American Journal of Biology and Life Sciences
2013; 1(1): 11-19
Published online December 10, 2013 (http://www.openscienceonline.com/journal/ajbls)

Single nucleotide polymorphism (SNP) markers discovery within *Musa spp* (plantain landraces, AAB genome) for use in beta carotene (Provitamin A) trait mapping

MMEKA, Ebelechukwu C[1, 2, 3, *], ADESOYE, Adenubi I[1], VROH, Bi I[2], UBAOJI, Kingsley I[4]

[1]Department of Botany, university of Ibadan, Ibadan, Nigeria
[2]Bioscience Centre, International Institute of Tropical Agriculture, IITA, Ibadan, Nigeria
[3]Department of Botany, Nnamdi Azikiwe University, Awka, Anambra State, Nigeria
[4]Department of Biochemistry, Nnamdi Azikiwe University, Awka, Anambra State, Nigeria

Email address

ebelemmeka@yahoo.com (E.C Mmeka)

To cite this article

MMEKA, Ebelechukwu C, ADESOYE, Adenubi I, VROH, Bi I, UBAOJI, Kingsley I. Single Nucleotide Polymorphism (SNP) Markers Discovery within Musa *spp* (Plantain Landraces, AAB Genome) for Use in Beta Carotene (Provitamin A) Trait Mapping. *American Journal of Biology and Life Sciences*. Vol. 1, No. 1, 2013, pp. 11-19.

Abstract

The need for biofortification of staple crops has become the concern of Modern Nutritional Genomics in producing staple crops with the ability of accumulating high micronutrients. Beta-carotene (Provitamin A) biosynthesis pathway is a stepwise process which is coded by two main genes; *phytoene synthase (PSY)* and *lycopene beta-cyclase (LYB)*. These two important genes *LYB 7 and PSY 1 1* were amplified, sequenced and analyzed within various diploid and triploid plantain landraces to identify SNPs present and to study their potentials for mapping the quantitative trait loci in those genes brought about by climatic variation within plantain landraces over time and also reveal their phylogenetic relationship. A total of 7 SNPs were found in the two genes from 68 genotypes used for PSY 11 and 35 plantain genotypes for *LYB 7* gene. A frequency of 1 SNP per 80bp was detected in *LYB 7* and I SNP per 32bp in *PSY 11* gene of 160 bp sequences each. The allelic diversity was 0.0286 in *LYB 7* and 0.0147 to 0.0294 in *PSY* 11. The phylogenetic relationship revealed that the genotypes for *LYB 7* gene follow two distinct lines of evolution whereas in *PSY 11* gene, the genotypes are similar with Zue and Sel advancing a bit. From the research, the high frequency of SNPs detected showed that it is sufficient in mapping the quantitative trait loci associated with high beta-carotene synthesis and could be incorporated in plantain breeding program.

Keywords

LYB 7, PSY 11, Indels, *Musa Species*, ORF, Quantitative Trait, SNPs, Transition, Transversion

1. Introduction

Musa species belong to the family *Musaceae* together with the genus *Ensete.* They are the most important food crops, which rank fourth in the world (after rice, wheat and maize). The major edible cooking bananas are triploid, that is with 2n=3x=33 chromosomes, and are mostly male and female sterile issued from the hybridization of two wild diploid ancestors namely *Musa acuminata* (AA genome) and *M. balbisiana* (BB genome). The triploids are grouped into three major types: AAA (Cavendish or dessert bananas), AAB (plantain), and ABB (cooking or dessert bananas). Nearly 30 million tons of bananas are produced yearly in Africa, mostly by smallholders and consumed locally. Bananas are rich in potassium, manganese, vitamins and fiber but low in fat and help to reduce risk of colorectal cancer (Hippolyte and Roux, 2005; Wong, 2011)

Evaluation of genetic diversity and genetic structure in crops has important implications for plant breeding

programs and the conservation of genetic resources. With changes in the environmental conditions, the challenges of pests and diseases, nutrient content as well as genetic vulnerability have posed a threat to Musa production (Kumar, 2010). So, there is a need to breed for improved Musa plants with these agronomic traits so as to alleviate poverty and food insecurity especially in the developing countries. Global genetic diversity studies have shown to be found in landraces, subspecies and wild–related ancestors of the current commercial cultivars. Different plantain landraces have large genetic broad base and many plantain genotypes have evolved to adapt to various environments (Adesoye *et al.,* 2012; Ren, 2013). In view of this, IITA conserves more than 100 plantain landraces segregating for several agronomic traits in different environments.

Moreover, *Musa* breeding is difficult and slow and requires integration of molecular techniques to achieve a faster result. Marker-assisted selection (MAS) has become a routine procedure in many breeding programs of major crops. Linkage Disequilibrium (LD) between DNA markers and quantitative trait loci, QTLs provides the basic principle of MAS that marker alleles are not randomly associated with QTL alleles (Utomo and Linscombe, 2009). *Musa* has been successfully studied with different molecular markers such as RFLP (Gawel *et al.,* 1992; Carreel *et al.,* 1994; Jarret *et al.,* 1994), AFLP (Noyer *et al.,* 2005; Ude *et al*., 2002a, b; Wong *et al.,* 2002); SSR (Grapin *et al*., 1998; Lagoda *et al*., 1998; Creste *et al.,* 2004) and SNP (Hearne *et al.,* 2009; Adesoye *et al*., 2012). In association with ploidy level, all these molecular methods led to clearer representations of the *Musa* complex (Perrier *et al*., 2009).

Single nucleotide polymorphisms (SNPs) refer to specific and defined positions at a chromosomal site which the DNA sequence of two genotypes differ by a single base. This might be the result of a transition (purine to purine or pyrimidine to pyrimidine change) or transversion event (purine to pyrimidine interchange) or a small deletions or insertions (indels). SNP markers are effective in detecting genetic diversity. They are the most abundant occurring at a frequency of about one SNP in 1000 nucleotides in genomic DNA and can be used to directly detect alleles responsible for a trait of interest (Wang *et al*., 1998; Ren *et al*., 2013). The data from DNA microsatellites or SSR and SNP are known to be consistent with each other and independent of the techniques used to produce them and is reproducible between laboratories but SNPs produce more reliable data than microsatellites (Magnid, 2006). Single-nucleotide polymorphisms (SNPs) are the markers of choice for those crops where massive sequence data are available, such as ESTs from diverse germplasm. In SNP technique, primers are designed, aiming to span single alleles at a locus independently thereby allowing the identification of a specific allele (Bhat, 2005). They are co-dominant markers. Due to their abundance and distribution through-out the genome, they are preferred for mapping, marker-assisted breeding and map-based cloning (Nageswara-Rao and Soneji, 2008).

Nevertheless, Carotenoids have been reported to have anticancer and antioxidant properties (Handelman, 2001). They help to prevent eye problems, skin disorders, heart diseases, and enhance immunity. Of all the carotenoids synthesized in plants, β-carotene is the most active and the best plant source of vitamin A. In the body, β-carotene is converted into retinol (vitamin A) by central cleavage by an oxygenase, yielding two molecules of retinal as against α-carotene that releases only one molecule with vitamin A activity (Mayer, 2007). Beta-carotene biosynthesis is a stepwise pathway of which geranylgeranyl-diphosphate (GGDP) is the key intermediate. The pathway splits into two mutually exclusive branches, only one of which leads to β-carotene which is of interest. The genes lycopene *β-cylase* and *phytoene synthase*, have been reported to be the most important in beta-carotene biosynthesis (Warburton, 2010).

Furthermore, the knowledge of the biosynthetic pathway of some genes has led to the production of biofortified crops with the ability of producing and accumulating the desired micronutrients in the edible portions of the plant's own biosynthetic (vitamins) or physiological (mineral) capacity (Mayer *et al.,* 2008). Biofortified crops may be obtained through conventional breeding (interbreeding with wild relatives) or genetic engineering. An example of known vitamin A biofortified crop is the golden rice which was produced by insertion of two genes responsible for the synthesis of β-cartene (Mayer, 2007).

Furthermore, Vitamin A deficiency (VAD) severely affects the immune system. According to WHO, an estimated 127 million children are affected by vitamin A deficiency, with 250,000 to 500,000 children becoming blind every year; half of whom die within a year (WHO database of vitamin A deficiency; http://www.who.int/vmnis/vitamina/data/en/index.html). *Musa spp* have been shown to have met the minimum nutritional requirement by FAO (2009). Improving the beta-carotene content of plantains will help to increase the micronutrients of plantain consumed by people especially in the developing countries like Africa where there are high cases of malnutrition, blindness and have the highest percentage of plantain consumption (FAOSTAT, 2009). This research work aims at generating SNP markers at beta carotene biosynthetic pathway in diploid and triploid plantain. The SNP markers generated will be used for association mapping, and the SNP tightly linked to beta-carotenoid trait will be recommended and used to assist breeding in *Musa*.

2. Materials and Methods

2.1. Plant Material and DNA Extraction

Sixty-eight different plantain germplasms used were obtained from the International Institute of Tropical Agriculture (IITA) plantain field germplasm. The DNA was

extracted from the young emerging (cigar) leaves following the DNA miniprep Extraction protocol of Doyle and Doyle (1990).

2.2. Primer Design

PCR primers were designed using primer program and the sequence of partial and complete cDNA. *PSY 11 (Phytoene synthase), LYB 7 (Lycopene β-cyclase)* genes sequences used are available at NCBI (2011) under accession number HM59159 and HM59169 respectively. Sequences of PSY *11* and *LYB 7* genes were PCR amplified for the sixty-eight selected plantain genotypes using the following forward (F) and reverse (R) primers:

PSY 11F (Phytoene synthase):	5′-GTGTGGTGTAGGAGGACAGATGAG-3′
PSY 11R:	5′-GCTCTCTTTGTGAAGTTGTTGTA-3′
LYB 7F (Lycopene β-cyclase):	5′-TGAGCTTCCCATGTATGACCC-3′
LYB 7R	5′-ACTGGGAGTGGACCACCCAT-3′

2.3. PCR and Sequencing

PCR reaction mixes were prepared for each sample by mixing 5 µl of Red Taq ready-mix (Sigma) which comprises of PCR buffer, dNTP, $MgCl_2$ and sterile water, 0.5 µl each of Forward and Reverse primers and 4 µl of genomic DNA (15 ng/µl) making a total of 10 µl for one reaction. The cycling parameters for the PCR were set at an initial 94^{o}C for 3 minutes, followed by 37 cycles of 94^{o}C for one minute, annealing step for 45 minutes of 60^{o}C – 51^{o}C of 1^{o}C touchdown for *PSY 11* and *LYB 7* forward and reverse primers, 72^{o}C for 30 minutes and a polishing step of 72^{o}C for 5 minutes. Products were separated by 2% agarose gel electrophoresis run at 100 Volts to check for efficiency of amplification and to ensure that only a single band product of the expected size was present. PCR products were then transferred to ABI plates and sequenced three times with the forward primers used in the PCR amplification.

2.4. SNP Marker Discovery

The approach to the discovery of single nucleotide polymorphisms (SNPs) is through direct sequencing of PCR products of the *Musa* genotypes on ABI ROBOT sequencer. Polymorphisms between the sequences were identified by sequence alignment using ClustalW. The SNPs were detected by aligning different *Musa* accessions using the software CLUSTALW2 available at European Bioinformatics Institution, EBI http://www.ebi.ac.uk/Tools/Services/ web_Clustal2). The files downloaded were then aligned with the software GeneDoc (www.psc.edu/biomed/genedoc) to show the conserved sites and mutation sites for SNP detection.

Furthermore, the sequence trace outputs from the sequencer were evaluated by eye to identify possible areas of heterozygous sequence. To increase the quality of sequences by minimizing false positives due to sequencing artifacts, potential SNPs were resequenced. Areas with ambiguous bases and baseline noise were removed usually at the beginning and at the end of the sequence. Since SNP analysis requires the highest sequence quality, the SNPs were detected from the Open Reading Frame (ORF), from the conserved sequence site in all the genotypes mostly about some distances from the beginning of the sequence spanning through the middle of the sequence. Sequences at the end of the gel at low resolution are cut off as it is prone to more oddly peaks and unidentifiable nucleotide bases.

2.5. Genetic Diversity and Data Analysis

To each SNP marker, two alleles are generally present because of its biallelic nature. Allelic frequencies for each marker were estimated by counting to determine the level of heterozygosity (h). This will help to measure the ability of a marker to adequately differentiate a genotype from another genotype effectively if it is present in only at that genotype. So low level of heterozygosity indicates the presence of a SNP at one or limited number of genotypes at a locus. Power of discrimination (PD) is the measure of the ability of the marker at different loci. It is used to measures the possibility of each marker to identify a locus was calculated using the formula (Kloosteman *et al.,* 1993).

$$PD = 1 - \sum_{i=1}^{g} f_i^2$$

Where f_i is the frequency of the ith genotype and the sum is over all genotypes.

The level of similarity among the different genotypes for a gene was estimated using online tool software, MrBayes (http://www.phylogeny.fr/version 2_cgi.index.cgi). Based on the genetic similarity matrix, the genotypes were clustered by the Maximum Likelihood (PhyML) with Bayesian analysis using the program Tool from Phylogenetic tree Service Tool of MrBayes-PhyML (Dereeper et al., 2008; http://www.phylogeny.fr/version 2_cgi.index.cgi) to show the evolutionary relationship.

3. Results

3.1. Single Nucleotide Polymorphism Detection

After editing and trimming the chromatogram PCR sequence products for the two genes (*LYB 7* and *PSY 11*), the good quality sequence were aligned for all the good genotypes. From the multiple sequence alignment of ClusterW2, the conserved regions are the shaded portions

forming blocks with few gaps indicating SNPs and Indels (Insertions and Deletions) (Fig 1 and 2). The SNP detection was obtained from the Open Reading Frame (ORF) regions of the sequences using ORF finder from NCBI website (http://www.ebi.ac.uk/Tools/Services/ web_Clustal2).

The shaded portions are the Conserved sequences among the Musa genotypes while the gaps (unshaded) are the SNPs.

For *LYB 7* gene, within the 160bp quality sequences each from 35 different diploid and triploid plantain and wild *Musa* genotypes used for SNP analysis (Fig. 1), two SNPs were detected at mutant loci 398 and 430 as shown in the Table 1. No Indel was found for LYB gene. *Dwa* and *Losb* genotypes have one mutant (SNP) locus each. The SNP base pairs are in bold italics.

The *PSY 11* gene for beta-carotene biosynthetic pathway is less polymorphic. From the 160 bp sequences of the multiple alignment of 68 genotypes (Fig. 2), five SNP sites were obtained (Table 2) within the ORF region at loci 11, 27, 92, 137 and 154. The base pair in bold italics indicates the mutant alleles (SNPs). Mont has the highest number of SNP loci (2) and Agb, B.v2, Sel and Zue have one each. No Indel was detected within the 160bp sequences analysed. In *LYB* 7, 50% each of SNPs detected was as a result of transition (purine-purine or pyrimidine-pyrimidine interchange) and transversion (purine-pyrimidine interchange) events. In *PSY 11*, 60% of SNPs detected was as a result of transition and 40% due to transversion.

The frequency of occurrence of the SNPs (Table 3) at each SNP locus was at 1 SNP per 80bp (1.25%) for *LYB 7* gene and one SNP in every 32bp (3.125%) for *PSY 11* gene. An average of 0.057 SNPs were found at each sequence of *LYB 7* with maximum number of 1 SNP per sequence and locus respectively. There is an average of 0.0735 SNP per sequence with a maximum number of 2 SNPs per locus and 2 SNPs per sequence for *PSY 11* gene.

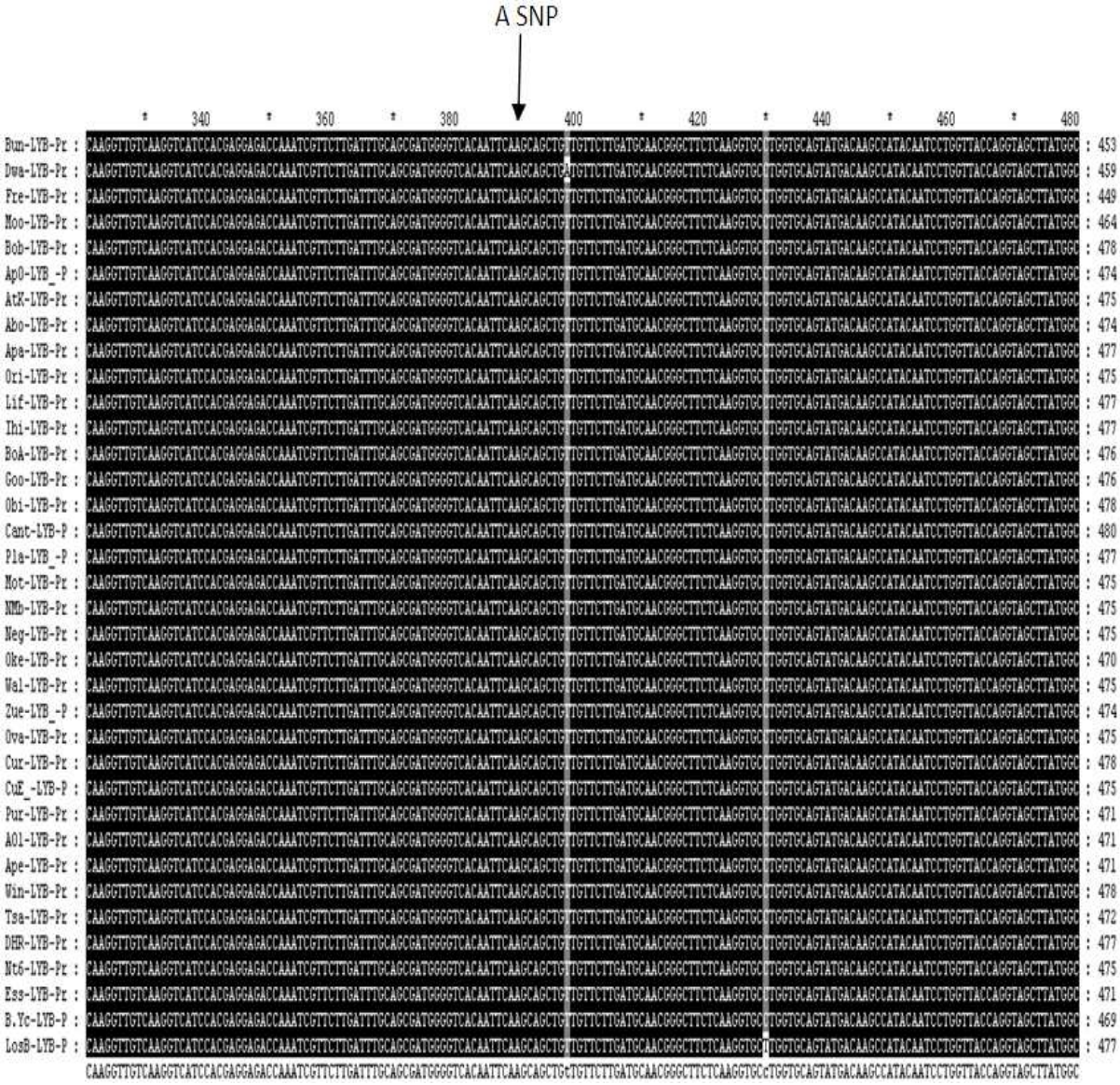


Fig 1. *Multiple sequence alignment obtained for Lycopene beta-cyclase (LYB 7) using GeneDoc Software (www.psc.edu/biomed/genedoc).*

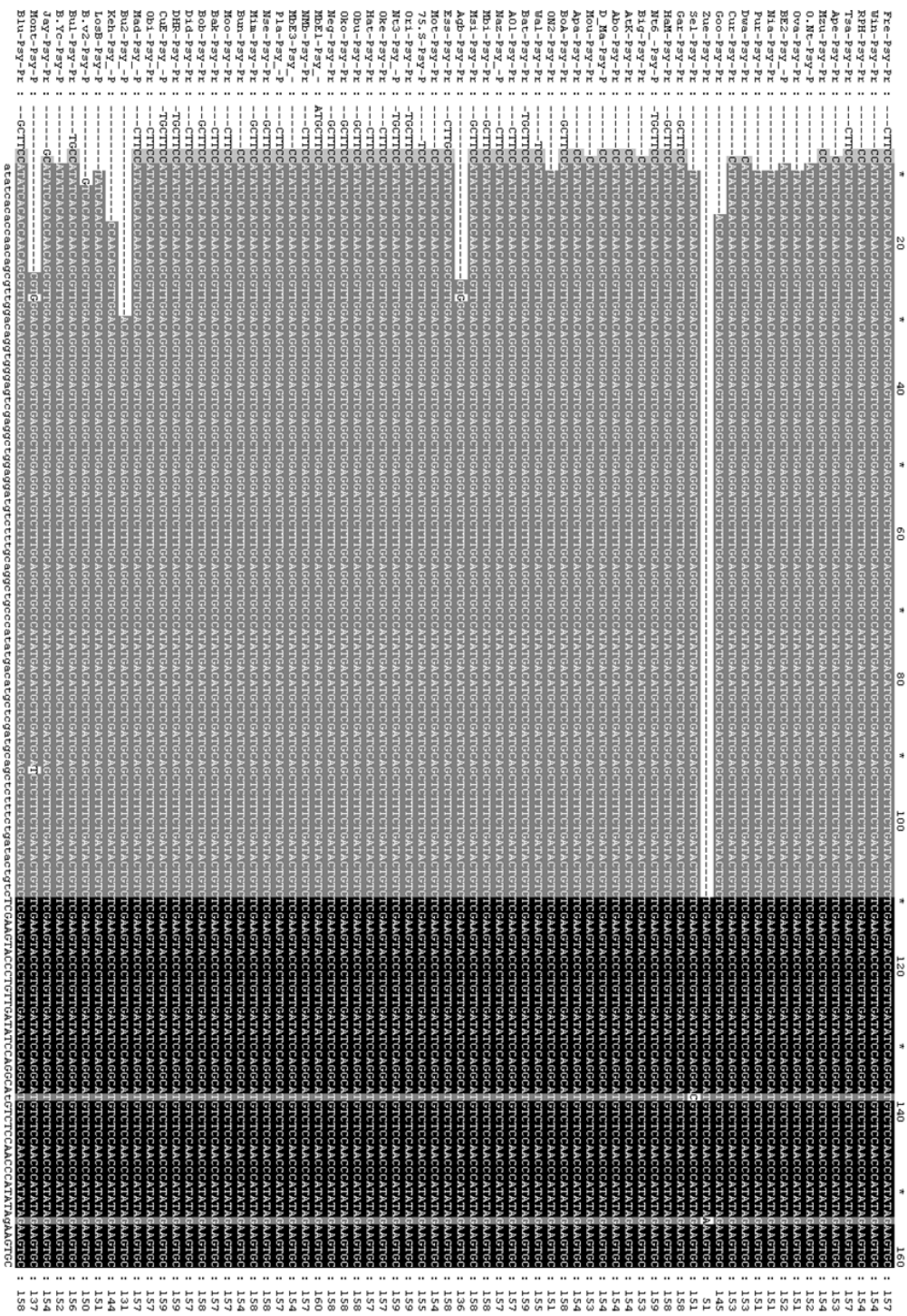


Fig 2. *Multiple sequence alignment obtained for Phytoene synthase (PSY 11) using GeneDoc Software (www.psc.edu/biomed/genedoc)*

Table 1. *SNPs detected at two mutant sites of Lycopene beta-cyclase 7 (LYB7) genes along the 160bp amplicon sequence of the genotypes.*

SNP site Genotypes	398	430	No of Mutant
Dwa	*A*	C	1
Losb	T	*T*	1
No of Mutant per loci	1	1	
SNP TYPE	V	T	2

V= transversion T=transition

Table 2. *SNPs detected at various six mutant sites of Phytoene synthase 11 (PSY11) genes along the 160 bp amplicon sequence of the genotypes.*

SNP site Genotypes	11	27	92	137	154	No of Mut
Agb	A	*G*	G	T	G	1
Mont	A	*G*	*T*	T	G	2
B.v2	*G*	T	G	T	G	1
Sel	A	T	G	*C*	G	1
Zue	A	T	G	T	*A*	1
No of Mutant per loci	1	2	1	1	1	
SNP TYPE	T	V	V	T	T	6

V= Transversions T=Transitions

Table 3. *SNPs found along the beta-carotene pathway for Lycopene beta-cyclase and Phytoene synthase genes*

Gene		Average per sequence	Average per loci	Minimum per loci	Maximum per loci	Minimum per sequence	Maximum per sequence	Total base sequence analysed	% variation
LYB 7	SNPs	0.057	0.013	0	1	0	1	160 (1 per 80bp)	1.25
PSY 11	SNPs	0.0735	0.025	0	2	0	2	160 (1 per 32bp)	3.125

3.2. Allelic and Genetic Diversity

The SNPs are biallelic in nature so can have two possible alleles at a locus. The level of heterozygosity (*h*) ranged from one mutant allele for *LYB 7* gene, 1.47% for one mutant allele to 2.94% for two mutant alleles in *PSY 11*. The power of discrimination (PD) of a mutant allele of a genotype at a locus calculates the ability to distinguish between genotypes at a locus based on the frequency of mutant alleles and from one locus to another. The power of discrimination of the mutant allele at each SNP locus were estimated (Table 4) which varied from 0.8163 for *PSY 11* (27) to 0.9998 for *PSY11* (11, 92, 137, 154) with an average of 0.9734.

Table 4. *Features and Frequencies of the new SNPs*

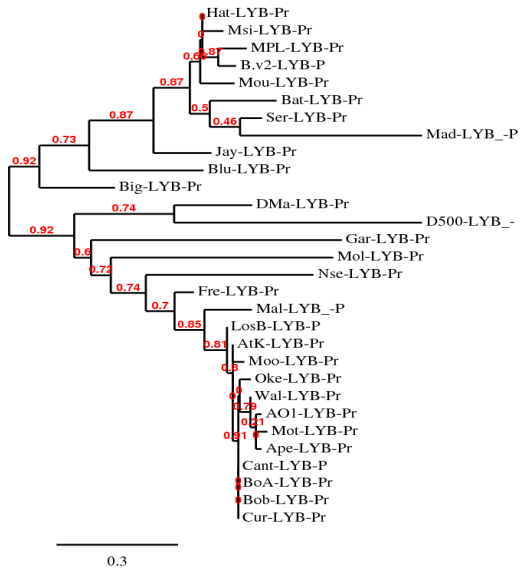
Gene	SNPs	SNP Type	Observed Heterozygosity, h (%)	Frequency of SNP allele	Power of discrimination, *PD* (%), $[1-\sum_{i=1}^{g} f_i^2]$
LYB7	LYB7 398*	T/A	2.86	1	99.92
	LYB7 430	C/T	2.86	1	99.92
PSY 11	PSY11 11	A/G	1.47	1	99.98
	PSY11 27	T/G	2.94	2	81.63
	PSY11 92	G/T	1.47	1	99.98
	PSY11 137	T/C	1.47	1	99.98
	PSY11 154	G/A	1.47	1	99.98

*relative to the nucleotide sequence number (see fig. 1 and 2) h is a measure of allelic diversity of the SNP. PD = Power of discrimination

3.3. Phylogenetic Relationship

The phylogenetic tree for the different genes using MrBayes software of Maximum likelihood with Bayesian analysis showed that the genotypes are all monophyletic originating from a common ancestor.

For *LYB 7* gene (Figure 3), Hat (AAB), Msi (AAB), MPL (BB), Bv2 (BB), Mou (AAB), Bat (AAB), Ser (AAB), Mad (AAB), Jay (BB), Blu (ABB), Big (AAB) genotypes follow the same line of evolution. Big diverging first from the node, followed by Blu, then Jay. The triploids Hat, Msi and Mou are sister taxa of which the diploids MPL (BB) and B.v2 (BB) must have one of their parents because of the relatedness. Bat, Ser and Mad are closely related though Mad genotype for *LYB 7* gene is the most advanced. In the second line of evolution, DMa (AAB) and D500 (AAB) are related and distinct from the rest, followed by Gar (AAB), then Mol (AAB), Nse (AAB), Fre (AAB), Mal (AA), LosB (BB), Atk (AAB), and Moo (AAB). Oke (AAB), Cant (AAB), BoA (AAB), Bob (AAB) and Cur (AAB) (sister taxa) are closely related from which Wal (AAB), A01 (AAB), Mot(AAB) and Ape (AAB) diverged from and are the most recent of their group. The diploids Mal (AA) and LosB (BB) might have been distant parents to the triploids Atk, Moo, Oke, Cant, BoA, Bob, Cur, Wal, A01, Mot and Ape.



Source: http://www.phylogeny.fr/version 2_cgi.index.cgi

Fig 3. *Phylogenetic tree for LYB 7 gene (p=0.3)*

Source: http://www.phylogeny.fr/version 2_cgi.index.cgi

Fig 4. *Phylogenetic tree for PSY 11 gene (p=0.01)*

For *PSY 11* gene (Fig. 4), all the genotypes are almost uniform though sel (AA) and Zue (AAB) are the most advanced within the group. All the genotypes are related at 99% (p=0.01) level of significance.

4. Discussion

The high frequency of transition to transversion found in the SNPs for *PSY 11* and *LYB 7* gene (table 1 and 2) along the beta-carotene biosynthetic pathway showed that there are more transition events to transversion for the SNPs discovered which is in conformation with the reports of (Hearne *et al.,* 2009; Shaheen *et al*., 2010).

SNPs of Dwa 398 and Losb 430 were the only mutant allele found along *LYB* gene sequence coding for beta-carotene. So these two SNPs might be potential markers for high beta-carotene production. Also for PSY gene, Agb and Zue having one SNP each are triploids. SNP found in Agb is also found in the wild plantain (Mont) so can have high beta-carotene content. Zue has a SNP not present in the wild plantains so may be a potential marker for high carotene production. Wild plantain genotypes like Mont, Yv1, MPL and triploid plantains with a fair level of SNP loci can be integrated into breeding work of *Musa*. This is done by cultivating the triploids in close links with those diploids (wild plantains) found to have high SNP loci (De Langhe *et al*., 2010).

The SNP types C/T, T/G,G/T,A/C,C/A,A/T,T/A,G/A,A/G have effects on the amino acids resulting from gene translation. Welsch *et al* (2010) have shown that the SNP at the 57bp of *PSY2* gene is associated with yellow root color in cassava. So these SNPs identified can have synonymous effect when it affects the gene product or non-synonymous or silent effect when it has no effect on the gene product.

Comparing the Phylogenetic analysis among the Musa genotypes for the two genes, the phylogenetic tree was able to depict the line of evolution of some diploids together with triploid plantain landraces assumed to have originated from them (Adesoye *et al*., 2012). The triploids were generally observed to be related with little difference due to the problem of homologues i.e. non- allelic, versions of genes residing on homologous chromosomes (Henry and Edwards, 2009) and also because they are vegetatively propagated unlike the diploids that have more genes interchange due to sexual reproduction. For *LYB 7* (Fig 3), the genotypes used followed two distinct line of evolution with A01, Mot, and Ape been the most recent in terms of the *LYB* gene. Whereas in the *PSY* (Fig. 4) phylogenetic tree, the genotypes are similar to each other though Sel and Zue have advanced a bit. The uniformity of the genotypes for *PSY 11* confirmed the role of *PSY* gene as a single gene controlling the presence or absence of any carotene (Mayer, 2007; Warburton, 2010; Welsch *et al.,* 2010; Welsch, 2011).

In conclusion, different SNPs were identified which explored the polymorphisms at the conserved sequences of *Musa* genes. The SNPs also identified the genotypes therefore solving the problems of Homonymy and synonym especially in *Musa* genotypes. So SNP Markers would help in reducing genotyping costs and provide a rapid and easy way to establish fingerprint of each landrace use in genotype identification and for the different biosynthetic pathway mobility. This research will find important application in breeding, agronomic practice and ecosystem research.

Acknowledgement

We are grateful to Bioscience laboratory, International Institute of Tropical Agriculture (IITA), Ibadan, Nigeria for the opportunity to carry out this research work.

References

[1] Adesoye, A., Mmeka, E. and Vroh, B. (2012). Single Nucleotide Polymorphism Markers Discovery in *Musa Spp* (Plantain Landraces, AAB Genome) and its Potentials for Use in Gibberellic Acid and Parthenocarpy Trait Mapping. *J Plant Mol Biol Biotechnol 3*(1): 9-21

[2] Bhat, K.V. (2005). *DNA fingerprinting and cultivar identification.* National Research Center on DNA fingerprinting, N.B.P.G.R, New Delhi -011 012

[3] Carreel, F., Fauré, S., Gonzalez De Leon, D., Lagoda, P., Perrier, X., Bakry, F., Tézenas du Montcel, H., Lanaud, C. and Horry, J.P. (1994). Evaluation de la diversité génétique chez les bananiers diploïdes (*Musa sp*.). *Genetics Selection Evolution 26*:125-136.

[4] Creste, S., A. Tulmann , N., Vencovsky, S.O. Silva and Figueira, A. (2004). Genetic diversity of *Musa* diploid and triploid accessions from the Brazilian banana breeding program estimated by microsatellite markers. *Genetic Resources and Crop Evolution 51*:723-733.

[5] De Langhe, E., Hribova, E., Carpentier, S., Dolezel, J., and Swennen, R. (2010). Did backcrossing contribute to the origin of hybrid edible bananas? Annals of Botany Page 1 of 9 doi:10.1093/aob/mcq187, available online at www.aob.oxfordjournals.org

[6] Dereeper, A., Guignon, V, Blanc, G., Audic S, Buffet S, Chevenet F, Dufayard, J.F, Guindon, S, Lefort V, Lescot M, Claverie J.M, Gascuel, O. (2008). Phylogeny.fr: robust phyogenetic analysis for the non-specialist. Nucleic Acids Res, 1; 36 (Web server issue) W 465-9 (http://www.phylogeny.fr/version 2_cgi.index.cgi)

[7] Doyle J J, Doyle J L, 1990. Isolation of plant DNA from fresh tissue. Focus 12:13-15

[8] EBI (2011). European Bioinformatics Institution (http://www.ebi.ac.uk/Tools/Services/ web_Clustal2).

[9] FAOSTAT (2009). *Acting together against banana diseases.* Food and Agricultural Organization Statistics. Available online at http://www.fao.org

[10] Gawel, N.J., Jarret, R and Wittemore, A.P. (1992). Restriction fragment length polymorphism (RFLP)-based phylogenetic analysis of *Musa. Theoretical and Applied Genetics 84*:286-290.

[11] GeneDoc (2011). www.psc.edu/biomed/genedoc

[12] Grapin, A., Noyer,J.J., Carreel, F., Dambier, D., Baurens, F.C., Lanaud, C and Lagoda, P.J.L. (1998). Diploid *Musa acuminata* genetic diversity assayed with sequence tagged microsatellite sites. *Electrophoresis* 19:1374- 1380.

[13] Handelman, G.J. (2001). The evolving role of carotenoids in human biochemistry. *Nutrition 17*:818-822.

[14] Hearne, S. J., Lorenzen , J., Town, C., and Zhuang, E. (2009). EST derived genomic resources for Musa. *IITA Report* .2008. www.IITA.org.

[15] Henry R, Edwards K. 2009. New tools for single nucleotide polymorphism (SNP) discovery and analysis accelerating plant biotechnology. Plant Biotechnology Journal 7(4): 311

[16] Hippolyte, I. and Roux,N. (2005). Molecular Markers for Allele Mining *Proceedings of a workshop, 22–26 August 2005, MS Swaminathan Research Foundation, Chennai,* India M. Carmen de Vicente, M and Glaszmann, J *editors* International Plant Genetic Resources Institute, Generation Challenge Programme. P26.

[17] Jarret, R., Gawel, N., Wittemore, A.P., and Sharrock, S. (1994). RFLP-based phylogeny of *Musa* species in Papua New Guinea. *Theoretical and Applied Genetics* 84:579-584

[18] Kloosteman , A. D., Budowle, B. and Daselaar, P.(1993). PCR-amplification and detection of the human DIS80 VNTR locus. Amplification conditions, population genetics and amplification in forensic analysis. *International Journal for Legal Medicine*, 7: 257-264.

[19] Kumar, P. (2010). Biotech Bananas. Biotech Protecting the Banana Plant from Diseases. Available online at http://www.biotecharticles.com/

[20] Lagoda, P.J.L., Noyer, J.L., Dambier, D., Baurens, F.C., Grapin, A. and Lanaud, C. (1998). Primer note – Sequence tagged microsatellite site (STMS) markers in the Musaceae. *Molecular Ecology* 7:659-663

[21] Magnid, C. R. (2006). Comparing techniques Review: AFLP Data Interpretation and Analysis. Sequence-tagged site markers, RFLPs, RAPDs/Arbitrary techniques, SNPs, Diversity Arrays, Microsatellite. Dr. Cs. Forestales Depto. de Silvicultura Facultad de Cs. Forestales Universidad de Chile Santiago de Chile

[22] Mayer, J. E., Pfeiffer, W.H. and Beyer, P. (2008). Biofortified Crops to Alleviate Micronutrient Malnutrition. *Journal of Current Opinion in Plant Biology* 11:166-170.

[23] Mayer, J.E. (2007). Golden Rice, Golden Crops, Golden Prospects. *Rev. Colomb. Biotecnol. 9* (1): 22-34.

[24] Nageswara-Rao, M. and Soneji, J. R. (2008). Molecular Genetic Markers: What? Why? Which one for Exploring Genetic Diversity? Retrieved online from www.scienceboard.net/community/perspectives.210.html

[25] National Center for Biotechnology Information, NCBI. (2011). http://www.ncbi.nlm.nih.gov

[26] Noyer, J.L., Causse, S., Tomekpe, K., Bouet, A. and Baurens, F.C. (2005). A new image of plantain diversity assessed by SSR, AFLP and MSAP markers. *Genetica 124*: 61–69.

[27] Perrier, X., Bakry, F., Carreel, F., Jenny, C., Horry, J., Lebot, V. and Hippolyte, I (2009). Combining Biological Approaches to Shed Light on the Evolution of Edible Bananas *Ethnobotany Research & Applications 7*:199-216. Available online at www.ethnobotanyjournal.org/vol7/i1547-3465-07-199.pdf

[28] Ren J, Sun D, Chen L, You FM, Wang J, Peng Y, Nevo E, Sun D, Luo MC, Peng J. (2013). Genetic diversity revealed by single nucleotide polymorphism markers in a worldwide germplasm collection of durum wheat. *Int J Mol Sci.* 14(4):7061-88.

[29] Shaheen, T., Zafar, Y. and Mehboob-ur, R. (2010). Detection of Single Nucleotide Polymorphisms in the conserved ESTs regions of Gossypium arboretum. *Electronic Journal of Biotechnology, 13*: 1-10. Available online at http://www.ejbiotechnology.info.

[30] Ude, G., Pillay, M., Nwakanma D. and Tenkouano, A. (2002b). Genetic diversity in *Musa acuminata* Colla and *Musa balbisiana* Colla and some of their natural hybrids using AFLP markers. *Theoretical and Applied Genetics* 104:1246-1252.

[31] Ude, G., Pillay, M., Nwakanma D. and Tenkouano, A. (2002a). Analysis of genetic diversity and sectional relationships in *Musa* using AFLP markers. *Theoretical and Applied Genetics 104*:1239-1245.

[32] Utomo, H. S. and Linscombe, S. D. (2009). Current Patents and Future Development Underlying Marker –Assisted Breeding in Major Grain Crops. *Recent Patents on DNA & Gene Sequences 3*: 53-62.

[33] Wang, D. G., Fan J. B., Siao, C. J, Berno, A., Young, P., Sapolsky, R., Ghandour, G., Perkins, N., Wincheter, E., Spencer, J., Kruglyak, L., Stein, L., Hsie, L., Topaloglou, T., Hubbell, E., Robinson ,E., Mittmann, M., Morris M.S., Shen, N. P., Kilburn, D., Rioux, J., Nusbaum, C., Rozen, S., Hudson, T. J, Lipshutz, R., Chee, M., Lander E. S. (1998). Large-scale identification, mapping and genotyping of SNP in the human genome. *Science*, *280*:1077-1082.

[34] Warburton, M. L. (2010). Identification and Use of Rare Natural Genetic Sequence Variation to Improve Levels of β-Carotene in Maize. *ISB News Reports* June 2010.

[35] Welsch, R., Arango, J., Bar. C., Salazar. B., Al-Babili. S., Beltra. J., Ba. R., Salazar. B., Chavanrriaga. P., Ceballos. H., Tohme. J., and Beyer. P. (2010). Provitamin A Accumulation in cassava (*Manihot esculenta*) Roots Driven by a Single Nucleotide Polymorphism in a Phytoene Synthase Gene. *Plant Cell*. 22(10) 3348-3356.

[36] Welsch. R. (2011). A Single Nucleotide Polymorphism in a Phytoene Synthase Gene can Simplify Provitamin A Biofortification of Cassava. *ISB News Reports* March 2011. 1-3

[37] WHO database of vitamin A deficiency; http://www.who.int/vmnis/vitamina/data/en/index.html).

[38] Wikipedia encyclopedia (2011). Banana. Retrieved online on 18th May, 2011 from http://en.wikipedia.org/wiki/Banana

[39] Wong. C., Kiew, R., Argent, G., Set, O. Lee , S.K and Gan, Y.Y. (2002). Assessment of the validity of the sections in *Musa* (Musaceae) using AFLP. *Annals of Botany 90*:231-238.

[40] Wong. J. (2011)."The Pursuit of Happiness (A.K.A. It Appears That The Writer Wrote About Bananas After Eating A Few Too Many)". The Science Creative Quarterly, University of British Columbia. Available online at http://www.scq.ubc.ca/the-pursuit-of-happiness-aka-it-appears-that-the-writer-wrote-about-bananas-after-eating-a-few-too-many/.