Isolation, identification and *in silico* analysis of alpha-amylase gene of *Aspergillus niger* strain CSA35 obtained from cassava undergoing spoilage

Oghenetega J. Avwioroko[a,*], Akpovwehwee A. Anigboro[b], Nnanna N. Unachukwu[c], Nyerhovwo J. Tonukari[b]

[a] *Biochemistry Division, Department of Chemical Sciences, College of Natural Sciences, Redeemer's University, Ede, Osun State, Nigeria*
[b] *Department of Biochemistry, Faculty of Science, Delta State University, P.M.B. 1, Abraka, Nigeria*
[c] *Bioscience Center, International Institute for Tropical Agriculture (IITA), Ibadan, Oyo State, Nigeria*

ABSTRACT

In this investigation, a gene (*CDF_Amyl*) encoding extracellular α-amylase in *Aspergillus niger* strain CSA35 associated with cassava spoilage was amplified using specific primers and characterized *in silico*. The gene had a partial nucleotide sequence of 968 bp and encoded a protein of 222 aa residues with a molecular weight and isoelectric point of 25.13 kDa and 4.17, respectively. Its catalytic site was located in the active site domain. BLASTp analysis showed that the protein primary sequence of the α-amylase gene had 98% and 99% homologies with the α-amylase of *A. niger* and *A. oryzae* RIB40, respectively. The gene is more closely related to α-amylase genes from fungi than to bacterial, plant, or animal α-amylase genes. Restriction mapping of the gene showed it can be digested with restriction enzymes like *Nco*I, *Pst*I, *Sma*I, and BcLI among others but not with *Eco*RI and *Eco*RV. Its protein product had a hydrophobicity score of − 0.43 but no transmembrane helix. The CDF_Amyl protein was subcellularly localized in the secretory pathway, an indication of its release into extracellular space after secretion. Also, the 3D structure of the CDF-Amyl protein was barrel-shaped with domains characteristic of α-amylases. The encoded α-amylase Vmax is 6.90 U/mg protein and Km is 6.70 mg/ml. It was concluded that the unique characteristics of the CDF_Amyl gene and its deduced protein could find applications in biotechnological, food and pharmaceutical industries where cloning and further modification of this gene would be required for product development and improvement.

1. Introduction

α-Amylases (EC 3.2.1.1.) hydrolyse polysaccharides such as starch, glycogen etc via random breakdown of their internal α − 1, 4-glucosidic bonds to produce oligosaccharides of various sizes [16,31]. They are among the most sought-after enzymes because of their high relevance in present-day biotechnology. α-Amylases have wide applications in starch processing, brewing industries, detergent, textile, pharmaceutical industries etc [31,41]. Although they can be obtained from almost all living organisms, the enzymes from microbial sources (basically bacteria and fungi) tend to be more suitable for industrial applications.

α-Amylases are grouped as family 13 of the glycosyl hydrolases. They have similar structures and catalytic sites and do not vary in their catalytic mechanism. A typical α-amylase has a three dimensional (3D) structure resembling a barrel shape with three characteristic domains –A, B and C [14,27]. Situated in loops at the C-termini of β-strands in the A domain are highly conserved residues of amino acid which partake in catalysis and binding of substrates [27]. The B domain is involved in binding of Ca^{2+} and also participates in binding of substrate. The C-domain's role, a second protrusion in front of or behind the A-domain, however, is not fully known; but a study by Holm et al. [21] on genetic aberrations in the C-domain of *Bacillus stearothermophilus* α-amylase suggests that the domain plays a role in enzyme activity.

Amylases vary in molecular weights; usually from about 10–210 kDa [40]. The presence of carbohydrate moieties in some organisms' amylases tends to raise their molecular weights; for example, glycoproteins have been observed in *Aspergillus oryzae, Lipomyces kononenkoae, Bacillus stearothermophilus* and *Bacillus subtilis* strains. The isoelectric point (pI) of amylases ranges from 3.25 to 10.1 [12]. The pI refers to the pH at which an enzyme's overall electric charge is zero [10]. Knowledge of enzymes' pI values is used for electro focussing.

Over the years, a number of application softwares or programmes have been developed and are being used for easy study of gene and

```
                              GATGCCTACCATGGC      15
                               D   A   Y   H   G                   5
         TACTGGCAGCAGGATATGtaagtcgatttctttaaatatctacctgtcat        65
          Y   W   Q   Q   D   M                                   11
         cttttacatcaatatgaactaacttgatggtttttagaTACTCTCTGAACG       115
                                           Y   S   L   N          15
         AAAACTACGGCACTGCAGATGACTTGAAGGCGCTCTCTTCGGCCCTTCAT       165
          E   N   Y   G   T   A   D   D   L   K   A   L   S   S   A   L   H   32
         GAGAGGGGGATGTATCTTATGGTCGATGTGGTTGCTAACCATATGgttcg       215
          E   R   G   M   Y   L   M   V   D   V   V   A   N   H   M   47
         tggtcctttgcaactgacttcgcggatatggttcatttcagtactgacaa      265
         tgagtaatatcagGGCTATGATGGAGCGGGTAGCTCAGTCGATTACAGTG       315
                      G   Y   D   G   A   G   S   S   V   D   Y   S   59
         TGTTTAAACCGTTCAGTTCCCAAGACTACTTCCACCCGTTCTGTCTCATT       365
          V   F   K   P   F   S   S   Q   D   Y   F   H   P   F   C   L   I   76
         CAAAACTATGAAGATCAGACTCAGGTTGAGGATTGCTGGCTAGGAGATAA       415
          Q   N   Y   E   D   Q   T   Q   V   E   D   C   W   L   G   D   N   93
         CACTGTCTCCTTGCCTGATCTCGATACCACCAAGGATGTGGTCAAGAATG       465
           T   V   S   L   P   D   L   D   T   T   K   D   V   V   K   N   109
         AATGGTACGACTGGGTGGGATCATTGGTATCGAACTACTCCagtaagata       515
          E   W   Y   D   W   V   G   S   L   V   S   N   Y   S   123
         tttctccctcattctacaacttggctgatcgatgatacttacgaaatcaG       565
         TTGACGGCCTCCGTATCGACACAGTAAAACACGTCCAGAAGGACTTCTGG       615
          V   D   G   L   R   I   D   T   V   K   H   V   Q   K   D   F   W   140
         CCCGGGTACAACAAAGCCGCAGGCGTGTACTGTATCGGCGAGGTGCTCGA       665
          P   G   Y   N   K   A   A   G   V   Y   C   I   G   E   V   L   D   157
         CGGTGATCCGGCCTACACTTGTCCCTACCAGAACGTCATGGACGGCGTAC       715
           G   D   P   A   Y   T   C   P   Y   Q   N   V   M   D   G   V   173
         TGAACTATCCCATGTATggttcctccaaccatgagccttcttgcaagtct       765
         L   N   Y   P   M   Y                                   179
         catctcctaacgaaacggctaaaaccagttacTATCCACTCCTCAACGCC       815
                                         Y   P   L   L   N   A    185
         TTCAAGTCAACCTCCGGCAGCATGGACGACCTCTACAACATGATCAACAC       865
          F   K   S   T   S   G   S   M   D   D   L   Y   N   M   I   N   T   202
         CGTCAAATCCGACTGTCCAGACTCAACACTCCTGGGCACATTCGTCGAGA       915
           V   K   S   D   C   P   D   S   T   L   L   G   T   F   V   E   218
         ACCACGACAAC                                            926
         N   H   D   N                                          222
```

Fig. 1. Partial nucleotide sequence and deduced amino acid sequence of the cassava degrading fungi amylase gene (CDF_Amyl). Four introns (non coding regions) were found in the *CDF_Amyl* gene sequence. The introns were denoted in lower case letters (italicized) while the exons (5) in upper case letters.

Table 1
BLASTp result of CDF_Amyl protein-protein matches and source organisms.

Matched protein/organism	Max score	Total score	Query cover	E-value	Identity	Accession number
Alpha-amylase A type − 1/2 [*Aspergillus oryzae* RIB40]	462	462	100%	4e − 159	99%	XP_003189619.1
Chain A, Orthorhombic Crystal Structure Of *Aspergillus niger* Alpha-Amylase	461	461	100%	7e − 159	98%	2GUY_A
Alpha-amylase, putative [*Aspergillus flavus* NRRL3357]	461	461	100%	1e − 158	98%	XP_002374124.1
Alpha-amylase [*Aspergillus kawachii*]	461	461	100%	1e − 158	98%	BAD01051.1
Taka-amylase A (Taa-G1) precursor [*Aspergillus oryzae*]	460	460	100%	3e − 158	98%	AAA32708.1
Alpha-amylase [*Aspergillus awamori*]	460	460	100%	3e − 158	98%	BAD06002.1
Alpha amylase Precursor [*Aspergillus awamori*]	459	459	100%	4e − 158	98%	Q02905.1
Alpha amylase Precursor [*Aspergillus shirousami*]	455	455	100%	3e − 157	98%	P30292.1
Taka-amylase A precursor [*Aspergillus oryzae*]	450	450	100%	2e − 154	97%	BAA00336.1
Alpha-amylase [*Aspergillus sojae*]	499	499	100%	7e − 154	95%	BAM28635.1

protein sequences in a new field of molecular biology presently referred to as bioinformatics or computational biology ([46]; [47]). These bioinformatics tools make characterisation of such nucleotide or protein sequences less stressful, time conservative and cost effective aside their highly precise or accurate outcomes. Some of these computational tools are utilized in an array of *in silico* analyses ranging from determination of evolutionary relatedness of genes or enzyme proteins (phylogeny), molecular weights, isoelectric points (pI), transmembrane

helices, hydrophobicity, subcellular localization of proteins, three-dimensional (3D) structure of proteins, conserved domains, detection of introns, to *in silico* determination of restriction enzyme cleavage sites in gene sequences ([25,7,8]; Blum et al. [9]; [33,24]).

Food spoilage generally has been closely linked with the action of some microorganisms capable of degrading some nutrient materials in food substances [1,39]. Cassava (*Manihot esculenta*) and its products, being notable rich sources of carbohydrate, water and some minerals
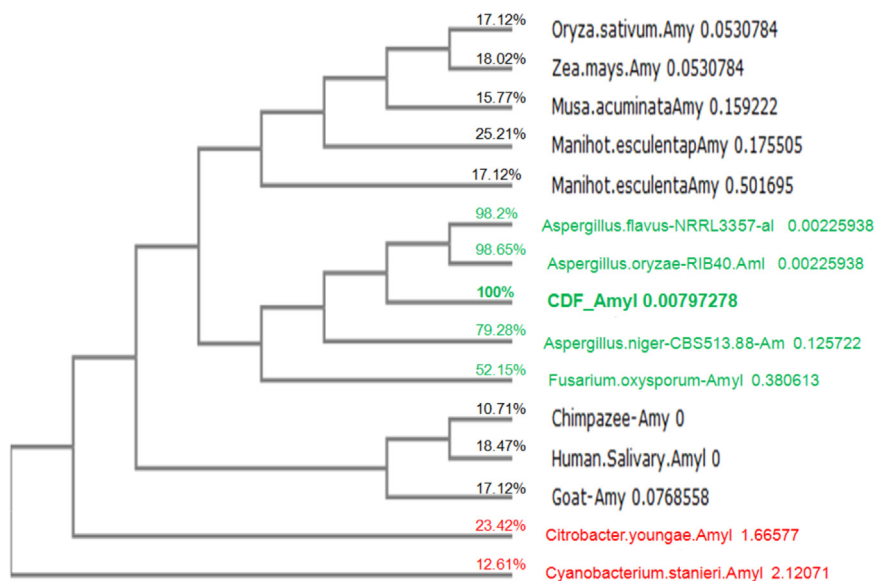
Fig. 2. Evolutionary relationship between cassava-degrading fungal amylase (CDF_Amyl) and those in plants, fungi, bacteria, lower animals and human in NCBI GenBank.

[6], support the growth of a number of microbes, including fungi [3,42]. Hence, the spoilage of cassava tubers or their processed products after a long period of exposure to air or dirts, are often times associated with visible growth of *Aspergillus* species on or around them with characteristic colour(s) and odour [4]. These saprophytic fungi are believed to grow and survive on the cassava tubers by secreting starch hydrolyzing enzymes such as amylases, amidst others, thereby breaking down cassava starch to simple sugars for nutrition [28,40].

The present study amplified and characterized the α-amylase gene of *Aspergillus niger* strain CSA35 using *in silico* analyses. The fungi, which grow on decaying cassava tubers and peels, had been identified in a previous study. The characterisation of the fungal cassava starch-degrading α-amylase gene and its deduced protein sequence would facilitate the potential industrial applications of the α-amylase gene and its protein product.

2. Material and methods

2.1. Preparation of cassava flour

Peeled cassava (*M. esculenta*) tubers were processed into flour according to the method described by Oyewole and Sanni [34].

2.2. YPD agar preparation

Preparation of yeast extract-peptone dextrose (YPD) agar was carried out as described by Tonukari et al. [44]. Thereafter, it was slightly cooled and poured into plates to solidify.

2.3. Growth of fungi

Aspergillus niger CSA35, previously identified using 18S rRNA sequence and reported to be associated with cassava spoilage in Nigeria [4], was inoculated into a sterile YPD agar in a Petri dish. It was allowed to grow and sporulate in the dark for one week and then subcultured in sterilized Cassava Starch Agar (CSA) which contained cassava flour (2%) as the main carbon source, $NaNO_2$ (1%) and agar powder (1.5%), for additional 7 days. To avoid growth of bacteria, 20% ampicillin (5 μl) was included in the subcultures. Thereafter, the hyphae were harvested.

2.4. Genomic DNA isolation

Genomic DNA was extracted from the fungi using a modified cetyl trimethyl ammonium bromide (CTAB) method [49] and purified using ethanol precipitation. The presence and quality of the isolated genomic DNA was ascertained via agarose gel electrophoresis using 0.8% agarose gel stained with ethidium bromide (5 μg/μl). Images of bands obtained in the gel were viewed and captured using a UV-light automated Gel Documentation machine (Enduro, USA).

2.5. PCR amplification of α-amylase gene

Alpha amylase gene was amplified by polymerase chain reaction (PCR) method as described by Zidani et al. [49] using the specific forward and reverse primers (AmyFP 5′-CATCTGGATCACCCCCGTTA-3′ and AmyRP 5′-AGACTTACGAAGCGAACCGT-3′, respectively) and the isolated fungal genomic DNA as template. The PCR mixture (25 μl) contained, 1.25 μl each of forward and reverse primers (5 mM each), 1.25 μl of 2.5 mM dNTP mix (dATP, dGTP, dTTP, dCTP) (Promega), $10 \times NH_4$ buffer (1.25 μl), 1.0 μl of 50 mM $MgCl_2$ (Promega), 1.0 μl of 5 unit/μl of Taq DNA polymerase, 7.25 μl of dd·H_2O and 7.25 μl of 10 ng/μl of genomic DNA. The PCR program was performed on a thermal cycler (Biometra) system as follow: Stage 1 (x1): 94 °C for 5 min to achieve the initial DNA denaturation; Stage 2 (x 35): 94 °C for 30 s, for denaturation; 60 °C for 30 s, for annealing; and 72 °C for 5 min, for extension. Stage 3 (x1): 72 °C for 5 min and 10 °C for infinity (∞). The presence of amplified α-amylase genes was established by agarose gel electrophoresis using 1.0% agarose stained with 3 μl of ethidium bromide (5 μg/μl) and viewed with Enduro Gel Doc machine.

2.6. Gene sequencing

The purified α-amylase amplicons were sequenced using an automated Applied Biosystem 3130xl Genetic Analyzer (ABI Sequencer, USA). The sequencing was carried out at the Bioscience Center, International Institute for Tropical Agriculture (IITA), Ibadan, Nigeria.

2.7. Bioinformatics characterisation of putative α-amylase gene

The nucleotide sequence of α-amylase obtained was translated to its corresponding protein primary sequence using the European
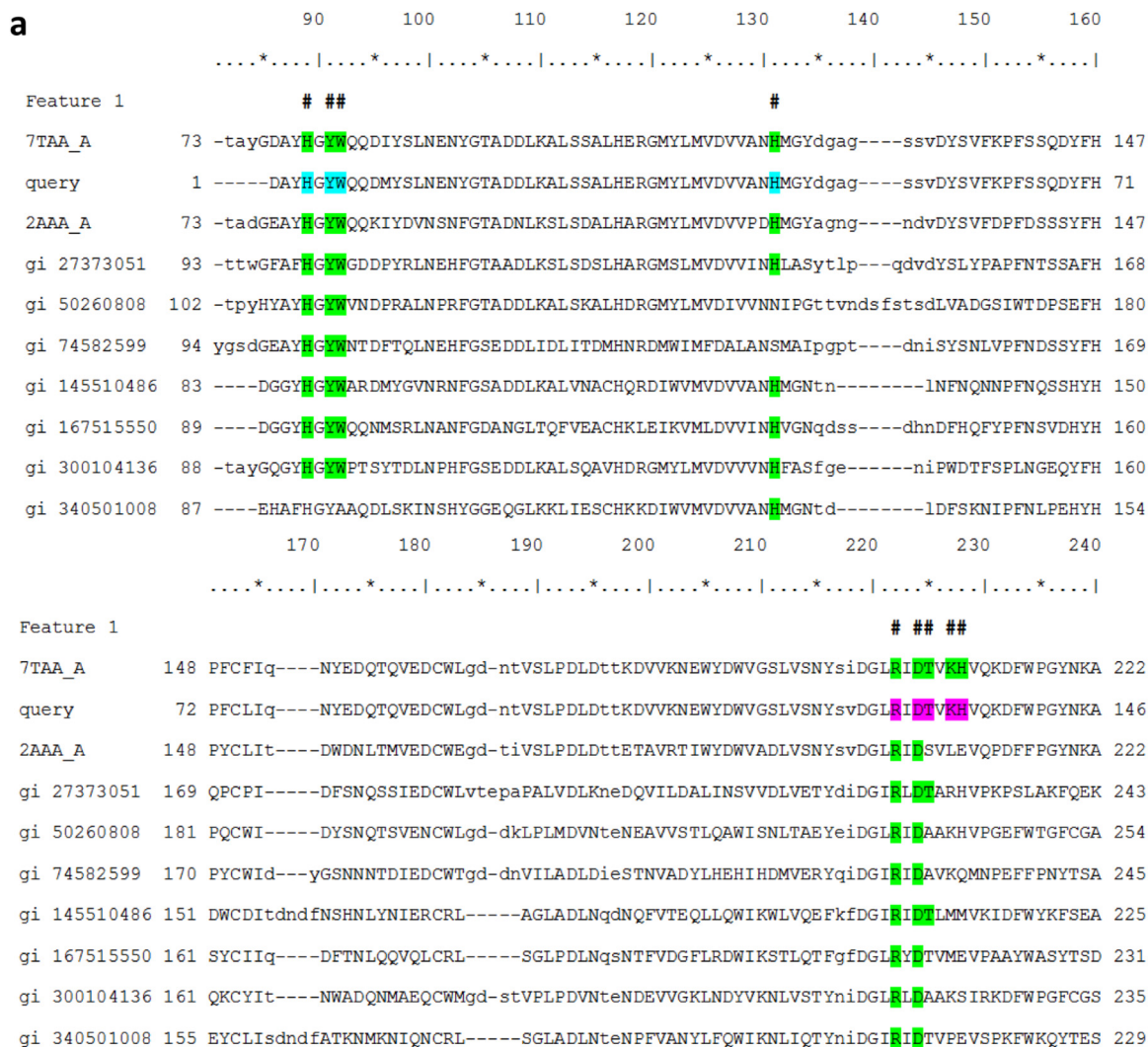
a.

```
                    90        100       110       120       130       140       150       160
         ....*....|....*....|....*....|....*....|....*....|....*....|....*....|....*....|

Feature 1              # ##                                    #

7TAA_A       73 -tayGDAYHGYWQQDIYSLNENYGTADDLKALSSALHERGMYLMVDVVANHMGYdgag----ssvDYSVFKPFSSQDYFH 147
query         1 -----DAYHGYWQQDMYSLNENYGTADDLKALSSALHERGMYLMVDVVANHMGYdgag----ssvDYSVFKPFSSQDYFH 71
2AAA_A       73 -tadGEAYHGYWQQKIYDVNSNFGTADNLKSLSDALHARGMYLMVDVVPDHMGYagng----ndvDYSVFDPFDSSSYFH 147
gi 27373051  93 -ttwGFAFHGYWGDDPYRLNEHFGTAADLKSLSDSLHARGMSLMVDVVINHLASytlp---qdvdYSLYPAPFNTSSAFH 168
gi 50260808 102 -tpyHYAYHGYWVNDPRALNPRFGTADDLKALSKALHDRGMYLMVDIVVNNIPGttvndsfstsdLVADGSIWTDPSEFH 180
gi 74582599  94 ygsdGEAYHGYWNTDFTQLNEHFGSEDDLIDLITDMHNRDMWIMFDALANSMAIpgpt----dniSYSNLVPFNDSSYFH 169
gi 145510486 83 ----DGGYHGYWARDMYGVNRNFGSADDLKALVNACHQRDIWVMVDVVANHMGNtn--------lNFNQNNPFNQSSHYH 150
gi 167515550 89 ----DGGYHGYWQQNMSRLNANFGDANGLTQFVEACHKLEIKVMLDVVINHVGNqdss----dhnDFHQFYPFNSVDHYH 160
gi 300104136 88 -tayGQGYHGYWPTSYTDLNPHFGSEDDLKALSQAVHDRGMYLMVDVVVNHFASfge------niPWDTFSPLNGEQYFH 160
gi 340501008 87 ----EHAFHGYAAQDLSKINSHYGGEQGLKKLIESCHKKDIWVMVDVVANHMGNtd--------lDFSKNIPFNLPEHYH 154

                    170       180       190       200       210       220       230       240
         ....*....|....*....|....*....|....*....|....*....|....*....|....*....|....*....|

Feature 1                                                              # ## ##

7TAA_A      148 PFCFIq----NYEDQTQVEDCWLgd-ntVSLPDLDttKDVVKNEWYDWVGSLVSNYsiDGLRIDTVKHVQKDFWPGYNKA 222
query        72 PFCLIq----NYEDQTQVEDCWLgd-ntVSLPDLDttKDVVKNEWYDWVGSLVSNYsvDGLRIDTVKHVQKDFWPGYNKA 146
2AAA_A      148 PYCLIt----DWDNLTMVEDCWEgd-tiVSLPDLDttETAVRTIWYDWVADLVSNYsvDGLRIDSVLEVQPDFFPGYNKA 222
gi 27373051 169 QPCPI-----DFSNQSSIEDCWLvtepaPALVDLKneDQVILDALINSVVDLVETYdiDGIRLDTARHVPKPSLAKFQEK 243
gi 50260808 181 PQCWI-----DYSNQTSVENCWLgd-dkLPLMDVNteNEAVVSTLQAWISNLTAEYeiDGLRIDAAKHVPGEFWTGFCGA 254
gi 74582599 170 PYCWId----yGSNNNTDIEDCWTgd-dnVILADLDieSTNVADYLHEHIHDMVERYqiDGIRIDAVKQMNPEFFPNYTSA 245
gi 145510486 151 DWCDItdndfNSHNLYNIERCRL-----AGLADLNqdNQFVTEQLLQWIKWLVQEFkfDGIRIDTLMMVKIDFWYKFSEA 225
gi 167515550 161 SYCIIq----DFTNLQQVQLCRL-----SGLPDLNqsNTFVDGFLRDWIKSTLQTFgfDGLRYDTVMEVPAAYWASYTSD 231
gi 300104136 161 QKCYIt----NWADQNMAEQCWMgd-stVPLPDVNteNDEVVGKLNDYVKNLVSTYniDGLRLDAAKSIRKDFWPGFCGS 235
gi 340501008 155 EYCLIsdndfATKNMKNIQNCRL-----SGLADLNteNPFVANYLFQWIKNLIQTYniDGIRIDTVPEVSPKFWKQYTES 229
```

Fig. 3. a. **Conserved domain amino acids of cassava-degrading fungi amylase (CDF_Amyl).** #Hash marks above the aligned sequences show the location of conserved feature residues. The amino acid residues present in the conserved domain of CDF_Amyl gene include 46 His (H), 128 Arg (R), 130 Gln (D), 154 Glu (E), and 220 His (H) amongst others. 7TAA_A denotes Taka Amylase from *Aspergillus oryzae* (GenBank Accession no. 7TAA_A); query = CDF_Amyl gene (from this study); 2AAA_A = alpha-amylases from *Aspergillus niger* (Accession: 2AAA_A); gi 27373051 = alpha-amylase from *Lipomyces kononenkoae* (GenBank: AAO12212.1); gi 50260808 = hypothetical amylase protein from *Cryptococcus neoformans* var. neoformans B-3501A (GenBank: EAL23458.1); gi 74582599 = alpha-amylase from *Schizosaccharomyces pombe* 972 h (UniProtKB/Swiss-Prot: O74922.1); gi 145510486 = hypothetical amylase protein from *Paramecium tetraurelia* strain d4-2 (NCBI Reference Sequence: XP_001441176.1); gi 167515550 = hypothetical amylase protein from *Monosiga brevicollis* MX1 (NCBI Reference Sequence: XP_001742116.1); gi 300104136 = glycoside hydrolase family 13 protein from *Schizophyllum commune* H4–8 (GenBank: EFI95542.1); gi 340501008 = hypothetical amylase protein from *Ichthyophthirius multifiliis* (GenBank: EGR27831.1). All amylases compared for conserve domain amino acids are from eukaryotic organisms. b. **Active and catalytic site regions of cassava-degrading fungi amylase (CDF_Amyl).** The CDF_Amyl belonged to the superfamily of amylases. The catalytic site region of the amylase is within the active site region.

Biotechnology Information (EBI) EMBOSS translation tool and National Center for Biotechnology Information (NCBI) BLASTX tool; the most suitable reading frame was selected and exons identified. Identifications of closely matched α-amylase sequences and conserved domains were carried out with the aid of NCBI BLASTp and CD tools available online at http://blast.ncbi.nlm.nih.gov/Blast.cgi and https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi respectively [30], while multiple sequence alignment and construction of phylogeny tree were done using EMBL EBI ClustalW2 and Phylogeny tools available at http://www.ebi.ac.uk/Tools. Restriction mapping was carried out on the cassava-degrading fungal amylase gene sequence (CDF_Amyl) using a suitable bioinformatics tool available online (http://www.bioinformatics.nl/tools/tacg4/form4.html). Prediction of CDF_Amyl molecular weight (Mw) and isoelectric point (pI) was done using ExPASy - Compute pI/Mw tool available online at http://web.expasy.org/compute_pi/, while the three-dimensional (3D) structure was determined using Phyre2

software available at http://www.sbg.bio.ic.ac.uk/phyre2/ [25]. The subcellular localization of gene protein product was predicted using MultiLoc2 software available at http://abi.inf.uni-tuebingen.de/Services/MultiLoc2 (Blum et al., 2009). *In silico* determinations of hydrophobic/hydrophilic nature and transmembrane helices of the deduced protein were carried out using the translated amino acid sequence of the gene as input data in specialized bioinformatics applications available at http://harrier.nagahama-i-bio.ac.jp/sosui/sosui_submit.html [20,32,33] and http://www.cbs.dtu.dk/services/TMHMM-2.0/ [24], respectively.

3. Results and discussion

Genes in their native forms generally comprise protein-coding nucleotide sequences called exons and non protein-coding sequences referred to as introns (de Conti et al., 2013). During posttranslational
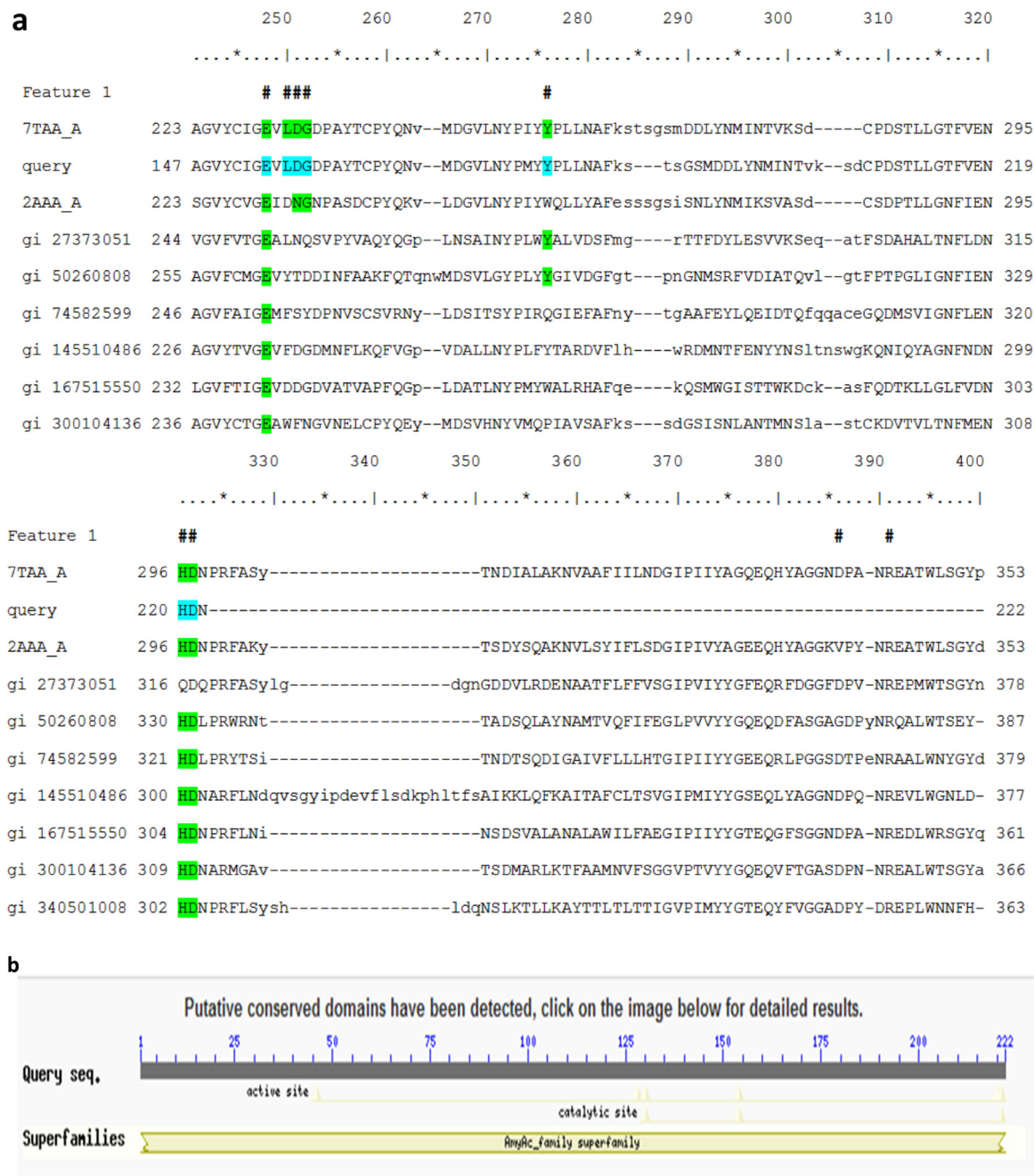
a

```
                    250        260        270        280        290        300        310        320
              ....*....|....*....|....*....|....*....|....*....|....*....|....*....|....*....|

Feature 1           #  ###                        #

7TAA_A        223 AGVYCIGEVLDGDPAYTCPYQNv--MDGVLNYPIYYPLLNAFkstsgsmDDLYNMINTVKSd-----CPDSTLLGTFVEN 295

query         147 AGVYCIGEVLDGDPAYTCPYQNv--MDGVLNYPMYYPLLNAFks---tsGSMDDLYNMINTvk--sdCPDSTLLGTFVEN 219

2AAA_A        223 SGVYCVGEIDNGNPASDCPYQKv--LDGVLNYPIYWQLLYAFesssgsiSNLYNMIKSVASd-----CSDPTLLGNFIEN 295

gi 27373051   244 VGVFVTGEALNQSVPYVAQYQGp--LNSAINYPLWYALVDSFmg----rTTFDYLESVVKSeq--atFSDAHALTNFLDN 315

gi 50260808   255 AGVFCMGEVYTDDINFAAKFQTqnwMDSVLGYPLYYGIVDGFgt---pnGNMSRFVDIATQvl--gtFPTPGLIGNFIEN 329

gi 74582599   246 AGVFAIGEMFSYDPNVSCSVRNy--LDSITSYPIRQGIEFAFny---tgAAFEYLQEIDTQfqqaceGQDMSVIGNFLEN 320

gi 145510486  226 AGVYTVGEVFDGDMNFLKQFVGp--VDALLNYPLFYTARDVFlh----wRDMNTFENYYNSltnswgKQNIQYAGNFNDN 299

gi 167515550  232 LGVFTIGEVDDGDVATVAPFQGp--LDATLNYPMYWALRHAFqe----kQSMWGISTTWKDck--asFQDTKLLGLFVDN 303

gi 300104136  236 AGVYCTGEAWFNGVNELCPYQEy--MDSVHNYVMQPIAVSAFks---sdGSISNLANTMNSla--stCKDVTVLTNFMEN 308


                    330        340        350        360        370        380        390        400
              ....*....|....*....|....*....|....*....|....*....|....*....|....*....|....*....|

Feature 1           ##                                                           #        #

7TAA_A        296 HDNPRFASy--------------------TNDIALAKNVAAFIILNDGIPIIYAGQEQHYAGGNDPA-NREATWLSGYp 353

query         220 HDN---------------------------------------------------------------------------- 222

2AAA_A        296 HDNPRFAKy--------------------TSDYSQAKNVLSYIFLSDGIPIVYAGEEQHYAGGKVPY-NREATWLSGYd 353

gi 27373051   316 QDQPRFASylg--------------dgnGDDVLRDENAATFLFFVSGIPVIYYGFEQRFDGGFDPV-NREPMWTSGYn 378

gi 50260808   330 HDLPRWRNt--------------------TADSQLAYNAMTVQFIFEGLPVVYYGQEQDFASGAGDPyNRQALWTSEY- 387

gi 74582599   321 HDLPRYTSi--------------------TNDTSQDIGAIVFLLLHTGIPIIYYGEEQRLPGGSDTPeNRAALWNYGYd 379

gi 145510486  300 HDNARFLNdqvsgyipdevflsdkphltfsAIKKLQFKAITAFCLTSVGIPMIYYGSEQLYAGGNDPQ-NREVLWGNLD- 377

gi 167515550  304 HDNPRFLNi--------------------NSDSVALANALAWILFAEGIPIIYYGTEQFSGGNDPA-NREDLWRSGYq 361

gi 300104136  309 HDNARMGAv--------------------TSDMARLKTFAAMNVFSGGVPTVYYGQEQVFTGASDPN-NREALWTSGYa 366

gi 340501008  302 HDNPRFLSysh---------------ldqNSLKTLLKAYTTLTLTTIGVPIMYYGTEQYFVGGADPY-DREPLWNNFH- 363
```

b

Fig. 3. (continued)

modifications, however, introns are spliced leaving exons that encode synthesis of functional proteins (de Conti et al. [13]). In this study, we observed the presence of four (4) introns and five (5) exons in the *Aspergillus niger* CSA35 α-amylase gene (*CDF_Amyl*) (Fig. 1). The gene has a partial nucleotide sequence of 926 bp and a deduced protein sequence of 222 amino acids (Fig. 1). Most fungi have been reported to harbor at least one fungal-type α-amylase gene that is evolutionarily intron-rich (da Lage et al. [11]).

The protein-protein BLAST (BLASTp) result obtained showed that the cassava starch-degrading partial α-amylase gene shares 98–99% identity with α-amylase proteins from *Aspergillus oryzae* RIB40, *A. niger*, *A. flavus* NRRL3357, *A. kawachii* (Table 1) and several other *Aspergillus* species and strains already deposited in NCBI Genbank. This is an indication that the amplified alpha amylase gene from *Aspergillus niger* CSA35 actually encodes α-amylase protein. Preliminary biochemical analysis of the encoded α-amylase protein secreted by the fungus

showed that it has a maximum specific activity (Vmax) of 6.90 U/mg protein and a Michaelis constant (Km) of 6.70 mg/ml using soluble starch as substrate [5].

The evolutionary relationship (phylogeny) between the cassava starch-degrading fungal partial α-amylase gene (CDF_Amyl) and α-amylase genes in plants (rice, maize, banana, and cassava plant), other fungi, bacteria, animals (Chimpazee, goat) and human saliva is as shown in Fig. 2. Phylogenies help to provide necessary information on biological diversity, for structuring classifications as well as for proper understanding of events that occurred during evolution [7,8]. The α-amylase gene amplified from the *Aspergillus niger* CSA35, associated with cassava spoilage in Nigeria, is more closely related to α-amylase genes from fungi than to bacterial, plant, or animal α-amylase genes (Fig. 2), as expected in the absence of horizontal gene transfers [23].

The amino acid residues present in the conserved domain of the fungal partial amylase gene (CDF_Amyl) are as shown in Fig. 3a. α-

Fig. 4. 3D-Protein structure of *Aspergillus* sp. CSA35 α-amylase. Image coloured by rainbow, N to C terminus.
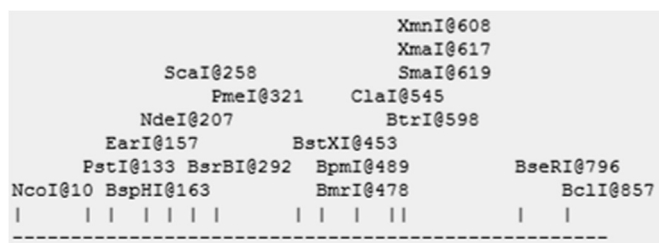

Fig. 5. Endonuclease restriction map of the cassava degrading fungi amylase gene (CDF_Amyl).

Table 2
Isoelectric point and molecular weight of CDF_Amyl.

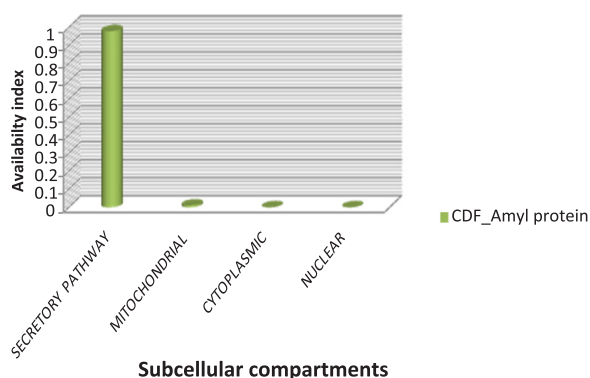| Amylase	Isoelectric point (pI)	Molecular weight (Dalton)	Number of amino acid residues
CDF_Amyl	4.17	25,128.71	222

Fig. 6. Subcellular localization of CDF_Amyl in Aspergillus sp. CSA35.

Amylase genes generally possess catalytic regions that comprise aspartate (D), glutamate (E) and histidine (H) residues that are involved in the enzymes' mechanism of starch hydrolysis [26]. In the present research work, the amino acid residues present in the conserved domain of the CDF_Amyl gene include histidine (H), arginine (R), aspartate (D),

and glutamate (E) amongst others (Fig. 3a). The catalytic site is present in the active site region (Fig. 3b). These findings agree with the report of Liu et al. [26] and affirm that the amylase gene (CDF-Amyl) belonged to the superfamily of α-amylases.

A typical structural characteristic of α-amylase proteins is their barrel-shape appearance [27]. The barrel shape is made up of curly polar α-helixes in the outer parts of the protein and β-sheets close to the middle of the protein. The 3D-protein structure of the cassava starch-degrading fungal amylase (CDF_Amyl) reported in this study (Fig. 4) is identical to that of α-amylase isolated from *Aspergillus oryzae* [27] and also agrees with the typical α-amylase structure described by de Souza and Magalhães [14] and a number of others already deposited in NCBI genbank.

As clearly depicted by the endonuclease restriction linear map shown in Fig. 5, a large array of restriction enzymes (such as *Pst*I, *Sma*I, *Nco*I, BcLI etc) can be used to cut the cassava-degrading fungal α-amylase gene (CDF_Amyl, 926 bp) when considering cloning and gene expression experiments. However, restriction enzymes like *Eco*RI and *Eco*RV will be unable to cut the CDF_Amyl gene (Fig. 5). These findings are in agreement with those of Yu [48] who observed that *Pst*I and *Sma*I can cut alpha-amylase gene (αAmy3 gene enhancer) isolated from rice (*Oryza sativa*). They are, however, in contrast to the report of Puspasari et al. [36] who digested raw starch–degrading α-amylase from *Bacillus aquimaris* MKSC6.2 using *Eco*RI and *Eco*RV during a gene expression study. The choice of restriction enzyme to use when considering digestion of the partial amylase gene (*CDF_Amyl*) in molecular biology studies, however, will not only depend on the choice of restriction enzyme but also on a number of other factors such as the size of gene fragment required, the ability of the choice restriction enzyme to cut both the potential vector of interest and the amylase gene (*CDF_Amyl*), as well as the goal or nature of the research being considered, etc [18,29,43].

The bioinformatics analysis also showed that the extracellular enzyme (alpha amylase) secreted by *Aspergillus niger* CSA35 has an isoelectric point (pI) of 4.17 and a molecular weight (Mw) of 25.13 kDa (Table 2). Molecular weights of amylases vary from about 10–210 kDa but most of these enzymes are in the range of 50–60 kDa. Amylases from *Bacillus caldolyticus* and *Chloroflexus aurantiacus* have the lowest (10 kDa) and highest (210 kDa) molecular weights, respectively [19,40]. The pI values are quite useful for understanding of enzyme behaviour against different purification procedures at different pH values. Amylase isoelectric points are generally in the range of 3.25–10.1 [12]. The pI value obtained in this study is also in accord with the report of Ahmad et al. [2] who opined that alpha amylase from *Aspergillus oryzae* is an acidic enzyme with optimum pH of 4.5 and low isoelectric point (pI = 4.2). The knowledge of the isoelectric point and the Mw would be of great advantage to laboratories or industries when considering small or large scale extraction of the CDF_Amyl protein.

In silico analysis of the subcellular localization of the α-amylase protein in *A. niger* CSA35 revealed that the enzyme (CDF_Amyl) is predominantly found in the secretory pathway within the cassava degrading fungi (Fig. 6). Soluble proteins are synthesized in the secretory pathway and thereafter delivered into the extracellular space where they are utilized for cellular functions [17]. Our report is in agreement with the reports of de Souza and Magalhães [14] and Deb et al. [15] that α-amylases are mainly extracellularly localized.

Many studies have shown that the amino acid sequence of proteins play a crucial role in determining solubility or hydrophobicity of expressed proteins [22,32,33]. In this study, it was observed that the CDF_Amyl protein had a hydrophobicity score of − 0.43 which indicates that the amino acid sequence is of a soluble protein (Fig. 7). Our present finding that the cassava starch-degrading fungal α-amylase (CDF_Amyl) is a soluble protein also agrees with the fact that the enzyme is localized in the secretary pathway (Fig. 6), since proteins secreted into extracellular space are mostly hydrophilic in nature [38].

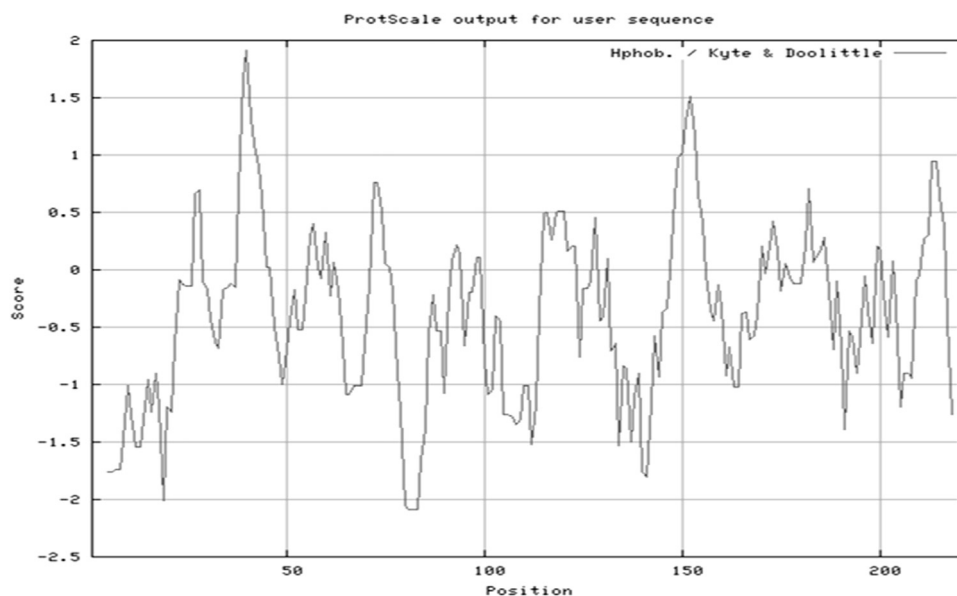The results of the *in silico* analysis also indicate that the amylase

Fig. 7. Hydrophobicity prediction of CDF_Amyl. The overall average hydrophobicity score is − 0.43 (below zero) which indicated that the protein is a soluble protein.
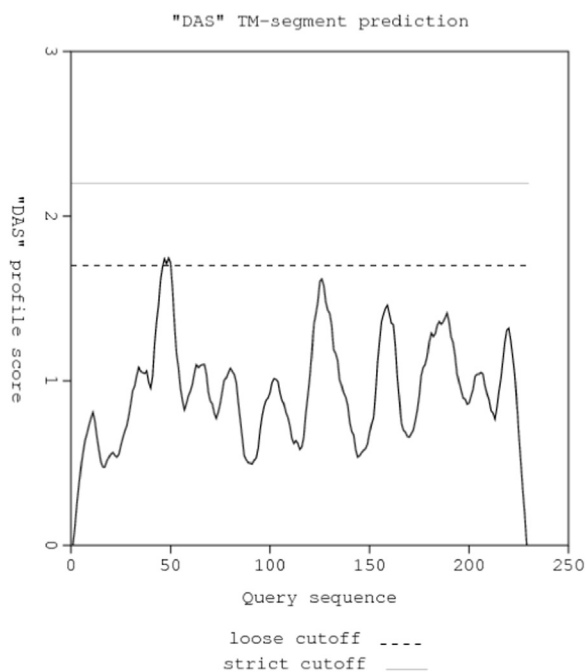

Fig. 8. Prediction of transmembrane segments in CDF_Amyl. No transmembrane helix was found in the CDF_Amyl protein which showed that the protein is a non TM-protein. DAS, dense alignment surface.

protein (CDF_Amyl) under study is not a transmembrane protein (Fig. 8). This is also in agreement with the previous findings of this study that the alpha amylase (CDF_Amyl) from A. niger CSA35 is not hydrophobic but a secreted, soluble protein. According to Phillips et al. [35], transmembrane proteins are typically hydrophobic in nature since they span through the non polar phospholipid bilayers of membranes and a large proportion of their structures are embedded in the lipids. Separation of such proteins from biomembranes is usually difficult except with use of nonpolar solvents, detergents, and/ or denaturing agents [37]. However, the CDF_Amyl protein, being a non-transmembrane and hydrophilic protein, is not bound to the fungal cell membrane but rather secreted into the fungi growth medium. This characteristic can be exploited during industrial isolation and purification of

the cassava starch-degrading fungal α-amylase to obtain the functional protein in its intact or undenatured form [45].

4. Conclusion

The present study findings strongly affirmed that the amplified amylase gene from A. *niger* CSA35 is a member of the α-amylase superfamily. The unique characteristics of the CDF_Amyl gene and its deduced protein could find applications in biotechnological, food and pharmaceutical industries where cloning and further modification of this gene would be required for product development and improvement.

Conflict of interests

The authors do not declare any conflict of interests.

Appendix A. Transparency document

Supplementary data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.bbrep.2018.03.006.

References

[1] C.O. Adebayo, B.I. Aderiye, O.B. Akpor, Assessment of bacterial and fungal spoilage of some Nigerian fermented and unfermented foods, Afri. J. Food Sci. 8 (3) (2014) 140–147.
[2] R. Ahmad, M. Mohsin, T. Ahmad, M. Sardar, Alpha amylase assisted synthesis of TiO₂ nanoparticles: structural characterization and application as antibacterial agents, J. Hazardous Mater. 283 (2015) 171–177.

[3] B. Andersen, J.C. Frisvad, I. Søndergaard, I.S. Rasmussen, L.S. Larsen, Associations between fungal species and water-damaged building materials, Appl. Environ. Microbiol. 77 (12) (2011) 4180–4188.

[4] O.J. Avwioroko, N.J. Tonukari, Isolation and molecular identification of *Aspergillus* species associated with the spoilage of cassava in Nigeria, Nig. J. Sci. Environ. 13 (1) (2014) 55–68.

[5] O.J. Avwioroko, N.J. Tonukari, S.O. Asagba, Biochemical characterization of crude α-amylase of Aspergillus spp. associated with the spoilage of cassava (Manihot esculenta) tubers and processed products in Nigeria, Adv. Biochem. 3 (1) (2015) 15–23.

[6] O.P. Bamidele, F.G. Ogundele, B.A. Ojubanire, M.B. Fasogbon, O.W. Bello, Nutritional composition of "gari" analog produced from cassava (Manihot esculenta) and cocoyam (Colocasia *esculenta*) tuber, Food Sci. Nutr. 2 (6) (2014) 706–711.

[7] D.A. Baum, Reading a phylogenetic tree: the meaning of monophyletic groups, Nature Educ. 1 (1) (2008) 190.

[8] D.A. Baum, S. Offner, Phylogenies and tree thinking, Am. Biol. Teacher 70 (2008) 222–229.

[9] T. Blum, S. Briesemeister, O. Kohlbacher, MultiLoc2: integrating phylogeny and Gene Ontology terms improves subcellular protein localization prediction, BMC Bioinformatics 10 (2009) 274.

[10] K. Chingin, J. Astorga-Wells, M. Pirmoradian Najafabadi, T. Lavold, R.A. Zubarev, Separation of polypeptides by isoelectric point focusing in electrospray-friendly solution using a multiple-junction capillary fractionator, Anal. Chem. 84 (15) (2012) 6856–6862.

[11] J.L. da Lage, M. Binder, A. Hua-Van, Š. Janeček, D. Casane, Gene make-up: rapid and massive intron gains after horizontal transfer of a bacterial α-amylase gene to Basidiomycetes, BMC Evolut. Biol. 13 (1) (2013) 40.

[12] T.M. da Silva, Fungal amylases: applications and functional properties, in: T.M.P. Maria de Lourdes, M. Rai (Eds.), 'Fungal Enzymes', CRC Press, Taylor & Francis Group, New York, 2013, pp. 152–172.

[13] L. De Conti, M. Baralle, E. Buratti, Exon and intron definition in pre-mRNA splicing, Wiley Interdisciplinary Rev. 4 (1) (2013) 49–60.

[14] P.M. de Souza, P.O. Magalhães, Application of microbial α-amylase in industry – a review, Braz. J. Microbiol. 41 (2010) 850–861.

[15] P. Deb, S.A. Talukdar, K. Mohsina, P.K. Sarker, S.M.A. Sayem, Production and partial characterization of extracellular amylase enzyme from Bacillus amyloliquefaciens P-001, SpringerPlus 2 (2013) 154–164.

[16] S. Delkash-Roudsaria, A. Zibaee, M.R.A. Mozhdehi, Digestive α-amylase of *Bacterocera oleae* Gmelin (Diptera: J. King Saud Univ. Sci. 26 (1) (2014) 53–58.

[17] H. Farhan, C. Rabouille, Signalling to and from the secretory pathway, J. Cell Sci 124 (2011) 171–180.

[18] P. Galka, E. Jamez, G. Joachim, P. Soumillion, QuickLib, a method for building fully synthetic plasmid libraries by seamless cloning of degenerate oligonucleotides, PLOS One 12 (4) (2017) e0175146.

[19] R. Gupta, P. Gigras, V.K. Goswami, B. Chauhan, Microbial amylase: a biotechnological perspective, Process Biochem. 38 (2003) 1599–1616.

[20] T. Hirokawa, S. Boon-Chieng, S. Mitaku, SOSUI: classification and secondary structure prediction system for membrane proteins, Bioinformatics 14 (1998) 378–379.

[21] L. Holm, A.K. Koivula, P.M. Lehtovaara, A. Hemminki, J.K.C. Knowles, Random mutagenesis used to probe the structure and function of Bacillus stearothermophilus alpha-amylase, Protein Engr. 3 (1990) 181–191.

[22] H.L. Huang, P. Charoenkwan, K.F. Kao, H.C. Lee, F.L. Chang, W.L. Huang, S.J. Ho, L.S. Shu, W.L. Chen, S.Y. Ho, Prediction and analysis of protein solubility using a novel scoring card method with dipeptide composition, BMC Bioinform. 13 (Suppl 17) (2012) S3.

[23] F. Husnik, J.P. McCutcheon, Functional horizontal gene transfer from bacteria to eukaryotes, Nat. Rev. Microbiol. 16 (2) (2018) 67.

[24] M. Ikeda, M. Arai, D.M. Lao, T. Shimizu, Transmembrane topology prediction methods: a re-assessment and improvement by a consensus method using a dataset of experimentally-characterized transmembrane topologies, In Silico Biol. 2 (1) (2002) 19–33.

[25] L.A. Kelley, M.J.E. Sternberg, Protein structure prediction on the web: a case study using the Phyre server, Nat. Protoc. 4 (2009) 363–371.

[26] Y. Liu, Y. Lei, X. Zhang, Y. Gao, Y. Xiao, H. Peng, Identification and phylogenetic characterization of a new subfamily of α-amylase enzymes from marine microorganisms, Mar. Biotechnol. 14 (3) (2012) 253–260.

[27] M.J.E.C. Maarel, B. Veen, J.C.M. Uitdehaag, H. Leemhuis, L. Dijkhuizen, Properties and applications of starch-converting enzymes of the α-amylase family, J. Biotechnol. 94 (2002) 137–155.

[28] R. Maheshwari, G. Bharadwaj, M.K. Bhat, Thermophilic fungi: their physiology and enzymes, Microbiol. Mol. Biol. Rev. 64 (3) (2000) 461–488.

[29] S. Manna, A. Harman, J. Accari, C. Barth, Altering the selection capabilities of common cloning vectors via restriction enzyme mediated gene disruption, BMC Res. Notes 6 (1) (2013) 85.

[30] A. Marchler-Bauer, Y. Bo, L. Han, J. He, C.J. Lanczycki, S. Lu, F. Chitsaz, M.K. Derbyshire, R.C. Geer, N.R. Gonzales, M. Gwadz, CDD/SPARCLE: functional classification of proteins via subfamily domain architectures, Nucl. Acids Res. 45 (D1) (2017) D200–D203, http://dx.doi.org/10.1093/nar/gkw1129.

[31] K. Metin, O. Koc, B.B. Ateşlier, H.H. Biyik, Purification and characterization of α-amylase produced by *Penicillium citrinum* HBF62, Afr. J. Biotechnol. 9 (45) (2010) 7692–7701.

[32] S. Mitaku, T. Hirokawa, Physicochemical factors for discriminating between soluble and membrane proteins: hydrophobicity of helical segments and protein length, Protein Eng. 12 (11) (1999) 953–957.

[33] S. Mitaku, T. Hirokawa, T. Tsuji, Amphiphilicity index of polar amino acids as an aid in the characterization of amino acid preference at membrane-water interfaces, Bioinformatics 18 (2002) 608–616.

[34] O.B. Oyewole, L.O. Sanni. Constraints in traditional cassava processing: A case of 'fufu' production. in: Transformation Alimentaire Du Manioc (Cassava Food Processing). Editors: Tom Agbor Egbe, AlainBrauman, Dany Griffon, Serge Treche. Institut Francais de Recherche Scientifique Pour le Development en Cooperation (ORSTOM), Paris, France. ISBN 2-7099-1279-1. Pp. 523–529.

[35] R. Phillips, T. Ursell, P. Wiggins, P. Sens, Emerging roles for lipids in shaping membrane-protein function, Nature 459 (7245) (2009) 379–385.

[36] F. Puspasari, O.K. Radjasa, A.S. Noer, Z. Nurachman, Y.M. Syah, M. Maarel, L. Dijkhuizen, Š. Janeček, D. Natalia, Raw starch-degrading α-amylase from Bacillus aquimaris MKSC 6.2: isolation and expression of the gene, bioinformatics and biochemical characterization of the recombinant enzyme, J. Appl. Microbial. 114 (1) (2013) 108–120.

[37] T. Rabilloud, Solubilization of Proteins in 2DE: An Outline. In Two-Dimensional Electrophoresis Protocols, Humana Press, Chicago, 2009, pp. 19–30.

[38] C. Rabouille, Pathways of unconventional protein secretion, Trends in cell biology 27 (3) (2017) 230–240.

[39] B. Ray, A. Bhunia, Fundamental Food Microbiology, CRC press, Florida, 2013.

[40] P. Saranraj, D. Stella, Fungal amylase: a review, Int. J. Microbiol. Res. 4 (2) (2013) 203–211.

[41] F. Siddique, I. Hussain, M.S. Mahmood, S.I. Ahmed, A. Iqbal, Isolation and characterization of a highly thermostable alpha-amylase enzyme produced by *Bacillus licheniformis*, Pak. J. Agri. Sci. 51 (2) (2014) 309–314.

[42] A.S. Sussman, Longevity and survivability of fungi, The fungi 3 (2013) 447–486.

[43] A. Taheri, S.J. Robinson, I. Parkin, M.Y. Gruber, Revised selection criteria for candidate restriction enzymes in genome walking, PLOS One 7 (4) (2012) e35117.

[44] N.J. Tonukari, E.E. Oliseneku, O.J. Avwioroko, E. Aganbi, O.C. Orororo, A.A. Anigboro, A novel pig feed formulation containing *Aspergillus niger* CSA35 pretreated-cassava peels and its effect on growth and selected biochemical parameters of pigs, Afr. J. Biotechnol. 15 (19) (2016) 776–785.

[45] O. Vit, J. Petrak, Integral membrane proteins in proteomics. How to break open the black box? J. Proteom. 153 (2017) 8–20.

[46] E.O. Voit, Computational Analysis of Biochemical Systems: A Practical Guide for Biochemists and Molecular Biologists, Cambridge University Press, United Kingdom, 2000.

[47] A.G.N. Wetie, I. Sokolowska, A.G. Woods, U. Roy, K. Deinhardt, C.C. Darie, Protein–protein interactions: switch from classical methods to proteomics and bioinformatics-based approaches, Cell. Mol. Life Sci. 71 (2) (2014) 205–228.

[48] S.M. Yu, Rice α-amylase transcriptional enhancers. U.S. Patent No. 7,094,958. Washington, DC: U.S. Patent and Trademark Office, 2006.

[49] S. Zidani, A. Ferchichi, M. Chaieb, Genomic DNA extraction method from pearl millet (*Pennisetum glaucum*) leaves, Afr. J. Biotechnol. 4 (8) (2005) 862–866.