KIMIKA Volume 27, Number 2, pp. 14-27 (2016)

ISSN 0115-2130 (Print); 2508-0911 (Online)
https://doi.org/10.26534/kimika.v27i2.14-27

Identification of Glucoamylase cDNA Sequence of *Saccharomycopsis (Syn. Endomycopsis) bubodii* 2066

Joel H. G. Tolentino*, Kevin L. Labrador, Jennifer P. Fronteras, Lani L. R. Bullo, Leslie P. M. Cancio, Joanne J.J. Añonuevo, Gabriel P.G. Eleria and Annabelle U. Novero

College of Science and Mathematics, University of the Philippines Mindanao, Davao City 8022

Saccharomycopsis (Syn. Endomycopsis) bubodii 2066 is an isolate from *bubod*, a starter used in making rice wine in northern Philippines. We have shown that the yeast has amylolytic activity on raw sago starch. In our attempt to identify the putative raw starch-digesting amylase in *S. bubodii*, we determined the cDNA sequence of a glucoamylase gene. One primer pair that was designed based on a glucoamylase of *Saccharomycopsis fibuligera* HUT7212 (GLU1, NCBI Accession Number L25641.1) produced a sequence of 1234 base pairs. To obtain a wider coverage, a primer walking strategy was carried out using four primer pairs designed based on GLU1 gene. The generated sequence of 1535 base pairs shows 98.7 to 100% homology when aligned with glucoamylase genes from four strains of *S. fibuligera* suggesting that this glucoamylase is highly conserved between the *Saccharomycopsis* species. This work further reports a gene sequence of glucoamylase derived from Philippine-isolated yeast. The sequence is deposited in GenBank and assigned the accession number KP068007.1. The gene may be heterologously expressed in Sac*charomyces cerevisiae* for possible utilization in the direct conversion of raw sago starch to bioethanol.

Keywords: *glucoamylase; cDNA; primer walking; Saccharomycopsis bubodii; Saccharomycopsis fibuligera; sago starch*

INTRODUCTION

Saccharomycopsis (Syn. Endomycopsis) bubodii 2066 was isolated from *bubod*, a starter culture used in making the rice wine locally called *tapuy* in Tublay, Benguet, Philippines (PNMCC Directory of Strains, 2012). This strain of filamentous yeast was studied taxonomically and was identified as a new species of the genus *Endomycopsis* (*Saccharomycopsis*) by Sakai and Caldo (Sakai and Caldo, 1985b). The genetic diversity of yeast isolates from Philippine rice wine was analyzed using modified Random Amplification of Polymorphic DNA (RAPD) with 20-mer Seoulin Research Institute Life Science (SRILS) uniprimers 1, 6 and 9. Dendrogram analysis by NTSYS based on banding patterns generated through these uniprimers showed *S. bubodii* 2066 as divergent yeast with only around 43% similarity with the *Saccharomycopsis fibuligera strains* (Lim *et al.*, 2006).

* Author to whom correspondence should be addressed; email: jgtolentino3@up.edu.ph

S. bubodii 2066 is suspected to possess enzymes that have amylolytic property not only on rice starch but also on other sources such as the starch from the Sago palm. Sago starch can be obtained from the trunk of Sago palm (*Metroxylon sagu* Rottb.) which is an indigenous plant in Mindanao (Flores, 2008). The palm offers a high starch yield, thus, a plant of prime economic importance (Flach, 1997).

Glucoamylase, GA (synonyms amyloglucosidase, glucogenic enzyme, starch glucogenase and γ-amylase; exo-1,4-*a*-D-glucan glucanohydrolase; EC 3.2.1.3) is an exo-acting glycoside hydrolase. It digests α-1,4 and α-1,6 linkages of starch, glycogen, and similar carbohydrates from the non-reducing end. It is significantly utilized for the commercial production of glucose (Pandey *et al*., 2000; Sauer *et al*., 2000; Kumar & Satyanarayana, 2009).

The yeast *Saccharomycopsis fibuligera* has been used as a source of starch-digesting enzymes particularly glucoamylase (Hostinova, 2002; Chi *et al*., 2009). To date at least five strains of this yeast are reported to secrete glucoamylase with the corresponding cDNA encoding the amylase being partially or completely elucidated (GenBank, NCBI; https://www.ncbi.nlm.nih.gov/genbank/). These strains are HUT7212 (Itoh *et al*., 1987), KZ (Hostinova *et al*., 1991), IFO 0111 (Hostinova *et al*., 2003), R64 (Natalia *et al*., 2011) and PD70 (https://www.ncbi.nlm.nih.gov/nuccore/JF751023.1). The accession numbers assigned by NCBI to the glucoamylase cDNAs are L25641.1, X58117.1, HQ415729.1, AJ311587.1 and JF751023.1, respectively.

This study showed that *Saccharomycopsis bubodii* 2066 is a potential source of raw-starch digesting amylase (RSDA). It further decoded the cDNA sequence of the glucoamylase, a putative RSDA, in *S. bubodii* 2066 by the primer walking strategy. The primers used were designed based on published glucoamylase cDNA of *S. fibuligera* strain HUT7212. The almost complete sequence was elucidated through reverse transcription reaction, polymerase chain reaction (PCR), DNA sequencing and application of readily available and accessible bioinformatics software.

MATERIALS AND METHODS

Cell Culture. *Saccharomycopsis bubodii* 2066 was purchased from the Philippine National Collection of Microorganisms (PNCM), BIOTECH, University of the Philippines Los Baños, Laguna. The yeast was cultured in YMP broth, which was composed of 0.3% yeast extract (Laboratorios Conda, Spain), 0.3% malt extract (Difco Laboratories, Detroit, Michigan) and 0.5% peptone (HiMedia Laboratories Pvt. Ltd., Mumbai, India), with 1% sago starch (Sago-Biotech Program, UPMindanao). It was maintained in YMP agar slants or plates. Liquid cultures were incubated at 30°C in a shaking incubator. Cells streaked on agar were grown at room temperature overnight and kept at 4°C for three months. All culture media were sterilized by autoclaving at 121°C for 15 min.

Screening for Amylolytic Activity. Screening for amylolytic activity was done initially using Lugol's Iodine Test on *S. bubodii* 2066 grown on YMP Agar with 1% raw sago starch. Formation of halo (zone of clearing) on the medium indicated starch hydrolysis by the isolate.

Production of Amylase. Twenty-five milliliters of YMP broth was inoculated with a loopful of the stock culture and incubated at 30°C for 24 h. This was used as the starter inoculum for amylase production. Subsequently, 225 mL of production medium (YMP broth without sago starch) in a 500-mL Erlenmeyer flask was sterilized. The sago starch was sterilized separately by dry sterilization in a convection oven at 180°C for 3 h. The liquid component of the medium was then aseptically added to the dry starch. The resulting medium was inoculated with 25 mL of the starter inoculum and incubated for 24 h at 30°C. The same procedure was done for the medium involving gelatinized starch except that the starch was added in the broth at the onset and gelatinization was achieved

during sterilization of the medium. The amylase was then harvested by centrifugation of the broth at 4°C and 10,000 x g for 10 min.

Amylase Activity Assay. Quantitative evaluation of amylolytic activity of *S. bubodii 2066* was measured by monitoring the increase in reducing sugar produced from the hydrolysis of starch by dinitrosalicylic acid (DNS) method. Five hundred microliters of 1% soluble starch was added with 50 µL of diluted enzyme (1:10 dilution) and incubated for 5 min at 30°C. The reaction was stopped by the addition of 1 mL of DNS reagent mix. The reducing sugar produced was quantified by a UV/VIS spectrophotometer (UV-1610A PharmaSpec, Shimadzu, Japan) at 500 nm wavelength using glucose as standard. The calibration curve was constructed from eight standard solutions of glucose with concentration range of 100 to 800 mg/L. All assays were done in triplicate. All assays were done in triplicate. One unit of enzyme activity is defined as the amount of enzyme required to produce 1µmol of glucose per min under assay conditions.

Protein Content Determination. The protein content of the enzyme solution was determined using the Bradford method. Three milliliters of Bradford dye reagent was added to 60 µL of protein sample. The solution was mixed and allowed to stand at room temperature for 5 min. The absorbance of the solution was then read using a UV/VIS spectrophotometer at 595 nm using bovine serum albumin (BSA) as standard. The calibration curve was constructed from seven standard solutions of BSA with concentration range of 50 to 350 mg/L. Analysis was done in triplicate. The milligram protein obtained was used in the computation of the specific activity of the enzyme.

Total RNA Extraction and Characterization. The total RNA was obtained using Ambion PureLink RNA Mini Kit (Life Technologies, USA) by following the manufacturer's protocol with some modifications. For the cell lysis and homogenization, Zymolyase-20T (Nacalai Tesque, Inc., Japan) was prepared in a digestion buffer (1.0 M Sorbitol, 0.1 M EDTA, pH 7.5, 0.1% β-mercaptoethanol) to a concentration of 5.0 mg/mL. Approximately 500 million yeast cells were harvested by centrifugation. The pellet was resuspended in the Zymolyase solution. The suspension was then incubated at 30°C for 1 h in a heat block. After incubation, lysis buffer with 1.0% β-mercaptoethanol was added to the tube and mixed thoroughly. Afterwards, the tube was centrifuged and the supernatant was collected. Ethanol (100%) was added to the lysate and the mixture was subjected to RNA purification according to the kit's protocol. The RNA isolated was characterized by running an aliquot in 1% agarose gel stained with 10X GelRed (Biotium, USA). The gel was electrophoresed (100 V) in a mini gel electrophoresis system with 1X SB (Sodium-Borate) tank buffer. The bands were imaged using the Compact Digimage System (Major Science, Taiwan), and analyzed using the software, UN-SCAN-IT gel v. 6.1 (Silk Scientific Co., USA). In the succeeding optimization and amplification steps, an aliquot of the RNA was stored in -20 °C; the rest were stored in – 80 °C.

Primer Design. The primers were designed using the software Primer-BLAST available at the NCBI website (http://blast.ncbi.nlm.nih.gov/Blast.cgi; Ye *et al.*, 2012). To check the compatibility and annealing position of the designed primers, they were aligned with the design template by using ClustalΩ software (http://www.ebi.ac.uk/Tools/msa/clustalo/; Sievers *et al.*, 2011; McWilliam *et al.*, 2013). The primers were purchased from Diamed Enterprise (Los Baños, Laguna).

First strand cDNA Synthesis and PCR. The synthesis of the first strand cDNA was done using SuperScript III Reverse Transcriptase (Invitrogen, USA), following the kit's protocol with minor modifications. The primer used was the gene-specific reverse primer FSSP-R, 5'-GAGGAACTCGAGCCA AAGCCTTGACCTTATTTC-3' (Natalia *et al.* 2011). One hundred nanograms of total RNA were used as template. Complementary DNA

synthesis was accomplished at 55°C for 60 min. The cDNA produced was used as the template for PCR.

The kit GoTaq PCR Core System I (Promega, USA) was used for the PCR experiments, following the manufacturer's protocol with slight modifications. Per reaction, 10 ng of template cDNA and a total volume of 25 μL PCR cocktail with the following final concentration/amount of components: 1X Green Buffer, 1.50 mM $MgCl_2$, 200 μM dNTPs, 0.2 μM each of forward and reverse primers, 1 U Taq Polymerase, were used. The reaction was run in a Veriti 96-well Thermal Cycler with the following thermocycle conditions: initial denaturation (94°C, 2 min) followed by 30 cycles of denaturation (95°C, 1 min), annealing (50°C, 1 min), and extension (72°C, 1 min), and final extension (72°C, 5 min). The same thermocycle conditions were used for the other primer pairs. The primers used in decoding the glucoamylase cDNA with their codes and coordinates in the design template are presented in Table 1.

Table 1. Forward (F) and Reverse (R) Primers Used in the Amplification of Glucoamylase cDNA.

Primer Notation	Sequence (5' to 3')	Coordinates on the Template[1]
P1-F	ATTGCTTATCCTGAAGGCCA	190 −209
P1-R	AGCCAAAGCCTTGACCTTAT	1557 − 1538
PWA-F	TTCCGTCTTTGCTGCTATTGT	21 − 41
PWB-R	TGTAGTACTCAACTGCCTTGG	376 − 356
PWC-R	TGCCTTGGTTTTCCTCCCAA	724 −705
PWD-F	TTTTGACGACGGCGACTTTG	789 − 808
PWE-F	TGGTCACATTCGGTGATTCC	1379 − 1398

[1] *glucoamylase gene from Saccharomycopsis fibuligera HUT7217 (NCBI Accesion No. L25641.1)*

Visualization and Quantification of PCR Products. Products obtained from PCR were visualized using 1.5% agarose gel containing 10X GelRed. Samples were first mixed with Gel Loading Buffer (Sigma-Aldrich, USA) before loading into the wells. A 1 kb ladder (Promega, USA) was used as marker. The gel was electrophoresed at 50 V for 1.5 h and the resulting bands were imaged and analyzed with the same systems used above for the total RNA.

Nucleotide Sequence Analysis. PCR products were sent to Macrogen, Inc., Korea for purification and sequencing. Two sets (runs) of samples per primer pair were sent, serving as duplicates in the sequencing procedure. After the sequencing data were received, the chromatograms were cleaned-up using FinchTV v.1.4 (http://www.geospiza.com/Products/finchtv.shtml). The two sequences generated from each primer were aligned using ClustalΩ to assess for consistency. The sequences derived from the forward and reverse primer pairs were assembled to form a contig sequence using BioEdit v7.2.5 (Hall, 1999). Every contig sequence was ran on NCBI-BLAST (Altschul *et al.*, 1997) using the megablast option to validate the gene's identity. The nucleotide sequence was converted into its amino acid sequence through the ExPASy Bioinformatics Resource Portal Translate Tool (http://web.expasy.org/translate/). Both the nucleotide and amino acid sequences were aligned with the published glucoamylase sequences in the NCBI database using ClustalΩ to determine percent similarity/homology.

RESULTS AND DISCUSSION

Amylolytic Activity of Saccharomycopsis bubodii 2066. Figure 1A shows the formation

of halo by *S. bubodii* 2066 on YMP Agar added with 1% sago starch after staining with Lugol's reagent. Starch forms a blue-black complex with iodine. The zone of clearing indicated hydrolysis of the raw sago starch in the medium to simple sugars by *S. bubodii 2066*, hence, it is an amylolytic organism. This result agrees with earlier reports of Sakai and Caldo (1985a), Limtong *et al.* (2002) and Takeuchi *et al.* (2006) that fermentation starters are the repository of microbial amylase producers. Further investigation was conducted to determine and compare quantitatively the amylolytic activity of *S. bubodii* in two kinds of substrate preparations namely, raw and gelatinized sago starch. As can be seen in Figure 1B, *S. bubodii* 2066 showed greater preference for raw sago starch over gelatinized starch as substrate with specific activity of almost 3-fold greater. This result strongly suggests that the production of RSDA is induced in *S. bubodii* 2066. This is due to the fact that the organism was isolated from *Bubod* which is essentially a raw starch preparation (uncooked rice cake). In a related study, another yeast strain identified as *Saccharomycopsis fibuligera* 2074 which is also obtained from *Bubod*, displayed the same preference for raw sago starch (Bullo, 2009). Hence, the source of the microorganism has a profound influence on its RSDA activity.

Additionally, a number of researches show that the preference of amylases for the kind of substrate preparation (whether raw or gelatinized) may also be influenced by the starch source. Like *S. bubodii* 2066 and *S. fibuligera* 2074, the endophytic fungus *Acremonium* sp. favored raw sago starch but manifested low activity on raw corn, potato and wheat starch (Marlida *et al.*, 2000). This possible influence of starch source is also observed for *Bacillus* sp. I-3 (Goyal *et al.*, 2005), *Aspergillus niger* AM07 (Omemu *et al.*, 1999) and *Penicillium* sp. X-1 (Sun *et al.*, 2006).

These findings led us to investigate the putative RSDA gene in *S. bubodii* 2066 which may be used in heterologous gene expression in *Saccharomyces cerevisiae* for bioethanol production from raw sago starch.

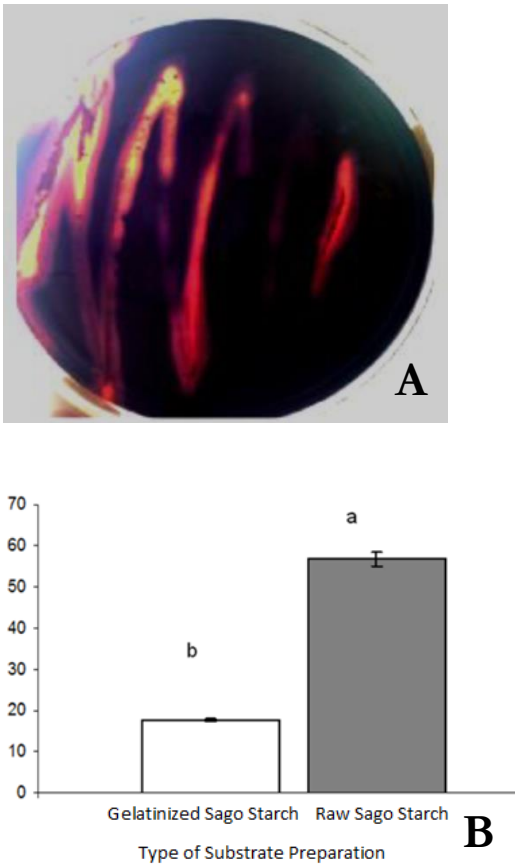

Figure 1. Amylolytic activity of Saccharomycopsis bubodii 2066 on sago starch. (A) Qualitative assessment by Lugol's reagent staining and (B) Specific activity measurement by Dinitrosalicylic (DNS) assay for reducing sugar.

Generation of Glucoamylase Gene Sequence by Primer Walking. Based on the growth pattern of *S. bubodii* established (data not shown), the cells were already in the log phase at 24 hours after the start of inoculation. Total RNA was isolated at this point to ensure RNA product of good quality. In Figure 2, the presence of two bands representing the 28S and 18S rRNAs confirm the successful isolation of total RNA from *S. bubodii* via the enzymatic method using Zymolyase. Zymolyase is an enzyme derived from *Athrobacter luteus* and was characterized to lyse yeasts cell membrane (Kitamura, 1972).

A number of primers were designed and tested to elucidate the sequence of putative RSDA gene(s) in *S. bubodii* 2066. One primer pair showed success: P1-F/P1-R (Table 1). These primers were designed using the glucoamylase gene, GLU1 (Accession No.

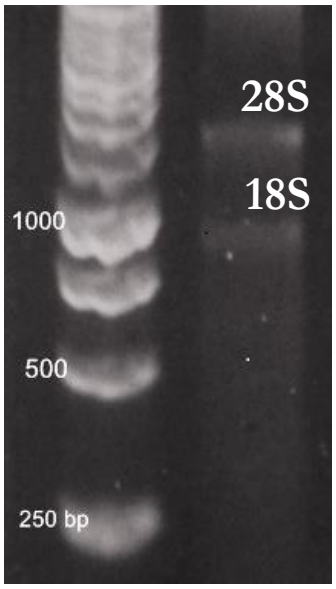

Figure 2. Agarose gel confirmation of total RNA extracted from Saccharomycopsis bubodii 2066.

L25641.1), as the template (Itoh *et al.*, 1987). The decision in selecting the primers from the list generated using the default settings of Primer-BLAST was based on two criteria: (a) the locations of the forward and reverse primers are at or close to the 5' end and 3' end of the template, respectively; and (b) both primers work not only on the design template, but also on glucoamylase templates that originated from other *S. fibuligera* strains.

After the first round of PCR followed by DNA sequencing, the expected size of 1400 base pairs was not achieved (Table 2 and Figure 3 lane E). Only 1234 base pairs (88% of the expected size or 80% of the design template size) were obtained. Alignment of the contig sequence generated showed very high homology with the template used in the design and three other glucoamylase sequences available in GenBank, NCBI.

Table 2. Primer Pairs and Their Expected and Actual Amplicon Sizes.

Primer Pairs		Estimated Size[1] (bp)	Actual Contig Size (bp)
P1-F	P1-R	1,400	1,234
PWA-F	PWB-R	350	359
PWA-F	PWC-R	700	501
PWD-F	P1-R	750	733
PWE-F	P1-R	150	177

[1] *PCR product size was estimated using the DNA ladder used.*

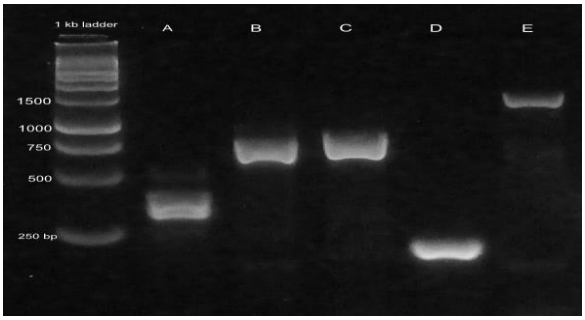

Figure 3. Agarose gel profile of PCR amplicons obtained from primer pairs used. (A) PWA-F and PWB-R (B) PWA-F and PWC-R, (C) PWD-F and P1-R, (D) PWE-F and P1-R, and (E) P1-F and P1-R (Refer to Table 1 for the primer coding).

The full-length sequence of the glucoamylase gene was difficult to obtain with just the P1-F/P1-R primer pair. Firstly, there are inherent limitations in PCR and Sanger sequencing associated with amplifying long gene fragments (> 1kb) which lead to errors in nucleotide base calls. Secondly, the 5' and 3' end of the gene was impossible to cover using the said primers. In order to increase the sequencing coverage and validate the sequences obtained, primer walking was done. Primer walking, also known as genome walking, is a DNA sequencing approach that comprises a number of PCR-based methods for the amplification of unknown genomic regions flanked by known sequences (Volpicella *et al.*, 2012; Li *et al.*, 2015).

Using both the contig sequence obtained from the first primer pair and the GLU 1 gene sequence, four more primer pairs were designed (Table 1) to: (a) partition the sequence length into several smaller, overlapping fragments; and (b) to decode the gene's 5' and 3' ends. The reverse primer for both primers PWD-F and PWE-F did not work (sequences not shown in the text) and it was decided to use P1-R instead which produced good quality PCR amplicons on the agarose gel (Figure 3 lanes C and D). The relative positions of the primer sequences relative to the sequence of the template used in the design are mapped in Figure 4.

Through first strand cDNA synthesis using the primer FSSP-R, PCR using a primer pair followed by DNA sequencing, a contig

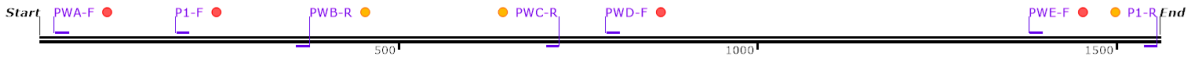

Figure 4. Map showing the location of primers relative to the design template used (1560 bp, represented by two thick horizontal lines). Refer to Table 1 for the sequences of the primers. Red or orange dot next to the primer code indicates the primer is forward or reverse, respectively. This map is generated using the SnapGene® Viewer free software (http://www.snapgene.com/products/snapgene_viewer/).

sequence was generated. Prior to DNA sequencing, all the PCR amplicons from all primer pairs were visualized in an agarose gel (Figure 3). The size of each amplicon can be estimated from the gel. The actual sizes of these amplicons after contig sequence generation are summarized in Table 2.

The two contig sequences obtained from the primer pairs closest to the 5' end (PWA-F/PWB-R and PWA-F/PWC-R) were aligned and put together first. The resulting contig sequence was aligned with the contig sequence generated from primer pair P1-F/P1-R, which served as the scaffold. This process was continued until the contig sequence from primer pair closest to the 3' end (PWE-F/P1-R) was considered. The whole process produced the longest possible sequence of 1535 bases. Aligning the overlapping fragments increased both the read depth at which base call was done and the total sequence length obtained. However, there was a need to truncate the chromatogram noise at the 5' and 3' ends of the alignment; hence, some bases at each end were not decoded: nineteen at the 5' end and six at the 3' end. Decoding these bases require the design of primers outside the open reading frame (ORF). Regardless, primer walking strategy generated a much more robust, longer sequence length.

Complementary DNA (cDNA) of S. bubodii 2066 Glucoamylase. The *S. bubodii* 2066 glucoamylase cDNA sequence decoded (1535 bp) is presented in Figure 5. The sequence is 98.4% of the expected full glucoamylase sequence of the open reading frame of GLU 1, primer design template used. The undecoded bases at each end are shaded in yellow in Figure 5. The sequence was deposited in GenBank, NCBI and was assigned an accession number KP068007.1.

Alignment of the sequence with four (including the design template sequence) out of five glucoamylase cDNAs from *S. fibuligera* strains obtained from GenBank showed high homology that ranges from 98.7% to 100% while alignment with a fifth glucoamylase sequence (accession #AJ311587.1; from strain IFO0111) showed homology of only 60.8%. This is illustrated further with a neighbor-joining (NJ) tree (Figure 6). The NJ tree was constructed in PAUP v.4.0b10 (Swofford, 2003) using the parameters for the best model obtained from jModelTest (Darriba *et al.*, 2012) to account for multiple hits. The tree shows *S. bubodii* 2066 clustering together with the four *S. fibuligera* strains. Surprisingly, the sequence is 100% homologous with the glucoamylase of *S. fibuligera* HUT7212. Based on the minimum spanning network (not shown), *S. fibuligera* IFO0111 separates from the rest for more than 500 nucleotide substitution steps, hence it was treated as the outgroup for rooting. Translation of the cDNA sequence into amino acid sequence produced the same magnitude of high sequence homology (98.4 to 100%) with the four suggesting that this gene is highly conserved in the *Saccharomycopsis* species.

Amino Acid Sequence of S. bubodii 2066 Glucoamylase. Although the glucoamylases from different *S. fibuligera* strains show high homology at the amino acid sequence level, they still exhibit differences in properties such as optimum pH and temperature, and molecular weight (Hostinova, 1998, 2002; Natalia *et al.*, 2011). Differences observed in the molecular weight can be due to the differences in post-translational modification particularly *N*-glycosylation of the enzyme in the different host strains (Itoh *et al.*, 1987; Gasperik *et al.*, 1991). Since the characterization of the glucoamylase from *S. bubodii* 2066 is yet to be investigated it cannot

```
2066        ----------------------T---TTCCGTCTTTGCTGCTATTGTTAGTGCTTTACCT   35
HUT7212     ---ATGAAATTCGGTGTTTTATT---TTCCGTCTTTGCTGCTATTGTTAGTGCTTTACCT   54
KZ          ---ATGAGATTCGGTGTTTTAAT---TTCCGTCTTTGTTGCTATTGTTAGTGCTTTACCT   54
R64         ------------------------------------------------------------   0
PD70        ------------------------------------------------------------   0
IFO0111     ATGATCAGATTGACAGTCTTTTTGACAGCAGTTTTTGCTGCAGTTGCTAGTT--------   52

2066        TTGCAAGAAGGTCCTTTGAACAAAAGAGCCTATCCTTCTTTTGAAGCTTATTCAAACTAT   95
HUT7212     TTGCAAGAAGGTCCTTTGAACAAAAGAGCCTATCCTTCTTTTGAAGCTTATTCAAACTAT   114
KZ          TTGCAAGAAGGTCCTTTGAACAAAAGAGCCTATCCTTCTTTTGAAGCTTATTCAAACTAT   114
R64         ---------------------------GCCTATCCTTCTTTTGAAGCTTATTCAAACTAT   33
PD70        ---------------------------GCCTATCCTTCTTTTGAAGCTTATTCAAACTAT   33
IFO0111     -GCGTCCCGGTGGAATTGGACAAAAGAAACACTGGACATTTCCAAGCTTATTCTGGATAC   111

2066        AAAGTTGACAGAACTGACTTGGAAACCTTCTTGGACAAACAAAAAGAAGTATCTTTATAC   155
HUT7212     AAAGTTGACAGAACTGACTTGGAAACCTTCTTGGACAAACAAAAAGAAGTATCTTTATAC   174
KZ          AAAGTTGACAGAACTGACTTGGAAACCTTCTTGGACAAACAAAAAGATGTATCTTTATAC   174
R64         AAAGTTGACAGAACTGACTTGGAAACCTTCTTGGACAAACAAAAAGATGTATCTTTATAC   93
PD70        AAAGTTGACAGAACTGACTTGGAAACCTTCTTGGACAAACAAAAAGATGTATCTTTATAC   93
IFO0111     ACAGTTGCCAGATCAAATTTCACTCAATGGATTCATGAGCAACCAGCTGTTTCTTGGTAT   171

2066        TATCTTTTACAAAACATTGCTTATCCTGAAGGCCAATTTAATAATGGTGTTCCTGGTACT   215
HUT7212     TATCTTTTACAAAACATTGCTTATCCTGAAGGCCAATTTAATAATGGTGTTCCTGGTACT   234
KZ          TATCTTTTACAAAACATTGCTTATCCAGAAGGCCAATTTAATGACGGTGTTCCCGGTACT   234
R64         TATCTTTTACAAAACATTGCTTATCCTGAAGGCCAATTTAATGATGGTGTTCCTGGTACT   153
PD70        TATCTTTTACAAAACATTGCTTATCCTGAAGGCCAATTTAATGATGGTGTTCCTGGTACT   153
IFO0111     TATCTTTTGCAAAACATTGATTATCCAGAAGGACAATTTAAATCTGCAAAGCCAGGCGTG   231

2066        GTTATTGCTTCTCCATCAACCTCTAATCCGGACTACTATTACCAATGGACCAGAGATTCC   275
HUT7212     GTTATTGCTTCTCCATCAACCTCTAATCCGGACTACTATTACCAATGGACCAGAGATTCC   294
KZ          GTTATTGCTTCTCCATCAACCTCTAATCCGGACTACTATTACCAATGGACCAGAGATTCC   294
R64         GTTATTGCTTCTCCATCAACCTCTAATCCGGACTACTATTACCAATGGACCAGAGATTCC   213
PD70        GTTATTGCTTCTCCATCAACCTCTAATCCGGACTACTATTACCAATGGACCAGAGATTCC   213
IFO0111     GTAGTTGCTTCTCCATCCACCTCAGAACCTGACTATTTTTATCAATGGACCAGAGACACT   291

2066        GCAATTACATTTTTGACAGTTCTTTCTGAACTAGAAGATAATAACTTC---AATACCACT   332
HUT7212     GCAATTACATTTTTGACAGTTCTTTCTGAACTAGAAGATAATAACTTC---AATACCACT   351
KZ          GCAATTACATTTTTGACAGTTCTTTCTGAACTAGAAGATAATAACTTC---AATACCACT   351
R64         GCAATTACATTTTTGACAGTTCTTTCTGAACTAGAAGATAATAACTTC---AATACCACT   270
PD70        GCAATTACATTTTTGACAGTTCTTTCTGAACTAGAAGATAATAACTTC---AATACCACT   270
IFO0111     GCCATTACATTTCTTTCGTTGATTGCCGAGGTTGAAGACCATAGCTTTAGCAATACCACC   351

2066        TTGGCCAAGGCAGTTGAGTACTACATTAACACCAGTTACAACCTTCAAAGAACCAGTAAC   392
HUT7212     TTGGCCAAGGCAGTTGAGTACTACATTAACACCAGTTACAACCTTCAAAGAACCAGTAAC   411
KZ          TTGGCTAAGGCAGTTGAGTACTACATTAATACCAGTTACAACCTTCAAAGAACCAGTAAC   411
R64         TTGGCTAAGGCAGTTGAGTACTACATTAATACCAGTTACAACCTTCAAAGAACCAGTAAC   330
PD70        TTGGCTAAGGCAGTTGAGTACTACATTAATACCAGTTACAACCTTCAAAGAACCAGTAAC   330
IFO0111     CTTGCCAAGGTCGTGGAATACTACATCAGCAACACCTACACTTTGCAAAGAGTTTCAAAC   411

2066        CCAAGTGGCAGCTTTGATGATGAAAATCATAAAGGCTTGGGAGAACCAAAATTTAACACA   452
HUT7212     CCAAGTGGCAGCTTTGATGATGAAAATCATAAAGGCTTGGGAGAACCAAAATTTAACACA   471
KZ          CCAAGTGGCAGCTTTGATGATGAAAATCATAAAGGCTTGGGAGAACCAAAATTTAACACA   471
R64         CCAAGTGGCAGCTTTGATGATGAAAATCATAAAGGCTTGGGAGAACCAAAATTTAACACA   390
PD70        CCAAGTGGCAGCTTTGATGATGAAAATCATAAAGGCTTGGGAGAACCAAAATTTAACACA   390
IFO0111     CCAAGTGGAAATTTCGACAGTCCTAACCACGACGGTTTGGGAGAACCAAAGTTCAATGTT   471

2066        GATGGTTCTGCATACACCGGAGCTTGGGGGAGACCGCAAAATGATGGTCCTGCTTTGAGA   512
HUT7212     GATGGTTCTGCATACACCGGAGCTTGGGGGAGACCGCAAAATGATGGTCCTGCTTTGAGA   531
KZ          GATGGTTCTGCATACACCGGAGCTTGGGGGAGACCGCAAAATGATGGTCCTGCTTTGAGA   531
R64         GATGGTTCTGCATACACTGGAGCTTGGGGGAGACCGCAAAATGATGGTCCTGCTTTGAGA   450
PD70        GATGGTTCTGCATACACTGGAGCTTGGGGGAGGCCGCAAAATGATGGTCCTGCTTTGAGA   450
IFO0111     GACGACACCGCCTACACAGCTTCTTGGGGCAGACCTCAAAATGATGGCCCAGCTTTAAGA   531
```

```
2066        GCTTATGCTATCAGTAGATACTTGAATGATGTCAATTCTTTAAATGAAGGTAAATTAGTA   572
HUT7212     GCTTATGCTATCAGTAGATACTTGAATGATGTCAATTCTTTAAATGAAGGTAAATTAGTA   591
KZ          GCTTATGCTATCAGTAGATACTTGAATGATGTCAATTCTTTAAATAAAGGTAAATTAGTA   591
R64         GCTTATGCTATCAGTAGATATTTGAATGATGTCAATTCTTTAAATGAAGGTAAATTAGTA   510
PD70        GCTTATGCTATCAGTAGATACTTGAATGATGTCAATTCTTTAAATGAAGGTAAATTAGTA   510
IFO0111     GCTTATGCCATTTCCAGATATTTGAATGCTGTGGCCAAACATAACAATGGCAAATTGTTG   591

2066        TTGACTGATTCAGGTGATATCAACTTTTCTTCAACTGAAGATATTTACAAAAATATCATC   632
HUT7212     TTGACTGATTCAGGTGATATCAACTTTTCTTCAACTGAAGATATTTACAAAAATATCATC   651
KZ          TTGACTGATTCAGGTGATATCAACTTTTCTTCAACTGAAGATATTTACAAAAATATCATC   651
R64         TTGACTGATTCAGGTGATATCAACTTTTCTTCAACTGAAGATATTTACAAAAATATCATC   570
PD70        TTGACTGATTCAGGTGATATTAACTTTTCTTCAACTGAAGATATTTACAAAAATATCATC   570
IFO0111     CTCGCCGGCCAAAACGGAATCCCTTATTCTAGTGCTTCTGACATTTATTGGAAAATTATT   651

2066        AAACCAGACTTGGAATATGTTATAGGGTACTGGGATTCTACTGGGTTTGATCTTTGGGAG   692
HUT7212     AAACCAGACTTGGAATATGTTATAGGGTACTGGGATTCTACTGGGTTTGATCTTTGGGAG   711
KZ          AAACCAGACTTGGAATATGTTATAGGGTACTGGGATTCTACTGGGTTTGATCTTTGGGAG   711
R64         AAACCAGACTTGGAATATGTTATAGGGTACTGGGATTCTACTGGGTTTGATCTTTGGGAG   630
PD70        AAACCAGACTTGGAATATGTTATAGGGTACTGGGATTCTACTGGGTTTGATCTTTGGGAG   630
IFO0111     AAACCAGACTTGCAACATGTCAGCACCCATTGGAGCACCTCTGGCTTTGATCTTTGGGAA   711

2066        GAAAACCAAGGCAGACACTTTTTTACAAGCTTGGTTCAACAGAAAGCCCTTGCTTATGCT   752
HUT7212     GAAAACCAAGGCAGACACTTTTTTACAAGCTTGGTTCAACAGAAAGCCCTTGCTTATGCT   771
KZ          GAAAACCAAGGCAGACACTTTTTTACAAGCTTGGTTCAACAGAAAGCCCTTGCTTATGCT   771
R64         GAAAACCAAGGCAGACACTTTTTTACAAGCTTGGTTCAACAGAAAGCCCTTGCTTATGCT   690
PD70        GAAAACCAAGGCAGACACTTTTTTACAAGCTTGGTTCAACAGAAAGCCCTTGCTTATGCT   690
IFO0111     GAAAATCAAGGAACTCATTTCTTCACTGCTTTGGTTCAACTCAAAGCTCTTAGCTACGGT   771

2066        GTCGATATTGCCAAAAGTTTTGACGACGGCGACTTTGCGAACACACTTTCTTCGACTGCT   812
HUT7212     GTCGATATTGCCAAAAGTTTTGACGACGGCGACTTTGCGAACACACTTTCTTCGACTGCT   831
KZ          GTCGATATTGCCAAAAGTTTTGACGATGGCGACTTTGCGAACACACTTTCTTCGACTGCT   831
R64         GTGGATATTGCCAAAAGTTTTGACGATGGCGACTTTGCAAACACACTTTCTTCGACTGCT   750
PD70        GTCGATATTGCCAAAAGTTTTGACGATGGCGACTTTGCGAACACACTTTCTTCGACTGCT   750
IFO0111     ATTCCTTTGAGTAAGACTTACAACGACCCTGGCTTTACTTCCTGGCTTGAAAAACAAAAA   831

2066        TCTACCCTCGAAAGTTATTTGAGTGGCAGTGATGGTGGATTTGTTAATACTGATGTTAAC   872
HUT7212     TCTACCCTCGAAAGTTATTTGAGTGGCAGTGATGGTGGATTTGTTAATACTGATGTTAAC   891
KZ          TCTACCCTCGAAAGTTATTTGAGTGGCAGTGATGGTGGATTTGTTAATACTGATGTTAAC   891
R64         TCTACCCTCGAAAGTTATTTGAGTGGCAGTGATGGTGGATTTGTTAATACTGATGTTAAC   810
PD70        TCTACCCTCGAAAGTTATTTGAGTGGCAGTGATGGTGGATTTGTTAATACTGATGTTAAC   810
IFO0111     GATGCCTTGAACTCAT------ACATCAACTCCTCTGGATTCGTCAACTCGGGTAAAAAA   885

2066        CACATTGTTGAAAACCCAGATTTGCTTCAACAAAACTCTAGACAAGGTCTAGATTCAGCC   932
HUT7212     CACATTGTTGAAAACCCAGATTTGCTTCAACAAAACTCTAGACAAGGTCTAGATTCAGCC   951
KZ          CACATTGTTGAAAACCCAGATTTGCTCCAACAAAACTCTAGACAAGGTCTAGATTCAGCC   951
R64         CACATTGTTGAAAACCCAGATTTGCTCCAACAAAACTCTAGACAAGGCCTAGATTCAGCC   870
PD70        CACATTGTTGAAAACCCAGATTTGCTCCAACAAAACTCTAGACAAGGTCTAGATTCAGCC   870
IFO0111     CATATTGTTGAAAGCCCACAAC---------TTTCTTCTAGAGGCGGTTTGGACAGTGCC   936

2066        ACATATATTGGCCCACTTTTGACTCATGATATTGG---TGAAAGCAGCTCAACTCCATTT   989
HUT7212     ACATATATTGGCCCACTTTTGACTCATGATATTGG---TGAAAGCAGCTCAACTCCATTT   1008
KZ          ACTTATATTGGCCCACTTTTGACTCATGATATTGG---CGAAAGCAGCTCAACTCCATTT   1008
R64         ACTTATATTGGCCCACTTTTGACTCATGATATTGG---TGAAAGCAGCTCTACTCCATTT   927
PD70        ACTTATATTGGCCCACTTTTGACTCATGATATTGG---TGAAAGCAGCTCAACTCCATTT   927
IFO0111     ACCTACATTGCTGCCTTGATCACCCATGACATTGGTGATGATGACACTTACACTCCTTTC   996

2066        GATGTTGACAATGAGTATGTTTTGCAATCATATTACTTGTTATTGGAGGATAACAAAGAC   1049
HUT7212     GATGTTGACAATGAGTATGTTTTGCAATCATATTACTTGTTATTGGAGGATAACAAAGAC   1068
KZ          GATGTTGACAATGAGTATGTTTTGCAATCATATTACTTGTTATTGGAGGATAATAAAGAC   1068
R64         GATGTTGACAATGAGTATGTTTTGCAATCATATTACTTGTTATTGGAGGATAATAAAGAC   987
PD70        GATGTTGACAATGAGTATGTTTTGCAATCATATTACTTGTTATTGGAGGATAATAAAGAC   987
IFO0111     AACGTGGATAATTCCTATGTGCTCAATTCCCTATACTACTTGTTGGTTGACAACAAAAAC   1056
```

```
2066        AGATACTCTGTTAACAGTGCTTATTCTGCTGGTGCAGCTATTGGCAGATACCCAGAAGAT  1109
HUT7212     AGATACTCTGTTAACAGTGCTTATTCTGCTGGTGCAGCTATTGGCAGATACCCAGAAGAT  1128
KZ          AGATACTCTGTTAACAGTGCTTATTCTGCTGGTGCAGCTATTGGCAGATACCCAGAAGAT  1128
R64         AGATACTCTGTTAACAGTGCTTATTCTGCTGGTGCAGCTATTGGCAGATACCCAGAAGAT  1047
PD70        AGATACTCTGTTAACAGTGCTTATTCTGCTGGTGCAGCTATTGGCAGATACCCAGAAGAT  1047
IFO0111     AGATACAAGATCAATGGCAACTACAAAGCAGGTGCTGCGGTTGGAAGATATCCAGAAGAC  1116

2066        GTTTACAATGGTGATGGTTCATCTGAAGGCAATCCATGGTTCTTAGCTACTGCCTATGCT  1169
HUT7212     GTTTACAATGGTGATGGTTCATCTGAAGGCAATCCATGGTTCTTAGCTACTGCCTATGCT  1188
KZ          GTTTACAATGGTGATGGTTCATCTGAAGGCAATCCATGGTTTTTAGCTACTGCCTATGCT  1188
R64         GTTTACAATGGTGATGGTTCATCTGAAGGCAATCCATGGTTCTTAGCTACTGCCTATGCT  1107
PD70        GTTTACAATGGTGATGGTTCATCTGAAGGCAATCCATGGTTCTTAGCTACTGCCTATGCT  1107
IFO0111     GTCTACAATGGCGTTGGAACTAGCGAAGGTAACCCATGGCAATTGGCTACTGCCTACGCT  1176

2066        GCCCAAGTTCCATACAAACTTGCTTATGATGCAAAGTCGGCCTCAAATGACATTACCATT  1229
HUT7212     GCCCAAGTTCCATACAAACTTGCTTATGATGCAAAGTCGGCCTCAAATGACATTACCATT  1248
KZ          GCCCAAGTTCCATACAAACTTGTTTATGATGCAAAGTCTGCCTCAAATGACATTACCATT  1248
R64         GCCCAAGTTCCATACAAACTTGTTTATGATGCAAAGTCTGCCTCAAATGACATTACCATT  1167
PD70        GCCCAAGTTCCATACAAACTTGTTTATGATGCAAAGTCTGCCTCAAATGACATTACCATT  1167
IFO0111     GGTCAAACTTTCTACACTTTGGCTTACAACTCTTTGAAAAATAAAAAGAACTTGGTTATA  1236

2066        AACAAGATTAACTACGATTTTTTTAACAAGTATATTGTTGATTTATCTACCATCAATTCT  1289
HUT7212     AACAAGATTAACTACGATTTTTTTAACAAGTATATTGTTGATTTATCTACCATCAATTCT  1308
KZ          AACAAGATTAACTACGATTTTTTTAACAAGTATATTGTTGATTTATCTACCATCAATTCT  1308
R64         AACAAGATTAACTACGATTTTTTTAACAAGTATATTGTTGATTTATCTACCATCAATTCT  1227
PD70        AACAAGATTAACTACGATTTTTTTAACAAGTATATTGTTGATTTATCTACCATCAATTCT  1227
IFO0111     GAAAAACTCAATTACGACCTTTACAACTCCTTTATTGCTGACTTGTCCAAGATTGACTCT  1296

2066        GCTTACCAGTCTTCTGATAGTGTCACCATTAAAAGTGGCTCTGATGAATTTAACACGGTT  1349
HUT7212     GCTTACCAGTCTTCTGATAGTGTCACCATTAAAAGTGGCTCTGATGAATTTAACACGGTT  1368
KZ          GGTTACCAGTCTTCTGATAGTGTCACCATTAAAAGTGGCTCTGATGAATTTAACACGGTT  1368
R64         GCTTACCAGTCTTCTGATAGTGTCACCATTAAAAGTGGCTCTGATGAATTTAACACGGTT  1287
PD70        GCTTACCAGTCTTCTGATAGTGTCACCATTAAAAGTGGCTCTGATGAATTTAACACGGTT  1287
IFO0111     TCTTATGCTTCCAAAGATAGTTTGACTTTGACTTATGGCAGCGACAACTATAAAAATGTT  1356

2066        GCTGATAATTTGGTCACATTCGGTGATTCCTTTTTGCAAGTCATTTTGGATCATATTAAT  1409
HUT7212     GCTGATAATTTGGTCACATTCGGTGATTCCTTTTTGCAAGTCATTTTGGATCATATTAAT  1428
KZ          GCTGATAATTTGGTCACATTCGGTGATTCCTTTTTGCAAGTCATTTTGGATCATATTAAT  1428
R64         GCTGATAATTTGGTCACATTCGGTGATTCCTTTTTGCAAGTCATTTTGGATCATATTAAT  1347
PD70        GCTGATAATTTGGTCACATTCGGTGATTCCTTTTTGCAAGTCATTTTGGATCATATTAAT  1347
IFO0111     ATCAAAAGTTTGCTACAATTTGGTGACTCTTTCTTGAAAGTTCTCCTTGACCATATTGAT  1416

2066        GATGATGGCTCCTTGAATGAACAACTTAACAGATATACCGGTTATTCCACCGGTGCCTAC  1469
HUT7212     GATGATGGCTCCTTGAATGAACAACTTAACAGATATACCGGTTATTCCACCGGTGCCTAC  1488
KZ          GATGATGGCTCCTTGAATGAACAACTTAACAGAAATACCGGTTATTCCACCAGTGCCTAC  1488
R64         GATGATGGCTCCTTGAATGAACAACTTAACAGATATACCGGTTATTCCACCGGTGCCTAC  1407
PD70        GATGATGGCTCCTTGAATGAACAACTTAACAGATATACCGGTTATTCCACCGGTGCCTAC  1407
IFO0111     GACAATGGCCAACTCACCGAGGAAATCAACAGATACACTGGTTTCCAAGCCGGCGCTGTC  1476

2066        TCTTTGACATGGAGCAGTGGTGCTCTTCTTGAAGCTATTAGACTTAGAAATAAGGTCAAG  1529
HUT7212     TCTTTGACATGGAGCAGTGGTGCTCTTCTTGAAGCTATTAGACTTAGAAATAAGGTCAAG  1548
KZ          TCTTTGACATGGAGCAGTGGTGCTCTTCTTGAAGCTATTAGACTTAGAAATAAGGTCAAG  1548
R64         TCTTTGACATGGAGCAGTGGTGCTCTTCTTGAAGCTATTAGACTTAGAAATAAGGTCAAG  1467
PD70        TCTTTGACATGGAGCAGTGGTGCTCTTCTTGAAGCTATTAGACTTAGAAATAAGGTCAAG  1467
IFO0111     TCCTTGACTTGGAGTAGTGGCAGTTTGCTTAGTGCAAACAGAGCTAGAAACAAATTGATT  1536

2066        GCTTTG------ 1535
HUT7212     GCTTTGGCTTAA 1560
KZ          GCTTTGGCTTAA 1560
R64         GCTTTGGCT--- 1476
PD70        GCTTTGGCT--- 1476
IFO0111     GAACTTCTTTGA 1548
```

Figure 5. Alignment of the 1535 base pairs glucoamylase cDNA sequence of Saccharomycopsis bubodii 2066 (in bold letters) with glucoamylase cDNAs of Saccharomycopsis fibuligera strains deposited in GenBank, NCBI. The start and stop codons are underscored. Remaining undecoded bases are shaded in yellow.

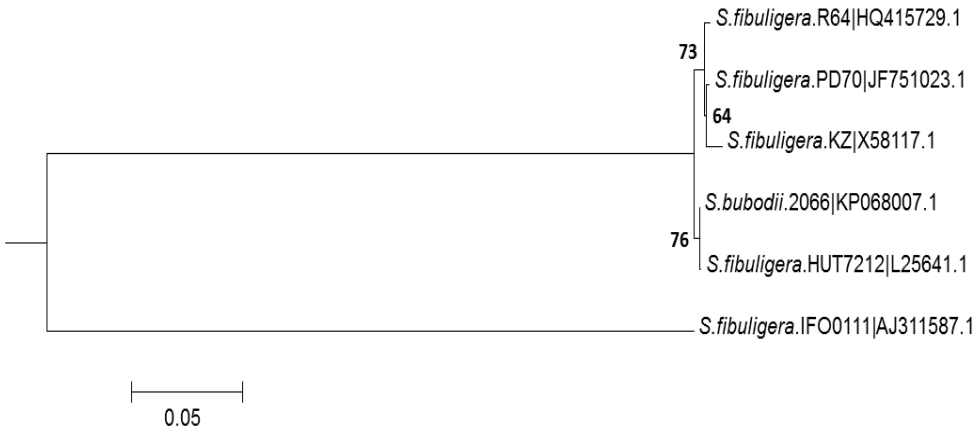

Figure 6. Neighbor-joining tree inferred from Saccharomycopsis glucoamylase gene, using transition model with three parameters (TIM3). Bootstrap (replicate = 1000) support values greater than 50 % are shown at their corresponding nodes. S. fibuligera IFO0111 served as the outgroup. Scale bar indicates 0.05 substitutions per nucleotide position.

be ascertained at this point that it is exactly the same (e.g. enzyme activity) as the glucoamylase (from strain HUT7212) where it showed 100% homology. It is possible that they are differently modified post-translationally (e.g. degree of *N*-glycosylation) due to differences in the host strain (De Barros *et al.*, 2009; Skropeta, 2009).

Other properties published on the glucoamylase from strain HUT7212 that can be attributed to the glucoamylase from *S. bubodii* 2066 (Figure 7) are the presence of four possible *N*-glycosylation sites (shaded yellow) and twenty hydrophobic amino acid segment (shaded green) at the amino terminal which resembles signal sequences found in various secretory protein precursors (Itoh *et al.*, 1987). Further, when this glucoamylase was aligned by Itoh *et al.* (1987) with glucoamylases from yeasts and fungi, five highly conserved segments (shaded light blue) are identified. The three-dimensional structure of this glucoamylase has been determined at 1.7 angstroms resolution by overexpression of the protein in *E. coli* (Sevcik *et al.*, 1998). The study revealed that the core of the enzyme is an $(\alpha/\alpha)_6$ barrel which is closely similar to that of the catalytic domain of *Aspergillus awamori* glucoamylase, the most thoroughly studied glucoamylase, where the active site is located at the narrower end of the barrel. Moreover, unlike that of *A. awamori* which has a starch binding domain (SBD), the presence of SBD in *S. fibuligera* HUT7212 glucoamylase was not determined. However, a more recent study by the same research group employing an improved resolution and mutating some residues at the suspected binding site revealed a starch binding site near the catalytic domain.

SVFAAIVSALPLQEGPLNKRAYPSFEAYSNYKVDRTDLETFLDKQKEVSLYYLLQNIAYPEGQFNNGVPGTVIAS
PSTSNPDYYYQWTRDSAITFLTVLSELEDNNFNTTLAKAVEYYINTSYNLQRTSNPSGSFDDENHKGLGEPKFNT
DGSAYTGAWGRPQNDGPALRAYAISRYLNDVNSLNEGKLVLTDSGDINFSSTEDIYKNIIKPDLEYVIGYWDSTG
FDLWEENQGRHFFTSLVQQKALAYAVDIAKSFDDGDFANTLSSTASTLESYLSGSDGGFVNTDVNHIVENPDLLQ
QNSRQGLDSATYIGPLLTHDIGESSSTPFDVDNEYVLQSYYLLLEDNKDRYSVNSAYSAGAAIGRYPEDVYNGDG
SSEGNPWFLATAYAAQVPYKLAYDAKSASNDITINKINYDFFNKYIVDLSTINSAYQSSDSVTIKSGSDEFNTVA
DNLVTFGDSFLQVILDHINDDGSLNEQLNRYTGYSTGAYSLTWSSGALLEAIRLRNKVKAL

Figure 7. Amino acid sequence of glucoamylase of Saccharomycopsis bubodii 2066. Notable amino acids and amino acid sequences are (1) shaded yellow: possible N-glycosylation sites; (2) shaded green: twenty hydrophobic amino acid segment that resembles signal sequences found in various secretory protein precursors; (3) shaded light blue: highly conserved segments in glucoamylases from yeasts and fungi and (4) shaded red: glutamic acid residues that are directly involved in the catalytic activity of the enzyme (Itoh et al., 1987; Sevcik et al., 2006).

Also, two glutamic acid residues (shaded red) that are directly involved in the catalytic activity of the enzyme are identified (Sevcik *et al*., 2006). These information warrant further investigation on the glucoamylase from *S. bubodii* 2066.

CONCLUSION

In this study, *Saccharomycopsis (Syn. Endomycopsis) bubodii* 2066 was shown to exhibit amylolytic activity on raw sago starch indicating the yeast as potential source of raw-starch digesting amylase (RSDA). Further, an almost full gene sequence (98.4% coverage) of a glucoamylase, a putative RSDA, from *S. bubodii* 2066 was elucidated via the primer walking strategy. The sequence is 100% homologous with the cDNA ORF sequence of glucoamylase of *S. fibuligera* strain HUT7212 that was used in the primer design. It is also at least 98.6% homologous to three other glucoamylases of *S. fibuligera* strains. The surprisingly high homology obtained suggests that this particular glucoamylase gene is highly conserved within the genus *Saccharomycopsis*.

This work is the first step towards cloning and expression of a putative raw starch- digesting amylase from another source in *Saccharomyces cerevisiae* for the conversion of raw sago starch into bioethanol using a single microorganism.

ACKNOWLEDGEMENT

The authors express their gratitude to the following institutions for funding this research: Department of Science and Technology Philippine Council for Industry, Engineering and Emerging Technology Research and Development (DOST-PCIEERD) and the University of the Philippines Mindanao Office of Research (UPMin-OR).

REFERENCES

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. Nucleic Acids Res. 1997; 25:3389-3402.

Bullo LL. Production studies, purification and characterization of amylolytic enzyme from *Saccharomycopsis fibuligera* 2074 isolated from Philippine *bubod* starter [BSc thesis]. Davao City, Philippines: University of the Philippines Mindanao; 2009. 28 p.

Chi Z, Chi Z, Liu G, Wang F, Ju L, Zhang T. *Saccharomycopsis fibuligera* and its applications in biotechnology. Biotechnol Adv. 2009; 27:423-431.

Darriba D, Taboada GL, Doallo R, Posada D. jModelTest 2: more models, new heuristics and parallel computing. Nature Methods. 2012; 9(8):772.

De Barros MC, Silva RdN, Ramada MHS, Galdino AS, De Moraes LMP, Torres FAG, et al. The influence of N-glycosylation on biochemical properties of Amy1, an a-amylase from the yeast *Cryptococcus flavus*. Carbohydr. Res. 2009; 344:1682-1686.

Farhat IA, Protzmann J, Becker A, Valles-Pamies B, Neale R, Hill SE. Effect of the extent of conversion and retrogradation on the digestibility of potato starch. Starch/Stärke. 2001; 53:431-436.

Flach M. Sago palm. *Metroxylon sagu* Rottb. Promoting the conservation and use of underutilized and neglected crops. 13. Institute of Plant Genetics and Crop Plant Research, Gatersleben/International Plant Genetic Resources Institute, Rome, Italy. 1997.

Flores DM. The versatile sago (*Metroxylon sagu* Rottb.) and its green potential in Mindanao. Banwa. 2008; 5(1):8-17.

Gasperik J, Kovac L, Minarikova O. Purification and characterization of the amylolytic enzymes of *Saccharomycopsis fibuligera*. Int J Biochem. 1991; 23(1):21-25.

Goyal N, Gupta JK, Soni SK. A novel raw starch digesting amylase from *Bacillus* sp. I-3 and its use in the direct hydrolysis of raw potato starch. Enzyme Micro Technol. 2005;37: 723-734.

Hall TA. BioEdit: A user-friendly biological sequence alignment editor and analysis program for windows 95/98/NT. Nucleic Acids Symp Ser. 1999; 41:95-98.

Hostinova E. Amylolytic enzymes produced by the yeast *Saccharomycopsis fibuligera*. Biologia, Bratislava. 2002; 57/Suppl 11:247-251.

Hostinova E, Balanova J, Gasperik J. The nucleotide sequence of the glucoamylase gene GLA1 from *Saccharomycopsis fibuligera* KZ. FEMS Microbiol Lett. 1991; 67(1):103-8.

Hostinova E. Genetic and enzymatic characterization of glucoamylases from yeasts. Gen Physiol Biophys. 1998; 17:19-21.

Hostinova E, Solovicova A, Dvorsky R, Gasperik J. Molecular cloning and 3D structure prediction of the first raw-starch-degrading glucoamylase without a separate starch-binding domain. Arch Biochem Biophys. 2003; 411:189-195.

Itoh T, Ohtsuki I, Yamashita I, Fukui S. Nucleotide sequence of the glucoamylase gene *GLU1* in the yeast *Saccharomycopsis fibuligera*. J Bacteriol. 1987; 69(9):4171-4176.

Kitamura K, Yamamoto Y. Purification and properties of an enzyme, zymolyase, which lyses viable yeasts cells. Arch Biochem Biophys. 1972; 153:403-406.

Konsula Z, Liakopoulou-Kyriakides M. Hydrolysis of starches by the action of an α-amylase from *Bacillus subtilis*. Process Biochem. 2004; 39:1745-1749.

Kumar P, Satyanarayana T. Microbial glucoamylases: characteristics and applications. Crit Rev Biotechnol. 2009; 29:225-255.

Li H, Ding D, Cao Y, Yu B, Guo L, Liu X. Partially Overlapping Primer-Based PCR for Genome Walking. PLOS ONE. 2015; 10(3):1-9.

Lim EAY, Panes VA, Romero GO. Species Identification and Genetic Diversity Analysis by DNA Fingerprinting of Yeast Isolates from Philippine Rice Wine Starters. Philipp Agric Sci. 2006; 89(4):326-337.

Limtong S, Sintara S, Suwanarit P, Lotong N. Yeast diversity in Thai traditional fermentation starter (loog-pang). Kasetsart J. 2002; 5:149-158 (abstr.).

Marlida Y, Saari N, Hassan Z, Radu S. Improvement in raw sago starch degrading enzyme production from *Acremonium* sp. Endophytic fungus using carbon and nitrogen sources. Enzyme Microb Technol. 2000; 27:511-515.

McWilliam HW, Li W, Uludag M, Squizzato S, Park YM, Buso N, et al. Analysis tool web services from the EMBL-EBI. Nucleic Acids Res. 2013; 41:597-600.

Natalia D, Vidilaseris K, Satrimafitrah P, Ismaya WT, Permentier PH, Fibriansah G, et al. Biochemical characterization of a glucoamylase from *Saccharomycopsis fibuligera* R64. Biologia. 2011; 66(1):27-32.

Omemu AM, Akpan I, Bankole MO, Teniola OD. Hydrolysis of raw tuber starches by amylase of *Aspergillus niger* AM07 isolated from the soil. Afr J Biotech. 1999; 5:144-150.

Pandey A, Nigam P, Soccol CR, Soccol VT, Singh D, Mohan R. Advances in microbial amylases. Biotechnol Appl Biochem. 2000; 31:135-152.

Philippine Network of Microbial Culture Collections (PNMCC) Directory of Strains. Monsalud RG, editor-in-chief. 3rd Edition. Los Banos, Laguna: BIOTECH, UPLB, 2012.

Sakai H, Caldo GA. Microbiological studies on bubod, a fermentation starter in the Philippines, II. Isolation and identification of yeasts. Philipp Agric. 1985a; 68:139-144.

Sakai H, Caldo GA. Microbiological studies on bubod, a fermentation starter in the Philippines. III. A new yeast species, *Endomycopsis bubodii* Sakai and Caldo sp. nov. and some enzymological properties of *Endomycopsis* strains. Philipp Agric. 1985b; 68:181-188.

Sauer J, Sigurskjold BW, Christensen U, Frandsen TP, Mirgorodskaya E, Harrison M, et al. Glucoamylase: structure/function relationships, and protein engineering. Biochim Biophys Acta. 2000; 1543:275-293.

Sevcik J, Hostinova E, Solovicova A, Gasperik J, Dauter Z, Gasperik J, et al. Structure of the complex of a yeast glucoamylase with acarbose reveals the presence of a raw starch binding site on the catalytic domain. FEBS J. 2006; 273:2161-2171.

Sevcik J, Solovicova A, Hostinova E, Gasperik J, Wilson KS, Dauter Z. Structure of glucoamylase from *Saccharomycopsis fibuligera* at 1.7Å resolution. Acta Cryst. 1998; D54:854-866.

Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, et al. Fast, scalable generation of high quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol. 2011; 7(539):1-6.

Skropeta D. The effect of individual N-glycans on enzyme activity. Bioorg. Med. Chem. 2009; 17:2645-2653.

Sun H, Ge X, Zhang W. Production of a novel raw-starch-digesting glucoamylase by *Penicillium sp* X-1 under solid state fermentation and its use in direct hydrolysis of raw starch. Appl Microbiol Biotechnol. 2006; 28:550-559.

Swofford D. PAUP:Phylogenetic analysis using parsimony, version 4.0 b10. 2003.

Takeuchi A, Shimizu-Ibuka A, Nishima A, Mura Y, Okada K, Tokue S, et al. Purification and characterization of an a-amylase of *Pichia burtonii* isolated from the traditional starter "Murcha" in Nepal. Biosci Biotechnol Biochem. 2006; 11:3019-3024.

Volpicella M, Leoni C, Costanza A, Fanizza I, Placido A, Ceci LR. Genome Walking by Next Generation Sequencing Approaches. Biology. 2012; 1:495-507.

Ye J, Couloris G, Zaretskay I, Cutcutache I, Roaz S, Madden TL. Primer-BLAST: A tool to design target-specific primers for polymerase chain reaction. BMC Bioinformatics. 2012; 13:134.