Polymorphism of starch pathway genes in cassava

L.M. Vasconcelos[1], A.C. Brito[2], C.D. Carmo[1] and E.J. Oliveira[2]

[1]Centro de Ciências Agrárias Ambientais e Biológicas,
Universidade Federal do Recôncavo da Bahia, Cruz das Almas, BA, Brasil
[2]Núcleo de Recursos Genéticos e Desenvolvimento de Variedades,
Embrapa Mandioca e Fruticultura, Cruz das Almas, BA, Brasil

Corresponding author: E.J. Oliveira
E-mail: eder.oliveira@embrapa.br

ABSTRACT. The distribution and frequency of single nucleotide polymorphisms (SNPs) can help to understand changes associated with characteristics of interest. We aimed to evaluate nucleotide diversity in six genes involved in starch biosynthesis in cassava using a panel of 96 unrelated accessions. The genes were sequenced, aligned, and used to obtain values for nucleotide diversity (π), segregating sites (θ), Tajima's D test, and neighbor-joining (NJ) clustering. On average, one SNP per 147 and 171 bp was identified in exon and intron regions, respectively. Thirteen heterozygous loci were found. Three of seven SNPs in the exon region resulted in non-synonymous replacement or four synonymous substitutions. However, no associations were noted between SNPs and root dry-matter content. The parameter π ranged from 0.0001 (granule bound starch synthase I) to 0.0033 (α-amylase), averaging 0.0011, while θ ranged from 0.00014 (starch branching enzyme) to 0.00584 (starch synthase I), averaging 0.002353. The θ

diversity value was typically double that of the π. Results of the D test did not suggest any evidence of deviance of neutrality in these genes. Among the evaluated accession, 82/96 were clustered using the NJ method but without a clear separation of the root dry-matter content, root pulp coloration, and classification of the cyanogenic compound content. High variation in genes of the starch biosynthetic pathway can be used to identify associations with the functional properties of starch for the use of polymorphisms for selection purposes.

Key words: SNP; Starch biosynthesis; Breeding; Dry matter content; *Manihot esculenta* Crantz

INTRODUCTION

Cassava (*Manihot esculenta* Crantz) is a species that produces tuberous roots and is rich in carbohydrates (Narina et al., 2011). It is, therefore, considered an important source of energy for more than 800 million people, mainly in the tropics (FAO, 2015). Cassava is highly adaptable to diverse soil and climatic conditions, especially in regions with low rainfall and low-fertility soils (Herrera-Campo et al., 2011), making it an essential crop for communities living in such regions. Besides the roots, which can be consumed fresh, other cassava products, such as starch, can be used for applications in various industries (Mufumbo et al., 2011).

Cassava starch offers some advantages to the food industry compared to starches from other sources, such as cereals. The differential physicochemical properties of cassava starch, when prepared in an aqueous solution, result in very clear and high-viscosity pastes with a low gelatinization temperature and lower retrogradation compared with other sources of starch. Moreover, cassava starch is preferred by industries for some processed foods that require a mild flavor (Mufumbo et al., 2011).

With the growing demand for starch, linked to the significant increase in the world population, the development of cassava cultivars with better starch yield per unit area has become a priority (Saithong et al., 2013). Conversely, the genetic breeding of cassava depends, among other factors, on improving our understanding of the genetic diversity between germplasm and the possible adaptive and evolutionary mechanisms involved. Our understanding of this diversity can be verified by comparing the nucleotide sequences of improved varieties with those of their wild progenitors. Such studies are scarce in cassava, but have enormous potential to contribute to an understanding of the genetic relationships involved in starch biosynthesis (Beyene et al., 2013).

The physicochemical properties linked to starch quality and content in cassava are correlated with genes involved in starch biosynthesis in distinct phases and stages of root development (Saithong et al., 2013). In general, the synthesis of α-1,4-glucan involves three major steps in chloroplasts and amyloplasts, which are related to the supply of glucose 6-phosphate in plastids, ADP-glucose synthesis from glucopyranosyl phosphate (G1P), and starch synthesis from ADP-glucose. The key enzymes in the starch biosynthetic pathway are ADP-glucose pyrophosphorylase and starch synthases (SS). In the formation of the starch molecule, SS adds glucose units (G1P) to the end of non-reduced α-1,4-glucan linkages (Saithong et al., 2013). At least nine SS isoforms are responsible for amylopectin synthesis. The branching enzymes are also involved in amylopectin synthesis. In contrast, granule-bound

starch synthase (GBSS), found within the granules, is a key enzyme in amylose synthesis (Saithong et al., 2013).

Several genes interact and influence the quantity and composition of starch granules (Lee et al., 2009). Changes in the activity of enzymes involved in starch biosynthesis can trigger the formation of starch granules with different characteristics. Such changes can occur naturally or can be induced by mutations during attempts to meet specific market demands (Franco et al., 2001).

Studies on the diversity of genes involved in important biochemical pathways may help to improve understanding of the different evolutionary processes that populations have undergone over time, in addition to identifying the genetic basis of phenotypic diversity within species and the origin of phenotypic divergence between species (O'Connor and Mundy, 2009). Several approaches have been applied with regard to phenotypes of high economic interest and with single nucleotide polymorphisms (SNPs).

SNP markers have been used in several studies of genetic diversity in cassava accessions (Lopez et al., 2005; Kawuki et al., 2009; Oliveira et al., 2014a). Conversely, even with advances in molecular techniques, there is limited understanding of the evolution of complex traits and the genes involved in several metabolic pathways. Identifying SNPs and understanding how their variation modulates phenotypic expression can lead to important advances in genetic improvement of cassava, directly impacting food security, sustainability, and competitiveness of the crop by allowing the exploration and use of different starches in the industry.

The aim of this study was to compare the pattern of nucleotide diversity in six candidate genes involved in cassava starch biosynthesis, starch branching enzyme (Me-*Be*), granule bound starch synthase I (Me-*GBSSI*), starch synthase I (Me-*SSI*), sucrose transporter1 (Me-*SUT1*), α-amylase (Me-*Amy2*), and glucan water dikinase (Me-*GWD*) in unrelated cassava accessions.

MATERIAL AND METHODS

Plant material and DNA extraction

Ninety-six cassava accessions from the Cassava Germplasm Bank at Embrapa Mandioca e Fruticultura (Cruz das Almas, Bahia, Brazil) were used. Total DNA was extracted using the cetyltrimethylammonium bromide protocol, as described by Doyle and Doyle (1987). In order to assess DNA integrity and standardize its concentration, DNA quantification was carried out in 1% agarose gel followed by visual comparison with various DNA concentrations of phage λ.

Primer design

Primers specific for six genes involved in the metabolic pathway of cassava starch were designed, which were properly characterized and deposited in the National Center for Biotechnology Information (NCBI) (Table 1). The criteria used to select primers were: i) annealing temperature ranging from 55° to 60°C, ii) a maximum difference of 3°C in the annealing temperature between primer pairs, and iii) GC content ranging from 40 to 60%.

Table 1. List of primers used for the amplification of genes involved in the metabolic pathway of cassava starch.

Primer	ID sequence	Gene	Primer sequence[1]	AT[2] (°C)	Fragment (bp)
Me-*Be*	HM046986	Branching enzyme	F: TGTGGAGTCGAAGCATGAA R: GCTGAGATTGCAAGACCTT	60	336
Me-*GWD*	JN618458	Glucan water dikinase	F: TCCACACGACCAATCCATC R: ACCTTAGTCGTCCTTCCAA	62	289
Me-*GBSSI*	JF708949	Granule bound starch synthase I	F: AGCTGAAGTTGGTCCCTGG R: CATAGCGGGGAGACACTGT	60	197
Me-*SSI*	EF667960	Starch synthase I	F: TGAAGGCAGAGAGGAAGCT R: ATCACCAAAAGCACCACGG	64	537
Me-*SUT1*	DQ138374	Sucrose transporter 1	F: TTGCTGTGTTTGTGGTCGG R: TCCCGTCTTCATCCTCATCG	60	381
Me-*Amy2*	DQ017830	Alpha amylase	F: ATGTCTGGCTTCCTCCTCC R: GTTCCACCTTCAAAGATGCA	61	354

[1]Primers F: forward, and R: reverse; [2]AT: annealing temperature.

DNA amplification and sequencing

PCR was performed in a final volume of 50 μL, containing 10 ng DNA, 1X buffer (200 mM Tris-HCl pH 8.4, 500 mM KCl), 2.0 mM MgCl$_2$, 1 U Platinum High Fidelity Taq DNA polymerase (Invitrogen, Carlsbad, CA, USA), 2.0 μM each dNTP (Invitrogen), 0.3 μM each primer (Qiagen, Hilden, Germany), and ultra-pure sterile water (Merck Millipore, Watford, UK) to make up the final volume.

Amplification was performed with a cycle of 5 min at 94°C; followed by 30 cycles of 40 s at 94°C; 60 s annealing temperatures ranging from 60° to 64°C, depending on the primer; 60 s extension at 72°C; and a final extension of 5 min at 72°C in the 96-well thermocycler model Veriti® (Applied Biosystems, Foster City, CA, USA). Amplified products were separated by 2.5% agarose gel electrophoresis at 100 V in 0.5X Tris-acetate EDTA buffer for approximately 4 h and then stained with ethidium bromide solution (1.5 μL/100 mL). The amplified fragments were visualized under UV light and photographed on a Kodak Gel Logic-220 (Carestream Health, Rochester, NY, USA). PCR products were purified by QIAquick PCR Purification Kit (Qiagen) and then subjected to Sanger sequencing on ABI 3730XL equipment (Applied Biosystems). After this, sequencing searches were performed using the NCBI database to compare the level and identity of the cassava gene sequences with those of other species.

Analysis of sequencing data

Sequencing data were edited to obtain high-quality DNA sequences. Quality filters were used to trim the bases, with a criterion of Q-score ≥ 5 and a threshold of 50% for heterozygous peaks. Then, contigs and their consensus sequences were formed by aligning sequences using the SeqManPro software (DNASTAR, Inc., Madison, WI, USA). The consensus sequence was searched for mutations, observing possible joint SNPs and inspecting the quality of the electropherogram peaks across alignments.

After obtaining the consensus sequences, a BLASTn search was carried out using the Phytozome database version 11.0 in order to identify similar sequences of each gene and to compare regions containing introns and exons (identified *in silico*). The average number of

nucleotide differences between the two randomly sampled sequences (nucleotide diversity, π) was calculated according to the method described by Nei (1987):

$$\pi = n/(n-1)\sum_{ij} p_i p_j \pi_{ij}$$ (Equation 1)

where n is the number of analyzed sequences; p_i and p_j are the frequencies of the i and j sequences, respectively, in the sample; and p_{ij} is the proportion of nucleotide differences between sequences i and j. The θ diversity described by Watterson (1975) was estimated using the formula, $\theta = k / a_n$, where k is the number of segregating sites (SNPs) in the sequences, and

$$a_n = \sum_{i=1}^{n-1}\frac{1}{i}$$ (Equation 2)

Tajima's D test (Tajima, 1989) was also performed as a neutrality test. All analyses were performed using the Mega 6.0 software (Tamura et al., 2013). Nucleotide substitutions from the initiation codon were also predicted to verify amino acid changes.

Phylogenetic analyses

A multiple alignment of the sequences was carried out using the ClustalW program, and the sequenced genes were edited manually using the BioEdit software (Hall, 1999). The genetic relationships among accessions were calculated using a distance matrix obtained by the nucleotide substitution model Kimura 2-parameter (Kimura, 1980). Afterward, cassava accessions were clustered based on the neighbor-joining method. The consistency of the clustering patterns was evaluated by 1000 bootstrap repetitions. Phylogenetic trees were edited in the Figtree v.1.4.2 software (Rambaut, 2014).

Associations between molecular polymorphism and root dry-matter content

The sequences of 96 individuals for each gene were filtered for quality, in which SNP ³10% of missing data and minimum allele frequency <0.01 were discarded. The mixed linear model implemented in the GAPIT R package (Lipka et al., 2012) was used to test the association between the root dry-matter content and molecular polymorphism of cassava genes involved in starch biosynthesis, by using the equation $Y = X\beta + Z\mu + e$, where Y is the vector of genotypic observations, β is the vector of fixed effects (effects other than markers and population structure), μ is the vector of random genetic effects, e is the random vector residual effects, and X and Z are the incidence matrices for β and μ, respectively.

Population structure was estimated using principal component analysis and relatedness among individuals was calculated using the kinship matrix according to the method described by VanRaden (2008). Bonferroni correction was used to control for multiple testing ($P < 0.05$), and the P value obtained for each SNP was transformed into a logarithmic scale.

RESULTS AND DISCUSSION

Sequences and polymorphism analysis

The size of the fragments analyzed varied from 171 to 350 bp, and DNA sequence

analysis identified several polymorphisms within genes of the starch metabolic pathway in cassava (Figure 1). The amplified fragments were larger than the analyzed fragments, because after base trimming, all low-quality bases were excluded.

Eleven SNPs were identified in six investigated genes, in both intron and exon regions (Table 2 and Figure 1). On average, one SNP per 147 bp in coding regions and one SNP per 171 bp in non-coding regions were identified in these genes. In general, one SNP was identified for every 100-500 bp. Previously reported SNP frequencies in cassava were as follows: one SNP per 121 bp (Kawuki et al., 2009), one SNP per 509 bp (Lopez et al., 2005), and one SNP per 350 bp (Pootakham et al., 2014). One explanation for the difference in SNP frequency is the nature of the genetic material used for comparison. The results of the present study were very similar to those obtained by Kawuki et al. (2009), in which 74 germplasm accessions were used to compare polymorphisms at the DNA level. This reinforces the hypothesis that high levels of polymorphisms can be identified when using materials with wide genetic diversity, such as the 96 unrelated germplasm accessions used in the present study.
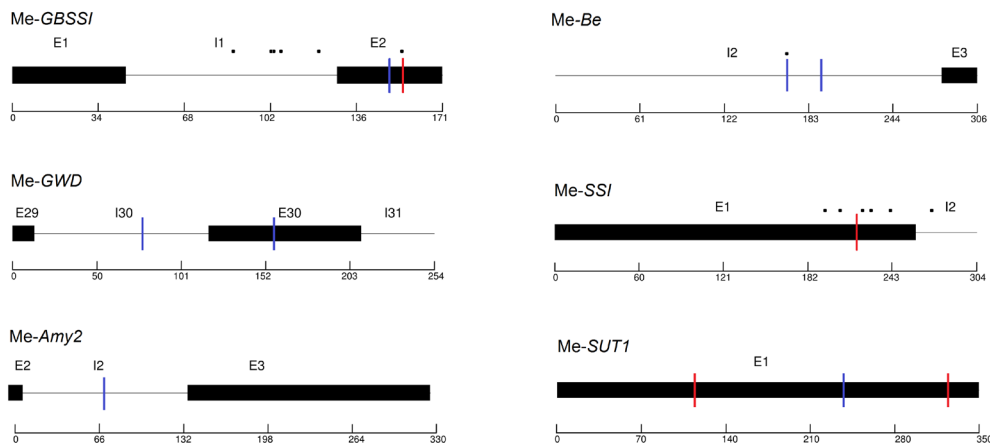

Figure 1. Mapping of gene regions based on the consensus sequences from different accessions of *Manihot esculenta* Crantz to granule bound starch synthase I (Me-*GBSSI*), glucan water dikinase (Me-*GWD*), α-amylase (Me-*Amy2*), starch branching enzyme (Me-*Be*), starch synthase I (Me-*SSI*), and sucrose transporter1 (Me-*SUT1*). Rectangles and lines represent exon and intron regions, respectively. The blue and red bars indicate transitions and transversions, respectively. The black dots represent the heterozygous loci.

Table 2. Distribution of single-nucleotide polymorphism (SNP) markers and heterozygous loci in genes of the cassava starch pathway in 96 accessions of *Manihot esculenta* Crantz.

Gene	Coding region			Non-coding region		
	Nº SNP	HL[1]	NN[2]	Nº SNP	HL	NN
Me-*Be*	0	0	26	2	1	280
Me-*GWD*	1	0	106	1	0	148
Me-*GBSSI*	2	1	87	0	5	84
Me-*SSI*	1	5	260	0	1	44
Me-*SUT1*	3	0	350	-	-	-
Me-*Amy2*	0	0	202	1	0	128

[1]HL: number of heterozygous loci; [2]NN: number of nucleotides. Me-*Be*: Starch branching enzyme; Me-*GWD*: Glucan water dikinase; Me-*GBSSI*: Granule bound starch synthase I; Me-*SSI*: Starch synthase I; Me-*SUT1*: Sucrose transporter1; Me-*Amy2*: α-amylase.

In addition to SNPs, 13 heterozygous loci were identified (Table 2 and Figure 1), which is to be expected when most *Passiflora* are outcross species with high inbreeding sensitivity. This tends to favor heterozygosity at different loci. Of the 26 SNPs identified by Kawuki et al. (2009) when analyzing nine cassava genes, only three loci exhibited no heterozygous genotypes, which confirms the existence of high heterozygosity, even in genic regions. Nevertheless, considering the number of heterozygous loci, an inverse situation was observed regarding the distribution of SNPs, since, on average, one heterozygous locus was identified for every 172 and 98 bp in coding and non-coding regions, respectively.

Regardless of the genomic region, the Me-*SSI* and Me-*GBSSI* genes presented the greatest number of heterozygous loci. In general, the appearance of heterozygous loci in cassava is due to the reproductive system of this species, which occurs almost predominantly by outcrossing (Narina et al., 2011). Maintenance of heterozygosity is preserved via clonal propagation, which ensures the permanence of this state across the different crop cycles.

The Me-*Be* gene presented two SNPs in transition form (Table 3) and one heterozygous locus in a 306-bp fragment, which is consistent with the various enzyme isoforms previously described (Lin et al., 2013). The product of the Me-*Be* gene has a basic function, i.e., branching of the amylopectin molecule in the starch metabolic pathway (Kharabian-Masouleh et al., 2012). However, it is still unclear which enzymes are responsible for the vast diversity of amylopectin molecule structures.

Table 3. Classification of molecular polymorphism in consensus sequences of genes involved in the cassava starch pathway in different accessions of *Manihot esculenta* Crantz.

Gene[1]	Transition	Transversion	Total changes	Total nucleotides
Me-*Be*	2	0	2	306
Me-*GWD*	2	0	2	254
Me-*GBSSI*	1	1	2	171
Me-*SSI*	0	1	1	304
Me-*SUT1*	1	2	3	350
Me-*Amy2*	1	0	1	330

[1]Me-*Be*: Starch branching enzyme; Me-*GWD*: Glucan water dikinase; Me-*GBSSI*: Granule bound starch synthase I; Me-*SSI*: Starch synthase I; Me-*SUT1*: Sucrose transporter1; Me-*Amy2*: α-amylase.

The Me-*GWD* gene presented two transition-type SNPs, one in the exon and one in the intron region from 106 and 148 bp, respectively (Figure 1 and Tables 2 and 3). This is consistent with the results reported by Carpenter et al. (2015) demonstrating the existence of polymorphisms in the *GWD* gene, whereby 32 SNPs were identified in a 690-bp fragment of the potato genome. The results of some studies suggest that the function of the *GWD* gene is ambiguous in plants, even though in crops such as maize and barley, these genes are responsible for starch synthesis, albeit at low levels (Mikkelsen et al., 2004).

Two SNPs were identified in the *GBSSI* gene, one was a transition and the other was a transversion type within the coding region, in addition to six heterozygous loci (Tables 2 and 3). Indeed, Aiemnaka et al. (2012) reported the presence of three loci in the *GBSSI* gene, which confer differences between waxy and non-waxy genotypes, even though only one SNP has been validated to distinguish these phenotypes. In other crops such as rice (*Oryza sativa*), Kasem et al. (2012) reported that no synonymous substitutions in exons of *GBSSI* from wild species can cause structural differences leading to changes in the activity of *GBSSI* or starch characteristics. Among these characteristics, retrogradation, amylose content, and viscosity

properties are the most commonly reported (Kharabian-Masouleh et al., 2012). *GBSSI* is thought to be responsible for amylose synthesis in plant storage organs, and its expression is associated with the generation of starch with no amylose (waxy), the specific functional properties of which make it highly suitable for certain industrial uses (Rolland-Sabaté et al., 2012).

The ME-*SSI* gene presented one transversion-type SNP and five heterozygous loci in the exon region of a 260-bp fragment (Table 2). Carpenter et al. (2015) identified a higher number of SNPs in potatoes (five SNPs in 315 bp) in a similar-sized fragment to that observed in the present study. There is evidence that some isoforms of this gene play specific roles in determining starch quality, such as the distribution of amylopectin branches (Yang et al., 2013).

Three SNPs (one transition and two transversions) were identified in the 350-bp coding region of the Me-*SUT1* gene. Some studies have shown that the Me-*SUT1* gene can be highly polymorphic, such as that of Li et al. (2011) in *Brassica napus* L., in which 142 SNPs were detected (120 SNPs in the 5'-UTR promoter region, 12 SNPs in exons, and 10 in introns). The product of this gene transports sucrose synthesized from leaves through the phloem to the storage tissue to synthesize three products (oil, starch, and protein) by carbohydrate metabolism (Li et al., 2011).

The Me-*Amy2* gene showed only one transversion in the intron region (Tables 2 and 3). Other studies in cassava have revealed high polymorphism in the isoamylase-1 gene, in which 27 and nine SNPs were identified in cassava and wild species, respectively, following analysis of a 954-bp fragment (Beyene et al., 2013). In general, more transitions (seven) than transversions (four) were observed. This type of replacement is commonly reported in molecular polymorphisms due to the biochemical structure of the nucleotides and the similar chemical properties of complementary base pairs. Similar observations have been reported in candidate genes in cassava (Kawuki et al., 2009).

Although SNPs are more prevalent in the non-coding regions of the genome, considering that these regions are subject to little or no selective pressure (González-Martínez et al., 2006), a higher prevalence of SNPs in the coding regions (seven SNPs in 1031 bp of the coding region and four SNPs in 684 bp of the non-coding region; Table 2) was noted in the present study. This bias may be due to the smaller size of the analyzed DNA sequences.

Of the seven SNPs identified in the coding regions of the different genes, three are related to non-synonymous substitutions and four to synonymous substitutions (Table 4). The percentage of non-synonymous mutations in this study (42.9%) is similar to that observed by Pootakham et al. (2014), when identifying SNPs via transcriptome sequencing of 16 cassava genotypes. This indicates that although non-synonymous mutations may result in changes in the protein sequence and consequently in function, these mutations do not cause physiological or genetic disorders, which could severely affect crop development. Furthermore, for synonymous mutations, most substitutions were observed at the third base position of the codon (57.14%) (Table 4). According to Jiménez-Gómez and Maloof (2009), SNPs tend to be located at the third position of the codon, which confirms the hypothesis that deleterious mutations occur at the first and second base, often resulting in non-synonymous substitutions that are eliminated through evolution.

The exon regions of Me-*SSI* and Me-*GWD* were highly conserved among cassava accessions, with only synonymous substitutions observed. This is in contrast to the *GBSSI* gene, which presented two non-synonymous substitutions (Table 4). Depending on the position in the protein, amino acid substitutions within the *GBSSI* gene (*Met/Tre* and *Val/Leu*) may modify the functional property of the enzyme, and hence the phenotypic characteristics

of starch. Previous studies on the *GBSSI* gene revealed the presence of a mutation (C/G) at position 78/79 of exon 11, which results in the production of contrasting starch-type phenotypes (waxy or non-waxy) (Aiemnaka et al., 2012). In other crops, such as sorghum, a mutation that changes amino acid 268 from glutamic acid (Glu) to histidine (His) results in non-waxy sorghum (Hamblin et al., 2007). In rice, other non-synonymous amino acid modifications within the *GBSSI* gene can lead to changes in starch characteristics (Kasem et al., 2012). In the present study, non-synonymous substitutions on SNPs 150 and 155 resulted in no phenotype change to waxy cassava.

Table 4. Types of nucleotide substitutions and classification of amino acids in genes of the cassava starch pathway in different accessions of *Manihot esculenta* Crantz.

Gene[1]	SNP position	Genetic codon	Amino acid[2]	Type of substitution[3]
Me-*GBSSI*	150	**A**TG/**A**CG	Met/Tre	NS
	155	**G**TG/**C**TG	Val/Leu	NS
Me-*SSI*	218	GC**A**/GC**T**	Ala/Ala	S
Me-*SUT1*	114	GC**C**/GC**A**	Ala/Ala	S
	237	TC**A**/TC**G**	Ser/Ser	S
	325	**G**GT/**C**GT	Gly/Arg	NS
Me-*GWD*	157	TA**T**/TA**C**	Tyr/Tyr	S

[1]Me-*Be*: Starch branching enzyme; Me-*GWD*: Glucan water dikinase; Me-*GBSSI*: Granule bound starch synthase I; Me-*SSI*: Starch synthase I; Me-*SUT1*: Sucrose transporter1; Me-*Amy2*: α-amylase. [2]Met: Methionine; Thr: threonine; Val: Valine; Ala: Alanine; Ser: Serine; Gly: Glycine; Arg: Arginine; Tyr: Tyrosine. [3]NS: non-synonymous; S: synonymous.

Identification of SNPs in genes related to phenotypes of wide commercial importance, such as those in cassava starch, has great potential for use in breeding programs, particularly as genetic markers for indirect selection. For example, studies on the isoamylase-1 gene revealed that the amino acid substitution of serine/leucine may be associated with changes in the quality of cassava root characteristics (Beyene et al., 2013).

Associations between molecular polymorphism and root dry-matter content

The significance of genome-wide association study between dry-matter content and SNPs was estimated by using an adjusted P value and the threshold for the association was set to -$\log_{10}$ (0.0021). However, even with the high efficiency of the kinship matrix (K) + structured association model (Q model) in the correction of spurious associations, in order to reduce the amount of false-positives and the high power for identifying important genetic associations, the mixed linear models, was not able to identify SNPs associated with dry matter content (Figure 2).

In general, the root dry-matter content in cassava tends to be subject to quantitative inheritance involving the action of several genes (Oliveira et al., 2014b). Therefore, when analyzing a panel of accessions with high genetic diversity, as well as the 96 cassava accessions used in the present study, it is unlikely for SNPs to explain the differences in dry-matter content. Indeed, Karlström et al. (2016) noted that although the slight difference in dry-matter content between waxy and non-waxy cassava clones (0.8%) was sufficient to detect statistical differences between the means for this trait, there is still a lack of knowledge regarding the genetic factors involved, including any knowledge to suggest the presence of pleiotropic effects or genes linked to *GBSSI*. Therefore, it is more likely that molecular differences within

genes involved in the starch metabolic pathway are directly related to characteristics of low complexity, such as those related to starch properties under heating and cooling conditions. Thus, molecular changes observed in genes involved in the starch metabolic pathway in cassava will be further validated for use in association studies of the paste properties.
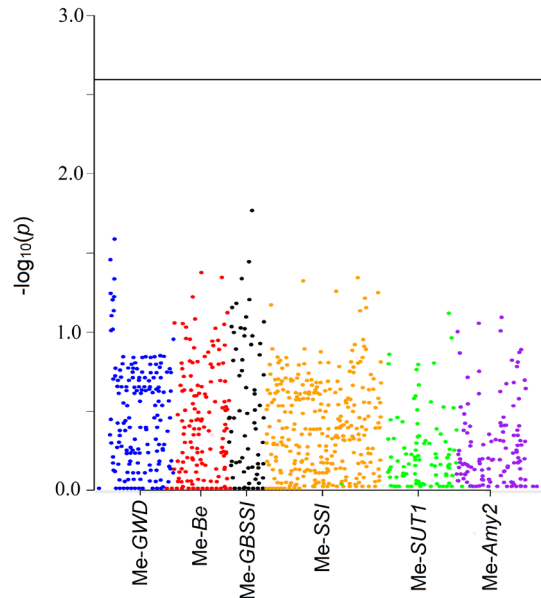

Figure 2. Manhattan plot showing the genome-wide marker associations with dry matter content to granule bound starch synthase I (Me-*GBSSI*), glucan water dikinase (Me-*GWD*), α-amylase (Me-*Amy2*), starch branching enzyme (Me-*Be*), starch synthase I (Me-*SSI*), and sucrose transporter1 (Me-*SUT1*). Black line indicates P value of 0.05 after Bonferroni correction.

Nucleotide diversity

Genes involved in the starch metabolic pathway presented π values that varied from 0.0001 (Me-*GBSSI*) to 0.0033 (Me-*Amy2*), averaging 0.0011, while the θ diversity reported by Watterson (1975) ranged from 0.00014 (Me-*Be*) to 0.00584 (Me-*SSI*), averaging 0.002353 (Table 5). Therefore, on average, the value of the θ diversity was twice that of the π diversity. In addition to the wide variation in these statistics, the difference between the parameters θ and π was approximately 40 and 28 fold, respectively, among the genes, indicating that some genes involved in the starch metabolic pathway in cassava experienced a high rate of mutation during the evolutionary process, or that some genes in this pathway were more likely to be affected by selection, such as Me-*Be*, Me-*GWD*, Me-*GBSSI,* and Me-*SUT1*. By analyzing contrasting rice varieties for the *Wx* and *OsC1* genes, which regulate the amylose content and apicule coloration, respectively, Choudhury et al. (2014) reported high π diversity for the *Wx* locus (the gene associated with *GBSSI*); however, this was similar to that observed between glutinous (π = 0.0053) and not glutinous (π = 0.0043) varieties. Those authors attributed this lack of clear differentiation between glutinous and non-glutinous rice phenotypes to the presence of other genes affecting the trait.

Table 5. Summary of gene polymorphisms observed in the starch metabolic pathway in accessions of *Manihot esculenta* Crantz considering different diversity indexes.

Gene[1]	Number of accessions	θ[2]	π[2]	D[2]
Me-*Be*	96	0.00014	0.00019	-1.33738
Me-*GWD*	92	0.00186	0.00021	-1.38928
Me-*GBSSI*	93	0.00165	0.00012	-1.07898
Me–*SSI*	87	0.00584	0.00255	-1.39486
Me-*SUT1*	87	0.00085	0.00019	-0.91084
Me-*Amy2*	49	0.00378	0.00334	-0.27819

[1]Me-*Be*: Starch branching enzyme; Me-*GWD*: Glucan water dikinase; Me-*GBSSI*: Granule bound starch synthase I; Me-*SSI*: Starch synthase I; Me-*SUT1*: Sucrose transporter1; Me-*Amy2*: α-amylase; [2]θ: Watterson's estimator of nucleotide diversity per site; π: nucleotide diversity; D: Tajima's D test.

Tajima's neutrality test (D) was used to assess sequence data that did not fit the neutral theory model. The observed D values were all negative and not significant (ranging from -0.278192 to -1.394866), showing that all genes presented neutral variation and that there was a balance between the effects of genetic drift and mutation in these accessions without evidence of strong selection intensities. In contrast, the process of species domestication tends to reduce the diversity of nucleotides in genes controlling specific and useful traits for humans during domestication. Therefore, genes associated with a particular trait under positive selection for the different steps of domestication and improvement tend to present "selection signatures" typically related to patterns of low π (Choudhury et al., 2014). In fact, the evaluation of only cultivated accessions (*M. esculenta*), revealed low π in the Me-*Be*, Me-*GWD*, Me-*GBSSI*, and Me-*SUT1* genes, which may represent evidence of a highly selective effect in these genes through evolution. Conversely, Beyene et al. (2013) presented a detailed description of the comparative SNP patterns in the isoamylase-1 (Meisa1) and dehydrogenase glyceraldehyde 3-phosphate (G3PDH) genes in 10 wild accessions (*M. esculenta* subsp *flabellifolia*) and 12 cultivated accessions. The results showed that the selection process may act in different ways in wild and cultivated species, since nucleotide diversity in the Meisa1 gene was higher in the cultivated species, whereas G3PDH was higher in the wild species.

Cluster analysis

The clustering profile generated from the analysis of nucleotide polymorphisms in starch genes revealed the formation of five groups (Figure 3). This included 82 clustered accessions and 14 other non-clustered accessions, which presented quite clear dissimilarities compared with the other clusters. In general, there was no relationship between the cluster formed and the root dry-matter content (Figure 4), which is similar to the observations from the individual analysis of the effects of SNPs with regard to this trait (Figure 2). Although the G1, G2, and GN clusters presented low dry-matter content in comparison with other clusters, there was no significant difference between these groups according to the Tukey's test (P < 0.05).

In addition to the root dry-matter content, we used historical information on cassava accessions for the root pulp color, classification of cyanogenic compounds content, and origin (local or improved variety). However, there were no relationships between the clustering of accessions by nucleotide polymorphisms in starch genes and the patterns of categorization mentioned above (Table 6).
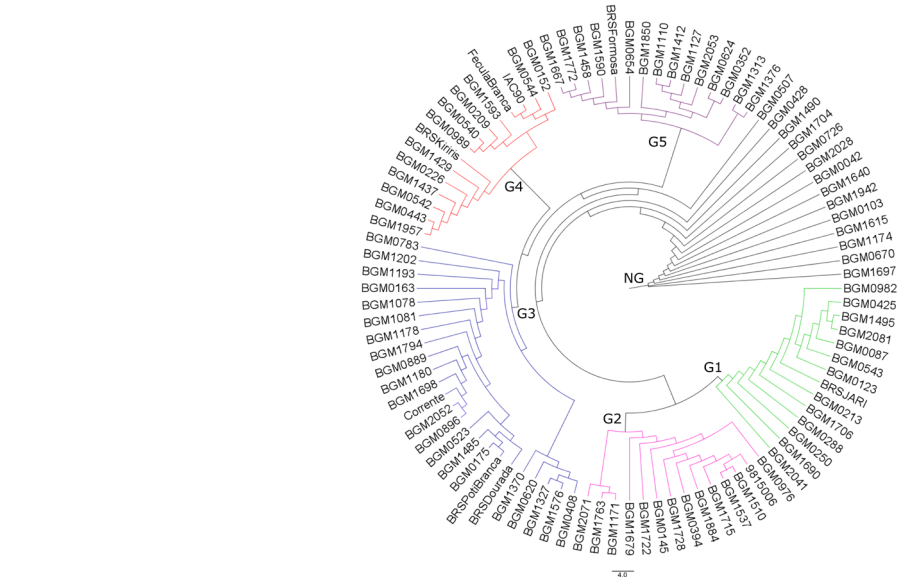
Figure 3. Clustering of 96 cassava accessions based on the neighbor-joining method and nucleotide polymorphism in starch genes. The clusters formed are represented by different colors. NG: not grouped accessions.
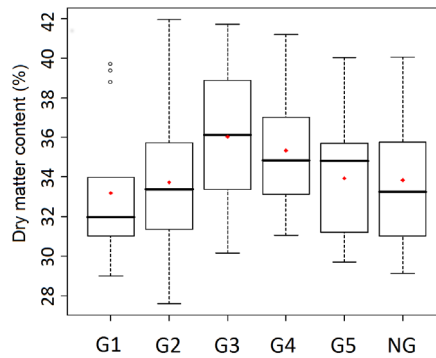

Figure 4. Box plot of the clusters formed by the analysis of 96 cassava accessions based on nucleotide polymorphisms in starch genes considering variation in the root dry-matter content.

Table 6. Summary of polymorphisms observed in the genes of the starch metabolic pathway in accessions of *Manihot esculenta* Crantz based on neighbor-joining clustering.

Cluster	Pulp color[1]				Cyanogenic compounds[2]				Origin of accessions[3]	
	Ye	Wh	Cr	WI	Bit	Swe	Int	WI	Land	Impro
G1	1	5	5	3	3	2	6	3	10	4
G2	-	6	3	5	4	-	7	3	12	2
G3	2	14	3	5	11	5	6	2	22	2
G4	-	11	1	3	4	-	7	4	12	3
G5	-	11	1	3	-	3	8	4	12	3
NG	2	7	-	5	8	1	2	3	12	2

[1]Ye: yellow, Wh: white, Cr: cream, WI: without information; [2]Bit: bitter, Swe: sweet, Int: intermediate; [3]Land: landrace, Impro: improved variety.

Due to the high level of genetic variation existing in the cassava crop, classification of accessions based on phenotypic traits has not been an easy task. For example, Oliveira et al. (2015) analyzed 474 cassava accessions based on starch yield, dry matter content in the roots, amylose content, and cyanogenic compounds content using affinity propagation algorithm. These authors identified nine clusters of diversity with high homogeneity of genetic distances between the clusters, especially for root dry-matter content and amylose content. Moreover, it was not possible to obtain a clear classification of the clusters based on the cyanogenic compound content.

SNP markers have also been used to classify cassava accessions in diverse clusters. In a recent study, Oliveira et al. (2014a) used two complementary approaches (discriminant analysis of principal component and Bayesian analysis) to understand the structure and genetic diversity of 1280 cassava accessions based on 402 SNPs. Those authors observed that both methodologies presented difficulties and controversies relating to the allocation of some accessions to specific clusters, and that prior information regarding the origin of accessions (local or improved variety) and geographic origins was not sufficient to clearly differentiate between clusters. Therefore, difficulty in the classification of cassava accessions based on these observations must be overcome through the integration of morphological, agronomic, and molecular tools. Specifically, at a molecular level, integration between the various types of genotyping platforms should be performed at a gene-specific level, or consideration of the whole genome should be prioritized for better classification and use of genetic resources of cassava.

One of the main objectives of this study was to compare the pattern of π in six genes involved in the starch metabolic pathway in cassava, and to evaluate its use for diversity studies. Therefore, the main results demonstrated the existence of several molecular polymorphisms in the Me-*Be*, Me-*GBSSI*, Me-*SSI*, Me-*SUT1*, Me-*Amy2*, and Me-*GWD* genes, which can be validated in the future as useful genetic markers, particularly for the analysis of intra-genic diversity of *M. esculenta* accessions. The average frequency of one SNP per 147 bp in the coding regions and one SNP per 171 bp in the non-coding regions of starch genes demonstrate the great potential for exploring these mutations in comparison with markers traditionally used in the molecular analysis of cassava, such as microsatellites.

Conflicts of interest

The authors declare no conflict of interest.

REFERENCES

Aiemnaka P, Wongkaew A, Chanthaworn J, Nagashima SK, et al. (2012). Molecular characterization of a spontaneous waxy starch mutation in cassava. *Crop Sci.* 52: 2121-2130. http://dx.doi.org/10.2135/cropsci2012.01.0058

Beyene D, Mukasa SB, Jansson C, Villiers EP, et al. (2013). Patterns of nucleotide diversity in *Meisa1* and G3pdh in wild

and cultivated cassava. *Ethiop. J. Sci.* 36: 65-72.

Carpenter MA, Joyce NI, Genet RA, Cooper RD, et al. (2015). Starch phosphorylation in potato tubers is influenced by allelic variation in the genes encoding glucan water dikinase, starch branching enzymes I and II, and starch synthase III. *Front. Plant Sci.* 6: 143. http://dx.doi.org/10.3389/fpls.2015.00143

Choudhury BI, Khan ML and Dayanandan S (2014). Patterns of nucleotide diversity and phenotypes of two domestication related genes (*OsC1* and *Wx*) in indigenous rice varieties in Northeast India. *BMC Genet.* 15: 71. http://dx.doi.org/10.1186/1471-2156-15-71

Doyle JJ and Doyle JL (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem. Bull.* 19: 11-15.

FAO. Faostat database (2015). Available at [http://www.faostat.org]. Accessed June 16, 2016.

Franco CML, Daiuto ER, Demiate IM, Carvalho LJCB, et al. (2001). Propriedades do amido. Séries: Séries: Culturas de Tuberosas Amiláceas Latino Americanas. São Paulo: Fundação Cargill.

González-Martínez SC, Ersoz E, Brown GR, Wheeler NC, et al. (2006). DNA sequence variation and selection of tag SNPs at candidate genes for drought-stress response in *Pinus taeda* L. *Genetics* 172: 1915-1926. http://dx.doi.org/10.1534/genetics.105.047126

Hall TA (1999). Bioedit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98NT. *Nucleic Acids Symp. Ser.* 41: 95-98.

Hamblin MT, Fernandez MGS, Tuinstra MR, Rooney WL, et al. (2007). Sequence Variation at candidate loci in the starch metabolism pathway in sorghum: prospects for linkage disequilibrium mapping. *Crop Sci.* 47: 125-134. http://dx.doi.org/10.2135/cropsci2007.01.0054tpg

Herrera-Campo BV, Hyman G and Belloti A (2011). Threats to cassava production: known and potential geographic distribution of four key biotic constraints. *Food Secur.* 3: 329-345. http://dx.doi.org/10.1007/s12571-011-0141-4

Jiménez-Gómez JM and Maloof JN (2009). Sequence diversity in three tomato species: SNPs, markers, and molecular evolution. *BMC Plant Biol.* 9: 85. http://dx.doi.org/10.1186/1471-2229-9-85

Karlström A, Calle F, Salazar S, Morante N, et al. (2016). Biological implications in cassava for the production of amylose-free starch: impact on root yield and related traits. *Front. Plant Sci.* 7: 604. http://dx.doi.org/10.3389/fpls.2016.00604

Kasem S, Waters DL and Henry RJ (2012). Analysis of starch gene diversity in the wild relatives of *Oryza sativa. Trop. Plant Biol.* 5: 286-308. http://dx.doi.org/10.1007/s12042-012-9111-z

Kawuki RS, Ferguson M, Labuschagne M, Herselman L, et al. (2009). Identification characterisation and application of single nucleotide polymorphisms for diversity assessment in cassava (*Manihot esculenta* Crantz). *Mol. Breed.* 23: 669-684. http://dx.doi.org/10.1007/s11032-009-9264-0

Kharabian-Masouleh A, Waters DL, Reinke RF, Ward R, et al. (2012). SNP in starch biosynthesis genes associated with nutritional and functional properties of rice. *Sci. Rep.* 2: 557. http://dx.doi.org/10.1038/srep00557

Kimura M (1980). A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.* 16: 111-120. http://dx.doi.org/10.1007/BF01731581

Lee GA, Koh HJ, Chung HK, Dixit A, et al. (2009). Development of SNP-based CAPS and dCAPS markers in eight different genes involved in starch biosynthesis in rice. *Mol. Breed.* 24: 93-101. http://dx.doi.org/10.1007/s11032-009-9278-7

Li F, Ma C, Wang X, Gao C, et al. (2011). Characterization of Sucrose transporter alleles and their association with seed yield-related traits in *Brassica napus* L. *BMC Plant Biol.* 11: 168. http://dx.doi.org/10.1186/1471-2229-11-168

Lin Q, Facon M, Putaux JL, Dinges JR, et al. (2013). Function of isoamylase-type starch debranching enzymes ISA1 and ISA2 in the *Zea mays* leaf. *New Phytol.* 200: 1009-1021. http://dx.doi.org/10.1111/nph.12446

Lipka AE, Tian F, Wang Q, Peiffer J, et al. (2012). GAPIT: genome association and prediction integrated tool. *Bioinformatics* 28: 2397-2399. http://dx.doi.org/10.1093/bioinformatics/bts444

Lopez C, Piégu B, Cooke R, Delseny M, et al. (2005). Using cDNA and genomic sequences as tools to develop SNP strategies in cassava (*Manihot esculenta* Crantz). *Theor. Appl. Genet.* 110: 425-431. http://dx.doi.org/10.1007/s00122-004-1833-3

Mikkelsen R, Baunsgaard L and Blennow A (2004). Functional characterization of α-glucan,water dikinase, the starch phosphorylating enzyme. *Biochem. J.* 377: 525-532. http://dx.doi.org/10.1042/bj20030999

Mufumbo R, Baguma Y, Kashub S, Nuwamanya E, et al. (2011). Amylopectin molecular structure and functional properties of starch from three Ugandan cassava varieties. *J. Plant Breed. Crop Sci.* 3: 195-202.

Narina SS, Jasti M, Buyyarapu R and Bhattacharjee R (2011). *Manihot.* In: Wild Crop Relatives: Genomic and Breeding Resources (Chittaranjan K, eds.). Springer Berlin Heidelberg, 133-155.

Nei M (1987). Molecular evolutionary genetics. Columbia University Press, New York.

O'Connor TD and Mundy NI (2009). Genotype-phenotype associations: substitution models to detect evolutionary associations between phenotypic variables and genotypic evolutionary rate. *Bioinformatics* 25: i94-i100. http://

dx.doi.org/10.1093/bioinformatics/btp231

Oliveira EJ, Ferreira CF, da Silva Santos V, de Jesus ON, et al. (2014a). Potential of SNP markers for the characterization of Brazilian cassava germplasm. *Theor. Appl. Genet.* 127: 1423-1440. http://dx.doi.org/10.1007/s00122-014-2309-8

Oliveira EJ, Santana FA, Oliveira LA and Santos VS (2014b). Genetic parameters and prediction of genotypic values for root quality traits in cassava using REML/BLUP. *Genet. Mol. Res.* 13: 6683-6700. http://dx.doi.org/10.4238/2014. August.28.13

Oliveira EJ, Santana FA, Oliveira LA and Santos S (2015). Genotypic variation of traits related to quality of cassava roots using affinity propagation algorithm. *Sci. Agric.* 72: 53-61. http://dx.doi.org/10.1590/0103-9016-2014-0043

Pootakham W, Shearman JR, Ruang-Areerate P, Sonthirod C, et al. (2014). Large-scale SNP discovery through RNA sequencing and SNP genotyping by targeted enrichment sequencing in cassava (*Manihot esculenta* Crantz). *PLoS One* 9: e116028. http://dx.doi.org/10.1371/journal.pone.0116028

Rambaut A (2014). FigTree v1.4.2, a graphical viewer of phylogenetic trees. Available at [http://tree.bio.ed.ac.uk/software/figtree]

Rolland-Sabaté A, Sánchez T, Buléon A, Colonna P, et al. (2012). Structural characterization of novel cassava starches with low and high-amylose contents in comparison with other commercial sources. *Food Hydrocoll.* 27: 161-174. http://dx.doi.org/10.1016/j.foodhyd.2011.07.008

Saithong T, Rongsirikul O, Kalapanulak S, Chiewchankaset P, et al. (2013). Starch biosynthesis in cassava: a genome-based pathway reconstruction and its exploitation in data integration. *BMC Syst. Biol.* 7: 75. http://dx.doi.org/10.1186/1752-0509-7-75

Tajima F (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123: 585-595.

Tamura K, Stecher G, Peterson D, Filipski A, et al. (2013). MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* 30: 2725-2729. http://dx.doi.org/10.1093/molbev/mst197

VanRaden PM (2008). Efficient methods to compute genomic predictions. *J. Dairy Sci.* 91: 4414-4423. http://dx.doi.org/10.3168/jds.2007-0980

Watterson GA (1975). On the number of segregating sites in genetical models without recombination. *Theor. Popul. Biol.* 7: 256-276. http://dx.doi.org/10.1016/0040-5809(75)90020-9

Yang Z, Wang Y, Xu S, Xu C, et al. (2013). Molecular evolution and functional divergence of soluble starch synthase genes in cassava (*Manihot esculenta* crantz). *Evol. Bioinform. Online* 9: 239-249. http://dx.doi.org/10.4137/EBO. S11991