World Journal of Microbiology and Biotechnology (2018) 34:95
https://doi.org/10.1007/s11274-018-2478-8

ORIGINAL PAPER

Introduction of novel thermostable α-amylases from genus *Anoxybacillus* and proposing to group the *Bacillaceae* related α-amylases under five individual GH13 subfamilies

Arzu Coleri Cihan[1] · Emine Derebay Yildiz[2] · Ergin Sahin[1] · Ozal Mutlu[3]

Received: 17 February 2018 / Accepted: 12 June 2018 / Published online: 15 June 2018

Abstract

Among the thermophilic *Bacillaceae* family members, α-amylase production of 15 bacilli from genus *Anoxybacillus* was investigated, some of which are biotechnologically important. These *Anoxybacillus* α-amylase genes displayed ≥ 91.0% sequence similarities to *Anoxybacillus* enzymes (*ASKA, ADTA* and *GSX-BL*), but relatively lower similarities to *Geobacillus* (≤ 69.4% to *GTA, Gt-amyII*), and *Bacillus aquimaris* (≤ 61.3% to *BaqA*) amylases, all formerly proposed only in a Glycoside Hydrolase 13 (GH13) subfamily. The phylogenetic analyses of 63 bacilli-originated protein sequences among 93 α-amylases revealed the overall relationships within *Bacillaceae* amylolytic enzymes. All bacilli α-amylases formed 5 clades different from 15 predefined GH13 subfamilies. Their phylogenetic findings, taxonomic relationships, temperature requirements, and comparisonal structural analyses (including their CSR-I-VII regions, 12 sugar- and 4 calcium-binding sites, presence or absence of the complete catalytic machinery, and their currently unassigned status in a valid GH13 subfamiliy) revealed that these five GH13 α-amylase clades related to familly share some common characteristics, but also display differentiative features from each other and the preclassified ones. Based on these findings, we proposed to divide *Bacillaceae* related GH13 subfamilies into 5 individual groups: the novel *a2* subfamily clustered around *α*-amylase B2M1-*A* (*Anoxybacillus* sp.), the *a1, a3* and *a4* subfamilies (including the representatives E184aa-*A (Anoxybacillus* sp.), *ATA* (*Anoxybacillus tepidamans*), and *BaqA*,) all of which were composed from the division of the previously grouped single subfamily around *α*-amylase *BaqA*, and the undefinite subfamily formerly defined as *xy* including *Bacillus megaterium* NL3.

Keywords *Anoxybacillus* · *Bacillaceaei* · α-Amylases · Glycoside hydrolases · Novel GH13 subfamilies

Introduction

α-Amylases (EC 3.2.1.1) are endo-acting enzymes catalyzing the hydrolysis of α-1,4-glucosidic linkages in starch, glycogen and related polysaccharides using the retaining mechanism (Janeček et al. 2014). The tertiary structures of α-amylases bear three domains designated as A, B, and C. With being the most conserved one, domain A contains the active site of the enzymes in a typical $(\beta/\alpha)_8$ TIM-barrel structure (Janeček et al. 1997, 2014). The Asp, Glu and Asp residues, positioned on β-strands of the barrel domain, mostly create the catalytic triad consisting of catalytic nucleophile, proton donor, and transition-state stabilizer functions, respectively (Matsuura et al. 1984; Janeček 2002). There are also seven conserved sequence regions (CSRs), CSR-I to CSR-VII, that are characteristic for the α-amylase family and function in catalytic or substrate-binding activities (Janeček 2002).

According to the actual sequence-based classification system in the Carbohydrate-Active enZYmes database (CAZy; http://www.cazy.org/), (Cantarel et al. 2009), α-amylases represent the largest group among Glycoside Hydrolyses (GHs), (Lombard et al. 2014). α-Amylase specificity is

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s11274-018-2478-8) contains supplementary material, which is available to authorized users.

✉ Arzu Coleri Cihan
arzucoleri@gmail.com; acihan@science.ankara.edu.tr

[1] Faculty of Science, Department of Biology, Ankara University, Tandogan, 06100 Ankara, Turkey

[2] Institute of Biotechnology, Ankara University, Besevler, 06100 Ankara, Turkey

[3] Faculty of Arts and Sciences, Department of Biology, Marmara University, Goztepe, 34722 Istanbul, Turkey

presented mostly in the GH13 family in addition to the GH57, GH119 and eventually in GH126 (Janeček et al. 2014, 2015). The family GH13 also forms the clan GH-H with the families GH70 and GH77 in the CAZy database. GH13 family is divided into totally 42 subfamilies, nevertheless the specific α-amylase activity is present only in the 15 subfamilies of GH13_1, 5, 6, 7, 15, 19, 24, 27, 28, 32, 36, 37, 39, 41, and 42 (Stam et al. 2006; Cantarel et al. 2009; Sarian et al. 2017). Many new subfamilies have also been proposed and awaiting to be assigned into a definite subfamily (Stam et al. 2006).

The starch is abundant in nature; therefore, the starch hydrolyzing amyloytic enzymes are also widespread among the living beings from eukaryotic to the prokaryotic origin, though displaying some differences in their enzyme activities and substrate specificities (Stam et al. 2006). Moreover, all the *Bacillaceae* family members frequently diverge in nature and human-made environments due to their endospore forming capabilities and their resistance to many harsh conditions such as high or low pH and temperature values. The temperature requirements of these bacilli also vary from mesophilic to thermophilic growth (Nazina et al. 2001; Cihan et al. 2014b). Since Cohn had first introduced the mesophilic genus of this family as genus *Bacillus* in the eighteenth century, its thermophilic members from the genera *Geobacillus* and *Anoxybacillus* could only be described after 2000's (Cohn 1872; Pikuta et al. 2000; Nazina et al. 2001). These explanations clarify why all the commercially available α-amylases are from mesophilic bacilli such as *Bacillus licheniformis* (MEGAZYME, E-BLAAM), *Bacillus amyloliquefaciens* (MEGAZYME, E-BAASS), and *Bacillus subtilis* subsp. *subtilis* str. 168 (PROZOMIX, PRO-E0403). The α-amylases of the mesophilic species *Bacillus aquimaris* MKSC 6.2 (*BaqA*), (Puspasari et al. 2013) and *Bacillus megaterium* NL3 (*BmaN1*), (Sarian et al. 2017) are also the well-characterized ones.

After the twentieth century, due to some beneficial features of the thermostable enzymes in the industrial processes, the isolation of novel thermophilic bacilli gained considerable attention. In the two phylogenetic diversity studies dealing with the *Bacillaceae* family revealed that *Anoxybacillus* species are more predominant in the extreme hot environments probably owing to their carbohydrate degrading abilities among the other genera members (Derekova et al. 2008; Cihan 2013). Many thermostable α-amylases from the genera *Anoxybacillus* and *Geobacillus* were characterized including: *Anoxybacillus flavithermus* (Bolton et al, 1997; Tawil et al. 2012; Agüloğlu Fincan et al. 2014; Ozdemir et al. 2015, 2016a), *Anoxybacillus amylolyticus* (Poli et al. 2006) and *Anoxybacillus caldiproteolyticus* D504 and D621 (Ozdemir et al. 2016b), *Anoxybacillus* spp. KP1, SK3-4 (*ASKA*), DT3-1 (*ADTA*), TSSC-1, IB-A, AH1, and YIM 342 (Chai et al. 2012; Kikani and Singh 2012; Hauli et al. 2013; Matpan and; Güven 2014; Acer et al. 2015; Zhang et al. 2015) in addition to *Geobacillus thermoleovorans* MTCC 4220 (*Gt-amyII*), CCB_US3_UF5 (*GTA*), YN, NP54 (Berekaa et al. 2007; Rao and Satyanarayana 2007; Mok et al. 2013; Mehta and Satyanarayana 2014), *G. thermoleovorans* subsp. *stromboliensis* Pizzo (*amyA*), (Finore et al. 2011), *Geobacillus thermodenitrificans* HRO10 (Ezeji and Bahl 2006), and *Geobacillus* sp. IIPTN (Dheeran et al. 2010).

Despite the accumulation of several α-amylase sequence data, isolation and characterization studies, it is worth mentioning that most of these endospore-forming bacilli-originated enzymes have not been validly assigned into any of the current 42 GH13 subfamilies until now, except the presence of α-amylases belonging to *Bacillaceae* in GH13_5, 19, 28 and 36 subfamilies. Nevertheless, some novel subfamilies containing these α-amylases from *Bacillaceae* members were proposed but still non-defined. In chronological order, the three undefined bacilli GH13 subfamilies sharing two conserved tryptophan residues (W200-W201, E184aa-*A* numbering) which are unexceptionally positioned between the loop 3 and β4 strand of the catalytic TIM-barrel structure, were proposed as (i) the mesophilic group including *Bacillus aquimaris* (*BaqA*), *B. coahuilensis, Bacillus* sp. SG-1 and NRRL B-14,911 α-amylases (Puspasari et al. 2013); (ii) the *ASKA* and *ADTA, BaqA, GTA, amyA*, and *Gt-amyII* amylases (Ranjani et al. 2015; Janeček et al. 2015; Sarian et al. 2017); and finally (iii) the non-defined *BmaN1* GH13 subfamily which includes the mesophilic enzyme of *Bacillus megaterium* NL3 having an atypical catalytic triad (Sarian et al. 2017). The thirdly proposed subfamily discriminated from the other two preceding ones by the absence of a complete catalytic machinary. Among these bacilli α-amylases, the X-ray crystal structures of *Anoxybacillus* sp. SK3-4 α-amylase (*TASKA*, PDP ID: 5A2B; it is a truncated form of *ASKA*) and *G. thermoleovorans GTA* α-amylase (PDB ID: 4E2O) were successfully unveiled (Mok et al. 2013; Chai et al. 2016).

In this study, a total of 15 isolates and reference strains from genus *Anoxybacillus* were screened for their α-amylase activity and some of them were found to exhibit very high level of thermostable enzyme activities, which possess potential in biotechnological processes and may satisfy the industrial demands. After the PCR amplification and sequence determination of these *Anoxybacillus* α-amylase genes, their putative protein sequences were subsequently used in BLAST search, phylogenetic analyses and amino acid sequence alignments along with their most closely related sequences belonging to *Bacillaceae* family members. Phylogenetic analyses clustered these 15 sequences together with the other thermostable *Anoxybacillus* α-amylases of *ASKA, ADTA* and *GSX-BL* and with some putative *Anoxybacillus* α-amylase sequences,

but not with genus *Geobacillus* or *Bacillus* related *GTA, BaqA, amyA* or *Gt-amyII* enzymes as proposed previously. When all 15 *Anoxybacillus* sequences and their homolog *Bacillaceae* family related α-amylase sequences were considered as a whole, the family branched into five separate clusters which exhibit a novel and three reorganized GH13 subfamilies in addition to the undefined "*x*y" labeled subfamily containing *BmaN1* (Sarian et al. 2017). These four representative sequences from the newly proposed or rearranged GH13 subfamilies were subjected to further secondary and tertiary structural comparison analyses via generally used *in silico* techniques based on their detailed domain and surface structures and maltose binding sites. The comparison of these new *Anoxybacillus* α-amylase sequences with a wide sequence collection, containing the other endo-spore forming bacilli sequences, depicted directly the evolutionary history of α-amylases from *Bacillaceae* family as all these bacilli-related clusters share some common features. Moreover, this approach allowed us to group *Bacillaceae* α-amylases under more accurate groups based on their taxonomicaly related genera, temperature optima, sequence features and their related phylogenetic analyses. In conclusion, we proposed a novel bacilli α-amylase GH13 subfamily in addition to the division of the previously proposed, but still unassigned GH13 subfamily into individual subfamilies, which contained *ASKA* and *ADTA, BaqA, GTA, amyA*, and *Gt-amyII* α-amylases (Ranjani et al. 2015; Janeček et al. 2015; Sarian et al. 2017) originated from different taxonomic groups: the genus *Anoxybacillus, Geobacillus*, or *Bacillus*.

Materials and methods

Bacterial isolates and reference strains

The bacterial isolates of *Anoxybacillus* spp. A321, A3210, D222b, E184aa, E184ab and E208a as well as the reference strains of *Anoxybacillus salavatliensis* DSM 22626^T, *Anoxybacillus gonensis* NCIMB 13933^T, *Anoxybacillus ayderensis* NCIMB 13972^T, *Anoxybacillus kestanbolensis* NCIMB 13971^T, *Anoxybacillus kamchatkensis* DSM 14988^T, *Anoxybacillus flavithermus* DSM 2641^T, *Anoxybacillus amylolyticus* DSM 15939^T, *Anoxybacillus thermarum* DSM 17141^T and *Anoxybacillus kamchatkensis* subsp. *asaccharedens* DSM 18475^T were used for the α-amylase assays, gene amplifications and phylogenetic analyses. The isolation procedures, characterization studies and 16S rRNA gene sequencing analysis of these bacilli were determined as previously described by Cihan et al. (2014a.)

Qualitative and quantitative α-amylase assays

The isolates and reference strains were screened for their amylolytic activity on Medium-I agar plates containing 1% soluble starch (Suzuki et al. 1976) upon incubation for 24 h at 55 °C. Then, the plates were treated with 0.2% I_2 dissolved in 2% KI solution, and the halo zones were measured around the colonies in order to determine the amylolytic enzyme producing ones. *Geobacillus stearothermophilus* DSM 22^T and *Bacillus amyloliquefaciens* DSM 7^T were used as the α-amylase producing references for comparison. Before determining the quantitative α-amylase activity for enzyme production, the bacilli were incubated in a modified Santos and Martins (2003) broth (1.0% tryptone, 0.5% yeast extract, 1.0% soluble starch), (Santos and Martins 2003) by shaking at 150 rpm during 72 h. The incubation temperature values (55–65 °C) and pH of the media (pH 7.0–8.0) were adjusted according to the bacteria. Extracellular α-amylase activity was carried out using 3,5-dinitrosalicylic acid (DNS) method with some slight modifications (Miller 1959). The standard reaction mixture contained 0.5 ml of each 0.2 M sodium phosphate buffer, 2% soluble starch and appropriately diluted enzyme solution. The effects of temperature and pH on α-amylase activities were determined as previously described (Ozdemir et al. 2016b). The reactions were carried out at the required optimal pH and temperature values of bacteria for 10 min and stopped by boiling 5 min after addition of 1 ml DNS. Finally, the absorbance was measured spectrophotometrically at 540 nm. One unit of α-amylase activity was defined as the amount of enzyme that catalyzed the liberation of reducing sugar equivalent to 1 µmol of maltose per min under the assay conditions. The millimolar extinction coefficient was calculated using maltose as the standard. Total amount of protein was determined by the Lowry method (1976) using Bovine Serum Albumin (Lowry and Tinsley 1976). The enzyme assays were performed at least three times.

Amplification of α-amylase genes

Cultures grown on Medium-I plates for 18 h at 55 °C were used for genomic DNA extraction (Fermentas K0512, Genomic DNA purification kit). The α-amylase genes were amplified by PCR using the protocol of Chai and colleagues with some modifications (Chai et al. 2012). The PCR conditions were optimized according to the primers' annealing temperature (T_m) as 55 or 58 °C and by adjusting the $MgCl_2$ concentration as 2.0 or 2.5 mM which varied for the bacilli used. The PCR products were purified with GeneJET PCR Purification kit (Fermentas K0702) and sequenced by using an ABI 3100 gene sequencer with a Bigdye cycle sequencing kit (Macrogen, Europe).

Sequence collections

The entire protein sequences, deduced from the *Anoxybacillus* nucleotide sequences, were taken into the query sequence analysis in the BLASTN and BLASTP programs (Altschul et al. 1997). In collecting the sequences, the criteria, proposed by Stam et al. (2006) were used for the identification of distinct subfamilies sufficiently. The caught sequences from the blast search contained the most similar sequences sharing high similarities, appearing at the top on the blast report, displaying a slow and progressive increase in *E*-values as well as belonging to a closely related taxonomic group from which the sequences were obtained (Stam et al. 2006). These sequences were also checked for the presence of (i) the $(\beta/\alpha)_8$-barrel architecture (Janeček 2002; Hostinová et al. 2010), (ii) the signal sequences, (iii) all the seven CSRs of the GH13 α-amylase family (Janeček 2002; Janeček et al. 2015), (iv) the catalytic triad, (v) the calcium ion binding sites from 1 to 4 (Mok et al. 2013; Chai et al. 2016), (vi) the substrate binding subsites (Chai et al. 2016), (vii) a pair of tryptophan (W200,W201, E184aa-*A*) between the CSR-V (loop 3) and CSR-II (β4 strand), (Puspasari et al. 2013), (viii) the consecutively repeated aromatic motifs of phenylalanine and tyrosine residues at the end of the C-terminal segment (Mok et al. 2013; Janeček et al. 2015), (ix) the signature residues with the invariable consecutive lysine-arginine (KR) at the terminus of domain C, and (x) the 4 residues involving in the formation of putative S1 and S2 transmembrane regions. Therefore, in addition to the 15 newly introduced amino acid sequences in this study, additional 78 sequences were retrieved directly from the Universal Protein Knowledgebase (UniProt Consortium 2017), the GenBank (Benson et al. 2014), and the Protein Data Bank (PDB), (Berman et al. 2000) or from the annotated whole genome projects found in the NCBI-genome databases (https://www.ncbi.nlm.nih.gov/genome/). The sequence sets also covered all the well-defined GH13 subfamilies in the CAZy database, having α-amylase specificity including subfamilies GH13_1, 5, 6, 7, 15, 19, 24, 27, 28, 32, 36, 37, 39, 41, 42 (Stam et al. 2006; Cantarel et al. 2009; Lombard et al. 2014), the unassigned cyclomaltodextrinase from *Flavobacterium* sp. No 92 (GH13_*??*; (Fritzsche et al. 2003)), and the formerly suggested bacilli related GH13 subfamilies the first one akin to *BaqA* α-amylase by Puspasari and colleagues (2013), the second one including *BaqA, ADTA, ASKA, GtamyII* and *GTA* α-amylases by Chai et al. (2012), Chai et al. (2016), Janeček et al. (2015), Sarian and colleagues (2017), and lastly the GH13_xy around *BmaN1* by Sarian et al. (2017). So, totally 93 identified enzymes and hypothetical proteins were studied with *in silico* techniques. The list of all the polypeptide sequences used for this study, their accession numbers from UniProt (UniParc), (UniProt Consortium 2017) and GenBank (Benson et al. 2014) databases, their amino acid lengths, the results of the specific α-amylase activity experiments, and their related classified or newly proposed GH13 subfamilies from *a1* to *a4* were presented in Supplementary file, Table SI.

Bioinformatics analysis

As in the case of previously published bacilli related new α-amylase subfamily establishment studies, similar, analogous or more comprehensive bioinformatic tools were used by comparing wider and more meaningful amylolytic enzyme sequence sets in this study (Chai et al. 2012, 2016; Puspasari et al. 2013, Ranjani et al. 2015; Janeček et al. 2015; Sarian et al. 2017). The conserved domain types and probable families of 15 new *Anoxybacillus* originated deduced protein sequences were searched with Pfam (http://pfam.xfam.org/; Finn et al. 2016) and NCBI-Conserved domain databases (https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi; Marchler-Bauer et al. 2017). The molecular weights and the isoelectric points of these sequences were predicted using Geneious R10 program (http://www.geneious.com; Kearse et al. 2012). The secondary structures including the α-helices and β-strands, the sugar binding pockets, and the transmembrane residues were predicted by Phyre2 server (Protein Homology/AnalogY Recognition Engine), (http://www.sbg.bio.ic.ac.uh/phyre/; Kelley et al. 2015). The signal peptides from transmembrane associated regions were defined using SignalP 4.0 server (http://www.cbs.dtu.dk/services/SignalP/; Petersen et al. 2011). The extracted sequences containing domains A and B of the compact $(\beta/\alpha)_8$ TIM-barrel structure in addition to domain C were subjected to a multiple sequence alignment using the CLUSTAL-OMEGA program (https://www.ebi.ac.uk/Tools/msa/clustalo/; Sievers et al. 2011). The similarities were maximized by a manual fine adjustment in order to align the individual CSRs and the borders of CSRs were determined from previous studies (Janeček 2002; Janeček et al. 2015). For the currently organized and suggested 4 novel α-amylase GH13 subfamilies from *a1* to *a4*, the individual sequence logos were created for their seven CSRs and two adjacent aromatic tryptophan motifs typical for the novel subfamilies, using the WebLogo 3.5.0 server (http://weblogo.berkeley.edu; Crooks et al. 2004). All phylogenetic analysis containing the GH13 family members was performed with using the Geneious R10 server. The evolutionary distance matrix and its related phylogenetic tree were constructed using the UPGMA algorithm (Sokal and Michener 1958) with the bootstrap values selected based on 1000 replications (Felsenstein 1985). Additionally, the phylogenetic analyses on 16S rRNA genes of the thermophilic bacilli members from genera *Anoxybacillus* and *Geobacillus* were carried out as described previously by Cihan et al. (2014a).

The three-dimensional (3D) structural models of four representative sequences belonging to each of the four suggested subfamilies were constructed using the SWISS-MODEL (http://swissmodel.expasy.org/; Biasini et al. 2014) program server. Models were compared with the template X-Ray structures of *Anoxybacillus* sp. SK3-4 α-amylase (*TASKA*, PDB ID: 5A2B), (Chai et al. 2016) and *G. thermoleovorans* CCB US3 UF5 α-amylase (*GTA* , PDB ID: 4E2O), (Mok et al. 2013). The predicted 3D models were further visualized and drawn with PyMOL software (The PyMOL Molecular Graphics System, version 1.7.4 Schrödinger, LLC; http://www.pymol.org) and ICM-BrowserPro version 3.8-5 (MolSoft LLC, La Jolla, CA, USA; http://www.molsoft.com/icm_browser_pro.html) programs. RMSD (Å) values were calculated using PyMOL. All computational analyses were performed using Intel® Core™ i5-4570 CPU@ 3.20 GHz processor and 16 GB RAM on Windows7 Enterprise 64x environment.

Results

Extracellular α-amylase activities

The α-amylase screening assay revealed that 15 bacilli were positive for the halo zones formed around their colonies on plates containing soluble starch, but especially the isolates E184aa, E184ab, E208a, D222b and the reference strains of *G. stearothermophilus, A. flavithermus, A. salavatliensis* and *A. amylolyticus* produced relatively higher clear zones in diameter. In the quantitative α-amylase assay, extracellular enzyme activities of the bacilli were found to vary from 0.8 U/g to 341.3 U/g (Fig. 1). The ones producing higher zones also displayed the highest amylolytic enzyme activities. It is also noteworthy that at 65 °C, pH 7.0, the measured enzyme activity values of E184aa, E184ab and D222b isolates were the best among the bacilli and the amount of α-amylases produced by these isolates were 1.5 to twofold higher than the control *G. stearothermophilus* DSM 22^T which was known as biotechnologically important.

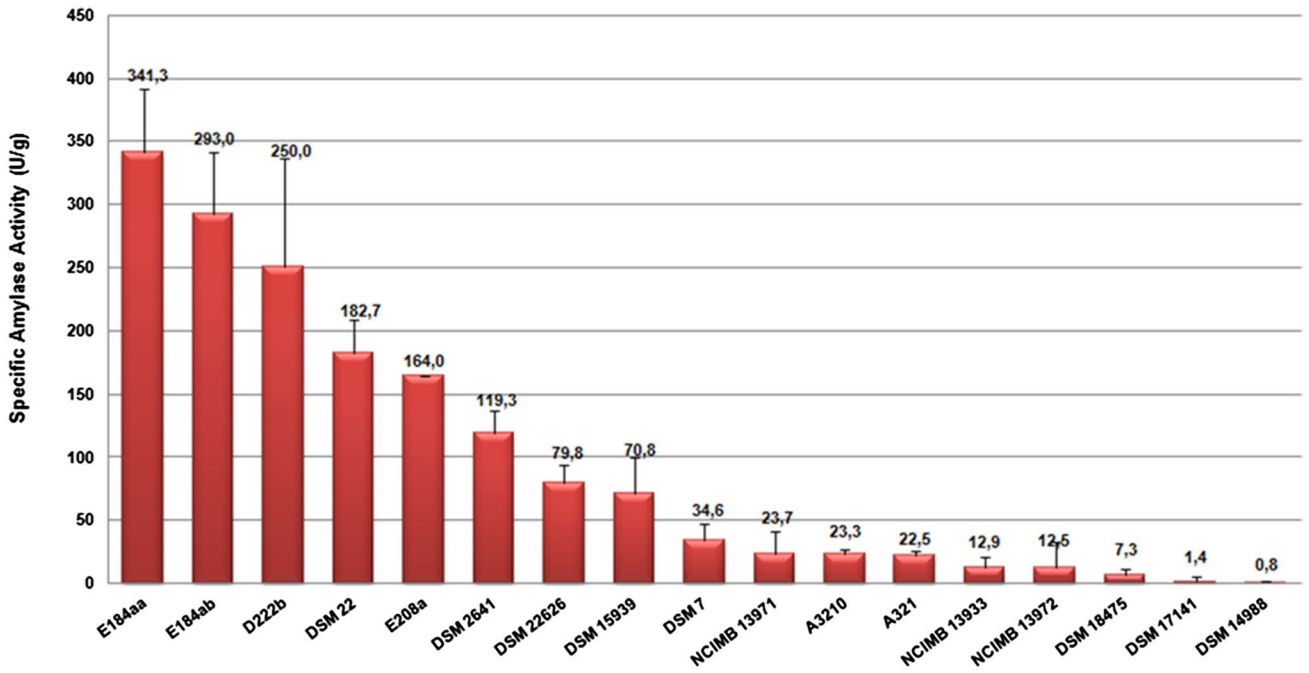

Fig. 1 Extracellular α-amylase activities of the bacilli when 1% of soluble starch was used as the substrate. The isolates and reference strains as well as their optimal temperature and pH values, used during determination of the extracellular α-amylase activities, are as follows: *Anoxybacillus* sp. E184aa (65 °C, pH 7.0), *Anoxybacillus* sp. E184ab (65 °C, pH 7.0), *Anoxybacillus* sp. D222b (65 °C, pH 7.0), *G. stearothermophilus* DSM 22^T (65 °C, pH 7.0), *Anoxybacillus* sp. E208a (65 °C, pH 7.0), *A. flavithermus* DSM 2641^T (65 °C, pH 7.0), *A. salavatliensis* DSM 22626^T (65 °C, pH 7.0), *(A) amylolyticus* DSM 15939^T (65 °C, pH 7.0), *(B) amyloliquefaciens* DSM 7^T (37 °C, pH 7.0), *A. kestanbolensis* NCIMB 13971^T (55 °C, pH 8.0), *Anoxybacillus* sp. A3210 (65 °C, pH 7.0), *Anoxybacillus* sp. A321 (65 °C, pH 7.0), *A. gonensis* NCIMB 13933^T (55 °C, pH 7.5), *A. ayderensis* NCIMB 13972^T (55 °C, pH 8.0), *A. kamchatkensis* subsp. *asaccharedens* DSM 18475^T (55 °C, pH 7.5), *A. thermarum* DSM 17141^T (60 °C, pH 7.0), and *A. kamchatkensis* DSM 14988^T (60 °C, pH 8.0)

Molecular characterization of the nucleotide sequences

The PCR amplicon sizes of the 15 *Anoxybacillus* were between 1515 and 1518 nucleotides which were compatible with the length of known α-amylase gene sequences, (1518 bp for only *A. flavithermus* and *A. amylolyticus*). The amplicons had the DNA G + C contents varying from 42.1 to 43.8 mol%. All the sequenced α-amylase genes were deposited in GenBank database (Benson et al. 2014) under the accession numbers of KY426431 (E184aa), KY426432 (E208a), KY426433 (E184ab), KY426434 (D222b), KY426435 (A3210), KY426436 (A321), KY426437 (*A. salavatliensis* DSM 22626^T), KY426438 (*A. gonensis* NCIMB 13933^T), KY426439 (*A. ayderensis* NCIMB 13972^T), KY426440 (*A. kestanbolensis* NCIMB 13971^T), KY426441 (*A. kamchatkensis* DSM 14988^T), KY426442 (*A. flavithermus* DSM 2641^T), KY426443 (*A. amylolyticus* DSM 15939^T), KY426444 (*A. thermarum* DSM 17141^T) and KY426445 (*A. kamchatkensis* subsp. *asaccharedens* DSM 18475^T). The nucleotide sequence query in BLASTN showed that the open reading frames (ORFs) of these novel sequences displayed 91.0–99.0% gene sequence similarities with α-amylases of *ASKA, ADTA*, and *GSX-BL*. The CDS (Conserved Domain Search) analysis, based on the annotations of the subfamily domain architectures, also revealed that they all belong to AmyAc_bac_CMD_like_2 (cd11339) Conserved Protein Domain Family, which is identified as the α-amylase catalytic domain in bacterial cyctomaltodextinases and the related proteins. Then the nucleotide sequences were converted into deduced amino acid sequences of polypeptides that composed of 504 to 505 amino acids in order to use in further phylogenetic analysis and *in silico* techniques.

The query results and the phylogenetic analysis

The BLASTP search query using 15 *Anoxybacillus* amino acid sequences revealed 48 similar putative or biochemically characterized α-amylases belonging to endospore-forming from *Bacillaceae* family. Sequence collection of GH13 α-amylases used in this study was given in the Supplementary file, Table SI. Interestingly, all of these closely related homologue α-amylases have not been classified properly into a definite GH13 subfamily yet. Therefore, these new *Anoxybacillus* α-amylase sequences were aligned not only with their 48 related unclassified homologues, but also along with other 30 currently classified prokaryotic and eukaryotic α-amylases under well-defined GH13 subfamilies. Surprisingly, the phylogenetic analysis of those 93 α-amylase amino acid sequences showed that all the 15 *Anoxybacillus* α-amylase sequences from this study are clustered with other *Anoxybacillus* α-amylases such as *ASKA, ADTA*, and *GSX-BL*, but not with *BaqA* or *GTA*. The 63 bacilli α-amylases, including the novel sequences were all originated from endospore-forming bacilli by creating 5 distinct phylogenetic clades, as can be seen in the evolutionary tree in Fig. 2. In the similarity matrices, deduced α-amylase protein sequences of the 15 *Anoybacilli* displayed 95.1–100% sequence homology to each other. While the protein sequences of A321, E184aa, E184ab, E208a, A3210 and *A. gonensis* α-amylases completely resembled to each other, the most distinct protein sequence was from *A. kestanbolensis* (95.1%-97.6%). Based on the protein sequence similarities and the evolutionary relations, four differently proposed subfamilies, which were not belonging to any of the currently classified GH13 subfamilies, were inferred from the phylogenetic analyses. The bootstrap values of *a1* to *a4* clades were higher than 70% which verified their brach positions in the phylogenetic tree. The sequence identity rates among the members were given in Supplementary Table SII. The firstly proposed subfamily *a1* was composed of 27 α-amylases from genus *Anoxybacillus* including our 15 sequences, *Anoxybacillus* sp. DT3-1, SK3-4, and GSX-BL enzymes and other 9 putative *Anoxybacillus* α-amylases annotated in whole genome sequences. The *a1* clade showed protein sequence similarities between 72.4% and 100% among the group, having the lowest similarity to BCO1 α-amylase. The *a1* clade was mostly related with a second *a2* α-amylase cluster containing other four hypothetical *Anoxybacillus* α-amylases deduced from annotated genome sequences of *Anoxybacillus* sp. B2M1, B7M1, P3H1B, and UARK-01 as well as with a third *a3* cluster composed of 19 endospore-forming, thermophilic bacilli genera from *Geobacillus* containing *GTA, Parageobacillus* in addition to genus *Anoxybacillus*. The members of *a2* and *a3* clade displayed sequence similarities ranging from 89.2 to 100% and from 74.9 to 100% within their own groups, respectively. All the α-amylases from mesophilic genus *Bacillus* formed a fourth *a4* cluster with 9 members including *BaqA* which showed 58.3–99.4% similarity values with each other. Finally, the unclassified but proposed four representative aberrant α-amylases from genus *Bacillus* clustered in a different clade (GH13_*xy* by Sarian et al. 2017) which diverged from other mesophilic bacilli from *a4* clade around the *BaqA* α-amylase.

Surprisingly, two of the α-amylases from *A. tepidamans* DSM 16325^T and *A. geothermalis* ATCC BAA 2555^T, sharing 75.1% similarity, were clustered together with the members of the genus *Geobacillus* from *a3* instead of their original *Anoxybacillus a1* and *a2* clades. *A. tepidamans* α-amylase (*ATA*) was most closely related to *P. thermantarcticus* M1 (80.8%), *G. thermoleovorans* MTCC 4220 (*Gt-amyII*), (76.7%) and *G. thermoleovorans* CCB US3 UF5 (*GTA*), (76.2%) α-amylases, whereas *A. geothermalis* enzyme was mostly related to to *Geobacillus* sp. LC300

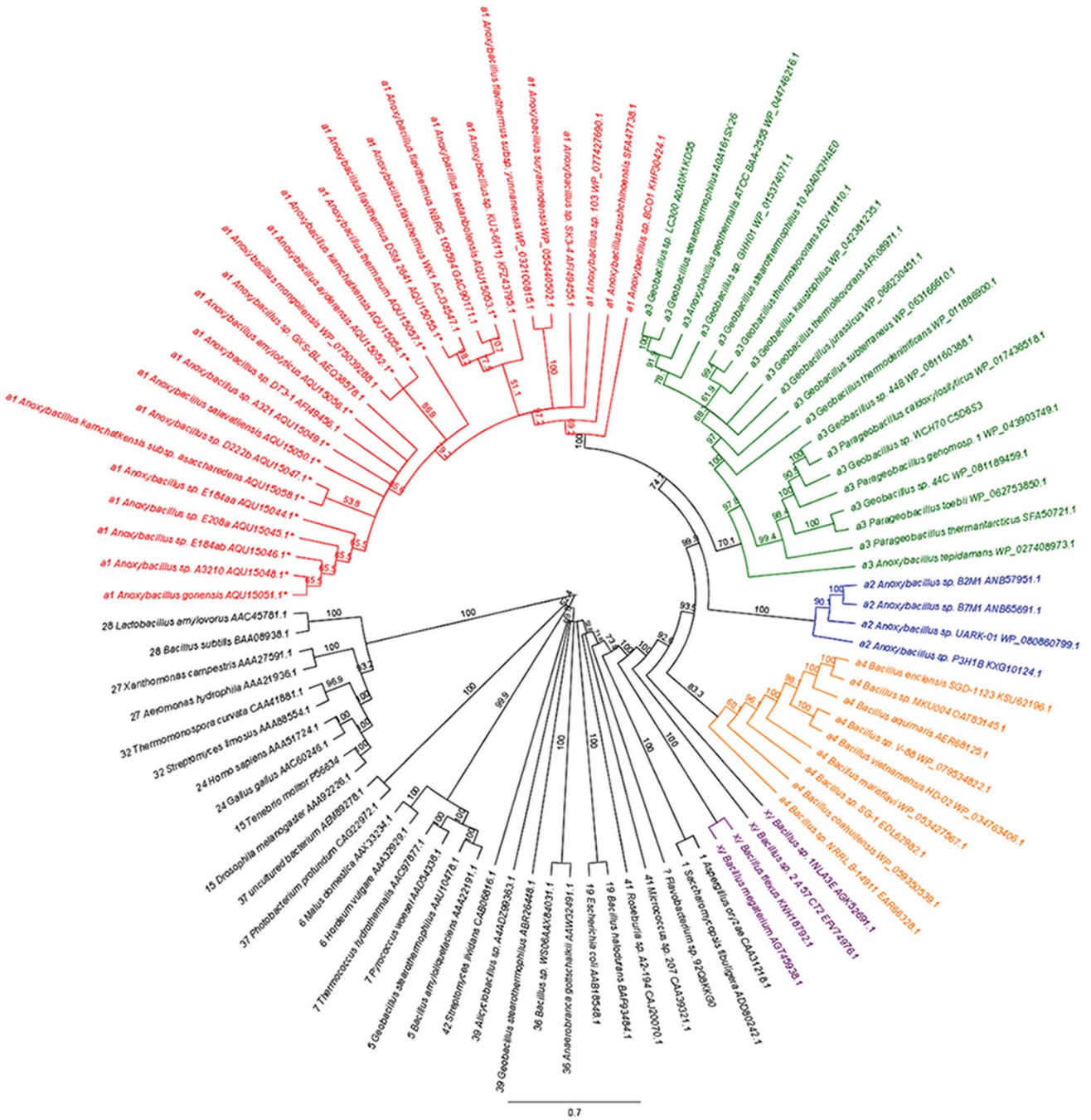

Fig. 2 The evolutionary tree of the GH13 α-amylase family obtained from 93 protein sequences. The newly proposed *a1* (red), *a2* (blue), *a3* (green) and *a4* (orange) members as well as the GH13_*xy* members (purple) all belonged to *Bacillaceae* family and none of these amylases have been assigned to any GH13 subfamilies till now. After GH13_ subfamily indication, mostly the GenBank accession numbers, but if not available, the UniProt (UniParc) numbers are given at the end of the species epithet. "*" symbol near the accession numbers display α-amylase gene sequences obtained from this study

(98.6%) and *G. thermoleovorans* CCB US3 UF5 (*GTA*), (94.0%) α-amylases were clustered within the *a3* clade. Interestingly, the results of 16S rRNA gene sequence analysis on all the described *Anoxybacillus* species revealed that members of this genus, harbouring 23 species and 2 subspecies, divide into two phylogenetically diverge branches that share high 16S rRNA gene homologies only within their related groups (Euzéby and Parte 1997; Gul-Guven et al. 2008; Deep et al. 2013). *A. tepidamans* DSM 16325^T (FN428691) and *A. geothermalis* ATCC BAA2555^T (KJ722458) species

were clustered together with the second group members of *A. vitaminiphilus* JCM 16594^T (FJ474084), *A. calidus* DSM 25220^T (FJ430012), *A. contaminans* DSM 15866^T (AJ551330), *A. voinovskiensis* DSM 17075^T (AB110008), *A. amylolyticus* DSM 15939^T (AJ618979), *A. rupiensis* DSM 17127^T (AJ879076) and *A. caldiproteolyticus* DSM 15730^T (FN428698). 16S rRNA gene sequences of this second group shared lower sequence homology (< 97.0%) with all the other type species from genus *Anoybacillus*, whereas there was a strict requirement for DNA hybridization analysis among each other for their species identifications. This second branch arised based on the analysis 16S rRNA genes, which encloses *A. tepidamans* and *A. geothermalis* species (98.2% homology to each other), was also seemed to be more closely related to genus *Geobacillus* than the first 16S rRNA gene group of *Anoxybacillus* group. In a recent study of Bezuidt et al. (2016) dealing with the comparative analysis on conserved core and flexible genes of 61 *Geobacillus, Anoxybacillus* and *Bacillus* genome sequences, *A tepidamans* PS2 (BioProject Accession: PRJNA214279) clustered within the genus *Geobacillus*, whereas other 12 *Anoxybacillus* species branched on their own clustures. They also suggested that *A. tepidamans* PS2 should be regarded as a species in *Geobacillus* based on their shared genes and Average Nucleotide Identity values (Bezuidt et al. 2016).

Beside these findings, the alignment of *Anoxybacillus* α-amylases with their other counterparts clearly displayed the presence and varieties of the seven CSRs from regions I to VII (Janeček 2002), the catalytic triad, the common $(\beta/\alpha)_8$-barrel fold containing domains A, B and C as well as a pair of tryptophans in the α3 helix between CSR-V and CSR-II (Mok et al. 2013; Puspasari et al. 2013) within these phylogenetic groups (Fig. 3). The aligned 93 sequences classified under 21 groups and, it was noteworthy that five of the clades including *a1, a2, a3* and *a4* (totally 59 sequences) and *xy* (representative 4 sequences) were all composed of endospore formers from *Bacillaceae* family and shared similarities in terms of these general features above in congruence with the results of phylogenetic analyses.

Since the 15 *Anoxybacillus* putative proteins obtained from this study were all clustered within the phylogenetic group *a1*, and the highest amylolytic enzyme activity was measured in *Anoxybacillus* sp. E184aa, E184aa-*A* (E184aa α-amylase) was chosen as the representative member of *a1* clade within the other 26 sequences. Among thermostable α-amylases from *Anoxybacillus* sp. B2M1, B7M1, P3H1B and UARK-01, only the amylolytic enzyme activities of B2M1 and B7M1 were known as positive (Filippidou et al. 2016), therefore, B2M1-*A* (B2M1 α-amylase) was selected as the representative sequence of *a2* clade for further structural investigations. *A. tepidamans* DSM 16325^T α-amylase (*ATA* α-amylase) was chosen as the representative member of the *a3* clade which contained 19 sequences belonging to three different genera from thermophilic bacilli. Finally, *Bacillus aquimaris* strain MKSC 6.2 α-amylase (*BaqA* α-amylase) was selected for being the representative member of the nine α-amylases from *a4* clade, all the members of which were from the mesophilic genus *Bacillus. A. tepidamans* DSM 16325^T (Coorevits et al. 2012), and *Bacillus aquimaris* MKSC 6.2 (Puspasari et al. 2013) were also known as amylolytic strains from previous studies.

The alignment results revealed the importance of CSR-V in the loop 3 which shares the conserved motif of LPDLx and a pair of tryptophans between loop 3 and β4 strand among the enzymes belonging to these *Bacillaceae* members (Janeček et al. 2015; Ranjani et al. 2015). All the thermostable α-amylases from *a1, a2* and *a3* clades had an alanine residue (A184, E184aa-*A* numbering) at the end of LPDLx signature, whereas mesophilic members contained an asparagine residue (N185, *BaqA*) at this site. The two consecutive tryptophans were also shared among these *Bacillaceae* members only with an exception in *(A) flavithermus* subsp. *yunnanensis* containing W200 and R201. By combining these phylogenetic analyses and alignment results, the specific CSRs logos defining the currently proposed or revised GH13 subfamilies of *a1, a2, a3*, and *a4* were created and presented in Fig. 4. Among these logos, the catalytic triad which consists of aspartic acid (D213$^\Delta$) residue serving as the catalytic nucleophile in CSR-II, glutamic acid (E242$^\Delta$) playing the proton donor role in CSR-III, the transition-state stabilizer aspartic acid (D310$^\Delta$) in CSR-IV, and the invariable arginine (R211) in CSR-II were all conserved except in members from *a2* subfamily (E184aa-*A* numbering). Nevertheless, the logo of *a2* was the most different among the common CSRs motifs of α-amylases from other well-known subfamilies as well as *a1, a3*, and *a4* groups. B2M1-*A*, B7M1-*A*, P3H1B-*A* and UARK-01-*A* members from *a2* clade, harboured an abnormal catalytic machinery as in the case of an atypical α-amylase GH13_*xy* subfamily (Sarian et al. 2017). Solely, the catalytic nucleophile aspartate (D215$^\Delta$, B2M1-*A*) was preserved in *a2* clade, but the proton donor of E242$^\Delta$ (E184aa-*A*) was replaced with a glycine G244$^\Delta$ (B2M1-*A*) residue in CSR-III (β5 strand) like in the *Bacillus* sp. 2 A57 CT2 from GH-13_*xy* (Sarian et al. 2017). But there was also an aspartate residue (D246, B2M1-*A*), shifted two positions downstream, which might be the potential catalytic nucleophile instead of usual glutamate. Moreover, in CSR-IV residing on β7 strand, the transition-state stabilizer aspartate replaced with a strange arginine (R312, B2M1-*A*) residue similar to the *Bacillus* sp. 1NLA3E α-amylase from aberrant GH13_*xy* subfamily. However, there was also an aspartate residue (D311$^\Delta$, B2M1-*A*), which was shifted one position upstream to this abnormal arginine. In addition, although the invariant arginine (R211, E184aa-*A*) in CSR-II was fully conserved among all GH13 subfamilies except almost all members of GH13_*xy*, it was

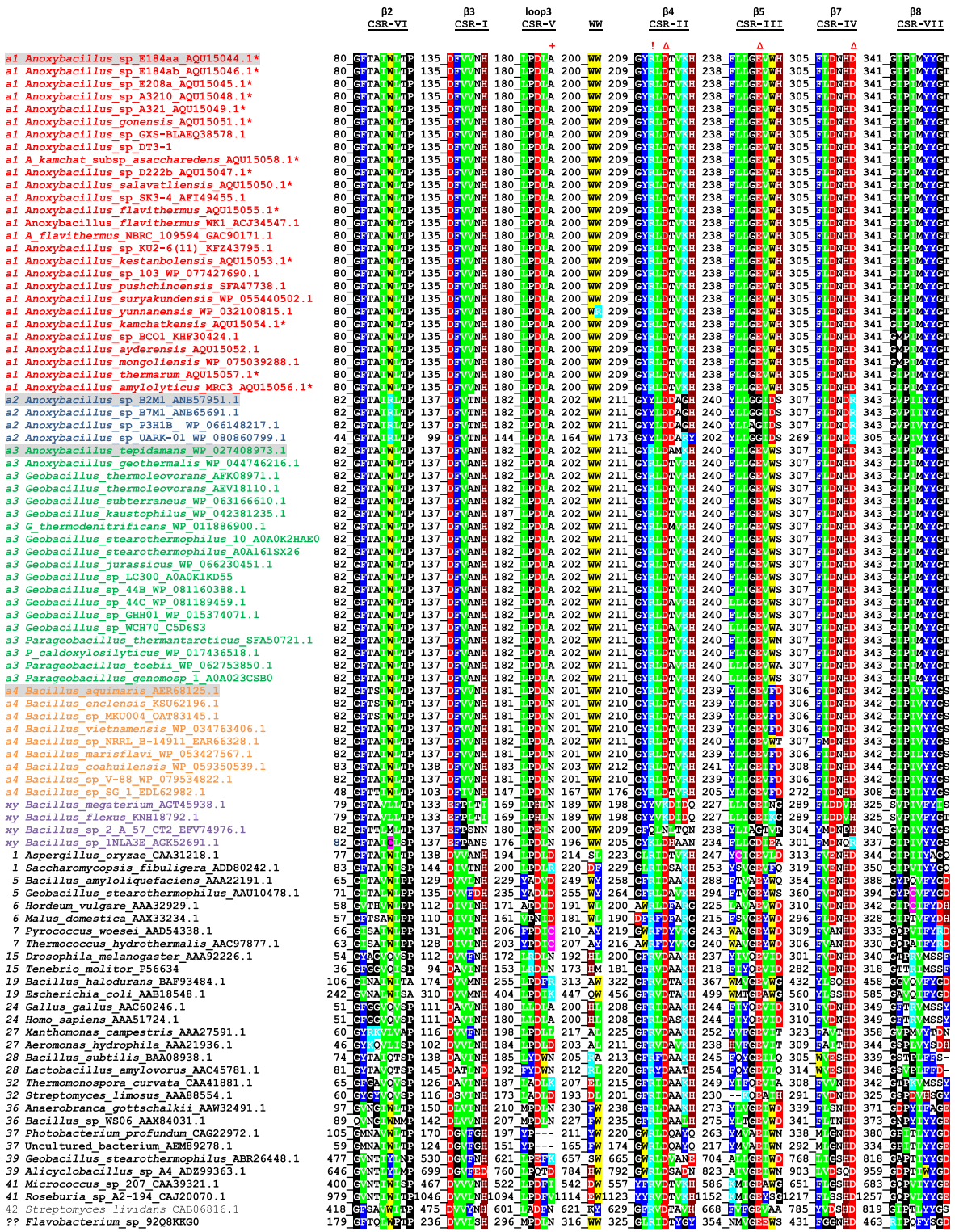

Fig. 3 The α-amylase sequence alignments of seven CSRs (I-VII) belonging to the newly proposed *a1, a2, a3* and *a4* novel GH13 subfamilies with GH13_*xy BmaN1* group, the unassigned cyclomaltodextrinase from *Flavobacterium* sp. No 92 (GH13_*??*) and the other well-defined GH13 subfamilies displaying α-amylase specificity including subfamilies GH13_1, 5, 6, 7, 15, 19, 24, 27, 28, 32, 36, 37, 39, 41, and 42. The representative members of *a1, a2, a3* and *a4* groups are highlighted with gray. +, characteristic alanine (A); $^{\Delta}$, the catalytic triad (D/E/D); !, invariable arginine (R). Colour code for the selected residues: W—yellow; F, Y—blue; V, L, I—green; D, E—red; R, K—cyan; H–brown; C—magenta; G, P—black

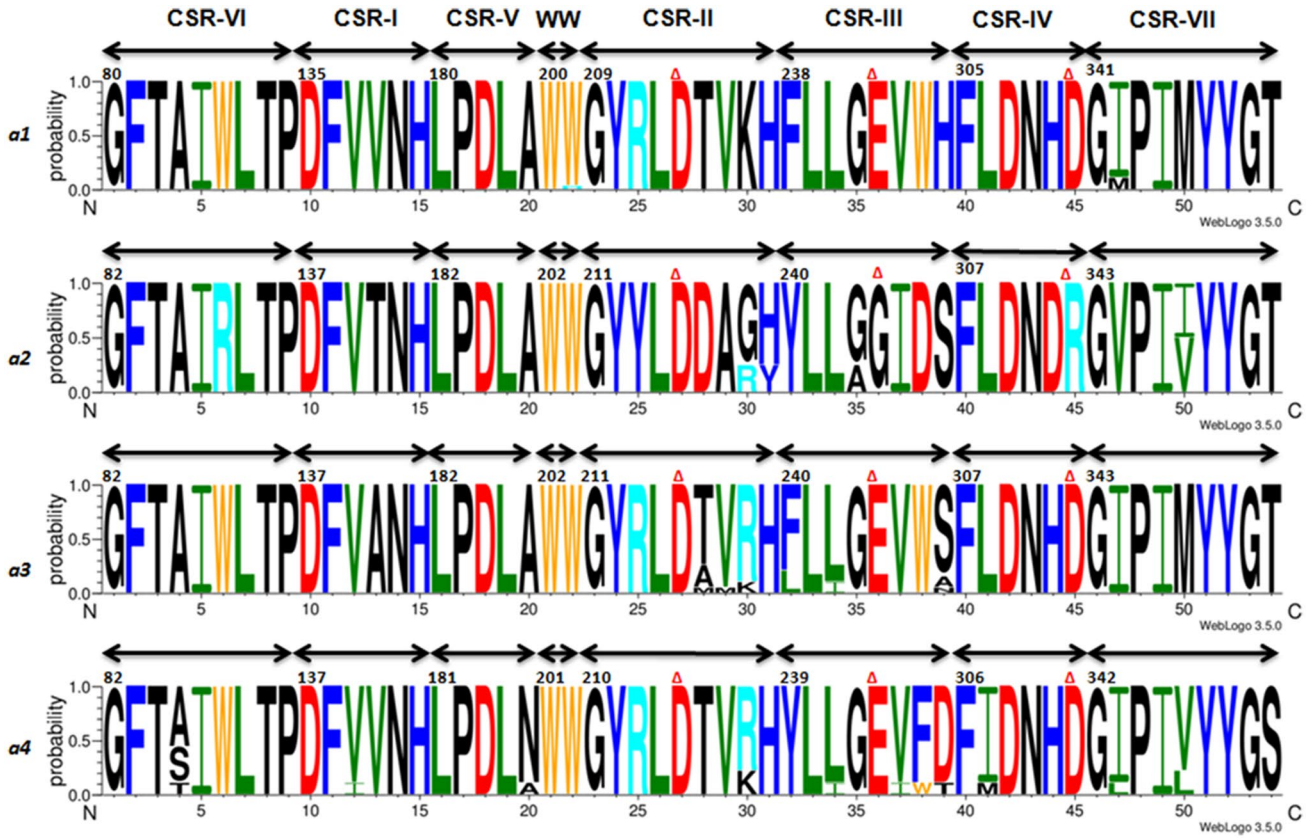

Fig. 4 Sequence logos made with regard to the CSRs of the four newly proposed GH13 subfamilies from *a1-a4* all belonging to *Bacillaceae* family. The positions of CSRs regions from I-VII, two consecutive tryptophans in addition to the catalytic triad ($^{\Delta}$) are also presented. The logos were created by using 27, 4, 19, and 9 protein sequences for the subfamilies *a1, a2, a3* and *a4*, respectively. The residues are numbered according to their representative α-amylase sequences: E184aa-A (*a1*), B2M1-*A* (*a2*), *ATA* (*a3*), and *BaqA* (*a4*)

changed to a tyrosine (Y213, B2M1-*A*) residue as in the case of *BmaN1* α-amylase of *(B) megaterium* NL3 belonging to the atypical subfamily.

In silico analyses on the secondary structures

The deduced polypeptides of 15 *Anoxybacillus* sequences contained 504 to 505 amino acids starting with a 23 residues long putative signal peptide. The predicted molecular weight and pI values of these enzymes ranged between 58.7 and 59.0 kDa and 5.98–6.26, respectively. From this point, totally 59 sequences from endo-spore forming bacilli which comprised the currently suggested *a1, a2, a3*, and *a4* clades, including the 15 *Anoxybacillus* α-amylases, were used for further *in silico* secondary and tertiary structure analyses. The characteristic features of these proposed GH13 subfamilies obtained from *in silico* techniques were also presented in Supplementary file, Table SII. The highest proline contents, predicted molecular weights and pI values of these enzymes were observed in *a3, a4* and *a2*, respectively. The only exception was the pI value of *ATA* which increased the maximum limits of *a3* from 7.91 to 9.31. All the members of *a1, a2, a3* and *a4* had also signal peptide sequences aligning between residues 1 and 23. The bacilli amylolytic enzymes were predicted to be transmembrane proteins as they shared two putative components embedded in the hydrophobic membrane: S1 and S2, which crossed the membrane from cytoplasmic to extracellular sides with four responsible amino acid residues (K330, Y346, I482 and F498, E184aa-*A* numbering) in their helix structure. An example of transmembrane helix image, created for E184aa-*A* α-amylase, was shown in Supplementary file, Fig. S1.

As deduced from topological alignment of primary and secondary (2D) structures of the representative members from the *a1-a4* clades (with *GTA* from *a3*), all the enzymes consisted of three domains typical for α-amylases from GH13 family (MacGregor et al. 2001). The catalytic domain A containing the exact $(\beta/\alpha)_8$-barrel structure, the domain B connecting the β3 strand, and α3 helix, and finally the domain C succeeding domain A, which contain eight antiparallel β-sheets were all shared by enzymes of these *a1, a2, a3* and *a4* clades. The predicted signal sequences, β-strand and α-helix numbering, the catalytic triad, the CSRs, and additionally the possible pockets for substrate-binding were

all displayed on the 2D structures of E184aa-*A*, B2M1-*A, ATA*, and *BaqA* obtained by Phyre2 server (Supplementary file, Fig. S2a-S2d).

When all these data obtained from CSRs sequence alignments, Phyre 2D topological alignments from this study as well as the previous experimental analyses of the most related α-amylases of *ASKA* (Chai et al. 2016), *GTA* (Mok et al. 2013) and *BaqA* (Puspasari et al. 2013) were considered, the entire sequences of 63 α-amylases belonging to *Bacillaceae* family were aligned eventually in order to compare the entire amino acid residues forming the signal sequences, the catalytic triad, the CSRs, the possible calcium and sugar binding residues, the possible sugar pockets for substrate specificities or sugar recognitions, the transmembrane helix regions, the tyrosine and phenylalanine repeats and the consecutive lysine and arginine residues at the end of C termini. However, only the sequence alignments of the 2 representatives of each proposed subfamilies that are E184aa-*A* and *ASKA* (*a1*), B2M1-*A* and P3H1B-*A* (*a2*), *ATA* and *GTA* (*a3*), *BaqA* and MKU004-*A* (*a4*) α-amylases were presented in Fig. 5 to demonstrate these characteristic properties.

In our phylogenetic investigations, *ASKA* and *GTA* thermostable α-amylases grouped under proposed *a1* and *a3* subfamilies. Their calcium and maltose binding sites as well as the other related residues were previously examined in detail by X-ray crystal structure analyses (Mok et al. 2013; Chai et al. 2016). Therefore, the amino acid residues associated with the 2 calcium binding sites of *GTA* and four sites of *TASKA* were screened for the newly suggested four subfamily members. As can be seen in the alignment of the representative sequences in Fig. 5, the residues involved in the formation of Ca1, Ca2 and Ca4 calcium binding sites, detected in the crystal structures of *GTA* and *TASKA*, were mostly conserved among the proposed *a1, a3* and *a4* subfamilies. Nevertheless, there were some amino acid substitutions, which were expected to have relatively minor effects on calcium binding like neutral (E/D) transitions or N/D changes from positive to negative charge the latter of which may increase the binding affinity. But the most critical changes from negative to positive charge were also detected as D/P, E/P, E/A, and E/Q mutations, pocessing high possibility of decreasing the sensitivity and affinity for calcium ion (Tien et al. 2014). The conversion of E173/Q174 (E184aa-*A*/*BaqA*) in Ca1 sites of *a4* members were observed as mutations that may have an effect on binding of calcium ion. The amino acid substitutions of N46/D48 (E184aa-*A*/in all *a2, a3* and *a4* members) in Ca2 sites, E109/D111 (E184aa-*A*/in all *a2, a3* and *a4* members) and E110/P112 (E184aa-*A*/B2M1-*A* and *ATA* numbering) residues in Ca3 sites, and additionally E283/A285 (E184aa-*A*/B2M1-*A*) and E283/D284 (E184aa-*A*/*BaqA*) in Ca4 sites were also noticed.

The secondary structure analysis of the four representative members using Phyre2 server also presented some hints, which might be associated with sugar binding pockets. These hints, which were not used in other studies before, were found to be very useful for further detailed structural analyses. When these hints were combined with the previous findings on maltose binding residues from *GTA* and *TASKA*, the conserved regions of CSR-II (β4), III (β5), and IV (β7) were thought to be important both in the catalytic activity and substrate binding. Whereas, the CSR-VI (β2), I (β3), V (loop 3), and VII (β8) regions probably play roles in enzyme specificity and substrate binding of these *a1* to *a4* subfamilies (Janeček 2002; Mok et al. 2013; Chai et al. 2016). Accordingly, the evaluation of 59 bacilli sequence alignments under 4 clades revealed 12 pockets having possible functions in enzyme activity and substrate specificity: (1) F pocket in β1 strand (in all), (2) W (in *a1, a3* and *a4*) or R (in *a2*) pockets in CSR-VI, (3) H-YW (in *a1, a3* and *a4*) or Q-KK (in *a2*) pockets downstream to CSR-VI, (4) H–Y (in *a1 and a3*) pocket in CSR-I, (5) W (in *a1, a2* and *a3*) or F (in *a4*) pockets upstream to CSR-V, (6) LF-L in CSR-V (in *a1*), LN-L (in *a2*) or LY-L (in *a3* and *a4*) pockets, (7) R-DTVKH (in *a1*), Y-DDAGH (in *a2*), R-DAMKH (in *a3*) or R-DTVRH (in *a4*) pockets in CSR-II, (8) E-W (in *a1* and *a3*), G-D (in *a2*) or E-F (in *a4*) pockets in CSR-III, (9) L-Y (only in *a2*) and V-F (solely in *a4*) pockets downstream to CSR-III, (10) HDTV (in *a1* and *a3*), DRTV (in *a2*) or HDME (in *a4*) pockets in CSR-IV, 11 I-Y (in all) pocket in CSR-VII and, 12) ED–NR (in *a1* and *a4*), KA–NH (in *a2*) or ND–NR (in *a3*) pockets downstream to CSR-VII (Fig. 5 and Supplementary file, Table SII).

In addition to these calcium and maltose binding sites, some specific residues found at the end of the sequences also took attention when 2D structures and the whole sequence alignments were compared. At the C-termini, two novel consecutive tyrosine residues (Y457, Y469, E184aa-*A* numbering) were detected in addition to the previously described repeated aromatic motifs of tyrosine (Y489, Y497, E184aa-*A* numbering) and phenylalanine (F481, F492, F495, E184aa-*A*) by Janeček et al. (2015), which could be evaluated as additional stop signals in all 59 bacilli sequences. Moreover, the residues in helix structures associated with the formation of predicted S1 and S2 regions crossing the cell membrane were nearly conserved among all the sequences (Fig. 5 and Supplementary file, Table SII). Residues that combined S1 region spanning from extracellular environment (S1-E: K330, E184aa-*A*) to cytoplasm (S1-C: Y346, E184aa-*A*) were all preserved in *a1*-*a4*, whereas the amino acid residues conjoining S2 region from extracellular to cytoplasm varied in S2-E: I482/I485 (in *a1*/*a2* and *a3*) or L485 (in *a4*), and S2-C: F498 (in *a1*), Y501 (in *a2* and *a3*) or L501 (in *a4*). Finally, all these α-amylase sequences from *Bacillaceae* family were found to be ended with two

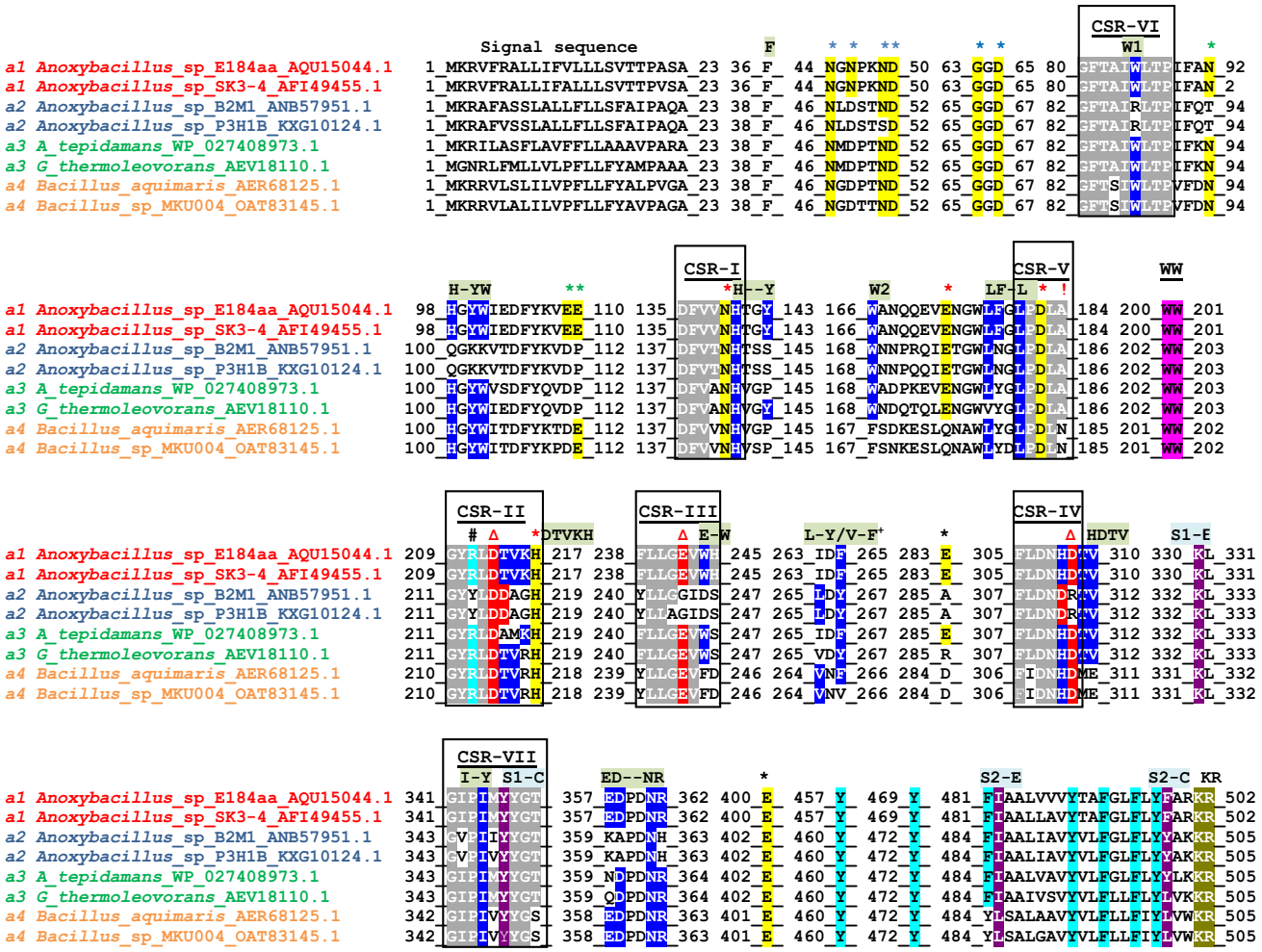

Fig. 5 The comparison of primary structures of four newly proposed subfamilies, presenting two representatives from each group. Colour code: (1) Conserved sequence regions (CSR I-VII) are highlighted in gray. (2) Residues for Ca1, Ca2, Ca3 and Ca4 binding sites are highlighted in yellow and marked by red, blue, green and black (*) asterisks, respectively. The conserved amino acids in calcium binding sites are indicated with yellow. (3) Residues involved in maltose binding are indicated by blue, and the related maltose binding pockets are abbreviated as F, W1, H-YW, H–Y, W2, LF-L, R-DTVKH, E-W, L-Y / V-F^{+} ($^{+}$; only found in *a2* and *a4*), HDTV, I-Y and ED–NR. (4) The catalytic triad are signified by red triangles ($^{\Delta}$). (5) The invariantly conserved position of the arginine in the CSR-II is highlighted in turquoise and marked by a hashtag (#). (6) Residue A in CSR-V, only found in thermophilic groups is indicated with (!). (7) The two adjacent characteristic tryptophans, positioned between CSR-V and CSR-II, are highlighted in pink. (8) The seven conserved tyrosine and phenylalanine residues at the C terminus are highlighted in turquoise. (9) The invariable KR residues at the C-termini are indicated in dark yellow. (10) The putative transmembrane regions and their related residues are indicated with violet and abbreviated as S1C:S1 region-cytoplasmic, S1E: S1 region-extracellular, S2C: S2 region-cytoplasmic, and S2E: S2 region-extracellular

conserved, consecutive and positively charged lysine and arginine (K501R502, E184aa-*A*) residues only with the exceptional RR or KK residues found in 5 of the α-amylases from *a3* and *a4*.

Tertiary structure predictions

The crystal structure analyses of *G. thermoleovorans GTA* α-amylase (PDB ID: 4E2O) and *Anoxybacillus* sp. SK3-4 α-amylase (*TASKA*, PDB ID: 5A2B) from the members of the rearranged subfamilies of *a1* and *a3* were already investigated in detail by Mok et al. (2013) and Chai et al. (2016), respectively. The 3D models of the representative E184aa-*A* (*a1*), B2M1-*A* (*a2*), and *BaqA* (*a4*) α-amylases were predicted by SWISS-MODEL and visualized by PyMOL and ICM-Browser-Pro using the best template as *TASKA* among nearly 240 candidates. Only in the case of *ATA* (*a3*), the best template model was preferred as *GTA* α-amylase in terms of higher sequence identity and lower RMSD values to *GTA* (79.60%, 0.085 Å) rather than *TASKA* (77.88%, 0.673 Å). The calculated RMSD values, obtained from the structural alignments of E184aa-*A*,

B2M1-*A* and *BaqA* models with *TASKA* template were 0.078 Å, 0.116 Å, and 0.153 Å, in addition to 0.085 Å RMSD value in the case of *ATA* model with *GTA* template. According to homology report, the sequence identities, coverage and QMEAN values of E184aa-*A* (96.92%, 0.90 and − 0.38), B2M1-*A* (69.45%, 0.89, -0.82), and *BaqA* (64.10%, 0.89, -1.73) to *TASKA* as well as *ATA* (79.60%, 0.88, -0.66) to *GTA* were all given in parenthesis, respectively. In Supplementary Fig S3a, the folded 3D structure models of E184aa-*A*, B2M1-*A, ATA* and *BaqA* were presented by giving the three domain structures arranged as domain A, B and C. The superimposed structures (green) of these four α-amylases both with *TASKA* (red) and *GTA*-II (blue) obviously pointed out the overall similarity of the catalytic $(\beta/\alpha)_8$-barrel structure exists in GH13 α-amylase family (MacGregor et al. 2001), (Supplementary file Fig. S3b). The surface views of these α-amylases were also displayed in Supplementary Fig. S3c, which depicted a big groove for the active site region associated with maltose binding.

The substrate binding sites

The four representative α-amylase models (green) were overlapped with *TASKA* (red) and the active site regions as well as the residues possibly associated with maltose binding were illustrated in Fig. 6. In all models, maltose bound to substrate interacting subsites of -1 and − 2 similar to *TASKA*. The substrate binding pockets, detected by the hints of the previous topological secondary structure analyses, revealed that these residues are directly interacting with sugar ring or the ones responsible for substrate specificity and stabilization. When the structural maltose binding region of E184aa-*A* α-amylase superimposed with *TASKA*, model, completely matched to the template as they were both proposed to be the members of *a1* subfamily. The catalytic triad of E184aa-*A*, consisting of Asp-Glu-Asp triad, were seemed to be acting on sugar ring in subsite − 2. Among these triplets, $D213^{\Delta}$ serve as a catalytic nucleophile, $E242^{\Delta}$ is the proton donor and $D310^{\Delta}$ act as transition state stabilizer. Moreover, invariable

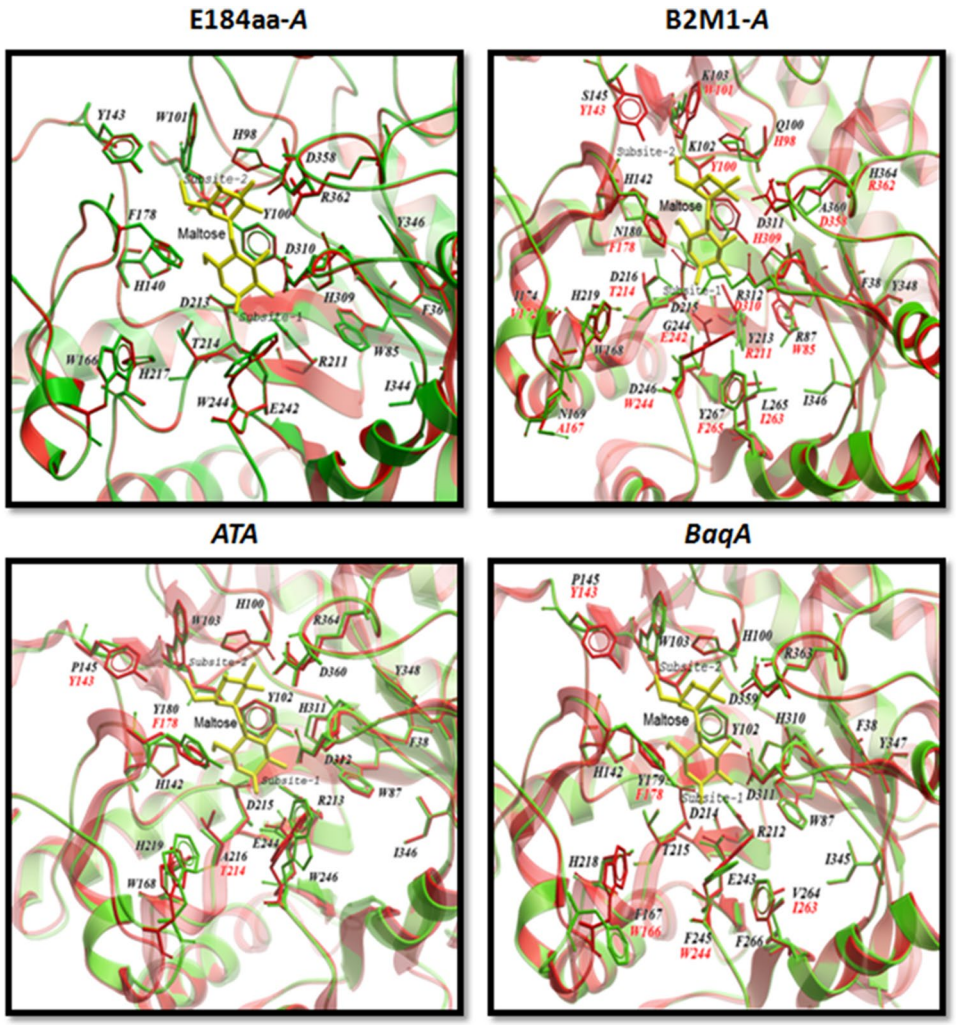

Fig. 6 Overlapping the active sites of the 3D models (green) of *Anoxybacillus* sp. E184aa (E184aa-*A*), *Anoxybacillus* sp. B2M1 (B2M1-*A*), *A. tepidamans* DSM 16325^{T} (*ATA*) and *B. aquimaris* MKSC 6.2 (*BaqA*) amylases with *TASKA* template (red, PDB ID: 5A2B). Model and template residue numbers are coloured in black and red, respectively. All the superimpositions with *TASKA*-maltose (yellow) were bound to active site region at subsites − 1 and − 2

arginine, positioned in the β4 strand, was located at R211 in E184aa-*A* α-amylase. These conserved catalytic residues and invariable arginine were all preserved in *ATA* ($D215^{\Delta}$, $E244^{\Delta}$, $D312^{\Delta}$ and R213) and *BaqA* ($D214^{\Delta}$, $E243^{\Delta}$, $D311^{\Delta}$ and R212) as a common feature of GH13 α-amylase family. But two of the catalytic residues ($E242^{\Delta}$/$G244^{\Delta}$ and $D310^{\Delta}$/$R312^{\Delta}$ transitions) in addition to the invariable arginine (R211/Y213) were replaced in the case of B2M1-*A* like *BmaN1* α-amylase. The various residues which may involve in maltose binding of E184aa-*A*, B2M1-*A, ATA* and *BaqA* α-amylases were listed in Supplementary Table SII.

In concordance with the secondary structure analysis of E184aa-*A* dealing with the detected pocket sites, the amino acid residues of H98, Y100 and W101 (H–YW pocket downstream to CSR-VI), H140 and Y143 (H–Y pocket in CSR-I), F178 (LF–L pocket in CSR-V), D358 and R362 (ED–NR pocket downstream to CSR-VII) were in close relationship with maltose by their side chains in subsite − 2, whereas amino acids of R211 and $D213^{\Delta}$ (R-DTVKH pocket in CSR-II), $E242^{\Delta}$ (E–W pocket in CSR-III), H309 and $D310^{\Delta}$ (HDTV pocket in CSR-IV) served for the catalytic activity directly in the subsite − 1 of the sugar binding groove. Moreover, there were some additional residues which were probably reside in the substrate binding subsite − 1, not being in contact with maltose, as T214 and H217 residues in the R-DTVKH pocket, the aromatic side chains of F36 and W85 in the HDTV pocket, and W166 with H217 in the R-DTVKH pocket. Additionally, W244 which seemed to render the E-W pocket with its bulky side chain, and finally, the residues of I344 and Y346 which appeared to form a loop may have an importance in the substrate specificity.

In substrate binding groove of B2M1-*A*, despite many residual variations, the overall catalytic system was conserved as *Anoxybacillus* sp. B2M1 is known to possess amylolytic activity (Filippidou et al. 2016). The amino acid residues of Q100, K102, K103 (Q-KK), H142, S145 (H–S), N180 (LN-L), A360 and H364 (KA–NH) bound to sugar ring at subsite − 2, and Y213, $D215^{\Delta}$, D216 (Y-DDAGH), $G244^{\Delta}$ (G-D), D311, and $R312^{\Delta}$ (DRTV) were all positioned in subsite − 1 of the maltose binding groove in B2M1-*A*. Other than the similar residues present in E184aa-*A*, the additional two loops between the residues L265, Y267 (downstream to CSR-III) and residues W168, N169, I174 (upstream to CSR-V) were only peculiar to B2M1-*A* (Supplementary Table SII). In the case of *ATA* and *BaqA* enzymes, most of the amino acid residues interacting directly or indirectly with sugar ring were more conserved similar to E184aa-*A* with some exceptions. Superimpositions of *ATA* and *BaqA* to *TASKA* revealed the Y143/P145, F178/Y180, T214/A216 amino acid residue modifications in *ATA* and aromatic Y143/P145, F178/Y179, W244/F245, W166/F167 residue changes in *BaqA* α-amylase. Besides, the residues of V264 and F266, (pocket downstream to CSR-III) and F167 were also only unique to *BaqA*.

Discussion

Thermostable α-amylases have been used in several industrial applications as they possess thermal stability to harsh industrial processes including elevated temperatures (Demirjian et al. 2001). Starch degradation, baking, brewing, production of glucose and fructose syrups, fruit juices, alcoholic beverages, papers, pharmaceuticals, α-amylase assay kits, detergents and textiles are the major areas of utilization for amylolytic enzymes in the industry (Klein et al. 1970; Vieille and Zeikus 2001; Van der Maarel et al. 2002; Gupta et al. 2003). *Anoxybacillus* species are thought to be widespread in thermal habitats rather than other *Bacillaceae* members (Deep et al. 2013), having heterogeneous intra-species 16S rRNA gene similarity values varying from 93.8 to 99.7% (from this study). Moreover, whole genome sequences of 27 *Anoxybacillus* were available on Genomes OnLine Database-GOLD v.6 (Mukherjee et al. 2017) and GenBank (Benson et al. 2014) databases now. We screened the starch hydrolysing activities of some newly isolated *Anoxybacillus* strains in this study, which could be suggested to novel starch hydrolysis applications. Although all the bacilli were found to be amylolytic, *Anoxybacillus* sp. E184aa, E184ab and D222b isolates stand out from the others by means of their α-amylase production capabilities. Additionally, α-amylase production capacity of *A. salavatliensis* was also experimentally proved to be similar with that of *A. flavithermus* (Bolton et al. 1997; Tawil et al. 2012; Agüloğlu et al. 2014; Ozdemir et al. 2015, 2016a) and *A. amylolyticus* (Poli et al. 2006) species which were already known as amylolytic enzyme producers. Totally, 15 novel *Anoxybacillus* α-amylase gene sequences were introduced to databases with this study and their preliminary BLASTP queries displayed (≥ 91.0%) gene sequence similarities to only well-known *ASKA, ADTA* and *GSX-BL* amylases originated from *Anoxybacillus* species, which were formerly proposed in a single subfamily within other *Bacillaceae* enzymes including *GTA, Pizzo, Gt-amyII* and *BaqA* from genus *Geobacillus* and *Bacillus* (Janeček et al. 2015; Ranjani et al. 2015; Chai et al. 2016; Sarian et al. 2017). Nevertheless, the protein sequence homologies of these *Anoxybacillus* amylases to *GTA* (≤ 69.4%) and *BaqA* (≤ 61.3%) were found to be relatively lower. As the putative protein and genome sequences were accumulated in databases, it is obvious that the accuracy of GH13 sequence-based classification system would increase and their related subfamilies might be classified under more meaningful groups. Thereby, the 15 *Anoxybacillus* sequences were analysed in detail with blast query and phylogenetic investigation by adding 30 α-amylase

sequences from formerly defined GH13 subfamilies and 48 endospore-forming bacilli sequences (hypothetical or experimentally characterized) with the recommendations of Stam et al. (2006). It was interesting that any of these 48 α-amylase sequences from *Bacillaceae* family still could not be validly assign to any of the defined GH13 subfamilies. The constructed phylogenetic tree in Fig. 2 has completely drawn the picture of the phylogenetic relations between the amylases from *Bacillaceae* family and the other described GH13 subfamilies. Bacilli amylases were divided into totally 5 distinct branches far from other 15 well-defined GH13 subfamilies. The reorganized *a1* and *a3* clades, composed from E184aa-*A* and *ATA*, and the newly proposed *a2* clade including B2M1-*A* representative were all thermostable members. Whereas the *a4* clade which appeared as akin to *BaqA* and the formerly proposed "*xy*" non-defined subfamiliy containing *BmaN1*, formed the other two mesophilic bacilli originated amylases belonging to *Bacillaceae* family. These bacilli amylases could easily be separated both from the other GH13 subfamilies and from the taxonomic genera within the *Bacillaceae* family except the *a3* clade. Only the clade *a3* contained species from different genera: *Anoxybacillus, Geobacillus* and *Parageobacillus*, all of which were thermophilic. This clade included two exceptional *Anoxybacillus* species, *A. geothermalis* and *A. tepidamans*. Nevertheless, it must be noted that these two species displayed higher sequence similarities to *a3* clade members than to *a1* and *a2 Anoxybacillus* amylases, and they shared high 16S rRNA gene homologies to each other compare to the other genera members (this study, Coorevits et al. 2012; Bezuidt et al. 2016; Filippidou et al. 2016). Moreover, a taxonomic revision of *A. tepidamans* species was recently proposed and this species was transferred from genus *Geobacillus* to *Anoxybacillus* (Schäffer et al. 2004; Coorevits et al. 2012). The differential situation of *A. tepidamans* α-amylase and its non-conserved calcium binding sites were also previously mentioned by Chai and colleagues (2016). Therefore, these explanations clearly elucidate why *A. geothermalis* and *A. tepidamans* α-amylases were positioned in the *a3* clade.

The phylogenetic tree shown in Fig. 2 and the comparison of the sequence alignments as well as the secondary and tertiary structure analyses of these 63 bacilli amylases definitely supported the findings below as presented in Fig. 5 and Supplementary file, Table SII. According to these results, five of the *Bacillaceae* α-amylase family clusters from *a1* to *xy* share some common features including high sequence homologies to each other and displaying slowly increased *E*-values within the related group as suggested by Stam et al. (2006). At least one or more of their members are known to possess amylolytic activity experimentally, including the thermostable representatives which are E184aa-*A* (*a1*), B2M1-*A* (*a2*), *ATA* (*a3*) and the mesophilic *BaqA* (*a4*) and *BmaN1* (*xy*) enzymes (Chai et al. 2012; Coorevits et al. 2012; Mok et al. 2013; Puspasari et al. 2013; Filippidou et al. 2016; Sarian et al. 2017). All the members contained signal peptide sequences (Chai et al. 2012; Mok et al. 2013) and predicted to be cellular components of the membrane that harboured residues in relation with cytoplasmic to extracellular sides. Although the two consecutive tryptophans in the loop 3 of domain A were preserved among *Bacillaceae* family, the 3D modelling analyses revealed that these aromatic residues were positioned far from the catalytic site (data not shown) and might play an unknown role instead of sugar binding as proposed before (Mok et al. 2013; Puspasari et al. 2013). The Phyre2 server also did not recognize two consecutive tryptophans as sugar binding pockets. When the surface view and overall fold of the 3D structure models were superimposed both with *TASKA* and *GTA*, they mostly covered the similar $(\beta/\alpha)_8$ TIM-barrel structure and three domain organization as arranged in the GH13 α-amylase family (MacGregor et al. 2001), (Supplementary Fig. S3). The 3D-modelling also supported the position of *ATA* and *GTA* α-amylases within the same *a3* subfamily because of the derived homology report displaying the highest sequence identity and the lowest RMSD values to each other. These bacilli sequences contained well-defined α-amylase family specific conserved regions from CSR-I to CSR-VII with various amino acid residue differences, and shared the characteristic LPDLx motif in their CSR-V regions (Janeček 2002; Ranjani et al. 2015). Besides the F481, Y489, F492, F495, Y497 (E184aa-A numbering) residues, involved in the conserved aromatic motifs at the C-terminus as previously reported (Mok et al. 2013; Janeček et al. 2015), Y457 and Y469 were also additionally found to be preserved among the *Bacillaceae* amylases. The end of their C-termini also contained consecutive lysine and arginine residues (K501-R502, E184aa-*A*).

Beside these general features, *Bacillaceae* family α-amylase clusters from *a1* to *xy* displayed significant differential characteristics from each other (Fig. 5, Supplementary Table SI and SII). The *a1* and *a2* members from genus *Anoxybacillus* in addition to the *a3* members from gerera *Anoxybacillus, Geobacillus* and *Parageobacillus* were all thermophilic groups of *Bacillaceae* family. The signature sequence of LPDLx motif in CSR-V was LPDLA in these thermophiles. In the case of mesophilic genus *Bacillus* counterparts of *a4* and *xy*, this residue changed to alanine in their LPDLN motif. The function of alanine residue (A184) of ASKA was already displayed by experimental mutation analysis (Chai et al. 2016) and its presence only in the thermophilic bacilli α-amylases indicates its importance in enzyme thermostability. The enzyme activity and stability of α-amylases from *a1* to *a3* clusters were also higher in elevated temperatures when compared with *a4* and *xy* members due to their temperature requirements. As there were some residue differences in the seven CSR regions of α-amylases

from *Bacillaceae* families, the created specific logos peculiar to each subfamily in this study would be useful for placing a novel bacilli α-amylase properly to its relevant subfamily in further studies. In 2017, a novel GH13_*xy* subfamily akin to *BmaN1* α-amylase, having catalytic activity despite its atypical catalytic residues, was described by Sarian et al. (2017). The catalytic triad including the traditional Asp, Glu and Asp (D213$^{\Delta}$, E242$^{\Delta}$, D310$^{\Delta}$, E184aa-*A*) residues as well as the invariable Arg (R211, E184aa-*A*) are all preserved in *a1, a3* and *a4* members like other defined GH13 subfamilies, but in the case of *a2* (D215$^{\Delta}$, G244$^{\Delta}$ and R312$^{\Delta}$, B2M1-*A*) and *xy* (K202$^{\Delta}$, E231$^{\Delta}$, and H294$^{\Delta}$, *BmaN1*) subfamilies, an irregular catalytic triad was identified in addition to the residue changes in invariable Arg position (Y213/Y200, B2M1-*A*/*BmaN1*). Although the amylolytic activities of both B2M1-*A* and *BmaN1* enzymes from *a2* to *xy* clusters were experimentally proved, they differed from the other three *Bacillaceae* related groups, by the lack of complete catalytic machinery and this invariable Arg position (Filippidou et al. 2016; Sarian et al. 2017). Moreover, the pI values of *a2* and *xy* members (> 9.0) were predicted higher than the others groups.

The previous studies confirmed that calcium ion could affect amylases by enhancing both their catalytic activities and structural stabilities (Declerck et al. 2000; Tao et al. 2008). Calcium had no effect on the enzyme activity of *GTA* α-amylase, but increased its thermostability (Mok et al. 2013). In the case of *ASKA* and *ADTA*, calcium ion increased both their enzyme activity and thermostability (Chai et al. 2012). The two calcium-binding sites in *GTA* and four sites in *TASKA* were already clarified with their crystal structure analyses. Therefore, when these residues involved in calcium binding were compared using sequence alignments of the *a1, a2, a3* and *a4* representatives, the possible Ca1, Ca2 and Ca4 binding sites were found to be mostly conserved among these bacilli amylases. In Ca2 of *a2, a3* and *a4* members, an Asn residue (N46, E184aa-*A*) was changed to Asp residue, from neutral to positive charge which may increase the affinity to calcium. In contrast, amino acid substitutions, having the possibility of decreasing the Ca^{2+} ion binding efficiency, was observed in *a2, a3* and *a4* members with Ca3 being the least conserved region among these bacilli. In the secondary structure alignment studies with Phyre2 server, hints for 12 substrate-binding pockets in relation with substrate specificities or sugar recognitions were also found out (Supplementary Fig. SIIa-d). The presence and roles of these pockets in maltose binding groove were confirmed by the *in silico* 3D models of E184aa-*A* (*a1*), B2M1-*A* (*a2*), *ATA* (*a3*), and *BaqA* (*a4*) α-amylases, which superimposed with both *GTA* and *TASKA* templates. Also, some additional residues for maltose binding, different from the previously reported ones were determined (Mok et al. 2013; Chai et al. 2016). The differences of the residue changes for 12 sugar pockets and 4 calcium-binding sites peculiar to the proposed subfamilies were also listed in Supplementary Table SII.

Novel thermostable α-amylases were introduced in this study that would be biotechnologically important and investigation of these *Anoxybacillus* α-amylases collectively with a great number of other bacilli originated enzymes provided a better look for the big picture of α-amylases from *Bacillaceae* family. Despite some other hypothesis, the evolution of α-amylase genes was thought to be occurred via divergent evolution (Jespersen et al. 1993; Janeček 2002). The number of the recognized α-amylase sequences from *Bacillaceae* family was considerably increased during the last decade and this survey described the basic story of divergent evolution of endospore-forming bacilli originated α-amylases. Thereby, we suggest the necessity that the still non-defined GH13 subfamily members containing the *ASKA, ADTA, GTA, Pizzo, Gt-amyII* and *BaqA* α-amylases which were formerly proposed in a single subfamily (Janeček et al. 2015; Ranjani et al. 2015; Chai et al. 2016; Sarian et al. 2017) should be classified into appropriately separated subfamilies and the examined 63 homologous *Bacillaceae* family related α-amylases could be grouped under more accurate and manageable GH13 subfamilies as they shared some similarities peculiar to the endospore-forming bacilli, but also contained significant differences that required dividing them into 5 individual subfamilies. This proposal was based upon the phylogenetic findings both on α-amylase and 16S rRNA genes, the genera level taxonomic origins as well as the temperature requirements of these amylolytic enzymes, the high sequence homologies of our 15 *Anoxybacillus* α-amylases to only *Anoxybacillus* related *ASKA, ADTA* and *GSX-BL* enzymes, their distant relatedness to *GTA, Gt-amyII* and *BaqA*, the separate branches and high boostrap values of *GTA* and *BaqA* in the cladogram, the comparisional sequence alignments and structural analyses including their 7 CSR regions, 12 sugar-binding and 4 calcium-binding sites, the presence or absence of the complete catalytic machinery in addition to the currently unassigned status of these bacilli α-amylases to a proper GH13 subfamiliy. Consequently, the proposed *Bacillaceae* family related subfamilies were the new *a2* group clustered around *α*-amylase B2M1-*A* from *Anoxybacillus* sp. B2M1, the *a1, a3* and *a4* subfamilies (including the representatives E184aa-*A, ATA*, and *BaqA*) all of which were composed from the division of the formerly grouped single subfamily clustered around *α*-amylase *BaqA*, and finally the *xy* subfamily previously designated by Sarian et al. (2017) that clustered around amylolytic emzyme *BmaN1* from *B. megaterium*.

Acknowledgements This study was funded by The Scientific and Technological Research Council of Turkey (TUBITAK) with the Project Number of 115R001. Therefore, we thank to TUBITAK for the financial support.

Compliance with ethical standards

Conflict of interest No conflict of interest declared.

References

Acer Ö, Pirinççiğlu H, Bekler FM, Gül-Güven R (2015) *Anoxybacillus* sp. AH1, an α-amylase-producing thermophilic bacterium isolated from Dargeçit hot spring. Biologia 70(7):853–862

Agüloğlu FS, Enez B, Özdemir S, Matpan BF (2014) Purification and characterization of thermostable a-amylase from thermophilic *Anoxybacillus flavithermus*. Carbohydr Polym 102:144–150

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25(17):3389–3402

Benson DA, Cavanaugh M, Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW (2014) GenBank. Nucleic Acids Res 42:D32–D37

Berekaa MM, Soliman NA, Abdel-Fattah YR (2007) Production, partial characterization and cloning of thermostable α-amylase of a thermophile *Geobacillus thermoleovorans* YN. Biotechnol 6:175–183

Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) The Protein Data Bank. Nucleic Acids Res 28(1):235–242

Bezuidt OK, Pierneef R, Gomri AM, Adesioye F, Makhalanyane TP, Kharroub K, Cowan DA (2016) The *Geobacillus* pan-genome: implications for the evolution of the genus. Frontiers in Microbiol 7:723

Biasini M, Bienert S, Waterhouse A, Arnold K, Studer G, Schmidt T, Kiefer F, Cassarino TG, Bertoni M, Bordoli L, Schwede T (2014) SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. Nucleic Acids Res 42(W1):W252–W258

Bolton DJ, Kelly CT, Fogarty WM (1997) Purification and characterization of the α-amylase of *Bacillus flavothermus*. Enzym Microb Technol 229(96):340–343

Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B (2009) The Carbohydrate-Active EnZymes database (CAZy): an expert resource for glycogenomics. Nucleic Acids Res 37:D233–D238

Chai YY, Rahman RNZRA., Illias RM, Goh K (2012) Cloning and characterization of two new thermostable and alkalitolerant α-amylases from the *Anoxybacillus* species that produce high levels of maltose. J Ind Microbiol Biotechnol 39(5):731–741

Chai KP, Othman NFB, Teh AH, Ho KL, Chan KG, Shamsir MS, Goh KM, Ng CL (2016) Crystal structure of *Anoxybacillus* α-amylase provides insights into maltose binding of a new glycosyl hydrolase subclass. Sci Rep 6:23126

Cihan AC (2013) Taxonomic classification of *Anoxybacillus* isolates from geothermal regions in Turkey by 16S rRNA gene sequences and ARDRA, ITS-PCR, Rep-PCR analyses. Polish J Microbiol 62(2):149–163

Cihan AC, Cokmus C, Koc M, Ozcan B (2014a) *Anoxybacillus calidus* sp. nov., a thermophilic bacterium isolated from soil near a thermal power plant. Int J Syst Evol Microbiol 64:211–219

Cihan AC, Koc M, Ozcan B, Tekin N, Cokmus C (2014b) *Thermolongibacillus altinsuensis* gen. nov., sp. nov. and *Thermolongibacillus kozakliensis* sp. nov., novel aerobic thermophilic long bacilli isolated from hot springs. Int J Syst Evol Microbiol 64:187–197

Cohn F (1872) Genus *Micrococcus*. Beitr Biol Pflanz 1:127–224

Coorevits A, Dinsdale AE, Halket G, Lebbe L, De Vos P, Landschoot AV, Logan NA (2012) Taxonomic revision of the genus *Geobacillus*: emendation of *Geobacillus, G. stearothermophilus, G. jurassicus, G. toebii*, G. thermodenitrificans and *G. thermoglucosidans* (nom. corrig., formerly 'thermoglucosidasius'); transfer of *Bacillus thermantarcticus* to the genus as *G. thermantarcticus* comb. nov.; proposal of *Caldibacillus debilis* gen. nov., comb. nov.; transfer of *G. tepidamans* to *Anoxybacillus* as *A. tepidamans* comb. nov.; and proposal of *Anoxybacillus caldiproteolyticus* sp. nov. Int J Syst Evol Microbiol 62:(7):1470–1485

Crooks GE, Hon G, Chandonia JM, Brenner SE (2004) WebLogo: A sequence logo generator. Genome Res 14:1188–1190

Declerck N, Machius M, Wiegand G, Huber R, Gaillardin C (2000) Probing structural determinants specifying high thermostability in *Bacillus licheniformis* α-amylase. J Mol Biol 301:1041–1057

Deep K, Poddar A, Das SK (2013) *Anoxybacillus suryakundensis* sp. nov, a moderately thermophilic, alkalitolerant bacterium isolated from hot spring at Jharkhand, India. PLoS One 8(12):e85493

Demirjian D, Moris-Varas F, Cassidy C (2001) Enzymes from extremophiles. Curr Opin Chem Biol 5:144–151

Derekova A, Mandeva R, Kambourova M (2008) Phylogenetic diversity of thermophilic carbohydrate degrading bacilli from Bulgarian hot springs. World J Microbiol Biotechnol 24:1697–1702

Dheeran P, Kumar S, Jaiswal YK, Adhikari DK (2010) Characterization of hyperthermostable α-amylase from Geobacillus sp. IIPTN. Appl Microbiol Biotechnol 86:1857–1866

Euzéby JP, Parte AC (1997) LPSN. http://www.bacterio.net/anoxybacillus.html. Accessed 01 Aug 2017

Ezeji TC, Bahl H (2006) Purification, characterization, and synergistic action of phytate-resistant α-amylase and α-glucosidase from *Geobacillus thermodenitrificans*. HRO10 J Biotechnol 125:27–38

Felsenstein J (1985) Confidence limits on phylogenies: An approach using the bootstrap. Evolution 39:783–791

Filippidou S, Jaussi M, Junier T, Wunderlin T, Jeanneret N, Palmieri F, Palmieri I, Roussel-Delif L, Vieth-Hillebrand A, Vetter A, Chain PS (2016) *Anoxybacillus geothermalis* sp. nov., a facultatively anaerobic, endospore-forming bacterium isolated from mineral deposits in a geothermal station. Int J Syst Evol Microbiol 66(8):2944–2951

Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, Potter SC, Punta M, Qureshi M, Sangrador-Vegas A, Salazar GA, Tate J (2016) The Pfam protein families database: towards a more sustainable future. Nucleic Acids Res 44:D279–D285

Finore I, Kasavi C, Poli A, Romano I, Oner ET, Kirdar B, Dipasquale L, Nicolaus B, Lama L (2011) Purification, biochemical characterization and gene sequencing of a thermostable raw starch digesting a-amylase from *Geobacillus thermoleovorans* subsp. *stromboliensis* subsp. nov. World J Microbiol Biotechnol 27:2425–2433

Fritzsche HB, Schwede T, Schulz GE (2003) Covalent and three-dimensional structure of the cyclodextrinase from *Flavobacterium* sp. no. 92. Eur J Biochem 270:2332–2341

Gul-Guven R, Guven K, Poli A, Nicolaus B (2008) *Anoxybacillus kamchatkenses* subsp. *asacchared.ens* subsp. nov., a thermophilic bacterium isolated from a hot spring in Batman. J Gen Appl Microbiol 54(6):327–334

Gupta R, Gigras P, Mohapatra H, Goswami VK, Chauhan B (2003) Microbial α-amylases: A biotechnological perspective. Process Biochem 38:1599–1616

Hauli I, Sarkar B, Mukherjee T, Mukhopadhyay SK (2013) Isolation and identification of a novel thermo-alkaline, thermostable, SDS and chelator resistant amylase producing *Anoxybacillus* sp. Adv Appl Sci Res 4:202–212

Hostinova E, Janecek S, Gasperik J (2010) Gene sequence, bioinformatics and enzymatic characterization of α-amylase from *Saccharomycopsis fibuligera* KZ. Protein J 29:355–364

Janeček Š (2002) How many conserved sequence regions are there in the α-amylase family? Biologia (Bratisl) 57(11):29–41

Janeček Š, Svensson B, Henrissat B (1997) Domain evolution in the α-amylase family. J Mol Evol 45:322–331

Janeček Š, Svensson B, MacGregor EA (2014) α-Amylase: an enzyme specificity found in various families of glycoside hydrolases. Cell Mol Life Sci 71(7):1149–1170

Janeček Š, Kuchtová A, Petrovičová S (2015) A novel GH13 subfamily of α-amylases with a pair of tryptophans in the helix α3 of the catalytic TIM-barrel, the LPDlx signature in the conserved sequence region V and a conserved aromatic motif at the C-terminus. Biologia 70(10):1284–1294

Jespersen HM, MacGregor EA, Henrissat B, Sierks MR, Svensson B (1993) Starch-and glycogen-debranching and branching enzymes: prediction of structural features of the catalytic (β/α) 8-barrel domain and evolutionary relationship to other amylolytic enzymes. J Protein Chem 12(6):791–805

Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Mentjies P, Drummond A (2012) Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28(12):1647–1649

Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJ (2015) The Phyre2 web portal for protein modeling, prediction and analysis. Nat Protoc 10(6):845–858

Kikani BA, Singh SP (2012) The stability and thermodynamic parameters of a very thermostable and calcium-independent α-amylase from a newly isolated bacterium, *Anoxybacillus beppuensis* TSSC-1. Process Biochem 47(12):1791–1798

Klein B, Foreman JA, Searcy RL (1970) New chromogenic substrate for determination of serum amylase activity. Clin Chem 16(1):32–38

Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B (2014) The carbohydrate-active enzymes database (CAZy) in 2014. Nucleic Acids Res 42:D490–D495

Lowry RR, Tinsley IJ (1976) Rapid colorimetric determination of free fatty acids. J Am Oil Chem Soc 53(7):470–472

Maarel MJ van der, Veen B van der, Uitdehaag JC, Leemhuis H, Dijkhuizen L (2002) Properties and applications of starch-converting enzymes of the alpha-amylase family. J Biotechnol 94(2):137–155

MacGregor EA, Janeček Š, Svensson B (2001) Relationship of sequence and structure to specificity in the α-amylase family of enzymes. Biochim Biophys Acta 1546(1):1–20

Marchler-Bauer A, Bo Y, Han L, He J, Lanczycki CJ, Lu S, Chitsaz F, Derbyshire MK, Geer RC, Gonzales NR, Gwadz M, Hurwitz DI. Lu F, Marchler GH, Song JS, Thanki N, Wang Z, Yamashita RA, Zhang D, Zheng C, Geer LY, Bryant SH (2017) CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. Nucleic Acids Res 45(D1):D200–D203

Matpan Bekler F, Güven K (2014) Isolation and production of thermostable amylase from thermophilic *Anoxybacillus* sp. KP1 from Diyadin hot spring in Ağri. Turkey Biol (Bratisl) 69(4):419–427

Matsuura Y, Kusunok M, Harada W, Kakudo M (1984) Structure and possible catalytic residues of Taka-amylase A. J Biochem 95:697–702

Mehta D, Satyanarayana T (2014) Domain C of thermostable α-amylase of *Geobacillus thermoleovorans* mediates raw starch adsorption. Appl Microbiol Biotechnol 98(10):4503–4519

Miller GL (1959) Use of dinitro-salicylic acid reagent for determination of reducing sugars. Anal Chem 31(3):426–428

Mok SC, Teh AH, Saito JA, Najimudin N, Alam M (2013) Crystal structure of a compact a-amylase from *Geobacillus thermoleovorans*. Enzyme Microb Technol 53(1):46–54

Mukherjee S, Stamatis D, Bertsch J, Ovchinnikova G, Verezemska O, Isbandi M, Thomas AD, Ali R, Sharma K, Kyrpides NC, Reddy TBK (2017) Genomes OnLine Database (GOLD) v.6: data updates and feature enhancements. Nucleic Acids Res 45(D1):D446–D456

Nazina TN, Tourova TP, Poltaraus AB, Novikova EV, Grigoryan AA, Ivanova AE, Lysenko AM, Petrunyaka VV, Osipov GA, Belyaev SS, Ivanov MV (2001) Taxonomic study of aerobic thermophilic bacilli: descriptions of Geobacillus subterraneus gen. nov., sp. nov. and Geobacillus uzenensis sp. nov. from petroleum reservoirs and transfer of Bacillus stearothermophilus, *Bacillus thermocatenulatus, Bacillus thermoleovorans, Bacillus kaustophilus, Bacillus thermoglucosidasius* and *Bacillus thermodenitrificans* to *Geobacillus* as the new combinations *G. stearothermophilus, G. thermocatenulatus, G. thermoleovorans, G. kaustophilus, G. thermoglucosidasius* and *G. thermodenitrificans*. Int J Syst Evol Microbiol 51(2):433–446

Ozdemir S, Okumus V, Ulutas MS, Dundar A, Akarsubası AT, Dumonted S (2015) Isolation of a novel thermophilic *Anoxybacillus flavithermus* SO-13, production, characterization and industrial applications of its thermostable α-amylase. J Bioprocess Biotech 5:237

Ozdemir S, Okumus V, Ulutas MS, Dundar A, Akarsubasi AT, Dumontet S (2016a) Production and characterization of thermostable α-amylase from thermophilic *Anoxybacillus flavithermus* sp. nov. SO-19 Starch-Stärke 68:1244–1253

Ozdemir SC, Cihan AC, Kilic T, Cokmus C (2016b) Optimization of thermostable alpha-amylase production from *Geobacillus* sp. D413″. J Microbiol Biotechnol Food Sci 6(1):689–694

Petersen TN, Brunak S, von Heijne G, Nielsen H (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat methods 8(10):785–786

Pikuta E, Lysenko A, Chuvilskaya N, Mendrock U, Hippe H, Suzina N, Nikitin D, Osipov G, Laurinavichius K (2000) *Anoxybacillus pushchinensis* gen. nov., sp. nov., a novel anaerobic, alkaliphilic, moderately thermophilic bacterium from manure, and description of *Anoxybacillus flavithermus* comb. nov. Int J Syst Evol Microbiol 50(6):2109–2117

Poli A, Esposito E, Lama L, Orlando P, Nicolaus G, de Appolonia F, Gambacorta A, Nicolaus B (2006) *Anoxybacillus amylolyticus* sp. nov., a thermophilic amylase producing bacterium isolated from Mount Rittmann (Antarctica). Syst Appl Microbiol 29(4):300–307

Puspasari F, Radjasa OK, Noer AS, Nurachman Z, Syah YM, Maarel M, Dijkhuizen L, Janeček Š, Natalia D (2013) 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 Microbiol 114(1):108–120

Ranjani V, Janeček Š, Chai KP, Shahir S, Rahman RNZRA., Chan KG, Goh KM (2015) Protein engineering of selected residues from conserved sequence regions of a novel *Anoxybacillus* α-amylase. Sci Rep 4:5850

Rao JLUM., Satyanarayana T (2007) Purification and characterization of a hyperthermostable and high maltogenic α-amylase of an extreme thermophile *Geobacillus thermoleovorans*. Appl Biochem Biotechnol 142:179–193

Santos EO, Martins MLL (2003) Effect of the medium composition on formation of amylase by *Bacillus* sp. Braz Arch Biol Technol 46(1):129–134

Sarian FD, Janeček Š, Pijning T, Ihsanawati Nurachman Z, Radjasa OK, Dijkhuizen L, Natalia D, van der Maarel MJEC. (2017) A new group of glycoside hydrolase family 13 α-amylases with an aberrant catalytic triad. Sci Rep 7:44230

Schäffer C, Franck WL, Scheberl A, Kosma P, McDermott TR, Messner P (2004) Classification of isolates from locations in Austria and Yellowstone National Park as *Geobacillus tepidamans* sp. nov. Int J Syst Evol Microbiol 54(6):2361–2368

Sievers F, Wilm A, Dineen DG, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Söding J, Thompson JD, Higgins DG (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol 7(1):539

Sokal RR, Michener CD (1958) A statistical method for evaluating systematic relationships. Univ Kans Sci Bull 38:1409–1438

Stam MR, Danchin EG, Rancurel C, Coutinho PM, Henrissat B (2006) Dividing the large glycoside hydrolase family 13 into subfamilies: towards improved functional annotations of α-amylase-related proteins. Protein Eng Des Sel 19:555–562

Suzuki Y, Kishigami T, Abe S (1976) Production of extracellular α-glucosidase by a thermophilic *Bacillus* species. Appl Environ Microbiol 31(6):807–812

Tao X, Jang MS, Kim KS, Yu Z, Lee YC (2008) Molecular cloning, expression and characterization of alpha-amylase gene from a marine bacterium *Pseudoalteromonas* sp. MY-1. Indian J Biochem Biophys 45:305–309

Tawil G, Viksø-Nielsen A, Rolland-Sabaté A, Colonna P, Buléon A (2012) Hydrolysis of concentrated raw starch: A new very efficient α-amylase from *Anoxybacillus flavothermus*. Carbohydr Polym 87(1):46–52

Tien J, Peters CJ, Wong XM, Cheng T, Jan YN, Jan LY, Yang H (2014) A comprehensive search for calcium binding sites critical for TMEM16A calcium-activated chloride channel activity. eLife 3:e02772

UniProt Consortium (2017) UniProt: the universal protein knowledgebase. Nucleic Acids Res 45:D158–D169

Vieille C, Zeikus GJ (2001) Hyperthermophilic enzymes: sources, uses, and molecular mechanisms for thermostability. Microbiol Mol Biol Rev 65:1–43

Zhang F, Yang X, Geng L, Zhang Z, Yin Y, Li W (2015) Purification and characterization of a novel and versatile α-amylase from thermophilic *Anoxybacillus* sp. YIM 342 Starch-Stärke 68:446–453