Available online at www.scholarsresearchlibrary.com

Scholars Research Library

Annals of Biological Research, 2013, 4 (2):105-114
(http://scholarsresearchlibrary.com/archive.html)

Annals of Biological Research
Scholars Research Library
ISSN 0976-1233
CODEN (USA): ABRNBW

Occurrence of two isoforms of granule-bound starch synthase II (GBSS II) gene in Cassava (*Manihot esculenta* Crantz)

Jelili T. Opabode[1*], Olufemi O. Oyelakin[2,3], John A. Akinkunmi [3,6], Oluyemisi A. Akinyemiju[1] and Ivan L. Ingelbrecht[4,5]

[1]*Department of Crop Production and Protection, Obafemi Awolowo University, Ile-Ife, Nigeria*
[2]*Biotechnology Centre, Federal University of Agriculture, Abeokuta, Nigeria*
[3]*Cassava Improvement Programme, International Institute of Tropical Agriculture, Nigeria*
[4]*Central Biotechnology Laboratory, International Institute of Tropical Agriculture (IITA), Ibadan, Nigeria; c/o Lambourn Ltd, Carolyn House, 26 Dingwall Road, Croydon CR9 3EE, UK*
[5]*Department of Plant Biotechnology and Genetics, Ghent University, K L Ledeganckstraat 35, B-9000 Ghent, Belgium*
[6]*Department of Agricultural Extension and Rural Development, Obafemi Awolowo University, Ile-Ife, Nigeria*

ABSTRACT

Cassava (Manihot esculenta) is a major source of starch in tropical and subtropical countries. The genomic organization of the granule-bound starch synthase II (GBSS I) in cassava was examined to increase starch biosynthesis knowledge and facilitate the production of modified starches in cassava. Six genomic clones encoding fragment of granule-bound starch synthase II (GBSS II) gene were produced by PCR amplification of genomic DNA from a primer pair. Sequence analyses revealed that the clones were of two types: pOYE303-4 and pOYE303-5. The insert in pOYE303-4 (accession no. HM038440) was 1690 nucleotides long and encoded a polypeptide of 122 amino acids with molecular weight of 118.52 kDa and calculated pI of 9.67. In contrast, pOYE 303-5 (accession no. HM046981), encoded 118 amino acids with a molecular weight of 13.22 kDa and pI of 9.59. HM038440 and HM046981 had 99.4% identity at nucleotide and amino acid levels. Phylogenetic analysis showed that the two isolated gbss II sequences were not the same but belonged to the same dicot group and closely related to R. communis and V. vinifera gbss II proteins. In silico identification of introns revealed different number of introns in the two sequences: four in HM038440 and two in HM046981. The results of sequence analyses, phylogenetic study and in silico identification of intron suggested that cassava genome contains two isoform of GBSS II.

Key words: Cassava, granule-bound starch synthase II, *Manihot esculenta*, amylose, starch

---

INTRODUCTION

Cassava is an important root crop in tropical and subtropical regions [1]. Depending on the cultivar, the starch content of cassava roots ranges from 65 -91% of its total root dry weight depending on the cultivar [2]. Cassava starch contains amylose and amylopectin polymers. Low-amylose and amylose-free starches have wider industrial uses than amylose-rich starch due to different physicochemical properties between the two kind of starches. Visser et al. [3] observed greater granule melting temperature, reduced retrogradation and enhanced adhesive properties in low-amylose potato compared with amylose-containing starch. The goal of increasing the industrial uses of this crop has triggered interest in producing cassava plant with modified starch characteristics. Blennow [4] forecast that

development of low-amylose cassava variety would give the crop advantage of a share in existing huge market for modified starch estimated at US$10 million.

Conventional breeding of cassava suffered setback because of the highly heterozygous nature of the crop that prevents backcross scheme and poor flowering with limited seed set of many varieties [5]. Starch biosynthesis is polygenic and there is limited natural variation with regards to starch metabolism in cassava compare to other crops such as maize. Only two starch mutants have been reported in cassava: first, a natural mutation in a *gbss* gene resulting in production of amylose-free starch and second, a gamma irradiation-induced mutation in isoamylase gene resulting in high-amylose starch [6,7]. Because most mutations affecting starch traits are recessive, transferring agronomically important mutations into the appropriate genetic background through conventional breeding is difficult.

In the field, cassava is typically propagated clonally by stem cuttings. This propagation strategy is ideal for a bio-engineering approach to crop improvement as gene segregation through outcrossing is limited [8]. Transgenic cassava producing amylose-free starch has been produced by antisense inhibition of granule-bound starch synthase I [9]. In addition, a transgenic cassava plants with an enhanced quantity of starch have been produced [10]. Successful manipulation of starch characteristics in cassava requires a more thorough characterization of the various cassava starch biosynthesis genes.

It has been established that granule-bound starch synthase (GBSS) is responsible for the biosynthesis of amylose. There are two types of GBSS in higher plants based on molecular mass and localization. The GBSS I with molecular mass of 58-60 KDa is tightly bound to starch granules and offers the largest proportion of total GBSS activity [11, 12]. GBSS II is present in starch granules as well as in the soluble fraction of plastids [13]. In contrast to GBSS I, the role of GBSS II is ambiguous, however GBSS II mutant studies in pea revealed that GBSS II might play an important role for determing amylopectin structure and starch morphology [13]. Cassava GBSS II gene has been cloned from a cDNA library [14,15]. The *gbss* II clone encoded a 751 amino acid polypeptide that showed homology to potato (61%) and pea (59%) *gbss* II. The derived amino acid sequence of cassava *gbss* II exhibited low sequence homology to cassava *gbss* I (35% identity). Southern analysis of cassava genomic DNA revealed that *gbss* II was a single copy gene that was localized in linkage group T of the male derived cassava genetic map. Cassava *gbss* II is highly expressed in young leaves, and exhibited much lower expression in developing tubers. These results are indicative of the differing and complementary role to *gbss* I that *gbss* II plays in leaf and tuber starch production [14, 15]. Multiple isoform of GBSS II have been reported in higher plants such as pea, soyabean rice, wheat, sorghum, maize and cowpea [16, 17, 18]. In cassava, only one isoform has been reported from cDNA cloning and genomic organization of GBSS II has not been described. In addition, the location of introns in the *gbss* II has not been identified and limited phylogenetic relationships conducted on *gbss* II sequences in higher plants. A good knowledge of genomic organization, location of introns, phylogenetic relationships and number of isoform of GBSS II gene is important for effective manipulations of the activity of this gene through proper gene construct design for down regulation or over expression of the starch gene for enhanced starch properties. This communication described partial genomic organization of two granule-bound starch synthase II (GBSS II) gDNAs, HM038440 and HM046981, cloned by PCR amplification of genomic DNA with a primer pair. It further identified introns from the two GBSS II clones *in silico* .

MATERIALS AND METHODS

Location of the study

The study was conducted at the Central Biotechnology Laboratory, International Institute of Tropical Agriculture (IITA), Ibadan between 2006 and 2010. The study was a collaborative research between IITA and Obafemi Awolowo University, Ile-Ife, Nigeria with the goal of developing transgenic cassava plants with modified starch.

Plant materials and DNA extraction

Extraction of genomic DNA from young leaves of cassava was carried out using Dellarporta *et al*. [19] procedures. About 0.5-1.0 g of fresh young leaves was harvested into labeled 1.5 ml eppendorf tubes. The fresh leaves were ground in a mortar containing 50 -100 ml liquid nitrogen with a pestle. About 800 µl of extraction buffer (100 Mm Tris-HCl, pH 8.0,50 Mm EDTA, pH 8.0, 500 Mm NaCl, 1% PVP) and 20 µl 0.7% beta mercarpto-ethanol, which have been preheated at 65 ^{0}C was added to the tubes. These were mixed with a pipette tip until all tissue became dispersed in the buffer. About 100 µl of 20% sodium diodecyl sulphate (SDS) was added and mixed thoroughly for

one minute. Thereafter, the mixture was incubated at 65^0C (with intermittent mixing 6 times) for 15 minutes. The mixture was then removed from 65^0C condition and cool to room temperature for two minutes. About 300 µl of ice-cold 5M Potassium acetate was added to the mixture followed by gentle inversion six times. Thereafter, the mixture was incubated on ice for 20 minutes and centrifuged at 3,000 g in microfuge tube for ten minutes. The supernatant was carefully transferred to two new 1.5 ml eppendorf tubes. One volume of ice-cold iso-propanol (approx. 700 µl) was added and mixed by inverting gently 8-10 times. This was followed by incubation at -80^0C for one hour and centrifugation at 3,000 g for one minute. The supernatant was tipped off and the last drop of iso-propanol was removed by placing tubes face down on paper towel. Pellets were suspended in 250 µl of high salt TE and 4 µl RNase was added. This was followed by incubation at 37^0C for one hour. Iso-propanol (500 µl) was added and mixed by gentle inversion 8-10 times. Then, the mixture was kept at -80^0C for one hour and centrifuged at 3,000 g for 10 minutes. Supernatant was decanted and the last drop of iso-propanol was removed by placing tubes face down on paper towels. The pellets were washed twice by 70% ethanol. Pellets were allowed to dry by leaving it on paper towel for one hour. Depending on the size of the pellet, about 100-200 µl sterile distilled water was added. Tubes were stores at 4^0C overnight to dissolve the DNA pellet. The supernatant was transferred to eppendorf tube and store at -20^0C for further use. The quality of the DNA was determined by running 2 µl of the DNA alongside a molecular weight marker (11.0 kb plus pst 1 lamba DNA) on 0.8% agarose gel electrophoresis in 1xTAE (Tris acetate EDTA) buffer at 500 volts for one hour. The concentration of the DNA was measured by taking absorbance of 2 µl of the DNA at 260 and 280 nm wavelengths on a Nanodrop spectrophotometer.

Primer design

A cassava granule-bound starch synthase II gene-specific primer pair was designed from cDNA sequences (accession AF173900) earlier deposited in *GenBank* database from nucleotide +3 to +369. The downloaded sequences were used as template to design primer specific for the gene using Lasergene sequence analysis software (DNASTAR Inc, Madison, USA). The composition of the primer pairs designed and used for the study were as follows: Forward 5'- GGCATTTATAGGATCACTTCC-3' and Reverse5'-GAGTTTTCCCTGTTCATGAG-3'. Synthesis of the primers was done by Integrated DNA Technologies Incorporation (Iowa, USA).

PCR amplification, cloning and DNA sequencing

Amplification was carried out in 50 µl-reaction volume, which composed of 1µl of 10x buffer, 0.5 µl of $MgCl_2$ (25 mM), 1 µl each of primer F and R (1µM), 0.5 µl of dNTPs (2.5 mM), 1 µl template DNA (500 ng), 4.8 µl of H_2O and 0.2 U of Taq DNA polymerase (Bioline, USA). The PCR amplification profile consisted of initial denaturation at 94^0C for one minute and 30 cycles of amplification (94 ^{0}C for 30 seconds, 50^0C for 30 seconds, 72^0C for 45 seconds) with a final cycle of 5 minutes at 72^0C. The PCR amplification was carried out in a Peltier thermal cycler (PTC 2000, MJ Research, India).The PCR fragments were purified and cloned into pDRIVE vector (QIAGEN, California, USA). The presence of the insert in the recombinant plasmid was confirmed by restriction digestion. Both strands of the DNA inserts were sequenced and any sequence ambiguities were resolved by re-sequencing. The three independent PCR clones were sequenced from each PCR reaction. The sequences were manually edited and vector sequences removed. DNA sequencing was performed by Iowa State University, USA.

Sequence analyses

BLAST searches were conducted at NCBI website (http://www.ncbi.nlm.nih.gov/). The genomic sequences of *gbss* II have been submitted to *GenBank* under the accession numbers HM038440 and HM046981. Sequences analyses was conducted by CLC DNA WORKBENCH software version 6.1. Putative introns were identified *in silico* by (i) alignments between genomic and cDNA (AF13900) sequences using ClustalW2 programme at www.ebi.ac.uk/ (European Molecular Biology Laboratory website) and (ii) confirmed by gene prediction method using GENSCAN web server [20].

Phylogenetic analysis

Sequence searches of the non-redundant and unfinished genome databases at NCBI were conducted with the isolated *gbss* II ORFs as query sequence. The ClustalX package was used to create an alignment of the sequences that was then submitted to a neighbour-joining analysis to generate a branching pattern. The phylogenetic tree was displayed using the CLC TREEVIEW program. Eighteen *gbss* II sequences of higher plants (and their accession numbers) included in the study were: *Ipomoea batatas* (AF068834), *Oryza sativa* (AF383878), *Sorghum bicolor* (EU307275), *Zea mays* (XM019296), *Hordeum vulgare* (AY133249), *Colocasia esculenta* (AY225862), *Triticum aestivum* (AJ269503), *Phaseolus vulgaris* (AB127938), *Glycine max* (NM001248867), *Manihot esculenta* (HM038440), *Manihot esculenta* (HM046981), *Ricinus communis* (XM002531810), *Vitis vinifera* (X87988), *Vigna unguiculata*

(FJ561745), *Vigna radiata* (FJ561747), *Solanum tuberosum* (X87988), *Lotus japonicus* (FJ561743) and *Ammarantus cruentus* (DQ178026).

RESULTS

Amplification, isolation and characterization of *gbss* II clones

In this study, GBSS II gene in cassava was cloned and partially characterized, and six introns in GBSS II genomic organization were identified. In the preliminary experiment, PCR amplifications were recorded in all tested annealing temperatures and DNA concentrations. However, out of the three annealing temperatures (45^0C, 50^0C and 55^0C) and two template DNA concentrations (100 ng/µl and 500 ng/µl) tested for amplification of *gbss* II, the band was distinct at annealing temperature of 50^0C using DNA template concentration of 500 ng/µl (data not shown). These conditions were used for enhanced amplification of *gbss* II in 50-µl reaction volume for fragment purification and cloning purposes (Fig. 1). The amplification profile used for the study was 30 cycles, when the number of the cycles was reduced to 25 cycles no discrete bands were observed probably because of non-specific primer hybridization due to the high nucleotide homology between and within *gbss* II. The observed product sizes were 1690 and 1698 bp against the expected 363 bp. The PCR products were purified and cloned into pDRIVE vector. Recombinant plasmid was extracted from the six selected clones. The presence of inserts in the recombinant plasmid was confirmed through release by restriction digestion with EcoR1 and by PCR amplification.

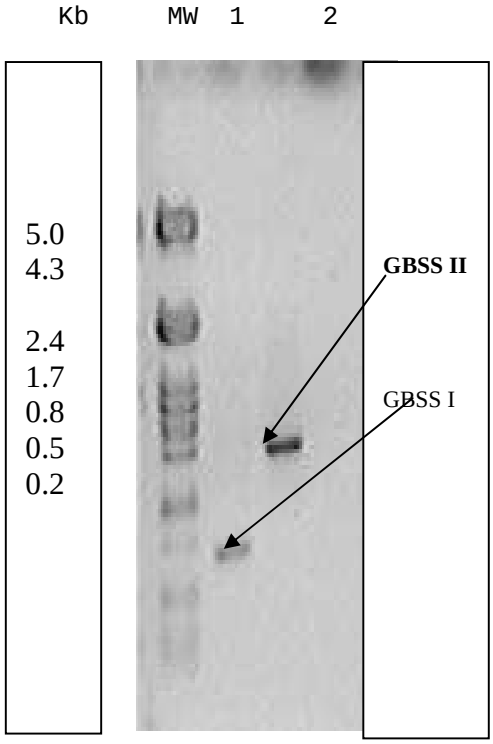

Fig.1: Amplification of GBSS I and GBSS II by PCR at 55^0C. MW= DNA marker, 1= GBSS I, 2=GBSS II, molecular weight markers are at the left in kb

Six independent clones obtained with the primer pairs designed for amplification of +3 to +369 portion of *Manihot esculenta gbss* II gene (accession AF173900) were sequenced. Surprisingly, sequence comparison showed that the inserts belonged to two groups. The first group had a complete nucleotides and amino acids identity and was represented by pOYE303-4. Similarly, the second group had a complete nucleotide and amino acids identity and was represented by pOYE303-5. The inserts in the two clones (pOYE303-4 and pOYE303-5) were subjected to further analyses. The sequences of the inserts in pOYE303-4 and pOYE303-5 have been submitted to *GenBank* and defined as HM038440 and HM046981, respectively. HM038440 and HM046981 had 99.4% homology at nucleotide level (Fig.2). The insert in pOYE303-4 (accession no. HM038440) was 1690 nucleotides long and encoded a polypeptide of 122 amino acids with molecular weight of 118.52 kDa and calculated pI of 9.67 (Table 1).

POYE303-4 was rich in serine (13.11%), leucine (12.29%) and phenyalanie (9.83%). The predicted secondary structure of POYE303-4 consists of 36.07% alpha helix, 10.66% extended strand and 53.28% random coil. Blast search with HM038440 retrieved *Manihot esculenta gbss* II cDNA (AF173900). The identity of HM038440 to *gbss* II of castor bean and pea is 87% (Table 1). The second clone, POYE 303-5 (accession no. HM046981) encoded 118 amino acids with a molecular weight of 132.2 kDa and pI of 9.59 (Table 1). POYE303-5 rich in phenyalanine (12.29%) and leucine (15.25%). The predicted secondary structure of POYE303-5 consist of 44.92% alpha helix, 11.86% extended strand and 43.22% random coil. Blast search with HM046981 retrieved *Manihot esculenta gbss* II cDNA (AF173900). The identity of HM046981 to *gbss* II of castor bean is 84% (Table 1).

```
HM038440        AFIGSLPFIIQTKAESSVLLHDKNLQRSRFSVFPCRSQNSFNLAVSLSLSFKPVRATGKE
60
HM046981        AFIGSLPFIIQTKAESSVLLHDKNLQRSRFSVFPCRSQNSFNLAVSLSLSFKPVRATGKE
60
                ************************************************************

HM038440        GVSGDGSEDTLQATIEKSKKVLALQRDLLQKVIYWPLFFFSRAKSGVSEADFDIMLPLNF
120
HM046981        GVSGDGSEDTLQATIEKSKKVLALQRDLLQKVIYWPLFFFFLELN----LEFLKQILTSC
116
                ************************************    .     :*   :  .

HM038440        EF 122
HM046981        CH 118
```

Fig.2: Alignment of deduced amino acid sequences of HM038440 and HM046981. The sequences were aligned by ClusatlW2 programme. Asterisks indicate 100% conserved, double dots similar, and single dots related amino acids: dashes depict gaps. Numbers refer to amino acids.

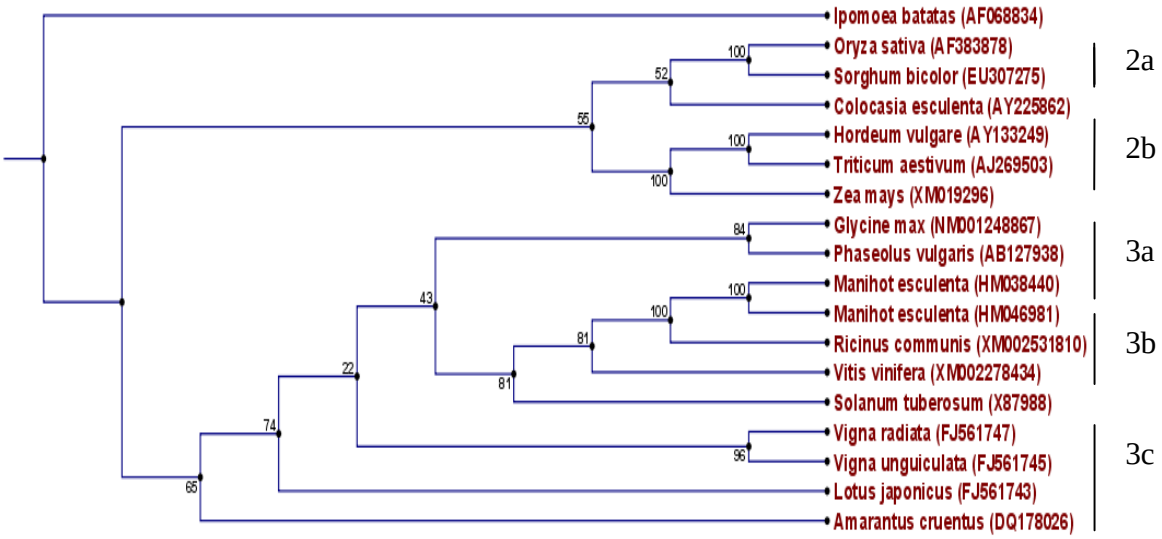

Fig. 3:Phylogenetic relationship of GBSS II sequences of higher plants. The tree was constructed on the basis of available GBSS II sequences in the GenBank and from this study (*Manihot esculenta*) using neighbourhood-joining method from bootstrapped data sets. The number of bootstrap replicates is indicated next to each branch.

Phylogenetic analysis of *gbss* II

The phylogenetic analysis of a total of 18 *gbss* II proteins indicate that *gbss* II in higher plants are grouped into three groups with six branches (Fig.3). The groups are sweet potato, monocots and dicots. The analysis revealed that duplication had occurred in *gbss* II proteins before the divergence of monocots and dicots groups and sweet potato

retained the ancient copy of *gbss* II gene. In the monocots, duplication of *gbss* II proteins took place and led to the formation of two branches, 2a and 2b. The dicots group has three branches (3a, 3b, 3c) following two duplication events. The two isolated cassava *gbss* II sequences (HM038440 and HM046981) belonged to 3b branch in the dicot group which also contain *R. communis* and *V. vinifera gbss* II proteins.

```
gbssIIcDNA      GGCATTTATAGGATCACTTCCTTTTATTATCCAAACCAAAGCAGAAAGTTCT--------
52
HM038440        GGCATTTATAGGATCACTTCCTTTTATTATCCAAACCAAAGCAGAAAGTTCTGTCCTTCT
60
                **************************************************
gbssIIcDNA      ----------------------------------------->----------------
        INTRON I
HM038440        CCATGACAAAAACCTACAGCGATCCAGATTCTCCGTTTTCCCATGTAGATCACAAAACTC
120

gbssIIcDNA      ------------------------------------------------------------
HM038440        TTTTAATTTAGCCGTTTCGTTATCTTTGAGTTTTAAGCCTGTAAGAGCTACAGGTAAGGA
180

gbssIIcDNA      --------------------AGAGGATACACTTCAAGCCACCATCGAGAAAAGCAAGAA
91
HM038440        AGGCGTTAGTGGTGATGGGTCAGAGGATACACTTCAAGCCACCATCGAGAAAAGCAAGAA
240
                                    **************************************
gbssIIcDNA      AGTTCTCGCCTTGCAAAGGGACCTACTTCAGAAGGTGATATACTGGCCACTATTTTTTTT
151
HM038440        AGTTCTCGCCTTGCAAAGGGACCTACTTCAGAAGGTGATATACTGGCCACTATTTTTTTT
300
                ************************************************************
gbssIIcDNA      TTCTAGAGCTAAATCTGGAGTTTCTGAAGCAGATTTTGACATCATGTTGCCATTAAATTT
211
HM038440        TTCTAGAGCTAAATCTGGAGTTTCTGAAGCAGATTTTGACATCATGTTGCCATTAAATTT
360
                ************************************************************
gbssIIcDNA      TGAATTTTAGATGATCATATGGGGTTTTTTTGGGTGGCATTTTATGTTACCCGTCGAAGG
271
HM038440        TGAATTTTAGATGATCATATGGGGTTTTTTTGGGTGGCATTTTATGTTACCCGTCGAAGG
420
                ************************************************************
gbssIIcDNA      AGAAATTGAATATGATCGAGTGTTACCCCTAAAAGAGGATTAGGAAGCTGATTGTAAGAA
331
HM038440        AGAAATTGAATATGATCGAGTGTTACCCCTAAAAGAGGATTAGGAAGCTGATTGTAAGAA
480
                ************************************************************
gbssIIcDNA      AATATC-------------------------->------------------------
337         INTRON II
HM038440        AATATCGTTGCCTGAAAAAGCTATGTAATTGATGATAAAGGATGTAAGTGGATATAATTA
540
                ******
gbssIIcDNA      ------------------------------------------------------------
HM038440        TTTTGCAAATTTTGGTAGCGTTATAAAAGGAATTCTCAATTAAGTAAAACAAAGCACATC
600

gbssIIcDNA      ----------------AGCTACAGGGATTGGACTAACATAGTGTAGACGTCGAGAAACC
380
HM038440        ATCTACAATAGGCCTATAGCTACAGGGATTGGACTAACATAGTGTAGACGTCGAGAAACC
660
```

```
                                          ******************************************
gbssIIcDNA      AAGTAACATGGATTGATTATATGTTTGCTCCTTAAACTGCATAAACTTTGTTATAATTGT
440
HM038440        AAGTAACATGGATTGATTATATGTTTGCTCCTTAAACTGCATAAACTTTGTTATAATTGT
720
               ************************************************************
gbssIIcDNA      AATCCAAAAAATTTAAACATATAATCAATAAGGAGCTAACATGGATTGTTAGCTTGTCTT
500
HM038440        AATCCAAAAAATTTAAACATATAATCAATAAGGAGCTAACATGGATTGTTAGCTTGTCTT
780
                ************************************************************
gbssIIcDNA      ATAATCCAAAAAAAAAAAAGAGATTTGCAAT----------------------------
531
HM038440        ATAATCCAAAAAAAAAAAAGAGATTTGCAATGTCTTTACCAATCTCCTTACAATCCAAAA
840
                ********************************
gbssIIcDNA      --------------------------------------------------------
INTRON III
HM038440        CAAGAGACCCAAACTCAGAAAGATCCAGATTATCTGACTTGAGAGCTTGATAGGCTTCAA
900

gbssIIcDNA      ------------------------------------------------------------CTT
534
HM038440        TATTATCCATCTCAATGACCGCACTAACAAAATTCCTCATACGAATCCAAGAAAGAGCTT
960
                                                                        ***
gbssIIcDNA      ATGCTTCTGCCTCTCTAGCTTGTAACAAACCATCAGACCAACCTGTTACAGCATGGAGGG
594
HM038440        ATGCTTCTGCCTCTCTAGCTTGTAACAAACCATCAGACCAACCTGTTACAGCATGGAGGG
1020
                ************************************************************
gbssIIcDNA      AGCACCCGGCATGATTGTGGAGGATAGCACTGCAACCAATTCGTCCACAGTGAGGAAAGA
654
HM038440        AGCACCCGGCATGATTGTGGAGGATAGCACTGCAACCAATTCGTCCACAGTGAGGAAAGA
1080
                ************************************************************
gbssIIcDNA      GTGCAGCATCAGCGTTCAGCTTGAGTAATGTTGAGGGAGCACGCCAGCTGGTGATGGCAG
714
HM038440        GTGCAGCATCAGCGTTCAGCTTGAGTAATGTTGAGGGAGCACGCCAGCTGGTGATGGCAG
1140
                ************************************************************
gbssIIcDNA      GCAAAGCAGTCTGCAAGAGATACTTCCAGAAT----------------------------
746
HM038440        GCAAAGCAGTCTGCAAGAGATACTTCCAGAATGTGTAGCAGCTAACAAGTTGACCCTCTC
1200
                ********************************
gbssIIcDNA      ------------------------------------------------------------
HM038440        CTGCCAACAATTTATGTTCTATGCTTCTAATTAAAGATTGATTACGAAAGACTAAACTTT
1260

gbssIIcDNA      --------------------------------------------------------
INTRON IV
HM038440        GATTCGATATGATGTATTGGGCCTTAAAAATTCTATGTATGTGGCCTTTTACCTGTGGAT
1320

gbssIIcDNA      ------------------------------------------------------------
```

```
HM038440        TATTGCGACCTATTATCCTTGGATAATTGATCTTTTAGCTTCTATTATTCTCTTCTATAA
1380

gbssIIcDNA      -TATATTTGCTTGATAAGAATGATGTAATCCTTACTAAATGTCAATATAATGGTGATTCT
805
HM038440        GTATATTTGCTTGATAAGAATGATGTAATCCTTACTAAATGTCAATATAATGGTGATTCT
1440
                 ************************************************************
gbssIIcDNA      AACATTATGTTATAATTTCAAAAGGGAAAAAAATTATATCATCTTCTTTCTTACCAACCA
865
HM038440        AACATTATGTTATAATTTCAAAAGGGAAAAAAATTATATCATCTTCTTTCTTACCAACCA
1500
                ************************************************************
gbssIIcDNA      AGTCTTTCTAGAGGTCAAAATCACTGTCATCTCATTAGCTCCTACTTCTGTGTGAGTAAT
925
HM038440        AGTCTTTCTAGAGGTCAAAATCACTGTCATCTCATTAGCTCCTACTTCTGTGTGAGTAAT
1560
                ************************************************************
gbssIIcDNA      CAAATCTAAATACCGTTATTGAAACATTCATCCTTTGCAGATTGCTGAAAGAAGGAAATT
985
HM038440        CAAATCTAAATACCGTTATTGAAACATTCATCCTTTGCAGATTGCTGAAAGAAGGAAATT
1620
                ************************************************************
gbssIIcDNA      GGTTTCTTCTATACAAAGTAGTGTTGGTGACCACGACACAAACAAAACTTCTCATGAACA
1045
HM038440        GGTTTCTTCTATACAAAGTAGTGTTGGTGACCACGACACAAACAAAACTTCTCATGAACA
1680
                ************************************************************
gbssIIcDNA      GGGAAAACTC 1055
HM0
```

Fig.4: Alignment between cassava genomic (HM038440) and cDNA (AF173900) sequences of GBSS II gene

***In silico* identification of introns**

The locations of four putative introns identified by sequence alignment between the *gbss* II isolated genomic sequence (accession no. HM038440) and *gbss* II cDNA sequence (accession (AF173900) deposited in the *GenBank* are shown in Fig. 4. The locations of the introns were nucleotide 53-203 (Intron 1), 487-619 (Intron II), 812-957 (Intron III) and 1173-1380 (Intron IV). Only two introns (53 -203 and 487 and 619) could be located in HM046981 using similar alignment procedure as applied to HM038440 (data not shown). The putative intron locations were further confirmed by gene prediction method using GENSCAN web server. The first intron was 201 bp long and located between nucleotide 53 and 201 . The first intron was 8.8% of the entire genomic sequence and made up of 31.5% T, 26.2% A, 20.1% C and 18.8% G. The second intron was 111 bp long and about 6.5% of HM038440. The second intron was located between 507 and 617 and made up of 46.8%A, 21.6%G, 37.0%T and 13.5%C. The third intron was 146 bp long and about 8.6% of HM038440. It was located between 812 and 957 nucleotide and 37.7% A, 11.6% G, 24.0% T and 24.7% C rich. The fourth intron was located between nucleotide 1172 and 1381 and made up of 26.7%A, 15.2%G, 30.1%T and 16.2%C. When the intron sequences were removed, the nucleotide sequences of HM038440 and *gbss* II cDNA were identical. All the introns of *gbss* II sequence contain the splice site sequences consistent with the consensus 5'-GT _____AG-3' [21]. The results of the present study provided evidences that there are more than one form of GBSS II gene in cassava and move the effort of genetic manipulation of GBSS II gene forward by identify the location of intronic sequences in the genomic organization of GBSS II gene.

DISCUSSION

The successful PCR amplification of *gbss* II indicated that the composition of oligonucleotide primer pairs used was adequate and specific for the gene. Six partial genomic clones of *gbss* II was isolated and sequenced. Three clones (first group, HM038440) were 100% identical at both nucleotide and amino acid levels. The other three clones (second group, HM046981) were also 100% identical at both nucleotide and amino acid levels. This occurrence of

two groups of clones suggested that GBSS II protein in cassava has two isoforms. This finding disagrees with previous report of one isoform of GBSS II in cassava by Munyikwa [14] and Munyikwa et al. [15]. In the present study, GBSS II gene sequence was derived from genomic DNA while Munyikwa [14] and Munyikwa et al. [15] derived GBSS II sequence from complimentary DNA. The complimentary DNA is less informative because it represents only the expressed portion of the gene while non-expressed part of the gene (intron) is excluded. In the case of genomic DNA (used in this study), both expressed and non-expressed parts of GBSS II gene are included in the sequence. To further support this finding, multiple isoform of GBSS II has been reported in pea, soyabean, rice, wheat sorghum, maize and cowpea [16, 17, 18]. To further confirm the identity of cassava GBSS I isoform, molecular studies have revealed that GBSSII proteins contain three conserved domains known as BOX I, BOX II and BOX III. From the primer pair location and through sequence analyses, HM038440 and HM046981 were located upstream the first conserved domains (BOX I) of GBSS II gene. BOX I contains 16 amino acids among which are KTGGL motif that is highly conserved among plants and bacteria starch synthases.

Phylogenetic analysis reveals relatedness of organisms based on the homology of their protein or DNA sequences. The phylogenetic study of higher plant GBSS II previously reported by Munyikwa et al.[15] and Ball and Morell [22] was expanded by including newly identified sequences from higher plants. The divergence of GBSS II into monocot and dicot groups has been documented [22]. The analysis showed that duplication plays a significant role in the emergence of groups and branches of GBSS II. Gene duplication events are the primary source of genetic novelty leading to speciation [23]. He explained further that after a duplication event, one daughter gene retains the preduplication function, while the other one, accumulates deleterious mutation and is eliminated, or, in some rare cases, survives by gaining a new function. The isolated cassava GBSS II sequences (HM038440 and HM046981) belongs to dicots group and are closely related to castor bean GBSS II gene. This is expected as cassava and castor plants belong to the same plant family Euphorbiaceae. Implication of this finding on future biotechnogical research on GBSS II gene is that any effective techniques for the gene modification in castor bean can be used for cassava with or without modification. Such techniques include GBSS II isoform discovery, gene cloning, downregulation and overexpression studies and functional characterization.

The observed amplification product size (1690 and 1698 bp) was larger than the expected amplification product size (363 bp) of the starch gene studied. Since genomic DNA was used as template DNA for PCR amplification, the difference in predicted and observed PCR product size could be attributable to the presence of introns in genomic DNA. An intron is any nucleotide sequence within a gene that is removed by RNA splicing to generate the final mature RNA product of a gene [24]. A total of six putative introns were identified in HM038440 and HM046981. This is expected to have wide applications on gene expression studies and starch modification. For instance, one of these introns is a component of a gene construct for cassava modification in our cassava transformation programme [25].The inclusion of introns in plant transformation constructs increased transformation efficiency [26,27]. The intron sequences can be used in the design of plant gene constructs to silence or over express starch genes in cassava or other plant genes for enhanced starch functionality [28]. The introns can also be used for expression studies of other genes in other plants as the significance of introns in gene expression has recently been established [24]. In addition to first reporting the presence of isoforms in cassava genome, future cassava biotechnological investigation will benefits from the result of this study in many ways. The nucleotides of the two isoforms of GBSS II gene (HM038440 and HM046981) can serve as effective sources of primer sequences for PCR amplification and gene cloning to obtain full-length sequences of the genes. Also the sequences of the two GBSS II isoforms reported in the study can be used to design probe for isolation of the gene from complementary and genomic DNA libraries. Furthermore, the sequences can be utilized to produce radioactive probe for nucleic acid hybridization to know the copy number and expression pattern of the genes

From the results of the present study on GBSS II gene in cassava, further studies are needed to compliment current knowledge and understanding on cassava GBSS II gene. These include isolation of the gene from complimentary or genomic DNA library, Southern hybridization, expression analyses of the GBSS II gene isoform by RNA blot analysis and functional characterization.

In summary, from a single primer pairs designed to amplify a portion (+3 and +369) of *gbss* II two different sequences (HM038440 and HM046981) were obtained. The two sequences had 99.4% and 84.6% nucleotides and amino acids identity. Phylogenetic analysis result showed that the two sequences belonged to dicots group but slightly different. In silico identification of introns revealed different number of introns in the two sequences: four in

HM038440 and two in HM046981. Taking together these results, it suggested that cassava genome encodes two isoforms of GBSS II gene.

CONCLUSION

The study described partial genomic organization of GBSS II in cassava. The study revealed that cassava genome contains two isoforms of GBSS II. The study also identified six introns in the genomic arrangement of GBSS II. The two isolated cassava *gbss* II sequences (HM038440 and HM046981) belonged to 3b branch in the dicot group which also contain *R. communis* and *V. vinifera gbss* II proteins. The introns will find applications in genetic modification of cassava for enhanced starch quality and in genetic manipulation of other crops.

Acknowledgement
The paper is a portion of the Ph.D Thesis of the first author. The Visiting Research Fellowship opportunity given to the first author by International Institute of Tropical Agriculture (IITA) is appreciated. The financial support of Obafemi Awolowo University Research Committee and National Biotechnology Development Agency (NABDA), Abuja is acknowledged.

REFERENCES

[1] Defloor, I., I. Dehing and J.A. Delcour. **1998**. *Starke,* 50:58-64
[2] Sanchez, T., E. Salcedo, H. Ceballos, D. Dufour and G. Mafla et al.. **2009**. *Starke*, 61:310.
[3] Visser, R.G., I. Somhorst, G.J. Kuipers, N.J. Ruys, W.J. Feenstra and E. Jacobsen. **1991**. *Mol. General Gen*., 225:289-296
[4] Blennow, A., **2003**. ***Starke***, 50:58-64.
[5] Ceballos, H., C.A. Iglesias, J.C. Pereze and A.G.O. Dixon. **2004.**. *Plant Mol. Biol*., 56: 503–516.
[6] Ceballos, H., T. Sanchez, N. Morante, M. Fregene and D. Dufour et al.. **2007**. *J. Agric. Food Chem*., 55:7469–7476.
[7] Ceballos, H., T. Sanchez, K. Denyer, A.P. Tofino and E.A. Rosero et al. **2008**. *J. Agric. Food Chem*., 56: 7215–7222.
[8] Taylor, N. Chavarriaga, P., Raemakers, K., Siritunga, D. and P. Zhang. **2004**. *Plant Mol. Biol*., 56:671-688
[9] Raemakers, K., M. Schreuder, L. Suurs, H. Furrer-Verhorst and J.P. Vincken. **2005**. *Mol. Breed*., 16:163-172.
[10] Ihemere, U., D. Ariaz-Garzon, S. Lawrence and R. Sayre. **2006.** *Plant Biotech. J*., 4:453-465
[11] Wang, S.J., K.W. Yeh and C.Y. Tsai.**1999**. *Physiol.Plant*., 106:253 – 261
[12] Opabode, J.T**.,** O. O.Oyelakin.,O. A. Akinyemiju and I. L. Ingelbrecht. **2011** *J. Plant Sci*. 6:174 -181
[13] Craig, J., J.R. Lloyd, K. Tomlinson, L. Barber and A. Edwards . **1998**. *Plant Cell*, 10: 413-426.
[14] Munyikwa, T. R. I. ,**1997**. Isolation and characterization of starch biosynthesis genes from cassava (*Manihot* esculenta Crantz). Ph. D. Thesis, Department of Plant Breeding, Wageningen Agricultural University, The Netherlands, 128pp.
[15] Munyikwa, T.R.I., E. Jacobsen and R.G.F. Visser. **1997**. *Afr. J. Root Tuber Crops*, 2: 116-120.
[16] Li, Z., F. Sun, S. Xu, X. Chu and Y. Mukai. **2003**. *Funct. Integr. Genomics*, 3, 76-81
[17] Isono, N., T. Senoura, M. Yoshikawa, Y. Sakurai, K. Watanabe, H. Ito and H. Matsui. **2004**. *J. Appl. Glycosci*., 51, 101-107
[18] Dry, I., A. Smith, A. Edwards, M. Bhattacharyya, P. Dunn and C. Martin. **1992**. *Plant J*. 2 , 193-202
[19] Dellaporta, S.L, J. Wood and J.B. Hicks. **1983**. *Plant Mol. Biol. Rep*., *1*:19-21
[20] Burge, C.B. and S. Karlin, **1998.** *Curr. Op. Struct. Biol.* 8: 346- 354
[21] Hanley, B.A. and M. A. Schuler. **1988**. *Nucl. Acid Res*., 16: 7159 – 7176.
[22] Ball, S.G. and M.K. Morell. **2003**. *Ann. Rev. Plant Biol*., 54:207 – 233
[23] Ohno, S., **1970**. Evolution by gene duplication, Springer-Verlag, Berlin, 125pp.
[24] Rose, A.B. **2008**. *Curr. Topics Microbiol.* Immunol., 326:277-290
[25] Opabode, J.T., **2010**. Development of transgenic cassava (*Manihot esculenta* Crantz) plants with potential for starch modification. Ph.D. Thesis, Obafemi Awolowo University, Nigeria, 172pp.
[26] Gonzalez, A.E., C. Schopke, N.J. Taylor, R.N. Beachy and C.M. Fauquet, **1998.** *Plant Cell Rep*., 17:827-831.
[27] Hankoua, B.B., N.J. Taylor, S.Y.C. Ng, I. Fawole and J. Puonti-Kaerlas, 2006. *Afr. J. Biotech*., 5:1700- 1712
[28] Chen, S., D. Hofius, U. Sonnewald and F. Bornke,. **2003.** *Plant J.*, 36:731-740.