SINET: Ethiop. J. Sci., 36(2):65–72, 2013
 ISSN: 0379–2897

PATTERNS OF NUCLEOTIDE DIVERSITY IN *MEISA1* AND *G3PDH* IN WILD AND CULTIVATED CASSAVA

Dereje Beyene[1], Settumba B. Mukasa[2], Christer Jansson[3], Etienne P de Villiers[4], Morag Ferguson[5], Ismail Yusuf Rabbi[5], Samuel Kyamanywa[2], James Sebuliba[2] and Yona Baguma[6]

[1] Department of Microbial, Cellular and Molecular Biology, Addis Ababa University, PO Box 1176, Addis Ababa, Ethiopia. E-mail: dereje.beyene@aau.edu.et; b.dereje@hotmail.com
[2] Department of Agricultural Production, Makerere University, PO Box 7062, Kampala, Uganda
[3] Lawrence Berkeley National Laboratory Berkeley, CA 94720 USA
[4] Biosciences eastern and central Africa-International Livestock Research Institute (BecA-ILRI) Hub, PO Box 30709, Nairobi, Kenya
[5] International Institute of Tropical Agriculture (IITA), PO Box 30709, BecA-ILRI Hub, Nairobi, Kenya
[6] National Crops Resources Research Institute (NaCRRI), PO Box 7084, Kampala, Uganda

ABSTRACT: The distribution and frequency of single nucleotide polymorphisms (SNPs) is an excellent tool for discerning evolutionary relatedness between cultivated and wild plant genomes. This type of information is scanty for the genus *Manihot,* and thus limiting systematic approaches in the genetic improvement of cassava. Here, we present a detailed description of the comparative patterns of SNPs in *Isoamylase1* (*Meisa1*) and *Glyceraldehyde-3-phosphate dehydrogenase* (*G3pdh)* in 10 accessions of wild (*Manihot esculenta* subsp. *flabellifolia*) and 12 accessions of cultivated cassava (*M. esculenta*). The results show that *Meisa1* is more variable in cultivated cassava than that in subspecies *flabellifolia*, where the 954 bp sequence region differs at 1 in 111 and 250 nucleotides of cultivated and wild species, respectively. Frequency analysis shows that SNP occurs once every 42 bp in cultivated and every 70 bp in wild. Tajima's D test statistics showed that *Meisa1* has been evolving under different selection pressures, diversifying in cultivated and purifying in wild. *G3pdh* is under diversifying selection in both populations. This may indicate the importance for *isoamylase1* in starch quality traits in cassava, a trait that is likely to have been the target for artificial selection by farmers and breeders, in addition to natural selection. This study also suggests that *G3pdh* may be a good marker for phylogeny study while *Meisa1* may be useful for intra and inter-cultivar diversity studies. The non-synonymous SNPs that changed the amino acid property were identified and the potential implication of the change in protein function was analyzed and discussed.

Keywords/phrases: ***Cassava, G3pdh, Isoamylase1, SNPs, Starch***

INTRODUCTION

Modern agriculture depends on the cultivation of a few highly productive crop species and varieties. It is obvious that crops suffer from domestication bottleneck characterized by genome-wide loss of genetic diversity (Tanksley and McCouch, 1997). Man driven selection for domestication traits and genetic drift through domestication bottlenecks are the two key processes that contribute towards shrinkage of genetic diversity (Eyre-Walker *et al*., 1998). For example, starch biosynthesis genes (*su1*, *SbeII* and *bt2*) experienced loss of genetic diversity in maize (Whitt *et al*., 2002). As a consequence, this reduction might limit the potential for crop improvement, but can be complemented by introgression of desirable traits from wild relatives.

Wild *Manihot* species are identified as gene reservoirs for enhancing agronomic important traits in cultivated cassava. For instance, wild *Manihot* have been targeted to improve micronutrient of cultivated cassava such as Iron and Zinc (Nassar *et al*., 2010). Interestingly, the inter-specific hybrids between cultivated and wild species show fairly regular meiosis and high fertility after backcrossing (Nassar, 2000). Nassar (2000) emphasized that ploidy level (2n = 36) is invariant in *Manhiot*, which contributes to regular meiosis and fertility of the hybrid. A phylogeography study using a single copy nuclear gene (*G3pdh*) marker showed that *Manihot esculenta* subsp. *flabellifolia* is a potential wild

progenitor of cultivated cassava (Olsen and Schaale, 1999). Genetic improvement of cultivated cassava partly depends on knowing the extent of genetic diversity in the crop and its wild relatives. The extent of genetic diversity can be measured by comparative nucleotide sequence analysis between cultivated cassava and its wild progenitors. These studies have been done for few economically important crops such as sunflower (Liu and Burke, 2006) and rice (Li *et al*., 2009).

Genetic diversity studies based on biochemical pathways may reflect different adaptive evolutionary processes that populations may have undergone (Tesfaye M. Baye *et al.,* 2009). *Isoamylase1* is one of the starch debranching enzymes that determine amylopectine architecture (Nakamura *et al*., 1996). Gene expression study on *Manihot esculenta Isoamylase1* (*Meisa1*) suggests that the gene is involved in the differentiation of fibrous root into storage roots (Dereje Beyene *et al*., 2010). Starch is generally composed of two fractions — amylopectin (70%–80%) and amylose (20%–30%) (Swinkels, 1985). *Isoamylase* mutation alters the number and spatial distribution of branches in amylopectin (James *et al*., 1995; Nakamura *et al*., 1996). On the other hand, N-glyceraldehyde-3-phosphate dehydrogenase (*G3pdh*) is a low copy, nuclear housekeeping gene involved in glycolysis, a central pathway for sugar metabolism (Strand *et al*., 1997). In our study, *G3pdh* included as a benchmark against which to compare variation in *Meisa1*, the locus is conserved in cultivated and variable in wild *Manihot* (Olsen and Schaale, 1999). Understanding the way in which sugar-metabolizing genes have evolved is one of biological research questions. In response, SNPs study on sugar metabolizing genes was designed. The aim of this study was to compare the pattern of nucleotide diversity between wild and cultivated cassava genotypes with respect to two genes: *Isoamylase1* (*Meisa1*, GU229751) and N-glyceraldehyde 3-phosphate dehydrogenase (*G3pdh*, AF136135).

MATERIALS AND METHODS

Plant material and DNA extraction

Twelve cultivated (*Manihot esculenta*) and 10 wild (*Manihot esculenta* subsp. *flabellifolia*) cassava genotypes were used (Table 1). Cultivated genotypes were collected from four geographic regions: North America (1), South America (1), Asia (1) and Africa (9) while the wild genotypes (10) were collected from South America. The germplasm were randomly selected from global cassava collections held in trust by the International Institute for Tropical Agriculture (IITA), Nigeria and International Centre for Tropical Agriculture (CIAT). Genomic DNA was extracted using a modified miniprep protocol (Dellaporta *et al.,* 1983).

Primer design and PCR

Nucleotide sequences of genes encoding *G3pdh* and *Isoamylase1* were obtained from the National Centre for Biotechnology Information (NCBI) (www.ncbi.nlm.nih.gov). These sequences were used for BLASTN similarity search against the cassava genome (Cassava Genome Project, 2009) to retrieve complete genomic sequence. The gene models and the exon-intron splice signal were predicted using Splign software (www.ncbi.-nlm.nih.gov/sutils/splign/splign.cgi; Fig. 1). Primers for PCR and sequencing were designed using Primers3 software (biotools.umassmed.-edu/bioapps/primer3_www.cgi) based on retrieved gene sequences and predicted gene models (Fig. 1; Table 2).

Table 1. ***Manihot*** **spp. populations used in this study.**

Cultivated Accessions		**Wild Accessions**	
Accession	***Origin***	**Accession**	**Origin**
BRA206	*S. America/Brazil*	4074	S. America/Brazil
USA7	*N. America/USA*	4107	S. America/Brazil
TAI1	*Asia/Thailand*	4127	S. America/Brazil
TME-05	Africa/Nigeria	4130	S. America/Brazil
Rwanagana	Africa/Rwanda	4131	S. America/Brazil
95/SE-00036	Africa/Uganda	4139	S. America/Brazil
I92/0067	Africa/Uganda	4140	S. America/Brazil
Bamunanika	Africa/Uganda	4145	S. America/Brazil
NASE-4	Africa/Uganda	4147	S. America/Brazil
Tsitakatromby	Africa/Madagascar	4149	S. America/Brazil
TMS30572	Africa/Kenya		
Liyayi	Africa/DR Congo		

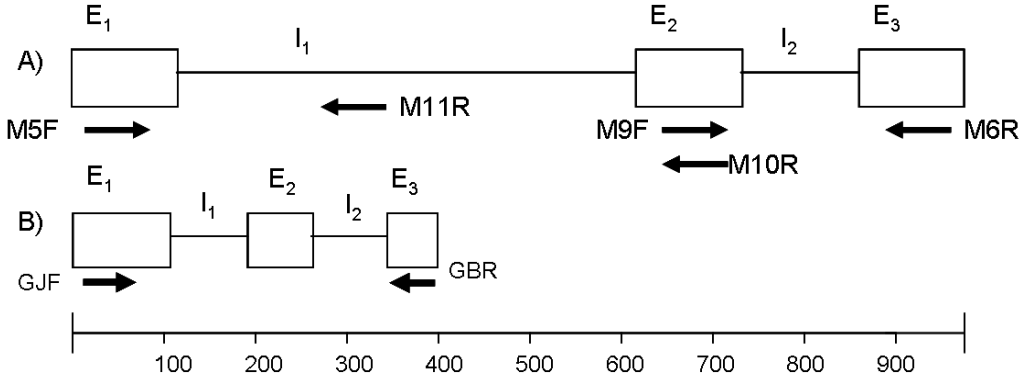

Fig. 1. The gene structures of the *Meisa1* and *G3pdh* in *Manihot*. The Exons designated by box and introns by line and numbered according to their positions. The arrows show the direction of the primes, tagged with name. Panel A for *Meisa1* gene that has three exons (E_1 63pb, E_2 74bp and E_3 75bp) and two introns (I_1 625bp and I_2 117bp) and panel B for *G3pdh* gene that has three exons (E_1 143bp, E_2 84bp and E_3 34bp) and two introns (I_1 101 bp and I_2 101 bp). The gene ruler in base pairs (bp).

Table 2. PCR cloning and sequencing primers, annealing temperature and $MgCl_2$ concentration for PCR amplification.

Gene/ID	Size	Name	Primer sequence	Ta	$MgCl_2$
Mesa1/	954bp	M5F[a]	5′-GGTTGTGGGAACACATTCAA	58°C	1.5 mM
GU229751		M6R[a]	5′-TCCCAAGAGGAGAACCAGTC		
		M11R	5′-AGCCAGCAGCAAATGGCAG		
		M9F	5′-TAGAAATGCATGTGGATGGC		
		M10R	5′-AGAAGCAAGATCAAAGCGG		
G3pdh/	363bp	GJF[a]	5′-TGTCGATCTCACTGTAAGGC	60°C	2.5 mM
AF136135		GBR[a]	5′-AAGCAATTCCAGCCTTGG		

[a] refers to primers used for PCR and sequencing; the rest of the primers were only used for sequencing; Ta = annealing temperature

PCR was conducted in 20 µl total reaction volume contained 1X PCR buffer (10 mM Tris-HCl (pH 9.0), 50 mM KCl, and 0.1 % Triton X-100), 1.5–2.5 mM $MgCl_2$, 0.3 µM of each primers, 200 µM of each dNTPs, 10–20 ng of genomic DNA template and 2 unit of GeneAmp® High Fidelity enzyme mix (Table 2). The amplification cycle consisted of initial denaturing (95°C, 5 min) and 30 cycles of denaturing (95°C, 30 sec), annealing (58°C–60°C, 30 sec) and extension (72°C, 1 min) and the third cycle final extension of 72°C for 10 minutes.

Sequencing

The PCR product of 5µl was prepared for sequencing using 10 units of Exonuclease I (Exo I) and 1 unit of Shrimp alkaline phosphatase (sAP) and incubated at 37°C for 1 hr followed by 85°C for 15 min as described by Werle *et al.* (1994). The resulting products were sequenced at Biosciences eastern and central Africa (BecA), Nairobi, Kenya platform using ABI BigDye terminator Cycling Sequencing chemistry according to manufacturer's recommendations and the signals visualized on ABI3730xl DNA analyzer (Applied Biosystem). The primers used during the sequencing are presented in Table 2. The primers were spaced at an interval of 400 bp length along *Meisa1* gene (Fig. 1). The sequencing was done on both strands with a minimum of three-fold coverage.

Sequence analysis

The trace sequences were imported into SeqMan™II (DNASTAR) and consensus sequences generated. The quality of trace files was checked by importing all genotype sequences into SeqMan™II (DNASTAR) and each potential SNPs site visually inspected. The sequences were aligned using ClustalW integrated into MEGA version 4 (Tamura *et al.*, 2007) and manually edited by comparison with trace chromatograms

generated by SeqMan™II. DnaSP v5.10 (Librando and Rozas, 2009) was used to calculate summary statistics: synonymous and non-synonymous substitutions, Tajima's D, estimates of single nucleotide polymorphism (π and θ, calculated on per site), and population subdivision (F_{ST} between wild and cultivated cassava excluding InDels). MEGA version 4 was used to generate files of variable sites, coding region alignment and amino acids prediction.

RESULTS

Nucleotide sequence polymorphism

Meisa1 length was 954 bp in both populations and 27 and 9 SNPs were identified in cultivated and wild species, respectively, translating to one SNP every 35bp in cultivated and one SNP every 106bp in wild. The structural-functional distributions of the SNPs were 6 exonic (1/35) and 21 in intronic (1/35) in cultivated, and 2 exonic (1/106) and 7 intronic (1/106) regions in the wild. Interestingly, Intron2 and Exon3 were conserved in *Meisa1* of both cultivated wild accessions (Table 3). For *G3pdh,* less SNPs were identified in the cultivated cassava than in the wild (Table 4). All observed SNPs in *G3pdh* were in the intronic region in the cultivated cassava while 4 of the 9 SNPs in the wild occurred in the exonic region. Transverstions were the most frequent point mutation in cultivated genotypes, while both transition and transversion mutations were in equal proportions in the wild (Table 3). A single InDel was identified at intron1 of *G3pdh* from the wild cassava.

Table 3. Nucleotide polymorphism and functional distribution in cultivated and wild cassava. Insertion and Deletion (InDel), Transition (T) and Transversion (TR) mutation types are indicated. Allelic positions are given in bracket.

Gene	Site and length	Allele (position)		Mutation Cv/Wt
		Cultivated	Wild	
Meisa1/	Exon1	A↔G (34)	------	T/--
954bp	63 bp	C↔T (46)	C↔T (46)	T/T
	Intron1	T↔C (82)	A↔T (76)	T/TR
	625 bp	A↔G (100)	------	TR
		A↔G (114)	------	TR
		A↔C (134)	------	T/--
		G↔T (167)	G↔T (167)	TR/TR
		T↔C (206)	T↔C (283)	T/T
		G↔T (219)	T↔C (572)	TR/T
		A↔T (278)	------	TR/--
		G↔T (307)	------	TR/--
		A↔G (347)	------	T/--
		G↔T (390)	------	TR/--
		G↔C (411)	------	TR/--
		A↔G (477)	------	T/--
		A↔C (563)	------	TR/--
		A↔T (578)	A↔T (578)	TR/TR
		G↔T (609)	G↔T (651)	TR/TR
		G↔T (625)	A↔C (752)	TR/TR
	Exon2	A↔C (700)	-----	TR/--
	74 bp	A↔G (713)	A↔T (427)	T/TR
	Intron2	A↔T (786)	------	TR/--
	117 bp	A↔C (787)	------	TR/--
		A↔G (845)	------	T/--
		G↔C (867)	------	TR/--
	Exon3	A↔T (883)	------	TR /--
	75 bp	T↔C (927)	------	T/--
G3pdh	Exon1	------	T↔C (41)	--/T
363bp	63	------	G↔A (43)	--/T
		------	T↔C (56)	--/T
	Intron1	A↔C (131)	G↔T (99)	TR/TR
	101 bp	G↔C (135)	G↔A (118)	TR/T
		-------	-- ↔T (147)	--/InDel
	Exon2 (84 bp)	------	G↔A (233)	--/T
	Intron2	G↔A (257)	G↔A (259)	T/T
	70bp	G↔C (297)	G↔T (294)	TR/TR
		---------	G↔C (299)	--/TR

The pooled *Meisa1* and *G3pdh* nucleotide polymorphisms per site (θw) were 0.081 and 0.005 for cultivated and wild cassava genotypes, respectively. The θw was equally distributed in different segments and pooled regions of *Meisa1* in wild (θw, 0.009) and cultivated (θw, 0.004) (Table 4). The θw of *G3pdh* identified in cultivated (0.004) and wild (0.011) showed that the nucleotide polymorphism per site equally distributed at different segments of wild *G3pdh* (0.011). Similarly, the nucleotide diversity per site (π) was estimated. The pooled nucleotide diversity per site was higher in cultivated (π, 0.012) than in wild (π, 0.005) (Table 4). Nucleotide diversity per site of cultivated *Meisa1* was 0.014, but in equal proportions in exon and intron (π, 0.0137). *G3pdh* nucleotide sequence diversity per site was 0.0053 and 0.012 in cultivated and wild genotypes, respectively. F_{ST} statistic averaged 0.268, suggesting that the wild and cultivated cassava gene pools are moderately differentiated.

Neutrality test statistics

Tajima's D-statistic was computed as described by Tajima (1989) to determine whether or not the observed DNA polymorphism has had a significant departure from neutral mutation hypothesis (Kimura, 1983). The test is based on differences between independent estimates of nucleotide polymorphism (θ), which is equal to 4Neμ; where, Ne is the effective population size, and μ is the mutation rate per generation. The neutrality tests were determined across loci, functional region (exons and introns) and also pooled across cultivated and wild genotypes (Table 4). The pooled statistics in cultivated and wild genotypes were 2.21 and −0.289, respectively, showing that cultivated cassava may be undergoing balancing/diversifying selection while the wild cassava is subjected to purifying selection or may have experienced a recent genetic bottleneck.

Coding SNPs (cSNPs) polymorphism and amino acid substitution types

Out of the six cSNPs identified in the cultivated *Meisal*, four were non-synonymous and two were synonymous substitutions (Table 5). One of the non-synonymous mutations resulted in alteration of polar amino acid type (Serine, *Ser*) with a non-polar type (Leucine, *Leu*). The BLASTp similarity search revealed that the substitution of *Leu* by *Ser* is rare; the position is preferably conserved by *Leu*. The synonymous substitutions (π_{Syn}, 0.02) were more than one-fold prevalent than non-synonynmous substitutions (π_{nonsyn}, 0.012). On the other hand, the wild *Meisa1* coding region was relatively less conserved than cultivated (Table 4). The synonymous substitution (π_{Syn}, 0.0052) was two-fold more prevalent than non-synonymous mutation (π_{nonsyn}, 0.0025) in wild *Meisa1*.

Table 4. Summary of Nucleotide polymorphisms.

Gene	Region	Nucleotide polymorphism information					Tajima's D
		N^a	S^a	f	π	θw	
Cultivated							
Meisa1	*Meisa1*	954	27	35	0.0141*	0.009*	2.27
	Exons	212	6	35	0.0137	0.009	1.77
	Introns	742	21	35	0.0137	0.009	2.04
G3pdh	*G3pdh*	363	4	90	0.0053*	0.004*	1.47
Pooled	Total	1,317	31	42	0.012	0.081	2.21
Wild							
Meisa1	*Meisa1*	954	9	106	0.0027*	0.004*	−1.210
	Exons	212	2	106	0.0032	0.004	−0.450
	Introns	742	7	106	0.0023	0.004	−1.360
G3pdh	*G3pdh*	317	9	35	0.012*	0.011*.	0.571
	Exons	149	4	37	0.010	0.010	0.078
	Introns	168	5	34	0.013	0.011	0.880
Pooled	Total	1,271	18	70.6	0.005	0.005	−0.289
Average*	Cv				0.0097	0.0065	
	Wt				0.0074	0.0075	

* Designates the averaged values where Cv cultivated and Wt wild genotypes; N^a=total number of sequences, S^a= total number of polymorphic sites, f=frequency of SNPs (N^a/S^a), π=nucleotide diversity per site between two sequences, θw=nucleotide polymorphism per site between two sequences and Tajima's D-statistic test.

Table 5. Nucleotide substitution types, amino acid information and nucleotide diversity (π_{syn} for synonymous (Syn) and π_{NonSyn} for non-synonymous (NonSyn) substitutions) in the cSNPs of *Meisa1* and *G3pdh*.

Genotype/ gene/Exon	cSNP Position	Genetic Codon	Amino acid	Class/ Type of amino acid	Substitution Types	π_{Syn}	π_{nonSyn}
Cultivated (*Meisa1*)						0.02	0.012
Exon1	34	**A**TT/**G**TT	*Ile/Val*	Non-polar	NonSyn		
Exon1	46	**T**TT/**C**TT	*Phe/ Leu*	Non-polar	NonSyn		
Exon2	700	GT**C**/GT**A**	*Val/Val*	Non-polar	Syn		
Exon2	713	**A**TG/**G**TG	*Met/Val*	Non-polar	NonSyn		
Exon3	883	CT**T**/CT**A**	*Leu*	Non-polar	Syn		
Exon3	927	T**C**G/T**T**G	*Ser/Leu*	Polar/Non-polar	NonSyn		
Wild (*Meisa1*)						0.0052	0.0025
Exon1	46	**T**TT/**C**TT	*Phe/Leu*	Non-polar	NonSyn		
Exon2	752	**A**GG/**C**GG	*Arg/Arg*	Polar	Syn		
Wild (*G3pdh*)						0.039	0.002
Exon1	41	TA**T**/TA**C**	*His/His*	Basic	Syn		
Exon1	43	G**A**G/G**G**G	*Glu/Gly*	Basic/Neutral	NonSyn		
Exon1	56	AT**C**/AT**T**	*Leu/Leu*	Neutral	Syn		
Exon2	223	GT**A**/GT**G**	*Val/Val*	Neutral	Syn		

Note: *Glu*= Glutamic acid, *Gly*= Glycine, *His*= Histidine, *Ile*= Isoleucine, *Leu*=Leucine, *Met*=Methionine, *Phe*= Phenylalanine, *Ser*= Serine, *Val*= Valine.

The exonic regions of *G3pdh* in cultivated genotypes were strongly conserved while the wild genotype experienced four mutations with three synonymous and one non-synonymous substitution (Table 4). Synonymous substitutions ($\pi_{Syn,}$ 0.039) were nineteen fold more prevalent than non-synonynmous substitutions ($\pi_{nonSyn,}$ 0.002).

DISCUSSION

The results of this study provide the most comprehensive, comparative nucleotide sequence analysis of two genes in wild (*M. esculenta* subsp. *flabellifolia*) and cultivated (*M. esculelenta*) cassava to date. The analysis was conducted on two sugar hydrolyzing genes *Meisa1* (GU229751) and *G3pdh* (AF136135); the pooled length was cultivated 1,317 bp and wild 1,271 bp and a total of 49 SNPs and one InDel (*G3pdh* of wild) were identified. The density of SNPs in cultivated cassava was higher than that in wild *i.e.*, 1/42 bp and 1/70 bp, respectively. The nucleotide substitution in cultivated cassava had an apparent mutation bias towards transversion than transition whereas no mutation bias was noted in wild. The SNPs discovered in cultivated *Meisa1* are equivalent to combined 26 polymorphic sites discovered in nine candidate genes controlling agronomically important traits in cultivated cassava (Kawuki, *et al*., 2009). The finding suggests that *Meisa1* is a useful marker in intra and inter cultivar diversity study and also developing array genotyping (Jaccoud *et al*., 2001; Comai, *et al*., 2004). The estimated SNPs frequency for *Meisa1* was 1/35 bp and 1/106 bp for cultivated and wild cassava, respectively. *G3pdh* was found to be conserved in cultivated than in wild cassava genotypes. The exons of cultivated cassava were conserved in cultivated than in wild type, whereas introns are highly variable (Olsen and Schaal, 1999; Liu and Burke, 2006). *G3pdh* is highly conserved and low copy; the gene is marker of choice for phylogeny study (Strand *et al*., 1997; Olsen and Schaal, 1999). The study identified *G3pdh* locus has little significance in intra and inter cultivar diversity survey.

The nucleotide polymorphism per site (θw) is roughly proportional to heterozygosity (Misawa and Tajima, 1997). The expected nucleotide differences between two randomly selected pairs of alleles can be estimated using θw. Thus, θw estimates suggest that two randomly chosen wild or cultivated cassava sequences vary on average 1 in ~133 or ~153 nucleotides (i.e., $1/0.0075 \approx 133$ and $1/0.0065 \approx 153$; Table 4), respectively. The finding is similar with higher diversity noted in wild sunflower than in cultivated sunflower (Liu and Burke, 2006). In terms of gene specific comparison of θw estimates two randomly chosen wild or cultivated cassava sequences vary 1 in ~250 or ~111 nucleotides and 1 in ~90 or ~250 nucleotides for *Meisa1* and *G3pdh*,

respectively. This suggests that in cultivated cassava, the diversity of *Meisa1* is higher and that of *G3pdh* lower than in the wild variety. The result for *Meisa1* is surprising, given experience with other crop plants (Whitt *et al*., 2002). The results suggest that cultivated cassava has more adaptability potential with respect to *Meisa1* locus. In terms of the extent of divergence between subsamples wild and cultivated, averaged *Fst* value was 0.268, indicating that the wild and cultivated cassava gene pools are moderately differentiated. This suggests that 26.8 % of the total genetic variation is distributed among wild and cultivated populations while 73.2% of the variation within the subpopulations.

The coding regions were strictly analyzed to find out possible mutations that can change the amino acid types. The amino acid substitutions (*Meisa1*, *Ser*/*Leu*) that altered the property and/or size of amino acids may be associated with alternation of starch traits. Previous study revealed that *Meisa1* is involved in differentiation of fibrous root into storage root of cultivated cassava (Dereje Beyene *et al*., 2010). This suggests that the gene engages in determining root quality traits. cDNA-based SNPs could be of great advantage for crop improvement. For instance, sorghum accessions with waxy starch (BTxArg1 and Tx2907) have a point mutation at positions 268 which codes for *His* (basic amino acid) instead of *Glu* (polar amino acid), *Glu* conserved in non-waxy sorghums (Hamblin *et al*., 2007). Similarly, in sugary maize, *su1-Ref* gene lost its enzymatic activity due to substitution of the hydrophobic residue *Tyr* (wild) with charged residue *Arg* (mutant) at position 578 (Dinges *et al*., 2001; Whitt *et al*., 2002). The mutation reduced the abundance of *Sugary1* both at the transcript and protein levels (Beatty *et al*., 1999). Moreover, mutations that change the amino acid sequences of a protein are associated with disease traits in humans (Bao and Cui, 2005). The BLASTp similarity search showed that non-synonymous substitution of *Leu* by *Ser* is rare. In conclusion, the finding provides an insight to design genetic association study between root quality traits and the corresponding cSNPs changed the amino acid type. Generally, the synonymous substitutions are more prevalent than non-synonymous substitutions. These suggest that purified selection prevail in coding regions; non-synonymous amino acid replacement may implicate lower fitness.

Tajima's D-statistic (Tajima, 1989) was used to test for deviation from neutral hypothesis theory (Kimura, 1983). Functional regions analysis might show differences in selection pressure types; this is because of differences in neutral mutation rates as well as the stochastic variation associated with evolutionary processes. Thus, species or population level analysis of the entire gene structure gives the general picture. In this case, *Meisa1* is experiencing different selection pressures in cultivated (D = 2.27) and wild (D = −1.21) translating to balanced/diversifying and purifying selection pressures, respectively. *G3pdh* is under diversifying selection in both cultivated and wild cassava genotypes.

CONCLUSIONS

Taken together, comparison of variations within *G3pdh*, a housekeeping gene involved in sugar metabolism in wild and cultivated cassava, indicated greater variation within wild cassava than cultivated cassava. This could be attributable to a lack of strong selection pressure, either natural or artificial, on the *G3pdh* locus, and a genetic bottleneck, during cassava domestication. *Meisa1* is more genetically diverse in the sample of cultivated cassava than in our sample of wild cassava. This suggests a strong effect of diversifying selection, either natural or artificial, on this locus as cassava has moved into different environments, and has been selected for different organoleptic preferences and purposes. The *Meisa1* could be used in intra and inter cultivar diversity study and developing array genotyping. On the other hand, the *G3pdh* is more applicable for taxonomic study. The functional effect of the cSNPs (*Ser*/*Leu*) identified in *Meisa1* that change the amino acid may be ascertained using candidate gene-based association study.

ACKNOWLEDGEMENTS

This work was supported by Swedish International Cooperation Agency/Department for Research Cooperation (Sida/SAREC) under BIO-EARN program. We are grateful for utilizing the ILRI/BecA laboratory platform in Nairobi, Kenya.

REFERENCES

1. Bao, L. and Cui, Y. (2005). Prediction of the phenotypic effects of non-synonymous single nucleotide polymorphisms using structural and evolutionary information. *Bioinformatics* **21**:2185–2190.

2. Beatty, M.K., Rahman A., Cao H., Woodman W., Lee M., Myers A.M. and James, M.G. (1999). Purification and molecular genetic characterisation of ZPU1, a pullulanase-type debranching enzyme from maize. *Plant Physiol.* **119**:255–266.
3. Cassava Genome Project (2009). Web addres at: http://www.phytozome.net/cassava.
4. Comai, L., Young, K., Till, B.J., Reynolds, S.H., Greene, E.A., Codomo, C.A., Enns, L.C., Johnson, J.E., Burtner, C., Odden, A.R. and Henikoff, S. (2004). Efficient discovery of DNA polymorphisms in natural populations by Ecotilling. *Plant J.* **37**:778–786.
5. Dellaporta, S.J., Wood, J. and Hicks, J.B. (1983). A plant DNA mini preparation. Verstion II. *Plant Mol. Biol. Rep.* **1**:19–21.
6. Dereje Beyene, Baguma, Y., Mukasa S.B., Sun. C. and Jansson, C. (2010). Characterization and role of *Isoamylase1* (*Meisa1*) gene in cassava. *African Crop Science Journal* **18**:1–8.
7. Dinges, J.R., Colleoni, C., Myers, A.M. and James, M.G. (2001). Molecular structure of three mutations at the Maize *sugary1* locus and their allele-specific phenotypic effects. *Plant Physiol.* **125**:1406–1418.
8. Eyre-Walker, A., Gaut, R.L., Hilton, H., Feldman, D.L. and Gaut, B.S. (1998). Investigation of the bottleneck leading to the domestication of maize. *Proc. Nat. Acad. Sci. USA* **95**:4441–4446.
9. Hamblin, M.T., Fernandez, M.G.S., Tuinstra, M.R., Rooney, W.L. and Kresovich, S. (2007). Sequence Variation at candidate loci in the starch metabolism pathway in sorghum: prospects for linkage disequilibrium mapping. *Crop Sci* **47**:S125–S134.
10. Jaccoud, D., Peng, K., Feinstein, D. and Kilian, A. (2001). Diversity arrays: A solid state technology for sequence information independent genotype. *Nucleic Acids Res.* **29(4):**e25.
11. James, M.G., Robertson, D.S. and Myers, A.M. (1995). Characterization of maize gene *sugary1*, a determinant of starch composition kernels. *The Plant Cell* **7**:417–429.
12. Kawuki, R.S., Ferguson, M., Labuschagne, M., Herselman, L. and Kim, D.J. (2009). Identification, characterization and application of single nucleotide polymorphism for diversity assessment in cassava (*Manihot esculenta*, Crantz). *Mol. Breeding* **23**:669–684.
13. Kimura, M. (1983). *The Neutral Theory of Molecular Evolution*. Cambridge University Press, London.
14. Li, X., Tan, L., Zhu, Z., Huang, H., Liu, Y., Hu. S. and Sun, C. (2009). Patterns of nucleotide diversity in wild and cultivated rice. *Plant Syst. Evol.* **218**:97–106.
15. Librando, P. and Rozas, J. (2009). DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25**:1451–1452.
16. Liu, A. and Burke, J. (2006). Patterns of nucleotide diversity in wild and cultivated Sunflower. *Genetics* **173**:321–330.
17. Misawa, K. and Tajima, F. (1997). Estimation of the amount of DNA polymorphism when the neutral mutation rate varies among sites. *Genetics* **147**:1959–1964.
18. Nakamura, Y., Umemoto, T., Takahata, Y., Komae, K., Amano, E. and Satoh, H. (1996). Changes in structure of starch and enzyme activities affected by sugary mutations in developing rice endosperm. Possible role of starch debranching enzyme in amylopectin biosynthesis. *Plant Physiol***. 97**:491–498.
19. Nassar, N.M.A. (2000). Cytogentics and evolution of cassava (*Manihot esculenta* Crantz). *Genetics and Molecular Biology* **23**:1003–1014.
20. Nassar, N.M.A., Barbosa, I.S., Haridassan, M., Ortiz, R. and Gomes, P.T. (2010). Cassava (*Manihot esculenta* Crantz) genetic resources: a case of high iron and zinc. *Genet. Resour. Crop Evol.* **57**:287–291.
21. Olsen K.M., Schaal B (1999) Evidence on the geographic origin of cassava: Phylogeography of *Manihot esculenta*. *Proc. Natl. Acad. Sci. USA* **96**:5586–5591.
22. Strand, A.E., Leebens-Mack, J. and Milligan, B.G. (1997). Nuclear DNA -based markers for plant evolutionary biology. *Molecular Ecology* **6**:113–118.
23. Swinkels, J.J.M. (1985). Composition and properties of commercial native starches. *Starch/Stärke* **37**:1–5.
24. Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **125**:585–595.
25. Tamura, K., Dudley, J., Nei, M. and Kumar, S. (2007). MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* **24**:1596–1599.
26. Tanksley, S.D. and McCouch, S.R. (1997). Seed banks and molecular maps: unlocking genetic potential from the wild. *Science* **277**:1063–1066.
27. Tesfaye M. Baye, Wilke, R.A. and Olivier, M. (2009). Genomic and geographic distribution of private SNPs and pathways in human populations. *Pre Med.* **6(6)**:623–641.
28. Werle, E., Schneider, C., Renner, M., Völker, M. and Fiehn, W. (1994). Convenient single-step, one tube purification of PCR products for direct sequencing. *Nucleic Acid Research* **22**: 4354–4355.
29. Whitt, S.R., Wilson, L.M., Tenaillon, M.I., Gaut, B.S. and Buckler IV, E.S. (2002). Genetic diversity and selection in the maize starch pathway. *Proc Natl Acad Sci USA* **99**:12959–12962.