Cloning and Sequence Analysis of the Nucleotide-Binding Domain of an α-Glucan, Water Dikinase Gene from Cassava (Manihot esculenta Crantz)

Jelili T. Opabode¹, Olufemi O. Oyelakin², Oluyemisi A. Akinyemiju¹ & Ivan L. Ingelbrecht³

¹ Department of Crop Production and Protection, Obafemi Awolowo University, Ile-Ife, Nigeria
² Central Biotechnology Laboratory, International Institute of Tropical Agriculture (IITA), Ibadan, Nigeria; c/o Lambourn Ltd, Carolyn House, 26 Dingwall Road, Croydon CR9 3EE, UK
³ IPBO, Department of Plant Biotechnology and Bio-informatics, Ghent University, Proeftuinstraat 86 N1, B-9000 Ghent, Belgium

Correspondence: Jelili T. Opabode, Department of Crop Production and Protection, Obafemi Awolowo University, Ile-Ife, Nigeria. E-mail: jopabode@yahoo.com

Received: January 29, 2013 Accepted: March 11, 2013 Online Published: November 15, 2013
doi:10.5539/jas.v5n12p56 URL: http://dx.doi.org/10.5539/jas.v5n12p56

Abstract
Several key genes involved in starch biosynthesis have been identified in cassava. However, while phosphorylation has been recognized as an essential step in starch metabolism in higher plants, the underlying gene(s) in cassava have not been isolated so far. To gain insights into starch phosphorylation in cassava, we produced three genomic clones encoding a fragment of an α-glucan, water dikinase (GWD), the primary enzyme required for starch phosphorylation. Sequence analysis showed that each of the clones contained the nucleotide-binding domain of the C-terminus of plant GWDs. The three genomic gwd clones had 98.8% homology at both nucleotide and amino acid levels and represented two distinct allelic sequences with two amino acid substitutions. The shorter clone, pOYE308-1, was 561 base pairs long and encoded a polypeptide of 106 amino acids with a predicted molecular weight of 180.3kDa. Two putative introns were identified in each of the gwd clones. Phylogenetic analysis of the cassava GWD sequences with other plant GWDs revealed that the cassava GWD belongs to the same group as that of castor bean, potato, tomato and tobacco. These resources add to the current knowledge base of starch metabolism in cassava and expand the molecular tool box for starch modification in this important tropical root crop.

Keywords: cassava, α-glucan, water dikinase, Manihot esculenta, phosphorylation, starch

Abbreviations: cDNA-Complementary DNA, GWD-α-glucan, water dikinase, ORF-Open Reading Frame

1. Introduction
Cassava is a starchy root crop and constitutes an important part of the diet for over 600 million people in sub-tropical regions (Defloor et al., 1998). The high starch content of cassava roots (20-40%, fresh weight), makes it also a desirable crop for industrial, non-food applications (Tonukari, 2004). Starch phosphorylation is an integral part of starch metabolism (Nielsen et al., 1994; Wischmann et al., 1999). The level of phosphorylation of starch is an important determinant of its physico-chemical properties such as pasting, gelling and stickiness (Hoover et al., 2001). A high phosphate content is desirable for many industrial applications (Blennow et al., 2002). There is need to better understand the role of starch phosphorylation in cassava as a potential avenue to manipulate starch metabolism and its physico-chemical and biological properties.

Cassava has limited natural variation with regards to starch metabolism compared to other crops such as maize. To date, only two starch mutants have been reported: a natural mutation in a granule-bound starch synthase (gbss) gene resulting in the production of amylose-free starch and a gamma irradiation-induced mutation in an isoamylase gene resulting in high-amyllose starch (Ceballos et al., 2007; Ceballos et al., 2008). Because most mutations affecting starch metabolism are recessive, moving useful genetic variants into the appropriate genetic background via traditional breeding is arduous. Barriers to conventional breeding and selection of cassava include its high heterozygosity, clonal propagation mode and poor flowering for most elite varieties and...
landraces (Ceballos et al., 2004). Recent progress in cassava bio-engineering whereby genes can be introduced into farmer-preferred varieties can provide an alternative approach for cassava improvement (Zainuddin et al., 2012).

In higher plants, starch is synthesized in the plastids of both photosynthetic and non-photosynthetic organs, including leaves, roots, and seeds. There appears considerable variation in the factors controlling starch metabolism in different organs and between different plant species (for a review, see Zeeman et al., 2010). Thus, the modification of starch properties in cassava will require adequate information on the genes involved from this crop. In cassava, the genes or cDNAs encoding the granule-bound starch synthase I and II involved in amylose synthesis (Salehuzzaman et al., 1993; Munyiwa et al., 1997), and the Starch Branching Enzymes (SBE) (Salehuzzaman et al., 1992; Baguma et al., 2003) have been characterized. Using crop bioengineering techniques, transgenic cassava plants producing amylose-free starch have been produced (Raemakers et al., 2005; Koehorst-van Putten et al., 2012). Transgenic cassava plants with enhanced starch production have also been reported (Ihemere et al., 2006). However, the genes underlying starch phosphorylation in cassava have not yet been characterized. Further manipulation of starch characteristics in cassava, particularly the level of phosphorylation, requires the molecular characterization of the underlying gene(s) (Opabode, 2010).

The primary enzyme required for starch phosphorylation, \( \alpha \)-glucan, water dikinase (GWD), formerly termed R1, was first discovered in potato (Lorberth et al., 1998). The GWD protein consists of a dikinase domain and a large (approximately 120kDa) N-terminal domain (Smith et al., 2005). It mediates starch breakdown through glucan phosphorylation (Hejazi et al., 2012; Stitt & Zeeman, 2012). GWD homologues have been reported in a variety of plant species including sweet potato, yam, maize, barley, and banana (Ritte et al., 2000; Smith et al., 2005). It has been suggested that GWD is ubiquitous and exerts a general function throughout the plant kingdom, even though some plants (e.g. maize and barley) synthesize storage starch with low or undetectable levels of phosphate (Mikkelsen et al., 2004).

Suppression of GWD synthesis in potato using antisense technology resulted in a 90% decrease in starch-bound phosphate (Lorberth et al., 1998; Viksø-Nielsen et al., 2001). Similarly, a mutation in a related gene in Arabidopsis, termed \( sex1 \), leads to a decrease in the phosphate content of leaf starch (Yu et al., 2001). Down-regulation of GWD by RNAi constructs in wheat resulted in increased vegetative biomass and 29% increased grain yield in pot trials (Ral et al., 2012). Also, transgenic Arabidopsis and maize lines engineered to reduce expression of GWD homologues displayed elevated leaf starch content (Weise et al., 2012).

The presence and cloning of GWD in cassava has not been reported so far. While 22 cDNA sequences of genes involved in starch and sucrose metabolism pathways in cassava have been characterized (Sakurai et al., 2007), GWD was not included. To further our understanding of starch biosynthesis in cassava and to facilitate the production of starch with modified phosphorylation levels, there is need to isolate and characterize \( \alpha \)-glucan, water dikinase as the primary enzyme required for starch phosphorylation. In this communication, we report three genomic clones encoding a fragment of an \( \alpha \)-glucan, water dikinase (GWD) gene from cassava. The sequence of the nucleotide-binding domain located at the C-terminus is presented and discussed in the context of GWD genes from other plant species.

2. Materials and Methods

2.1 Plant Materials and DNA Extraction

DNA was extracted from young leaves (0.5–1.0 g) of field-grown cassava genotype TMS 4(2)1425 as described by Dellaporta et al. (1983). DNA was resuspended at a concentration of 500 ng/μl. The quality of the DNA was verified by running 2 μl of the DNA alongside a molecular weight marker λPst I on 0.8% agarose gel in 1 x TAE (Tris Acetate EDTA) buffer at 500 V for one hour.

2.2 Primer Design

The ORF of Arabidopsis thaliana GWD 1 (NM_100952) sequence was used as query to search for following GWD homologues from higher plants: AtGWD 2 (AF312027) and AtGWD3 (NM_118578) from Arabidopsis thaliana; CrGWD (AY094062) from Citrus reticulate; GmGWD (AK285534) from Glycine max; StGWD (Y09533) from Solanum tuberosum; SgGW (EU908574) from Solanum lycopersicum and OsGWD (NM0010642343) from Oryza sativa. Multiple sequence alignments of the sequences were conducted. Nine primer pairs (two degenerate and seven non-degenerate) were designed from the conserved regions of the alignment using Lasergene sequence analysis software (DNASTAR Inc, Madison, USA). Details of the primer pairs are presented in Table 1. Primers were synthesized by Integrated DNA Technologies Incorporation (Iowa, USA).
Table 1. Primer pairs used for amplification of α-glucan, water dikinase homologues from cassava

<table>
<thead>
<tr>
<th>Primer name</th>
<th>sequence</th>
<th>Expected product size (bp)</th>
</tr>
</thead>
</table>
| gwd1        | F: GGAAAGTGAGGCHCAGAAGT  
               R: AGATCTGCACTGGAATGAAC | 635 |
| gwd2        | F: GACAGAACCCHTCTGCACT | 1169 |
| gwd3        | F: GATGAGAGACGATACTTCAG  
               R: CATTGGAATCAGATCGGAAGA | 335 |
| gwd4        | F: AGCTAAGAGATTGTTCATG  
               R: CTGACCTGTGTTATCCACTT | 652 |
| gwd5        | F: TCTGCCAGCTGTCCTTCAAC  
               R: GCAGCACTTACCTCTTCATC | 1150 |
| gwd6        | F: AGGGCTTCTCTTCACAGAAG | 481 |
| gwd7        | F: TCTGCCAGCTGTCCTTCAAC  
               R: GCAGCACTTACCTCTTCATC | 1150 |
| gwd8        | F: AGGGCTTCTCTTCACAGAAG  
               R: TCACCAACCCAAACACCTTC | 481 |
| gwd9        | F: TCTGCCAGCTGTCCTTCAAC  
               R: GCAGCACTTACCTCTTCATC | 1150 |
|             | F: AGGGCTTCTCTTCACAGAAG  
               R: TCACCAACCCAAACACCTTC | 481 |

H - A, T, C; M - A, C.

2.3 PCR Amplification and Cloning

Amplification of gwd was carried out in a 10 µl-reaction volume, composed of 1 µl of 10x buffer, 0.5 µl of MgCl₂ (25 mM), 1 µl each of primer F and R (1 µM), 0.5 µl of dNTPs (2.5 mM), 1 µl template DNA (500 ng), 4.8 µl H₂O and 0.2 µl of Taq DNA polymerase (5 U) (Bioline Inc., USA). The PCR amplification profile consisted of initial denaturation at 94°C for one minute and 30 cycles of amplification (94 °C for 30 seconds, 55°C for 30 seconds, 72°C for 45 seconds) with a final cycle of 5 minutes at 72°C. All PCR amplifications were carried out in a Peltier Thermal Cycler (PTC 2000, MJ Research, India). The PCR fragments were purified and cloned into pDRIVE vector (QIAGEN, CA, USA). The presence of the insert in the recombinant plasmid was confirmed by restriction digestion.

2.4 DNA Sequencing and Analyses

For each PCR reaction, two to three independent clones were sequenced. Both strands of the DNA inserts were sequenced and any sequence ambiguities were resolved by re-sequencing. The sequences were manually edited and vector sequences removed. DNA sequencing was performed by Iowa State University, USA. Putative introns were identified in silico using the GENSCAN web server (Burke and Karlin, 1998). All DNA sequences have been submitted to GenBank and their accession numbers are as follows: clone pOYE308-1 (accession no. HM046983), clone pOYE308-2 (accession no. HM046984) and pOYE308-3 (accession no. HM046985).

2.5 Phylogenetic Analysis

Sequence searches of the non-redundant and unfinished genome databases at NCBI were conducted with the entire AtGWD1, AtGWD2 and AtGWD3 ORFs as query sequences. From each selected protein sequence, we used exclusively the sequence aligning with the deduced amino acid sequences of cassava gwd clones (pOYE308-1, pOYE308-2 and pOYE308-3) for construction of the unrooted tree. The ClustalX package was used to create an alignment of the sequences that was then subjected to a neighbour-joining analysis to generate a branching pattern. The phylogenetic tree was displayed using the CLC TREEVIEW program.

3. Results
3.1 Amplification, Isolation and Characterization of Cassava Gwd Clones

Two classes of GWD proteins have been reported in *Arabidopsis* and other higher plants (Glaring et al., 2007). However, there is no information on the existence and number of GWD isoforms in cassava. To isolate GWD homologues in cassava, we designed nine primer pairs from consensus sequences of the two groups of GWD available at GenBank; *gwd 1 to gwd 4* from the first group; and *gwd 5 to gwd 9* from the second group (see Table 1) to conduct PCR assays on cassava total DNA (Figure 1). Overall, twenty-seven independent genomic clones were obtained and sequenced, three each for the nine GWD primer pairs. Only three clones (*pOYE308-1, pOYE308-2* and *pOYE308-3*) contained gene sequences homologous to previously reported *gwd* genes as indicated by BLAST searches. The three cassava GWD homologues were isolated using primer pairs *gwd 7* and *gwd 9* (Table 1). Nucleotide and deduced amino acid sequences showed that *pOYE308-2* and *pOYE308-3* had identical inserts while the insert in *pOYE308-1* showed 98.8% nucleotide identity with either *pOYE308-2* or *pOYE308-3* (Figure 3), resulting in a single amino acid substitution, leucine to phenylalanine, between *pOYE308-1* and *pOYE308-2*.

![PCR amplification of gwd gene sequences in cassava. M – λPstI DNA size markers. Lanes 1 to 9: PCR products amplified with primer gwd1 (lane 1); gwd2 (lane 2); gwd3 (lane 3); gwd4 (lane 4); gwd5 (lane 5); gwd6 (lane 6); gwd7 (lane 7); gwd8 (lane 8); and gwd9 (lane 9)](image)

The insert in clone *pOYE308-1* (accession no. HM046983) was 561 bp long encoding a polypeptide of 106 amino acids with a predicted molecular weight of 180.3 kDa. The partial GWD protein encoded by *pOYE308-1* was rich in alanine (9.4%), serine (9.4%), valine (9.4%) and glycine (7.8%). BLAST searches using this insert retrieved the *gwd* sequence of *Ricinus communis* (castor bean) as the most significant hit with 3e-56 E-value. The insert in *pOYE308-1* contained a conserved nucleotide-binding domain (Figure 3) present in GWD proteins and dikinases such as pyruvate phosphate dikinase (EC 2.7.9.1) and pyruvate water dikinase (EC 2.7.9.2) which is located at the C-terminus. Two putative introns were identified in *silico* between nucleotide 158-261 and 356-493 (Figure 2).
Figure 2. Partial genomic sequence of an α-glucan, water dikinase homologue in cassava derived from pOYE308-1 (Acc No. HM046983). The deduced amino acid sequence is indicated along with the two putative introns which are underlined.

Nucleotide –binding domain

Figure 3. Amino acid sequence alignment of the nucleotide-binding domain of cassava, castor bean and tobacco GWDs. Meseculenta (HM046983), Manihot esculenta GWD homologue 1; Meseculenta (HM046984), Manihot esculenta GWD Homologue 2; Rcommunis-GWD, Ricinus communis GWD, accession no. XM_002527856; Nsanderae-GWD, Nicotiana sanderae GWD, accession no. DQ021469
Clones pOYE308-2 (accession no. HM046984) and pOYE308-3 (accession no. HM046985) were 566 bp long each and also encoded a polypeptide of 106 amino acids with a predicted molecular weight of 181.9 kDa. Each of the clones was rich in alanine (9.4%), serine (9.4%), valine (9.4%) and glycine (7.8%). BLAST searches with the inserts of pOYE308-2 and pOYE308-3 also retrieved the gwd sequence of *R. communis* as their most significant hit with 7e-58 E-value. The inserts in pOYE308-2 and pOYE308-3 also contain the conserved nucleotide-binding domain of dikinases. Next, an alignment was made of the polypeptide sequences of the two distinct cassava GWD homologues and the corresponding region from *R. communis* and *Nicotiana sanderae*. As shown in Figure 3, the cassava GWD homologues comprise two distinct sequences which are 93% and 94% identical to GWD from *R. communis*. BLAST searches of the cassava genome database with the GWD sequences isolated in this study further identified two regions. The first region was on scaffold 12211 with E-value of 0 located between nucleotides 47,877 and 48, 437. The second region was on scaffold 02895 with E-value of 6.8e-18 located between nucleotides 27,5001 and 32, 500.

3.2 Phylogenetic Analysis

Next, a phylogenetic tree was constructed using the cassava GWD sequences in comparison to the homologous GWD sequences present in *Arabidopsis*, castor bean, sorghum, tobacco, citrus, tomato and potato. These sequences contain the highly conserved C-terminal ATP-binding domain of higher plant GWD sequences. The resulting domain tree, illustrated in Figure 4, shows the classification of GWD sequences into two main groups and further shows that the two isolated cassava GWD sequences belong to the subgroup comprising the GWD of *R. communis*.

![Figure 4. Phylogenetic relationship of GWD sequences of higher plants. The domain tree was constructed using selected GWD genomic sequences from GenBank and from this study (cassava) encoding the highly conserved C-terminal ATP binding domain using neighbour-joining method from bootstrapped data sets. The number of bootstrap replicates is indicated next to each branch. The abbreviations and GenBank accession numbers are as follows: Athaliana-GWD1, Arabidopsis thaliana GWD 1 sex1, accession no. NM_100952; Athaliana-GWD2, A. thaliana GWD homologue 2, accession no. AF312027; Athaliana-GWD3, A. thaliana GWD homologue 3, accession no. NM_118578; Nsanderae-GWD, Nicotiana sanderae GWD, accession no. DQ021469; Mesculenta (HM046983), Manihot esculenta GWD Homologue 1; Mesculenta (HM046984), Manihot esculenta GWD Homologue 2; Recommunis-GWD, Ricinus communis GWD, accession no. XM002527856; Osativa-GWD, Oryza sativa GWD, accession no. NM01064234; Slycopersicum-GWD, Solanum lycopersicum GWD, accession no. EU908574; Stuberorum-GWD, Solanum tuberosum GWD, accession no. YO9533; Hvilgare-GWD, Hordeum vulgare GWD, accession no. FN 179402.](image-url)

4. Discussion

In this study, the presence, isolation and cloning of a partial genomic clone encoding an α-glucan, water dikinase (GWD), a key starch phosphorylating enzyme, in cassava is reported for the first time. Three genomic clones containing GWD sequences were obtained from degenerate primers designed from two consensus sequences of
GWD sequences found in higher plants. BLAST searches confirmed the identity of the isolated GWD sequences. Phylogenetic analysis of the isolated cassava GWD sequences with the corresponding GWD domain of other plant species suggest that they belong to the same subgroup comprising the GWD from *R. communis*. Further analysis using other or complete GWD sequences from cassava should be carried to substantiate this finding.

Members of the GWD gene family have been suggested to be evenly distributed in the plant kingdom, including some plants whose storage starch contains a very low level of phosphate (Blennow et al., 2002). In *Arabidopsis*, three isoforms of GWD have been reported (Baunsgaard et al., 2005; Glaring et al., 2007). ESTs encoding GWDs are in *GenBank* for soybean, barley, maize, grape, rape, tomato, potato and wheat. GWDs in higher plants reportedly have four conserved domains: a chloroplast transit peptide, a starch binding, a phospho-histidine and a nucleotide binding domain (Baunsgaard et al., 2005). The identity of the cassava GWD sequences was established by the presence of a conserved nucleotide-binding domain that is common to both glucan water dikinase (GWD), pyruvate and phosphate dikinase (PPDK). The GWD family members have been reported to show regional sequence similarities to the dikinases in the so called PEP family, which includes PPDK, bacterial PPS and enzyme I (EI) of the PEP-dependent phosphotransferase system (Yu et al., 2001; Blennow et al., 2002; Mikkelsen et al., 2004). As a result, the overall reaction mechanism of GWD is similar to that of PPS and PPDK, but the GWD family appears to have arisen after divergence of the plant kingdom (Mikkelsen et al., 2004; Hejazi et al., 2012). In summary, the cloned cassava GWD sequences encode the C-terminus including the highly conserved nucleotide-binding domain. Our sequence data further suggests that cassava possesses at least two GWD isoforms.

The phylogenetic analysis grouped the GWD sequences from higher plants into two groups, in agreement with previous reports that have broadly classified higher plants GWD sequences into two subgroups: the first subgroup consists of AtGWD 1, AtGWD 2, AtGWD3 and many higher plants GWDs and the second group comprises many other higher plant GWDs (Mikkelsen et al., 2004; Baunsgaard et al., 2005; Glaring et al., 2007). The cassava GWD sequences obtained in this study belong to the second subgroup. As expected, cassava GWD was closely related to that of castor bean, potato, tobacco and tomato.

The putative introns in the *gwd* sequences were identified in silico through GENSCAN. The introns contain the splice site sequences consistent with the consensus 5′-GT_AG-3′ (Hanley & Schuler, 1988). Intron sequences are widely used in crop bio-engineering constructs to increase transformation efficiency (Gonzalez et al., 1998; Hankoua et al., 2006), in RNAi studies to silence genes or for overexpression studies in cassava or other plants (Chen et al., 2003; Rose, 2008).

5. Conclusion

In this study, we report the existence of GWD homologues in cassava. We also present partial genomic sequences encoding an α-glucan water dikinase, the phosphorylation and degradation enzyme in cassava for the first time. Comparative sequence analysis analysis suggests that the cassava genome encodes at least two isoforms of GWD, assuming that the sequences cloned here represent actively transcribed and functional genes. This study provides basic information to facilitate the isolation of the complete genomic sequence and expression analysis of *gwd* gene(s) in different cassava tissues. Such knowledge is expected to open up new opportunities to produce modified starches via bio-engineering of cassava.

Acknowledgements

This paper is a portion of the Ph.D Thesis of the first author. The Visiting Research Fellowship given to the first author by the International Institute of Tropical Agriculture (IITA), Ibadan is appreciated. The financial support of Obafemi Awolowo University’s Research Committee is acknowledged.

References


Copyrights
Copyright for this article is retained by the author(s), with first publication rights granted to the journal. This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (http://creativecommons.org/licenses/by/3.0/).