information about relationships that cannot be resolved without them and provide important evidence of relationship e.g. 9-bp indel at 475-483 and 36-bp indel at 498-534 (three gaps were coded), separated clade I from clade II and III. These finding agreed with previous reports that length differences among orthologous sequences provide enough information for phylogenetic reconstruction even at low taxonomic levels (Domon et al 2002; Ingram and Doyle 2003). Not only for inferring evolutionary history, the variables could also ease the identifying of the alleles using specific primers and gel electrophoresis and eliminating the need for sequencing many clones to obtain all variables. From the alignment or from indel matrix (Fig. 4.2) of banana *GBSS* sequence, the nucleotide length differences of the *GBSS* alleles can be determined. The allele II was the longest and allele I was the shortest.

4.5.3 Phylogenetic Utility of the *GBSS* Gene in *Musa*

The *GBSS* gene has demonstrated phylogenetic utility in allotetraploids, tef (Poaceae) (Ingram and Doyle 2003). In *Musa*, an analysis of all variable sites excluding gaps and recoded (Fig. 4.6) yield three lineages which correspondent to the genome constitution of triploids. The first allele was found in all *M. balbisiana* accessions including diploid 'NAN', 'PKW', and 'CMR' and may be regarded as 'B' allele. This allele was found in 'NWN' and 'NGA' in clade I, though the two contained different genome constitutions i.e. BBA and AAB. One exceptional taxa in this clade I was AAA 'GNN'. Possible hypothesis concerning the B origin of this clone is that it introgressed one B allele into its genome while possessing higher influence A-genome characteristics. The second allele in clade II, however, cannot be determined whether it derived from *M. acuminata* because none of the *GBSS* sequence from *M. acuminata* was obtained for the analysis due to PCR and cloning difficulties. As well, the third allele was found in only one clone from 'NWN' and was unclear if it was B- or A-like.

The results are relatively concordant with the cpDNA phylogeny regarding the different of B lineage and the others. In earlier analyses (Swangpol et al. in preparation), we found that cpDNA of B genome was definitely differ to A genome. Also, the placement of 'CMR' with 'NAN', and apart from 'PKW' was in agreement with those cp-sequence data and confirmed our postulation that this 'CMR' came from Thailand. The presence of three homeologous alleles in *Musa* denoted that the hybrids came from multiple origins which is in agreement with previous investigation using RFLP (Gawel and Jarret 1991a; Bhat et al. 1994) and cpDNA sequences (Swangpol et al, in preparation). The placement of

AAB 'NGA' clones higher in the topology of clade I and II could indicate more recent hybridization compared to other 'Namwa' clones.

Low resolution and support values of phylogenetic reconstruction of banana *GBSS* sequences were due particularly to missing characters and under sampling. Two sequences were ~200-bp shorter and the other two were ~500 bp shorter than the others. More accessions, including *M. acuminata* diploids, triploids, *M. balbisiana* triploids, and interspecific hybrid cultivars, should be obtained and investigated in order to clarify the A genome characteristics and to confirmed the presence of orthologous alleles in bananas.

4.5 CONCLUSION

The *GBSS* gene potentially provides good resolution and helps clarify progenitors and relationships among hybrid cultivars in bananas. The phylogenetic analyses of the gene revealed that there are three *GBSS* alleles in the triploid hybrids. This information could be useful to identify ploidy in the *Musa* cultivars, in corporation with the conventional chromosome counting, flow cytometry, and other techniques (Vandenhout et al. 1995; Jenny et al. 1997; Tenkouano et al. 1998; Lysák et al. 1999). Results from our research indicated that at least one allele could be designated as the B allele. The BBA hybrids, 'Namwa', possessed all three alleles, while 'AAB' hybrid, 'Nga Chang', contained two. However, this investigation is premature to discriminate ancestral *GBSS* alleles found in the hybrid cultivars.