Electronic Journal of Biotechnology 17 (2014) 137–147

Contents lists available at ScienceDirect

Electronic Journal of Biotechnology

Single nucleotide polymorphisms in partial sequences of the gene encoding the large sub-units of ADP-glucose pyrophosphorylase within a representative collection of 10 *Musa* genotypes

Muhilan Mahendhiran [a], Jorge Humberto Ramirez-Prado [b], Rosa Maria Escobedo-Gracia Medrano [b], Blondy Canto-Canché [a], Miguel Tzec-Simá [a], Rosa Grijalva-Arango [a], Andrew James-Kay [a,*]

[a] Unidad de Biotecnología, Centro de Investigación Científica de Yucatán, A.C., Calle 43, # 130, Col. Chuburná de Hidalgo, CP. 97200 Mérida, Yucatán, México
[b] Unidad de Bioquímica y Biología Molecular de Plantas, Centro de Investigación Científica de Yucatán, A.C., Calle 43, # 130, Col. Chuburná de Hidalgo, CP. 97200 Mérida, Yucatán, México

ARTICLE INFO

Article history:
Received 21 July 2013
Accepted 13 March 2014
Available online 6 May 2014

Keywords:
Haplotypes
Starch metabolism enzyme
Variants

ABSTRACT

Background: ADP-glucose pyrophosphorylase (AGPase) is a rate-limiting enzyme catalyzing the first step in the starch biosynthesis pathway in higher plants. To date, there are no reported variants or isoforms of the AGPase enzyme in bananas (*Musa* spp. family Musaceae) as is the case of other plants. In this study, genomic DNA sequences homologous to the gene encoding one of the large subunits of the enzyme were amplified from 10 accessions of the genus *Musa*, including representatives of wild ancestors (AA and BB genomes), dessert bananas (AA, AAA, AB and AAB genomes), plantains (AAB genome) and cooking bananas (ABB and AAA genomes), and studied in order to find single nucleotide polymorphisms (SNP) base variations in *Musa* accessions.
Results: In the 810-base pair amplicons of the AGPase large sub-unit (LSU) gene analyzed in ten *Musa* accessions, a total of 36 SNPs and insertions/deletions (indels) were found. The phylogenetic analysis revealed fifteen distinct haplotypes, which were grouped into four variants. Deep examination of SNPs in the 2nd exon in the LSU of AGPase showed that at seven locations, five SNPs altered their amino acid sequence.
Conclusions: This work reveals the possible number of AGPase enzyme isoforms and their molecular levels in banana. Molecular markers could be designed from SNPs present in these banana accessions. This information could be useful for the development of SNP-based molecular markers for *Musa* germplasm, and alteration of the allosteric properties of AGPase to increase the starch content and manipulate the starch quality of banana fruits.

1. Introduction

Musa spp. and cultivars are the largest herbs in the higher plant kingdom. The edible cultivars are parthenocarpic with diploid, triploid or tetraploid hybrids largely derived from crosses between different subspecies within the diploid wild species of *Musa acuminata* Colla and *Musa balbisiana* Colla. These two species represent the AA and BB genomes respectively, within the section Eumusa and originated within Southeast Asia [1,2,3,4]. *M. acuminata* and *M. balbisiana* have a highly conserved genome structure, which shows that these two genomes diverged relatively recently, ~4.6 Mya [5]. *Musa* cultivars including bananas and plantains are the fourth most important crop in terms of value after rice, wheat and maize, with an annual production of more than 102 million metric tons from a harvested area of approximately 10 million hectares [6,7,8,9]. They are cultivated in more than 120 countries throughout the tropical and subtropical regions of the world, are a staple year-round calorie source for more than 500 million people in the tropics, and are an essential source of income and employment [10,11,12].

In some African countries, the average consumption of bananas is between 200 and 250 kg per capita annually, and they are considered a major staple food [13]. Plantains (AAB), cooking bananas (ABB, AAA), East African cooking bananas (AAA) and unripe dessert bananas (AA, AAB, AAA) are cooked before consumption, and have a high resistant starch content [14,15]. However, mature dessert bananas are eaten raw, because they are sweet at ripeness and less

* Corresponding author.
E-mail address: andyj007@cicy.mx (A. James-Kay).
Peer review under responsibility of Pontificia Universidad Católica de Valparaíso.

Production and hosting by Elsevier

http://dx.doi.org/10.1016/j.ejbt.2014.04.004

starchy than plantains and cooking bananas [16]. These properties reflect differences in the physico-chemical properties of starch (amylose/amylopectin ratio), sugars and reducing sugar in ripe and unripe fruits [17,18].

Cull bananas (below marketable size) which are discarded by large banana plantations, constitute ~25% of the total banana crop and can be used as a source of industrial starch [19,20]. Starch is the most important storage carbohydrate found in all plant species and is used for food, animal feed and industrial purposes [21]. After cellulose, starch is the second most abundant biopolymer on earth [22]. The annual production of starch harvested from all crops is approximately 2,500 million tons [23].

Starch is the main carbohydrate in plant storage organs [24]. In higher plants (except cereal crops), starch is synthesized within plastids, and is composed of two D-glucose (Glc) molecules, comprising the homopolymers amylose and amylopectin [25]. In most plant species, amylose is a linear glucan that constitutes up to 20 to 30% of normal starch, whereas amylopectin is highly branched, representing ~70% of the total starch [22]. The functions of starch depend on the plant tissue in which it is formed. Transient starch synthesized in leaves during the day is generally rapidly turned over, thereby playing a key role in meeting the energy requirements of plant development. Conversely storage starch, which accumulates in starchy seeds (such as wild banana seeds), roots and tubers, provides long-term carbon storage for the next generation [24].

Thirty candidate genes encoding the starch biosynthetic enzymes have been identified in the maize genome, and 27 have been identified in the rice genome [18,26]. Adenosine diphosphate glucose pyrophosphorylase (AGPase, EC 2.7.7.27) is a key metabolic enzyme for starch biosynthesis. It catalyzes the first committed step, which is a rate-limiting reaction in the synthesis of starch in plants and glycogen in bacteria [27]. The AGPase enzyme catalyzes the synthesis of the nucleotide carbohydrate ADP glucose and the release of inorganic pyrophosphate (PPi) from glucose-1-phosphate (Glc-1-P) and ATP [28,29,30]. ADP glucose acts as a glucosyl donor via starch synthases (SSs; EC 2.4.1.21) and starch branching enzymes (SBEs; EC 2.4.1.18) for starch biosynthesis [31]. Plants contain multiple isoforms of this enzyme, most of which are encoded by multiple genes in cereals [18]. AGPase is sensitive to allosteric regulation by, the activator, 3-phosphoglyceric acid (3-PGA) and the inhibitor, inorganic phosphate (Pi); these regulators are the major photosynthetic metabolites in algae and plants [32,33]. AGPase is a heterotetrameric enzyme composed of two large sub-units (51 kD) and two slightly smaller sub-units (50 kD) encoded by paralogous genes that evolved by several rounds of duplication, whereas most of the bacterial AGPases involved in glycogen synthesis are composed of four identical subunits [22,34,35]. In general, the large subunits (LSU) are involved in allosteric regulation, and the small subunits (SSU) are involved in catalytic reactions [30,36,37]. However, in rice, potato and maize, both LSU and SSU are involved in catalytic reactions [26,36,38]. The AGPases participate in the metabolism of transient starch in chloroplasts and chromoplasts and of storage starch in amyloplasts. Based on sequence comparisons, the large subunits of AGPase are divided into four groups: stem/tuber, leaf, fruit/root, and endosperm AGPases. AGPase is located only in the plastids in dicots, but it is present in the plastids and cytosol in cereal monocots [39]. The cytosolic AGPase is present in the endosperm, and the plastidial AGPase is present in the endosperm and leaf tissues [40]. Genes encoding AGPase in angiosperms form a multigenic family and are considered a relevant model to study the evolution of paralogs [6].

Because the genomic constitutions of cultivated *Musa* spp. show different levels of ploidy, understanding the influence of ploidy on the evolution of paralogous genes involved in starch metabolism may lead to a molecular understanding of the biochemistry underlying the differences in starch metabolism among dessert bananas, plantains and cooking bananas. The aims of this study were to evaluate the AGPase large subunit gene to analyze the phylogenetic relationships in different *Musa* genomes, compare the evolutionary relationships between paralogous AGPase genes in *Musa* with those in both monocots and dicots, and identify the candidate polymorphisms in the AGPase gene that may be associated with known phenotypic traits.

2. Materials and methods

2.1. Sample collection

Plant material of ten accessions representative of bananas and plantain, including wild species belonging to the edible Eumusa and Australimusa sections (Table 1), was obtained from the International Transit Centre (ITC, http://www.crop-diversity.org/banana/#AvailableITCAccessions) hosted by the Katholieke Universiteit, Leuven, Belgium. These are maintained as field plants at the *Musa* germplasm collection of CICY located in the Experimental Station of the Instituto Nacional de Investigaciones Forestales Agrícolas y Pecuarias (INIFAP) research station at Uxmal (Lat. 20° 24′ 27.72″ N, Long. 89° 45′ 06.66″ W) in Yucatán, México.

Samples of "cigar leaves" from each accession were collected, dipped in a solution containing three parts citric acid (75 mg/100 ml H_2O) and one part ascorbic acid (25 mg/100 ml H_2O) and transported in a cooler from the field to the laboratory. The tissues were disinfected with sodium hypochlorite 6% (v/v) and ethanol 70% (v/v) for 1 min each, followed by three rinses with distilled water; the excess water was removed with a sterile-paper towel. Subsequently, 100 mg of leaf sample was weighed, wrapped in aluminum foil, frozen with liquid nitrogen and stored at -80°C until use.

2.2. Primer design

A pair of specific primers was designed for the gene encoding the large sub-unit of AGPase. Protein sequence information was obtained from the *M. acuminata*, dwarf banana accession (NCBI-GenBank **ABF69950.1**) corresponding to the cDNA. The homologous sequences of 18 closely related plant accessions (including *Musa*) from monocotyledonous plants and dicotyledonous plant species were retrieved by BLAST searches (blastx) from the NCBI database. The consensus sequences were chosen to design specific primers targeting the corresponding nucleotide sequences of the *Musa* AGPase gene. The sequences obtained for primer design are listed below:

AAB40724.1 (*Solanum lycopersicum*), **AAB91464.1** (*Cucumis melo*), **AAB91468.1** (*Citrullus lanatus* subsp. *vulgaris*), **AAB94012.1** (*Sorghum bicolor*), **AAC49941.1** (*S. lycopersicum*), **AAK27718.1** (*Cicer arietinum*), **AAK27719.1** (*C. arietinum*), **AAU10700.1** (*Oryza sativa* Japonica Group), **ABC26921.1** (*S. lycopersicum*), **ABE77800.1** (*Medicago truncatula*), **ABE85951.1** (*M. truncatula*), **CAA65541.1** (*Pisum sativum*), **CAB52196.1** (*Ipomoea batatas*), **NP_174089.1** (*Arabidopsis thaliana*), **NP_197423.1** (*A. thaliana*), **P12299.2** (*Triticum aestivum*), and **P55241.1** (*Zea mays*).

Pairwise alignments were used to select the highly conserved regions, which were used to create a multiple sequence alignment (MAS) FASTA file. Subsequently, the MAS file was uploaded into Bio-edit v.12.5. Specific primers were designed for highly conserved regions, and their properties were analyzed using Oligo Analyzer 1.0.2. The sequences of the forward and reverse primers are 5′-CCTCTCACTA GTACCAGAGCC-3′ and 5′-CAATTCACTTGGGTATTTGAG-3′, respectively. The expected size of the amplicon for this primer set is 810 bp.

2.3. DNA extraction, PCR amplification and electrophoresis

Total genomic DNA was extracted from 100 mg of frozen samples according to the protocol of Dellaporta et al. [41] with minor modifications. The DNA quality and concentration were determined

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

No	Section	Species/hybrid	Subspecies/subgroup	Genome	Accession name	Abbreviation	ITC[a]
1	Eumusa	*M. acuminata*	*banksii*	AA w[b]	banksii	Ban	0623
2		*M. acuminata*	*malaccensis*	AA w[b]	Pisang Kra	P.kra	1345
3		*M. acuminata*	–	AA cv[b]	Pisang lilin	P.li	1121
4		Dessert banana	Neypoovan	AB cv[b]	Safet velchi	S.ve	0245
5		*M. balbisiana*	–	BB w[b]	BB-ClCY	BB-Cicy	[c]
6		Dessert banana	Gros Michel	AAA cv[b]	Gros Michel	GMI	0484
7		East African cooking	Mutika/Lujugria	AAA cv[b]	Mbwazirume	Mbw	0084
8		Cooking banana	–	ABB cv[b]	Bluggoe	BLU	0767
9		Pacific Plantain	Maoli/Popolou	AAB cv[b]	Popolou	Pop	0335
10	Australimusa	*M. beccari*	–	w[b]	beccari	BEC	1070

[a] All the plant materials with an ITC were provided by the Biodiversity International Centre for *Musa* located at the Laboratory for Tropical Plant Improvement in the Katholieke Universiteit Leuven, Belgium.
[b] *M. acuminata* AA w wild type, AA cv cultivar.
[c] Collected from Teapa, Tabasco, México.

using a spectrophotometer according to the method described by Stulnig and Amberger [42] using the tools of Khirshyat 1.0 [43].

Samples were normalized to a concentration of approximately 10 ng/μl prior to PCR amplification. The PCR mixture (25 μl) contained 500 μM each dNTP (Invitrogen, USA), 0.5 μM each specific primer, ~10 ng of DNA, 1 × PCR buffer (Invitrogen, USA), 1.5 mM $MgCl_2$ (Invitrogen, Brazil), and 5 U *Taq*-DNA polymerase (Invitrogen, USA). The PCR reaction was performed in an Eppendorf Master Cycler (Life Sciences Biotech, USA) with initial denaturation at 94°C for 5 min, followed by 35 cycles of 94°C for 40 s, with an annealing temperature of 55°C for 50 s and, 72°C for 1 min, and a final extension at 72°C for 8 min. The amplified PCR products were electrophoresed on 1% w/v ultrapure agarose gels (Invitrogen, Spain) with 1× Tris Acetate EDTA (TAE) buffer; 0.1 μg/μl of 1 kb plus DNA ladder (Invitrogen, USA) was included as a size-standard marker. After electrophoresis, the gel was stained with ethidium bromide. The PCR fragments were excised using a UV transilluminator 2000 (BioRad, USA), and the DNA was extracted and purified using the Gene Clean II Kit (MP Biomedicals, LLC, France).

2.4. Cloning of DNA fragments and sequencing

The purified ~810 bp amplicons were cloned using the TOPO-TA Cloning Kit (Invitrogen, USA) according to the manufacturer's instructions. The plasmid DNAs were extracted and purified using a Qiagen Spin Miniprep kit (Qiagen, Germany) according to the manufacturer's protocol. The presence of inserts was confirmed by digestion of the plasmid DNA with the *Eco* R1 restriction enzyme (Invitrogen, USA), and the digestion products were analyzed on a 1.2% w/v agarose/EtBr gel with 1× TAE buffer; a 1 kb DNA ladder (Invitrogen, USA) was used to estimate the insert size. For each *Musa* accession, 10 clones were sequenced on a 3730XL DNA Analyzer at the University of Arizona Genetics Core Sequencing Facility Services, USA.

2.5. Construction of the phylogenetic tree

The sequences were edited to remove vector sequences at both ends, and low quality end regions (Q score <20). The forward and reverse primers were identified. The sequences were automatically aligned using Sequencher 5.1 (Gene Codes). A 1× depth coverage was employed for all isolates. The aligned chromatograms were manually curated to check whether the observed single nucleotide polymorphisms (SNPs) were miscalled bases (accepted if Q score >30). Subsequently, the Sequencher 5.1 alignment file was exported in the FASTA format. In addition to the *Musa* AGPase LSU gene sequences, 13 accessions of related species were retrieved from GenBank. The chosen sequences were aligned using ClustalW and manually adjusted. A distance matrix was computed, according to the Kimura 2 Parameter model, chosen as the best DNA model. An unrooted phylogenetic tree was constructed using the MEGA 5.05 program by Tamura et al. [44] based on the nucleotide sequence alignments under the Neighbor-Joining method. The ancestor reliability was tested by 1000 bootstrap replications of the data. The Dot Matrix analysis was performed using the two-sequence alignment option at the NCBI BLAST website. The analyses of haplotype diversity in the DNA sequence polymorphism data were generated from a Nexus formatted file in DnaSP 5.10.

The sequences grouped into 15 haplotypes divided among the four AGPase variants were deposited at NCBI-GenBank under the following accession numbers:

AJ_AGP1_varian1.sqn: KF418633–KF418649,
AJ_AGP2_variant2.sqn: KF418703–KF418721,
AJ_AGP3_variant3.sqn: KF418650–KF418695,
AJ_AGP4_variant4.sqn: KF418696–KF418702.

3. Results

3.1. Cloning and analysis of the partial sequences of the AGPase LSU gene in Musa

To investigate the variation within the gene encoding the LSU of AGPase in the genus *Musa*, 10 accessions representing nine members of the Eumusa sections and one Australimusa section were studied (Table 1). The designed primers (Fig. 1) amplified a fragment of 810 bp of the AGPase large subunit, which extends from exons 1 to 3 of *M. acuminata* (NCBI-GenBank Reference clone **AC186746.1**). This BAC clone contains the 3893 bp AGPase gene that includes 14 exons and 13 introns. 87 sequences from the amplicons of the AGPase LSU gene were aligned and assembled into 4 different "variants," 1 to 4. The most represented one was variant 3, which contained 46 clones. Variant 1 had 15 clones, variant 2 included 19 clones, and the least abundant one was variant 4, which had only 7 clones (Fig. 2 and Table 2). Variants 1, 2 and 3 covered 3 exons and 2 introns; however, variant 4 covered only 2 exons and 2 introns, because of the sequencing efficiency (Fig. 2). The third and longest variant (810 bp) had 90% sequence identity in exons 1, 2 and 3 to the reference clone (Fig. 2), and high identity in both introns. For each variant, the regions with highest identity to the reference genome were as follows: for variant 1, the 1st exon from 1 to 55 bp, the 2nd exon from 435 to 600 bp and the 3rd exon from 641 to 775 bp, with 90, 70 and 90% identity respectively; for variant 2, the 1st exon from 1 to 75 bp, the 2nd exon from 410 to 570 bp, and the 3rd exon from 699 to 775 bp, with 90, 80 and 80% identity, respectively; and for variant 3, the 1st exon from 1 to 65 bp, the 2nd exon from 379 to 552 bp, and the 3rd exon from 710 to 775 bp, with 90% sequence identity in exons 1 to 3; for variant 4, we were unable to sequence the 1st exon; the 2nd exon covered 503 to 661 bp, and the 3rd exon covered 711 to 775 bp, with 80% identity for both exons (Fig. 2). The Dot Matrix analysis (Fig. 3) shows the

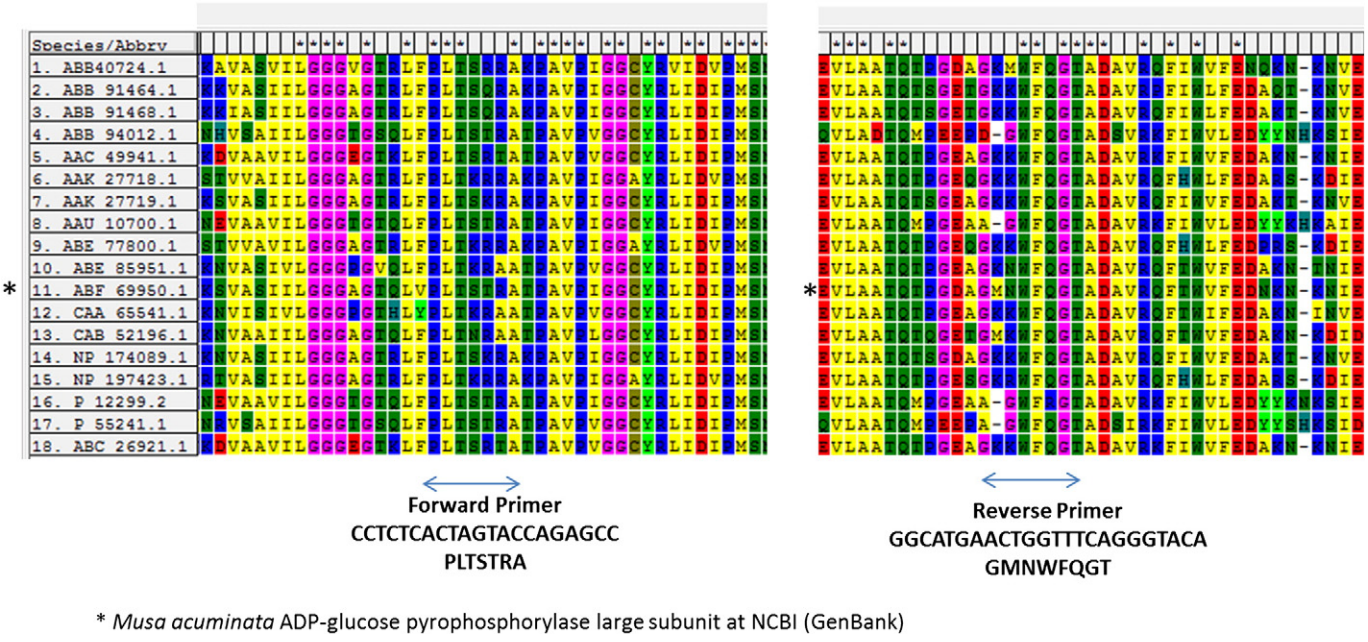

Fig. 1. Primer design for the AGPase large sub-unit (LSU) gene fragment. Alignment of AGPase LSU protein sequences obtained from GenBank. Eighteen plant species were used to design the primers. The locations of the forward and reverse primers on the sequences homologous to AGPase protein sequences are indicated. The consensus amino acid sequences based on the alignment were considered for the primer design to target the corresponding nucleotide region of the *Musa* AGPase gene. The protein sequences included in the alignment are **AAB40724.1** (*Solanum lycopersicum*), **AAB91464.1** (*Cucumis melo*), **AAB91468.1** (*Citrullus lanatus* subsp. vulgaris), **AAB94012.1** (*Sorghum bicolor*), **AAC49941.1** (*Solanum lycopersicum*), **AAK27718.1** (*Cicer arietinum*), **AAK27719.1** (*Cicer arietinum*), **AAU10700.1** (*Oryza sativa* Japonica Group), **ABC26921.1** (*Solanum lycopersicum*), **ABE77800.1**, **ABE85951.1** (*Medicago truncatula*), **ABF69950.1** (*Musa acuminata*), **CAA65541.1** (*Pisum sativum*), **CAB52196.1** (*Ipomoea batatas*), **NP_174089.1** (*Arabidopsis thaliana*), **NP_197423.1** (*Arabidopsis thaliana*), **P12299.2** (*Triticum aestivum*), and **P55241.1** (*Zea mays*).

sequence similarities in the alignments of DNA sequences, and the distribution of indels and possible frameshifts in the four variants relative to the *Musa* clone reference sequence. In Fig. 3, the X-axis represents the sequence of the reference *Musa* AGPase clone AC-186746.1, and the Y-axis represents the sequenced fragment for each of the *Musa* variants 1, 2, 3 and 4. The transversal lines in graphs a, b, c and d, denote regions of sequence identity to the reference *Musa* clone; gaps in the lines denote the absence of identity because of mismatches or indels. The deviations above and below the central transversal line indicate frameshifts because of indels. The graphs in a, b and d illustrate the various degrees of identity in the nucleotide sequences represented by differences in the lengths of the gaps in the transversal lines (especially on gaps spanning introns), whereas in graph c (variant 3), there are no such gaps because it has sequence

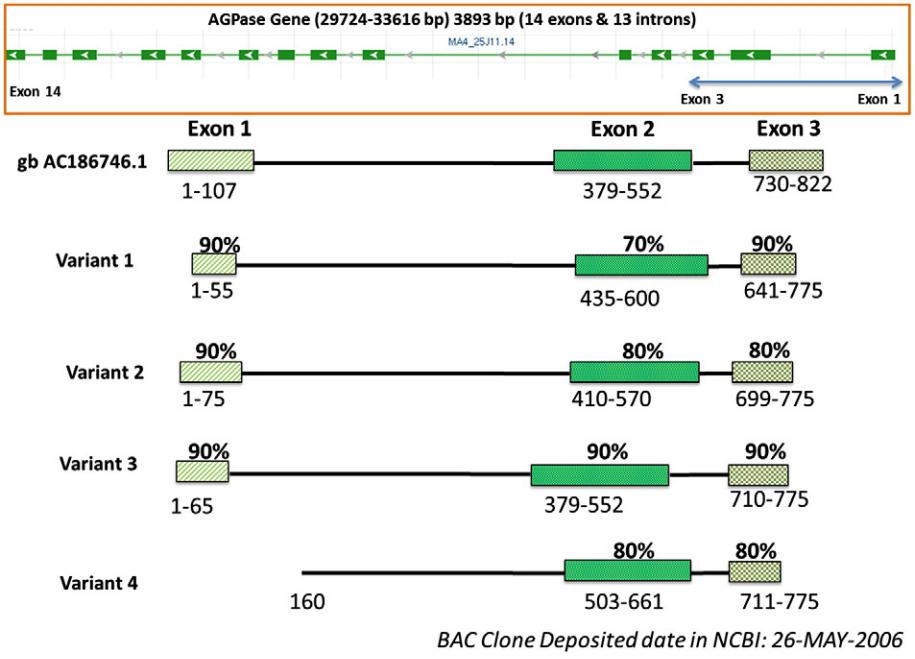

Fig. 2. Four AGPase gene fragment variants in *Musa*. Comparison of the *Musa* AGPase LSU gene variants with the reference clone **AC186746.1**. The PCR product amplified from the genomes of ten *Musa* accessions corresponded to position: -32,795–33,617 bp on the reference sequence, covering exons 1–3 of the AGPase LSU gene, except for variant 4, which lacks exon 1, and ends in exon 3.

Table 2
AGPase LSU gene variant distribution in *Musa*. Distribution of the AGPase large sub-unit (LSU) gene variants in 10 *Musa* accessions. Eighty-seven clones were classified into 4 variants.

No	Accession name	*ITC	Genome	Variant 1 (no. of clones)	Variant 2 (no. of clones)	Variant 3 (no. of clones)	Variant 4 (no. of clones)	No of clones
1	*M. a.* banksii	0623	AA	3	2	3	–	8
2	*M. a. malaccensis*	1345	AA	5	2	3	–	10
3	*M. a.* cv. Pisang lilin	1121	AA	–	6	3	–	9
4	Safet velchi	0245	AB	–	2	7	–	9
5	*Musa balbisiana*	–	BB	3	–	3	1	7
6	Gros Michel	0484	AAA	–	3	4	1	8
7	Mbwzruimbe	0084	AAA	–	–	9	–	9
8	Bluggoe	0767	ABB	2	1	4	2	9
9	Popolou	0335	AAB	–	3	6	–	9
10	*Musa beccari*	1070	–	2	–	4	3	9
	Total clones	–	–	15	19	46	7	87

identity spanning the exons and the introns (Fig. 2 and Fig. 3). The connecting regions of continuous dots are used for optimal alignment of the sequences. In Fig. 3, the frameshifts in variants 1, 2 and 4 indicate indels of variable sizes. Variant 1 had a large indel (~350 bp) in intron 1, and variant 4 had a large indel (~220 bp) in intron 2, whereas intron 2 of variant 2 had multiple small indels (~60 bp). Variant 1 had an inverted repeat in intron 1. The sequence data for variant 4 was incomplete at the 5′ end of exon 1, and its 2nd and 3rd exons had different levels of sequence identity with the *Musa* BAC clone, as shown in the graph (Fig. 3).

3.2. Distributions of the four variants identified in the region spanning exons 1 to 3 in the AGPase LSU gene in Musa accessions

The analysis of 87 sequences classifying them into 4 identified variants is summarized in Table 2. As stated above, variants 1, 2, 3 and 4 contained 15, 19, 46 and 7 clones, respectively. The least number of clones (7) was obtained from *M. balbisiana* (BB genome), and the largest number (10) was obtained from *M. acuminata* spp. *malaccensis* (AA); 8 or 9 clones were obtained from the other accessions (Table 2). Variant 1 was present in 5 out of ten accessions that included the two *M. acuminata* wild subspecies *M. acuminata* spp. *banksii* (AA) (3 clones) and, *M. acuminata* spp. *malaccensis* (AA) (5), *M. balbisiana* (BB CICY) (3), Bluggoe (ABB) (2) and, *Musa beccari* (2). Variant 2 was present in all accessions except *M. balbisiana* (BB CICY), Mbwazirume (AAA) and *M. beccari*. Variant 3 was shared by all the ten accessions. All 9 clones of Mbwazirume had only variant 3. Variant 3 shared the highest identity with the reference clone and was present in all the accessions (Fig. 3a and b; Table 2). Variant 4 was found in *M. balbisiana* (1), Gros Michel (1), Bluggoe (2) and *M. beccari* (3). The Popolou AAB genome had 6 clones out of 9 in variant 3, and 3 clones in variant 2.

3.3. Distribution of SNPs in the Musa AGPase LSU gene variants

Variable numbers of SNPs were detected in the 4 variants, as shown in Table 3; the distribution of SNPs was highest in variant 1. The

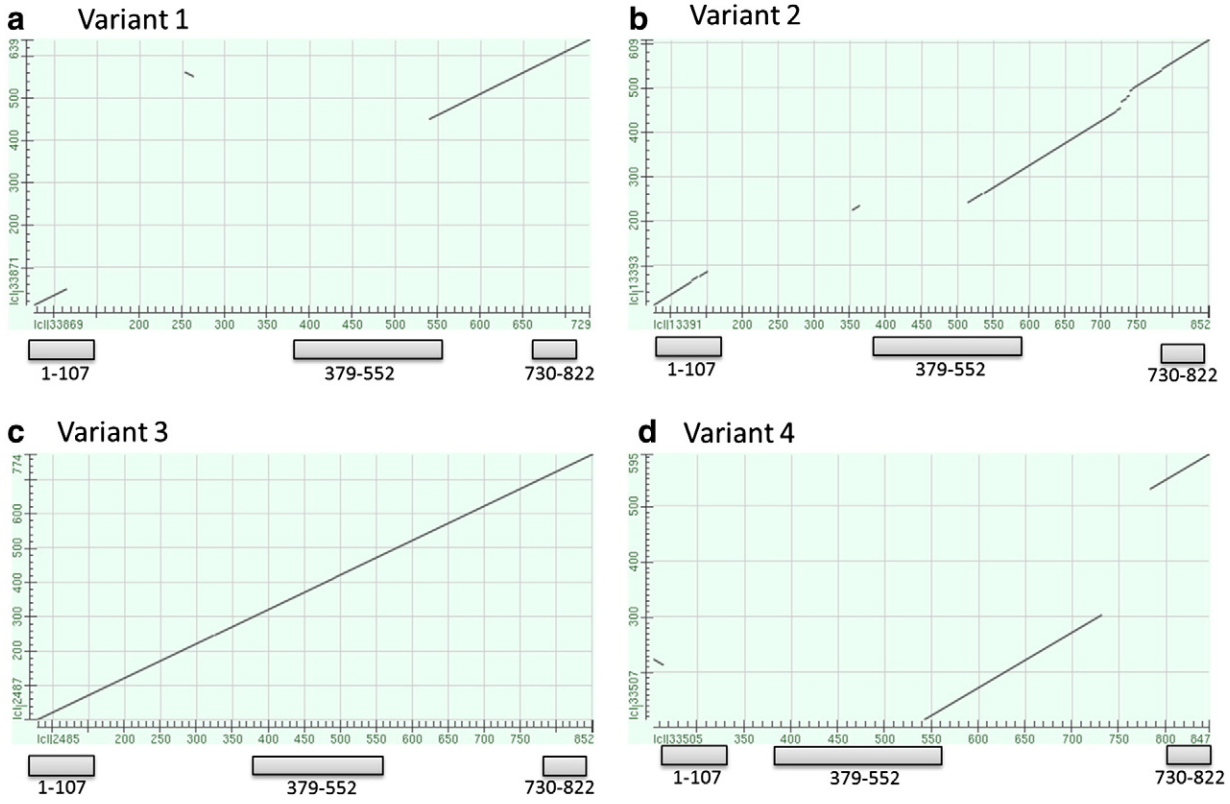

Fig. 3. Dot Matrix analysis of four AGPase gene fragment variants. Dot Matrix comparison of the nucleotide sequences of the *Musa* AGPase LSU gene variants with the *Musa* AGPase reference sequence (BAC clone **AC186746.1**). This matrix shows the distribution of indels in each variant. This analysis was performed using the Dotter program in NCBI database. The transversal line represents the similarity of the variants with the deposited *Musa* clone. Deviations from this represent indels and other changes. The perpendicular lines (in variants 1 and 4) represent inversions.

fragment for variant 4 was incomplete, but the fragments for variants 1, 2 and 3 contained the sequence differences (Table 3). Variant 2 contained two gaps when compared with the other variants (data not shown). In all 4 variants, 36 SNPs were detected, of which 14 and 22 SNPs were located in the exons and introns, respectively (Table 3).

3.4. Haplotype diversity analysis and phylogenetic reconstruction of Musa spp. based on exons 1 to 3 of the AGPase LSU gene

The analysis of 87 sequences from the 10 *Musa* accessions showed 36 SNPs in both exons and introns (Table 2). These were grouped into 4 different clades; each clade was grouped as a variant based on the 2nd exon that contained 175 nucleotides (Fig. 4). All variants showed 70–90% identity with the gene encoding the large subunit of *Musa* AGPase (Fig. 2). To avoid redundancy, the identical sequences were compressed to form 15 haplotypes, based on the nucleotide sequence divergence. For the reference sequence, intron 1 was the largest (~437 bp), and intron 2 was the smallest (~72 bp). Variant 1 contained haplotypes 8 to 10; variant 2 contained haplotypes 3 to 7; variant 3 contained haplotypes 1, 2, 11, 12 and 13, and; Variant 4 contained haplotypes 14 and 15 (Fig. 5 and Table 4). After in silico removal of the introns, all the variants were compared with the *M. acuminata* complete reference sequence (Accession no. **AC186746.1**). In addition, 5 sequences of monocotyledons, 7 of dicotyledons and one Bryophyta (*Physcomitrella patens*) were used to construct a phylogenetic tree; all shared the 2nd exon of the AGPase LSU gene (Fig. 6). The 15 haplotype gene fragment sequences and 13 corresponding orthologous outgroups were used for phylogenetic reconstruction.

The reconstruction of the phylogenetic relationships between the *Musa* taxa was analyzed based on the 2nd exon (175 bp fragment) of the AGPase LSU gene sequences corresponding to 32,898 to 33,071 bp in the *Musa* reference genome using the Neighbor-Joining (NJ) method (Fig. 5). The phylogenetic tree was divided into four clades with the 15 haplotypes with the 4 variants. Clade 1 contained variants 1 and 4 shared with the out-groups from the monocotyledonous plants *O. sativa*, *Z. mays* and *S. bicolor*. The Poales were very close to variants 1 and 4, whereas Clade 2 contained variant 2 and Clade 3 contained variant 3 along with the reference *M. acuminata* sequence (Accession no. **AC186746.1**), which is consistent with the 90% sequence similarity with variant 3, as shown in Fig. 2. Variant 3 had almost complete sequence identity with the reference sequence (Fig. 5). Clade 4 contained outgroups from the monocotyledonous plant *Spirodela polyrhiza*, the Bryophyte *P. patens* and the dicotyledonous plants, such as *Ricinus communis*, *A. thaliana*, *Populus trichocarpa*, *Vitis vinifera*, *Citrus sinensis*, *Ipomea batatas*, and *Glycine max*.

The corresponding AGPase LSU gene fragments for all these plants were obtained from the NCBI database and their accession numbers are shown in Fig. 5. The changes in the coding regions varied within the haplotypes. The aligned nucleotide sequences were translated into protein sequences. To distinguish between synonymous and non-synonymous changes in the coding regions, the amino acid sequences were compared to the *Musa* BAC Clone (**ABF69950.1**) that was used for primer design. The SNPs that caused protein sequence changes are shown in Table 4 and Fig. 6. SNPs were found at seven locations in the 2nd exon of the AGPase LSU gene region in the 15 haplotypes. The positions of the SNPs in the exons of the AGPase structural gene varied within the haplotypes. Seven SNPs produced amino acid substitutions in the AGPase polypeptide. For example, in the 30th position, haplotype 15 had histidine, which has a positively charged side chain, but the other 14 haplotypes had glutamine, which has a polar, uncharged, side chain. In the 31st position, haplotypes 14 and 15 had tyrosine, which has an aromatic side chain; the other haplotypes had phenylalanine, which also has an aromatic side chain. In the 38th position, haplotype 9 had cysteine, which has a polar, uncharged side chain, and the other haplotypes had arginine, which has a positively charged side chain. In the 41st position, variants 1 and 4 had tyrosine, which has an aromatic side chain, but variants 2 and 3 had serine, which has a polar, uncharged side chain. In the 44th position, variants 1 and 4 had phenylalanine, which has an aromatic side chain, and variants 2 and 3 had tyrosine, which also has an aromatic side chain. In the 45th position, variant 2 had isoleucine, which has a nonpolar, aliphatic side chain; but the other 3 variants had asparagine, which has a polar, uncharged side chain. In the 48th position, haplotype 5 had aspartic acid, which has a polar, negatively charged side chain, but all other 14 haplotypes had asparagine, which has a positively charged side chain. Of the 7 SNPs that caused amino acid changes, SNPs (31st and 44th positions) caused substitutions with similar or conserved amino acids, whereas the other 5 SNPs caused changes to different types of amino acids (Table 4).

The bioinformatics analysis using prfectBLAST software [45] revealed that the four variants were located on different chromosomes; therefore, these may be different isoforms (these variants are from a small 810 bp fragment of the AGPase LSU gene). The consensus sequences for each of the four AGPase LSU gene variants was blasted against the *M. acuminata* genome database, and the best hits were obtained. Variants 1 and 4 showed orthologous genes at different loci on chromosome 4. Variants 3 and 9 aligned with regions of chromosomes 9 and 6, respectively. The variant 2 sequence aligned with chromosomes 4 and 6 when it was blasted against the peptide sequences encoded by the *Musa* genome or against the entire genomic sequence, respectively. In addition, two other isoforms present on chromosomes 1 and 6 were deposited in the *M. acuminata* genome databases. However, using our analysis, we were unable to recover these 2 isoforms of the LSU gene in the *M. acuminata* genome. The *M. balbisiana* variety Pisang Klutuk Wulung (PKW) genome was recently sequenced [46]. The obtained LSU gene variant sequences were blasted against the published *M. balbisiana* PKW reference genome AGPase databases, and the following results were obtained: variants 1 and 4 had orthologous genes in chromosome 4, variant 2 was found on chromosome 6, and variant 3 aligned with chromosome 9 (consistent with the hits obtained using the *M. acuminata* genome databases).

Table 3
The occurrence of SNPs in the 4 AGPase LSU gene variants from 10 *Musa* accessions. Arrangement of single nucleotide polymorphisms (SNPs) in the introns and exons of AGPase LSU gene variants within *Musa* relative to the reference clone (**AC186746.1**).

Variant	Exon 1	Intron 1	Exon 2	Intron 2	Exon 3	Total SNPs in exons	Total SNPs in introns	Total SNPs
1	0	2	5	10	0	5	12	17
2	1	1	3	0	0	4	1	5
3	0	9	0	0	0	0	9	9
4	---	0	5	0	0	5	0	5

4. Discussion

AGPase is key enzyme for understanding the functional basis of protein sequence conservation [28]. The LSU of the AGPase enzyme is important for its allosteric properties [34]. This study investigated the polymorphisms in 810 bp sequence fragment of the AGPase LSU gene within wild and cultivated *Musa*. Ten different accessions of diploids and triploids were analyzed; wild species were analyzed based on their predicted contribution as ancestors of cultivars, and cultivars were analyzed based on their known culinary and dessert-food uses. The gene region covered by the amplicon spanned exons 1 to 3 within the allosteric regions of the AGPase LSU of *Musa* BAC clone **AC186746.1** and has sequence homology with the rice, maize and *Arabidopsis*

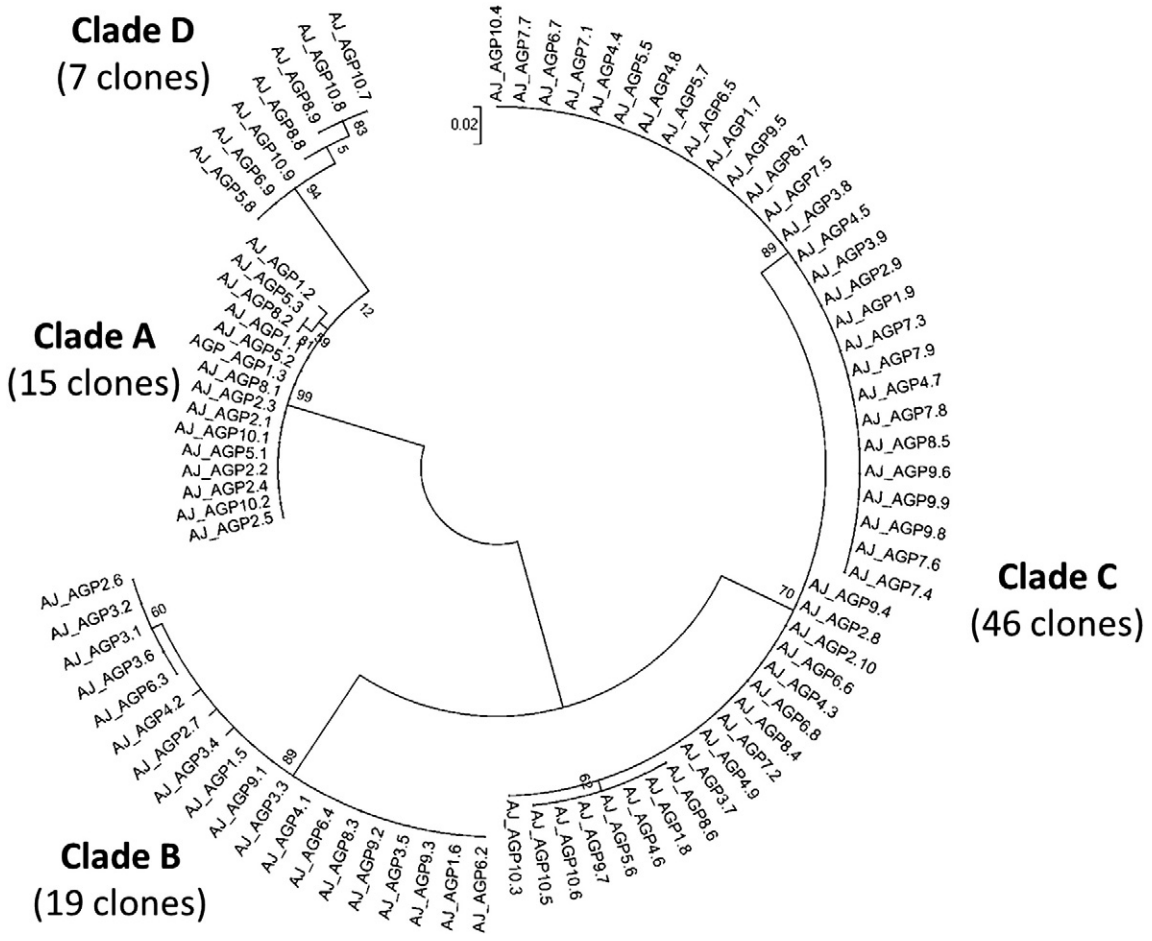

Fig. 4. Dendrogram of the 2nd exon region AGPase LSU gene fragments in 10 *Musa* accessions. One hundred and seventy-four nucleotides of the 2nd exon were used for creating the dendrogram using the Neighbor-Joining method, with the Kimura 2 parameter model. The substitution rate per site is 0.02.

genomes. Four variants were found within the ~810 bp fragment analyzed. Six complete sequences of the genes encoding LSU and two small subunits of the AGPase enzyme have been reported in five different chromosomes of the *Musa* genome database (http://banana-genome.cirad.fr/) [47].

The analysis of the amplicon sequences from 10 *Musa* accessions grouped the 87 sequences into 4 variants, in which the exons differed in their sequence identity relative to the reference *Musa* BAC clone (NCBI-GenBank). The exonic sequence identity varied because of SNP-indels in different proportions in each exon in all the variants. Furthermore, the Dot Matrix analysis confirmed the sequence similarity/dissimilarity of the variants relative to the BAC clone. Indels and direct and inverted repeats of the sequences have been reported in the coding and non-coding regions of the LSU genes of rice [48], maize [49,50] and other cereal plants [51]. The different proportions of indels shown by the variants may cause frame-shifts in the open reading frames. All the variants had single base substitutions in their DNA (nucleotides) sequence. The sequences blasted against the *M. acuminata* genome database indicated that these four AGPase gene variants might be alleles of the chloroplast LSU-1 and 2 genes, which are located in the *Musa* chromosomes 4, 6 and 9. Three pairs of paralogous genes encode the LSU and SSU subunits of AGPase in maize [52]. To our knowledge, this is the first report detailing the AGPase LSU gene variants present in the *Musa* genome.

Our results from the phylogenetic analysis of the AGPase LSU gene indicated the distribution of haplotypic diversity in *Musa* with the three related plant species of monocotyledons. The BAC *Musa* clone shared homology with variant 3 in clade 3; this variant was clearly distinct by its respective SNPs; whereas in clade 1, variants 1 and 4 shared homology with the AGPase LSU in three species of Poales [53] probably because these monocotyledonous plants are very close to *Musa*. Each of the variant consensus sequences was blasted against the *M. acuminata* and *M. balbisiana* genomes and the sequences obtained were clearly present in their respective variant subclade.

In this study, the Mbwazirume AAA genome of the subgroup Mutika/Lujugria contained only variant 3; all the clones analyzed for this accession did not have any SNPs and had 100% identity. This suggests that variant 3 encodes the functional version of the LSU of the AGPase enzyme for this East African cooking banana. During evolution, the *Musa* genotypes were exchanged by humans, suggesting that this accession was imported into African regions a long time ago, and has not been hybridized with other banana varieties [54]. Variant 3 was also the most frequent in the Pacific plantain Popolou AAB genome, although this accession also carries variant 2. Evidence suggests that the Pacific plantain was transported to Oceania a long time ago [55,56], and extremely long periods of somaclonal mutation occurred during its cultivation [57,58]. The two wild *acuminata* diploids *M. acuminata* spp. *banksii* and *M. acuminata* spp. *malaccensis* (Pisang kra) contained 3 variants (variants 1, 2 and 3) of the AGPase LSU gene among the 4 reported complete genes that encode LSU-1 and LSU-2. The latter accession is closely related to the *M. acuminata* spp. *malaccensis* double haploid whose genome has been sequenced [46]. However, the diploid genomic cultivars AA and AB, Pisang lilin and Safet velchi, respectively, lack the LSU-1 gene and contain only two variants in the LSU2 gene. According to Hippolyte [59], Mbwazirume (Mutika-Lujugria) has a binary banksii-zebrina origin. Our analysis indicated that the banksii genome encodes LSU-1 and LSU-2 but that the East African cooking banana, Mbwazirume (Mutika-Lujugria)

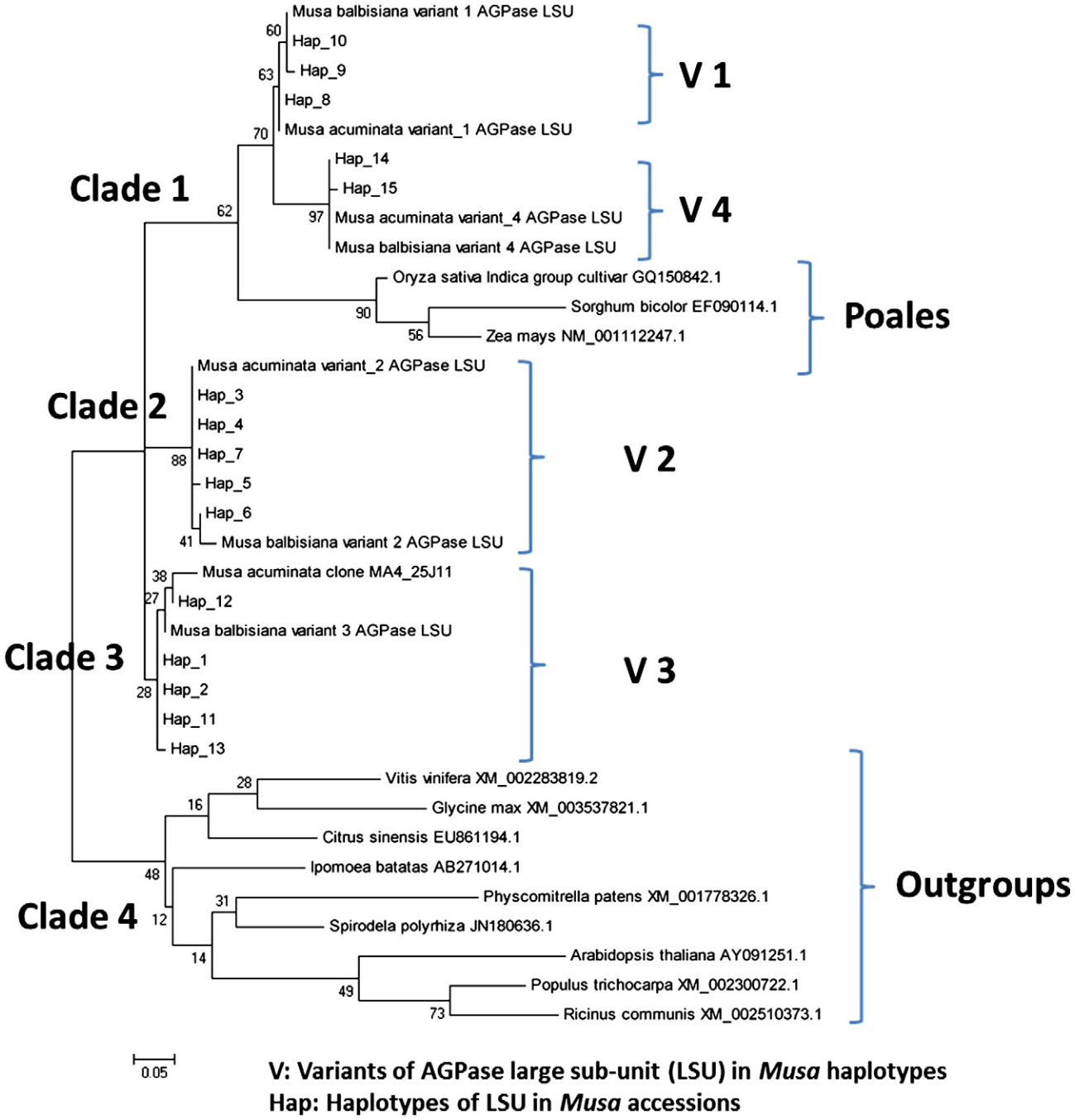

Fig. 5. Phylogenetic analysis of AGPase large sub-unit gene variants of *Musa*. A Neighbor-Joining tree showing the phylogenetic positions of the four variants in the 2nd exon of genes encoding the large subunits of AGPase and their relationships with monocots and dicots. The scale bar represents 0.05 nucleotide substitutions per site (1000 bootstrap replications were used in the phylogeny test). All nodes have 100% bootstrap support, and the Kimura 2 (K2) parameter model was used with 95% partial deletion in this tree. The coding sequence numbers of the AGPase LSU genes retrieved from the NCBI database and the GenBank accession numbers of the plant species are as follows: *Musa acuminata* **AC186746.1**, *Ipomea batatas* **AB271014.1**, *Arabidopsis thaliana* **AY091251.1**, *Physcomitrella patens* **XM_001778326.1**, *Sorghum bicolor* **EF090114.1**, *Citrus sinensis* **EU861194.1**, *Oryza sativa* Indica group cultivar **GQ150842.1**, *Spirodela polyrhiza* **JN180636.1**, *Zea mays* **NM_001112247.1**, *Vitis vinifera* **XM_002283819.2**, *Populus trichocarpa* **XM_002300722.1**, *Ricinus communis* **XM_002510373.1**, and *Glycine max* **XM_003537821.1**.

encodes only one version, LSU-2 (variant 3). Of six cultivars used in our analysis, four cultivars did not encode LSU-1 (variant 1) but did encode LSU-2. In contrast, variant 4, for which information on the 1st exon was lacking, was the least represented version in the accessions analyzed. It was almost absent in the A genomes, except for Gros Michel, which had one cloned sequence; this variant might represent a pseudogene of the AGPase LSU gene in *Musa*. To our knowledge, no AGPase LSU pseudogenes have been reported in banana, but have been reported in *A. thaliana* [60,61].

The genome of *Musa* is very similar to that of cereals, because the divergence of the Poales and Zingiberales belonging to the Commelinid order occurred before ~117 Mya [5]. Four isoforms of the AGPase LSU are present in rice [62], the model plant *A. thaliana* [38], potato [63] and tomato [64]. In our analysis, using the partial fragment of the AGPase LSU gene, we observed four variants (out of the six published in the *M. acuminata* genome).

From the AGPase LSU gene variants reported here, SNP-based molecular markers can be designed and used for the screening of core

Table 4
Association between the haplotypes/SNPs and variants within a region of exon 2. Variants 1 and 4 shared the same clades in the phylogenetic tree and variant 2 in clade 2 and variant 3 in clade 3. The SNPs or indels indicated in red were alterations in the open reading frame (ORF) at 7 positions in the 2nd exon of the four variants.

Variants	SNPs[av]	C/T	C/T	C/G	A/C	C/A	A/G	A/G/T	T/A	C/T	G/A	A/C	A/G	T/C	C/T	C/T	C/T	C/A	T/C	A/T	A/T	C/T	T/C	G/C/A	A/G	T/C	T/C	T/G
	Position	57	60	63	75	81	87	90	92	93	99	105	108	111	112	114	117	122	123	131	134	135	138	141	142	144	153	159
	Code	N1	N2	N3	N4	N5	N6	N7	N8	N9	N10	N11	N12	N13	N14	N15	N16	N17	N18	N19	N20	N21	N22	N23	N24	N25	N26	N27
V1	H8	C	T	C	A	C	A	A	T	T	A	C	A	C	C	T	T	A	T	T	A	C	T	C	A	T	C	T
	H9	C	T	C	A	C	A	A	T	T	A	C	A	C	T	T	T	A	T	T	A	C	T	C	A	C	C	T
	H10	C	T	C	A	C	A	A	T	T	A	C	A	C	C	T	T	A	T	T	A	C	T	C	A	C	C	T
V2	H3	C	C	G	C	A	G	G	T	C	G	A	G	T	C	C	C	C	T	A	T	C	C	G	A	T	C	T
	H4	C	C	G	C	A	G	G	T	C	G	A	G	T	C	C	C	C	T	A	T	C	C	G	A	T	C	T
	H5	C	C	G	C	A	G	G	T	C	G	A	G	T	C	C	C	C	T	A	T	C	C	G	G	T	C	T
	H6	C	C	G	C	A	G	G	T	C	G	A	G	T	C	C	C	C	T	A	T	C	C	C	A	T	C	T
	H7	C	C	G	C	A	G	G	T	C	G	A	G	T	C	C	C	C	T	A	T	C	C	G	A	T	C	T
V3	H1	T	T	G	C	A	A	A	T	C	G	A	G	T	C	C	C	C	T	A	A	C	T	G	A	T	T	T
	H2	T	T	G	C	A	A	A	T	C	G	A	G	T	C	C	C	C	T	A	A	C	T	G	A	T	T	T
	H11	T	T	G	C	A	A	A	T	C	G	A	G	T	C	C	C	C	T	A	A	C	T	G	A	T	T	T
	H12	T	T	G	C	A	A	A	T	C	A	A	G	T	C	C	C	C	T	A	A	C	T	A	A	T	T	T
	H13	T	T	G	C	A	A	A	T	C	G	C	G	T	C	C	C	C	T	A	A	C	T	G	A	T	T	T
V4	H14	C	C	C	A	C	A	A	A	T	A	C	A	T	C	C	T	A	C	T	A	T	T	C	A	T	C	G
	H15	C	C	C	A	C	A	T	A	T	A	C	A	T	C	C	T	A	C	T	A	T	T	C	A	T	C	G

V: variants.

collections, initially, to associate them with known starch phenotypes and subsequently, to analyze the diversity of the types of starch in the banana accessions. Most published studies on starch metabolism in monocots used the order Poales; however Musaceae is within the order Zingiberales. Therefore, this study may be useful for comparative genomics and as a model for the Zingiberaceae family (ginger, cardamom, galangal, myoga ginger and turmeric). However, the limitation of this work is that the functions of these variants are unknown; we cannot rule out the possibility that one or more of the variants are present in pseudogenes. The functionality of these variants has to be analyzed by evaluating their transcripts (mRNA). The primers used in this study were designed before the release of the banana genome sequence; therefore, it was not known whether these variants were chloroplastic or cytosolic. However, BLAST analysis against the *Musa* genome sequence indicated that they were chloroplastic. The *Musa* accessions analyzed in this study are diploids (wild and cultivars) and triploids (cultivars). Some variants are alleles, whereas others are paralogs. It might be possible to detect further isoforms in studies on the whole population. More isoforms might also be detected by sequencing more clones. Many cultivars are sterile, triploid, parthenocarpic and vegetatively propagated, and conventional breeding in the banana is difficult; therefore, the development of molecular markers for traits such as starch quality might be an important tool to improve this crop. The genome sequence of banana has recently been released, and this will help in many aspects of banana research, particularly in their genetic improvement.

AGPase is an allosteric enzyme, and alterations in its allostery affect the amount of starch synthesized [32]; therefore, it might be possible to manipulate the allosteric site of the large sub-unit to increase starch synthesis. In potato, a mutant variant of AGPase over-expressed using *Escherichia coli* (*glgC16*) accumulated up to 60% more starch than untransformed potatoes [12]. Bemiller and Whistler [19] showed that increasing the levels of 3-PGA and Pi, which are positive and negative regulators of AGPase, respectively, resulted in increased ADP-Glucose synthesis and this could be applied for the improvement of starch biosynthesis in higher plants. In rice, the AGPase genes were genetically engineered to increase the grain yield and biomass production [65]. Therefore, the AGPase enzyme could be manipulated to increase the starch content in the banana, which is an important fruit crop available throughout the year. This is important for plant biotechnology, because the demand posed by the rapid growth in the human population requires increased crop yield using the same or smaller cultivated areas. This requirement and the problems of global climate change and demand for high caloric content pose significant challenges for plant breeders.

In conclusion, SNPs were found in a fragment of the AGPase LSU gene in *Musa*. If and when starch phenotypes are associated with these variants, SNP-based molecular markers could be designed rapidly and easily for each variant of the AGPase LSU gene fragment in the banana genomes using techniques such as the mismatch amplification mutation assay (MAMA-PCR) methodology by Syverson and Bradeen [66]. This strategy might enable the determination of the total number

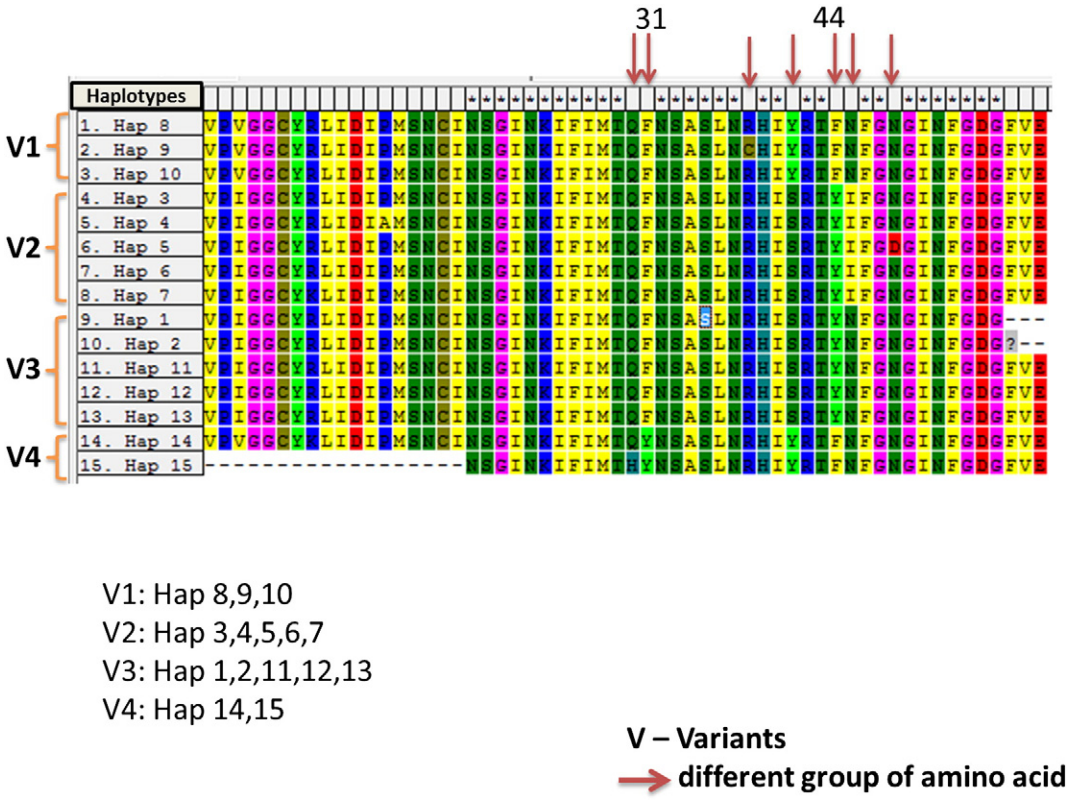

Fig. 6. Deduced peptide regions in the *Musa* haplotypes. The amino acid sequences encoded by the 2nd exon used for creating the dendrogram for the AGPase LSU are shown. The arrows highlight the positions of amino acid changes in the open reading frame.

of variants present and which of those are translated into proteins. The results of this study can be applied in basic science research and by banana breeders and biotechnologists interested in manipulating starch traits for food or industrial purposes. The SNP-based molecular markers can be linked to useful traits for accelerating crop improvement, and for basic molecular studies of orphan crops, such as the banana, and other modern crops.

Financial support

International Atomic Energy Agency (IAEA) Austria through research grant and Consejo Nacional de Ciencia y Tecnología (CONACyT), Mexico for the first author's scholarship. Project number: IAEA #RC 13198/R0 (Andrew James) and CONACyT #239810 (Muhilan Mahendhiran).

Acknowledgments

Ir. I. Van den Houwe, curator of the *Musa* International Transit Centre (ITC) at Katholieke Universiteit, Leuven, Belgium for supplying the banana accessions, INIFAP-CIRSE for the field facilities provided for the conservation of the ex situ *Musa* collection launched by the CICY at the INIFAP-Uxmal-Experimental Station-Yucatán, México, and MSc José Roberto Ku-Cauich and Yereni Minero Garcia for their technical support.

Author contribution

Proposed theoretical frame: AJK, JHRP; Conceived and designed the experiments: MM; Software development: JHRP; Contributed reagents/materials/analysis tools: AJK, RMEGM, JHRP, BCC, MTS, RGA; Wrote the paper: MM, JHRP, RMEGM, BCC, AJK; Performed the experiments: MM; Analyzed the data: MM, JHRP, RMEGM.

References

[1] Crouch JH, Vuylsteke D, Ortiz R. Perspectives on the application of biotechnology to assist the genetic enhancement of plantain and banana (*Musa* spp.). Electron J Biotechnol 1998;1:11–22. http://dx.doi.org/10.2225/vol1-issue1-fulltext-2.
[2] Simmonds NW. The evolution of the bananas. Longmans, Green and Co Ltd. 1962.
[3] Till BJ, Jankowicz-Cieslak J, Sági L, Huynh OA, Utsushi H, Swenne R, et al. Discovery of nucleotide polymorphisms in the *Musa* gene pool by Ecotilling. Theor Appl Genet 2010;121:1381–9. http://dx.doi.org/10.1007/s00122-010-1395-5.
[4] Youssef M, James AC, Rivera-Madrid R, Ortiz R, Escobedo-Gracia Medrano RM. *Musa* genetic diversity revealed by SRAP and AFLP. Mol Biotechnol 2011;47:189–99. http://dx.doi.org/10.1007/s12033-010-9328-8.
[5] Lescot M, Piffanelli P, Ciampi AY, Ruiz M, Blanc G, Leebens-Mack J, et al. Insights into the *Musa* genome: Synthetic relationships to rice and between *Musa* species. BMC Genomics 2008;9. http://dx.doi.org/10.1186/1471-2164-9-58.
[6] Aurore G, Parfait B, Fahrasmane L. Bananas, raw materials for making processed food products. Trends Food Sci Technol 2008;20:1–13. http://dx.doi.org/10.1016/j.tifs.2008.10.003.
[7] Christelova P, Valarik M, Hribova E, De Langhe E, Dolezel J. A multi gene sequence-based phylogeny of the Musaceae (banana) family. BMC Evol Biol 2011;11:1–13. http://dx.doi.org/10.1186/1471-2148-11-103.
[8] FAOStat: Food and agricultural commodities production. Food and Agricultural Organization Statistics; 2010 [http://faostat3.fao.org/faostat-gateway/go/to/home/E].
[9] Sharrock S, Lusty C. Nutritive value of banana. INIBAP annual report. Montpellier Francia: INIBAP; 1999. p. 28–31.
[10] Jain M, Swennen R. Banana improvement cellular, molecular biology, and induced mutations. Proceedings from a meeting held September 24–28, 2001, in Leuven, Belgium. Science Publishers, Inc. 2004.
[11] Pillay M, Ude G, Kole C. Genetics, genomics and breeding of bananas. USA: Science Publishers, CRC Press; 2012.
[12] Youssef M, James A, Mayo-Mosqueda A, Ku-Cauich RJ, Grijalva-Arango R, Escobedo-Gracia Medrano RM. Influence of genotype and age of explant source on the capacity for somatic embryogenesis of two Cavendish banana cultivars (*Musa acuminata* Colla, AAA). Afr J Biotechnol 2010;9:2216–23.
[13] Pillay M, Tenkouano A. Banana breeding — Progress and challenges. CRC Press, Taylor and Francis Group; 2011.
[14] Bello-Pérez LA, De Franscisco A, Acevado-Agama E, Gutierrez-Meraz F, Garcia-Suarez FJL. Morphological and molecular studies of banana starch. Food Sci Technol Int 2005;11:367–72. http://dx.doi.org/10.1177/1082013205058409.
[15] Bello-Perez LA, Sayago-Ayerdi SG, Mendez-Montealvo G, Tovar J. In vitro digestibility of banana starch cookies. Plant Foods Hum Nutr 2004;59:79–83. http://dx.doi.org/10.1007/s11130-004-0026-1.

[16] Mohapatra D, Mishra S, Singh CB, Jayas DS. Post-harvest processing of banana: Opportunities and challenges. Food Bioprocess Technol 2011;4:327–39. http://dx.doi.org/10.1007/s11947-010-0377-6.
[17] Eggleston G, Swennen R, Akoni S. Physicochemical studies on starches isolated from plantain cultivars, plantain hybrids and cooking bananas. Starch/Starke 1992;44:121–8. http://dx.doi.org/10.1002/star.19920440402.
[18] Yan HB, Pan XX, Jiang HW, Wu GJ. Comparison of the starch synthesis genes between maize and rice: Copies, chromosome location and expression divergence. Theor Appl Genet 2009;119:815–25. http://dx.doi.org/10.1007/s00122-009-1091-5.
[19] Bemiller J, Whistler R. Starch chemistry and technology. 3rd ed. USA: Elsevier; 2009.
[20] Zhang P, Whistler RL, Bemiller JN, Hamaker BR. Banana starch: Production, physicochemical properties, and digestibility—A review. Carbohydr Polym 2005;59:443–58. http://dx.doi.org/10.1016/j.carbpol.2004.10.014.
[21] Smith AM, Denyer K, Martin C. The synthesis of the starch granule. Annu Rev Plant Physiol Plant Mol Biol 1996;48:67–87. http://dx.doi.org/10.1146/annurev.arplant.48.1.67.
[22] Geigenberger P. Regulation of starch biosynthesis in response to a fluctuating environment. Plant Physiol 2011;155:1566–77. http://dx.doi.org/10.1104/pp.110.170399.
[23] Zeeman SC, Kossmann J, Smith AM. Starch: Its metabolism, evolution, and biotechnological modification in plants. Annu Rev Plant Biol 2010;61:209–34. http://dx.doi.org/10.1146/annurev-arplant-042809-112301.
[24] Streb S, Zeeman SC. Starch metabolism in *Arabidopsis*. The *Arabidopsis* book 2012;9. http://dx.doi.org/10.1199/tab.0160.
[25] James MG, Denyer K, Myers AM. Starch synthesis in the cereal endosperm. Curr Opin Biotechnol 2003;6:215–22. http://dx.doi.org/10.1016/S1369-5266(03)00042-6.
[26] Ohdon T, Francisco PBJR, Sawada T, Hirose T, Terao T, Satoh H, et al. Expression profiling of genes involved in starch synthesis in sink and source organs of rice. J Exp Bot 2005;56:3229–44. http://dx.doi.org/10.1093/jxb/eri292.
[27] Ballicora MA, Iglesias AA, Preiss J. ADP-glucose pyrophosphorylase, a regulatory enzyme for bacterial glycogen synthesis. Microbiol Mol Biol Rev 2003;67:213–25. http://dx.doi.org/10.1128/MMBR.67.2.213-225.2003.
[28] Georgelis N, Braun EL, Shaw JR, Hannah CL. The two AGPase subunits evolve at different rates in angiosperms, yet they are equally sensitive to activity-altering amino acid changes when expressed in bacteria. Plant Cell 2007;19:1458–72. http://dx.doi.org/10.1105/tpc.106.049676.
[29] Hannah CL, Shaw JR, Giroux MJ, Reyss A, Prioul JL, Bae JM, et al. Maize genes encoding the small subunit of ADP-glucose pyrophosphorylase. Plant Pathol 2001;127:173–83. http://dx.doi.org/10.1104/pp.127.1.173.
[30] Rosti S, Denyer K. Two paralogous genes encoding small subunits of ADP-glucose pyrophosphorylase in maize, bt2 and l2, replace the single alternatively spliced gene found in other cereal species. J Mol Evol 2007;65:316–27. http://dx.doi.org/10.1007/s00239-007-9013-0.
[31] Ventriglia T, Ballicora MA, Crevillen P, Preiss J, Romero JM. Regulatory properties of potato-*Arabidopsis* hybrid ADP-glucose pyrophosphorylase. Plant Cell Physiol 2007;48:875–80.
[32] Hannah C, Greene T. The complexities of starch biosynthesis in cereal endosperms. In: Kriz AL, Larkins BA, editors. Molecular genetic approaches to maize improvement. Heidelberg: Springer-Verlag Berlin; 2009. p. 287–301.
[33] Hwang SK, Hamada S, Okita WT. Catalytic implications of the higher plant ADP-glucose pyrophosphorylase large subunit. Phytochemistry 2007;68:464–77. http://dx.doi.org/10.1016/j.phytochem.2006.11.027.
[34] Boehlein SK, Shaw JR, Hannah LC, Stewart JD. Probing allosteric binding sites of the maize endosperm ADP-glucose pyrophosphorylase. Plant Physiol 2010;152:85–95. http://dx.doi.org/10.1104/pp.109.146928.
[35] Stark DM, Timmerman KP, Barry GF, Preiss J, Kishore GM. Regulation of the amount of starch in plant tissues by ADP-glucose pyrophosphorylase. Science 1992;258:287–92. http://dx.doi.org/10.1126/science.258.5080.287.
[36] Cross JM, Clancy M, Shaw JR, Greene TW, Schmidt RR, Okita TW, et al. Both subunits of ADP-glucose pyrophosphorylase are regulatory. Plant Physiol 2004;135:137–44. http://dx.doi.org/10.1104/pp.103.036699.
[37] Kavakli HI, Kato C, Choi SB, Kim KH, Salamone PR, Ito H, et al. Generation, characterization, and heterologous expression of wild-type and up-regulated forms of *Arabidopsis thaliana* leaf ADP-glucose pyrophosphorylase. Planta 2002;215:430–9. http://dx.doi.org/10.1007/s00425-001-0727-8.
[38] Crevillén P, Ventriglia T, Pinto F, Orea A, Mérida A, Romero JM. Differential pattern of expression and sugar regulation of *Arabidopsis thaliana* ADP-glucose pyrophosphorylase-encoding genes. J Biol Chem 2005;280:8143–9. http://dx.doi.org/10.1074/jbc.M411713200.
[39] Tetlow IJ, Emes MJ. Starch biosynthesis in higher plants: The enzymes of starch synthesis. In: Moo-Young M, editor. Comprehensive biotechnology. Elsevier; 2011. p. 47–65.
[40] Tetlow IJ, Morell MK, Emes MJ. Recent developments in understanding the regulation of starch metabolism in higher plants. J Exp Bot 2004;55:2131–45. http://dx.doi.org/10.1093/jxb/erh248.
[41] Dellaporta SL, Wood J, Hicks JB. A plant DNA minipreparation: Version II. Plant Mol Biol Rep 1983;1:19–21. http://dx.doi.org/10.1007/BF02712670.
[42] Stulnig TM, Amberger A. Exposing contaminating phenol in nucleic acid preparations. Biotechniques 1994;16:402–4.
[43] Youssef M. Khirshyat 1.0: A simple micro-program for some molecular biology protocols. Gene Genomes Genom 2012;6:102–5.
[44] Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 2011;28:2731–9. http://dx.doi.org/10.1093/molbev/msr121.
[45] Santiago-Sotelo P, Ramirez-Prado JH. prfectBLAST: A platform-independent portable front end for the command terminal BLAST+ stand-alone suite. Biotechniques 2012;53:299–300. http://dx.doi.org/10.2144/000113953.
[46] Davey MW, Gudimella R, Harikrishna JA, Sin LW, Khalid N, Keulemans J. A draft *Musa balbisiana* genome sequence for molecular genetics in polyploid, inter- and intra-specific *Musa* hybrids. BMC Genomics 2013;14:683. http://dx.doi.org/10.1186/1471-2164-14-683.
[47] D'Hont A, Denoeud F, Aury JM, Baurens FC, Carreel F, Garsmeur O, et al. The banana (*Musa acuminata*) genome and the evolution of monocotyledonous plants. Nature 2012;488:213–7. http://dx.doi.org/10.1038/nature11241.
[48] Mizuno K, Kawasaki T, Shimada H, Satoh H, Kobayashi E, Okumura S, et al. Alteration of the structural properties of starch components by the lack of an isoform of starch branching enzyme in rice seeds. J Biol Chem 1993;268:19084–91.
[49] Garwood DL, Shannon JC, Creech RG. Starches of endosperms possessing different alleles at the amylose-extender locus in *Zea mays*. Cereal Chem 1976;53:355–64.
[50] Liu F, Ahmed Z, Lee EA, Donner E, Liu Q, Ahmed R, et al. Allelic variants of the *Amylose extender* mutation of maize demonstrate phenotypic variation in starch structure resulting from modified protein–protein interactions. J Exp Bot 2012;63:1167–83. http://dx.doi.org/10.1093/jxb/err341.
[51] Regina A, Kosar-Hashemi B, Ling S, Li Z, Rahman S, Morell MK. Control of starch branching in barley defined through differential RNAi suppression of starch branching enzyme IIa and IIb. J Exp Bot 2010;61:1469–82. http://dx.doi.org/10.1093/jxb/erq011.
[52] Corbi J, Debieu M, Rousselet A, Montalent P, Le Guilloux M, Manicacci D, et al. Contrasted patterns of selection since maize domestication on duplicated genes encoding a starch pathway enzyme. Theor Appl Genet 2011;122:705–22. http://dx.doi.org/10.1007/s00122-010-1480-9.
[53] Corbi J, Dutheil JY, Damerval C, Tenaillon MI, Manicacci D. Accelerated evolution and coevolution drove the evolutionary history of AGPase sub-units during angiosperm radiation. Ann Bot 2012;109:693–708. http://dx.doi.org/10.1093/aob/mcr303.
[54] Perrier X, De Langhe E, Donohue M, Lentfer C, Vrydaghs L, Bakry F, et al. Multidisciplinary perspectives on banana (*Musa* spp.) domestication. Proc Natl Acad Sci U S A 2011;108:1131–8. http://dx.doi.org/10.1073/pnas.1102001108.
[55] De Langhe E, Vrydaghs L, De Maret P, Perrier X, Denham T. Why bananas matter: An introduction to the history of banana domestication. ERA 2009;7:165–78.
[56] De Langhe E. The establishment of traditional plantain cultivation in the African rainforest: A working hypothesis. In: Denham TP, Iriate J, Vrydaghs L, editors. Rethinking agriculture: Archaeological and ethnoarchaeological perspectives. California: Left Coast Press Inc.; 2007. p. 361–70.
[57] Neumann K, Hildebrand E. Early bananas in Africa: The state of the art. ERA 2009;7:353–62.
[58] Perrier X, Bakry F, Carreel F, Jenny C, Horry JP, Lebot V, et al. Combining biological approaches to shed light on the evolution of edible bananas. ERA 2009;7:199–216.
[59] Hippolyte I, Jenny C, Gardes L, Bakry F, Rivallan R, Pomies V. Foundation characteristics of edible *Musa* triploids revealed from allelic distribution of SSR markers. Ann Bot 2012;109:937–51. http://dx.doi.org/10.1093/aob/mcs010.
[60] Sulmon C, Gouesbet G, Ramel F, Cabello-Hurtado F, Penno C, Bechtold N, et al. Carbon dynamics, development and stress responses in *Arabidopsis*: Involvement of the apl4 subunit of ADP-glucose pyrophosphorylase (starch synthesis). PLoS One 2011;6:e26855. http://dx.doi.org/10.1371/journal.pone.0026855.
[61] Zou C, Lehti-Shiu MD, Thibaud-Nissen F, Prakash T, Buell CR, Shiu SH. Evolutionary and expression signatures of pseudogenes in *Arabidopsis* and rice. Plant Physiol 2009;151:3–15. http://dx.doi.org/10.1104/pp.109.140632.
[62] Bao J, Lu Y, Yang F, Zhang G, Shao Y, Corke H, et al. Nucleotide polymorphisms in *OsAGP* genes and their possible association with grain weight of rice. J Cereal Sci 2012;55:312–7. http://dx.doi.org/10.1016/j.jcs.2012.01.001.
[63] La Cognata U, Willmitzer L, Muller-Rober B. Molecular cloning and characterization of novel isoforms of potato ADP-glucose pyrophosphorylase. Mol Gen Genet 1995;246:538–48. http://dx.doi.org/10.1007/BF00298960.
[64] Park SW, Chung WII. Molecular cloning and organ-specific expression of three isoforms of tomato ADP-glucose pyrophosphorylase gene. Gene 1998;206:215–21. http://dx.doi.org/10.1016/S0378-1119(97)00588-X.
[65] Sakulsingharoj C, Choi SB, Hwang SK, Edwards GE, Bork J, Meyer CR, et al. Engineering starch biosynthesis for increasing rice seed weight: The role of the cytoplasmic ADP-glucose pyrophosphorylase. Plant Sci 2004;167:1323–33. http://dx.doi.org/10.1016/j.plantsci.2004.06.028.
[66] Syverson RL, Bradeen JM. A novel class of simple PCR markers with SNP-level sensitivity for mapping and haplotype characterization in *Solanum* Species. Am J Potato Res 2011;88:269–82. http://dx.doi.org/10.1007/s12230-011-9191-6.