EXPLORING NATURAL SELECTION SIGNATURES ON THE ALPHA-AMYLASE GENE OF NOVEL *BACILLUS LICHENIFORMIS* 208 STRAIN ISOLATED FROM A LOCAL HOT SPRING

ASAD, W.[1*] – KIRAN, T.[1,3] – KHAN, M. H.[2] – SALEEM, F.[1,3] – ASAD, S. B.[4] – RASOOL, S. A.[1] – SHAH, T. A.[5] – AZIZ, T.[6] – ALHARBI, M.[7] – ALASMARI, A. F.[7] – ALBEKAIRI, T. H.[7]

[1]*Department of Microbiology, University of Karachi, Karachi, Pakistan*

[2]*Karachi Institute of Radiotherapy and Nuclear Medicine (KIRAN), Karachi, Pakistan*

[3]*Department of Microbiology, Federal Urdu University of Arts, Science and Technology (FUUAST) Karachi, Pakistan*

[4]*Bond Life Centre, University OF Missouri, USA*

[5]*College of Agriculture Engineering and Food Sciences, Shandong University of Technology, Zibo, China*

[6]*Laboratory of Animal Health, Food Hygiene and Quality, Department of Agriculture, University of Ioannina, 47132 Arta, Greece*

[7]*Department of Pharmacology and Toxicology, College of Pharmacy, King Saud University, P. O. Box 2455, Riyadh 11451, Saudi Arabia*

**Corresponding author*
e-mail: wajeehaasad1@gmail.com, iwockd@gmail.com

(Received 20th Oct 2023; accepted 13th Mar 2024)

Abstract. This study emphasized the alpha-amylase gene from *Bacillus licheniformis* as this species has been witnessed as the most prolific bio-factories for the alpha amylase production on industrial scales with particular reference to its wide commercial applicability. In the present study, a gene encoding alpha-amylase from an indigenously isolated *B. licheniformis* 208 strain was PCR amplified and sequenced. The obtained 1313 bp long nucleotide sequence (GenBank KM27230) exhibited the closest match with its homologous genes in *B. licheniformis* ATCC 9945A and ATCC 14580 strains. Gene translation retrieved 437 amino acid long enzyme sequence and exhibits a molecular weight of 50231.8 KDa. Other computed parameters include isoelectric point (5.95), aliphatic index (67.4), instability index (27.45), and GRAVY (-0.627) values. The structure of alpha-amylase 208 was found to exhibit a 17 residues long active site, a conserved catalytic site with 3 residues, and a ligand-binding site comprised of 3 residues. Secondary structure configuration in the alpha-amylase 208 was predominated with α-helix and random coils followed by extended strands. The number of β-turns was found comparatively less. ITASSER assigned E.C number to the BLA 208 sequence was 3.2.1.1 with extracellular origin. Deduced tertiary structure of BLA 208 showed the 7 residues long active site comprising of histidine, arginine, aspartic acid, glutamic acid, tryptophan, histidine, and aspartic acid at position 122, 246, 248, 278, 280, 344, 345 respectively. Na, K, and Ca were observed as the ligand-binding sites existing as a metal triad. Natural selection analysis explains that the alpha-amylase gene in *B. licheniformis* species is under purifying or negative selection i.e. it stabilizes itself by removing the deleterious mutation. Recombination event analysis was done using GARD (Genetic algorithm for recombination detection). Although, it rather inferred the presence of two break points which were found insignificant when validated via Kishino-Hasegawa test which demonstrates topological incongruence.

Keywords: *B. licheniformis alpha-amylase (BLA), natural selection analysis, insilico characterization, secondary structure, 3D structure,*

© 2024, ALÖKI Kft., Budapest, Hungary

Introduction

By far, alpha-amylase (α-1,4-glucan-4-glucanohydrolase, EC 3.2.1.1) is one of the most studied enzymes of industrial interest (Aziz et al., 2024; Ullah et al., 2022; Nguyen et al., 2002). It catalyzes the endo fashion hydrolysis of α-1,4-glucosidic linkages in starch and thus merits application in several industries such as baking, brewing, food, textile, and detergent manufacturing (Rana et al., 2013). Bacterial colonization ability in any environment is related to nutrient acquisition, and it is also an important manifestation of bacterial adaptability. Amylase is well known for its ability to hydrolyze starch to produce glucose, maltose, maltodextrin, and other compounds (Khatoon et al., 2023; Ullah et al., 2023; Huang et al., 2022). The genes encoding the alpha-amylase enzyme are found universally across various domains, indicating that its synthesis is not confined to a specific domain. Genus *Bacillus* secured a central position for the alpha-amylase production on an industrial scale pertaining to its GRAS (Generally Recognized As Safe) status, high growth rates, and potential for the extracellular secretion of the enzyme into the fermentation medium (Asad et al., 2014; Feto, 2017). Alpha-amylase from *B. licheniformis* has been witnessed as the established model for encountering the questions having both technological and fundamental inferences regarding thermostability (Declerck et al., 2000). Despite being released from a mesophilic bacterial species, this amylolytic enzyme displays a great deal of thermal stability and is thus widely employed in biotechnological applications particularly in starch liquefaction where process temperature reaches as high as 110°C. A number of mutational, structural, and biochemical studies have been documented investigating BLA thermal properties (Chakraborty et al., 2011; Li et al., 2017; Wu et al., 2018).

The emergence of molecular techniques such as next-generation sequencing made it possible to feasibly establish the comparative account of closely related prokaryotic genomes for evolutionary studies. In general, comparative genomics can discuss two key groups of phenomena, macroevolutionary and microevolutionary. It tackles the macroevolutionary case studies between maximally diverse groups, such as the correlation between bacterial and archaeal phyla or horizontal transmission of gene between distantly related organisms. Microevolutionary processes, on the other hand require several, closely linked genomes to be contrasted, are arguably fundamental to our interpretation of evolutionary phenomena, such as the differential impact of selection on various forms of genomic sequences (Andersson et al., 2002; King Jordan et al., 2002).

The dN/dS ratio is the fundamental indicator of the selection pressure operating on protein-coding sequences. The lesser the dN/dS ratio, the greater would be the pressure of purifying selection that affects the given protein-coding gene. The impact of positive selection at a molecular level has been witnessed par expectation while mining the bacterial genomes (Bustamante et al., 2005; Novichkov et al., 2009). Details of selection signatures during genome-wide molecular selection analysis have demonstrated the process of microevolution particularly in pathogen–host interaction, and adaptation to a varying environment (such as antibiotic or physical stress) (Vitti et al., 2013; Kosiol and Anisimova, 2012). Positive selection analysis has been conducted on model organisms such as *E. coli* (Petersen et al., 2007), *Listeria* (Orsi et al., 2008), *Campylobacter spp* (Lefébure and Stanhope, 2009), or the genus *Streptococcus* (Suzuki et al., 2011), to date, the outcome of these studies, interprets the significance of positive selection as a part of natural selection in fixing the advantageous mutations and providing better adaptability to a wide range of environmental variations. Generally, spontaneous mutations are more likely to be deleterious than advantageous, thus

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2204_33733388
© 2024, ALÖKI Kft., Budapest, Hungary

numerous novel alleles are instantly subjected to negative selection and gotten expelled from the gene pool before they can accomplish perceptible frequency within the population. This process of excluding harmful mutations can be referred to as negative selection or as background selection (Vitti et al., 2013).

Temperature has a significant impact on protein evolution because it frequently leads to adaptive changes in protein-structure and function in species that live in thermally different environments. The majority of earlier studies on protein temperature adaptation, however, have been on proteins that can withstand extremely high temperatures. As a result, little is known about the kinds and scope of evolutionary change that happens to proteins when organisms adjust to slight temperature changes in their environment. Thus, the objective of the present study was to evaluate the type of selective pressure operating on the gene of one of the most studied enzymes i.e. alpha amylase from *B. licheniformis* species including one of the variants from an indigenously isolated *B. licheniformis* 208 strain from a local hotspring and to compare the properties of BLA 208 enzyme with the other closely related *B. licehniformis* strains. To the best of our knowledge, this study represents the first of its kind, and no prior reports relevant to the alpha-amylase enzyme from any microbial species have been documented.

Materials and methods

Isolation and culture maintenance of alpha-amylase producing strain

The amylolytic *B. licheniformis* 208 strain used in this study was isolated from a local hot spring of Karachi, Pakistan. Culture maintenance was done on LB agar slants at 50°C periodically after 15 days (Asad et al., 2014).

Isolation and amplification of BLA 208 gene

Genomic DNA of the *B. licheniformis* 208 strain was isolated using TRIZol reagent and attempts were made to characterize the alpha amylase gene. Purified DNA after extraction was preserved at -20°C until further use. For alpha amylase gene amplification, the set of primers were designed via PRIMER DESIGN and their default parameters (Ye et al., 2012) are presented in *Table 1*.

Table 1. *Primers for alpha amylase gene amplification*

Primer set I	Forward primer	CTTTATGCCCGATTGCTGCC
	Reverse primer	CCCGGCTGTGTATCATGGTT
Primer set II	Forward primer	GTCATCAACCACAAAGGCGG
	Reverse primer	GGAACGGTTTCCGGTGATGT

Amplification of the target sequence was done c-100 touch thermocycler (biorad cfx 96 real-time system) as per the following pcr program and conditions listed in *Table 2*.

APPLIED ECOLOGY AND ENVIRONMENTAL RESEARCH 22(4):3373-3388.
http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2204_33733388
© 2024, ALÖKI Kft., Budapest, Hungary

Table 2. *PCR conditions for alpha amylase gene amplification*

Step	Conditions
Pre-denaturation	95°C for 10 min
Amplification (45) cycles	
Denaturation	95°C for 15 s
Annealing	60°C for 1 min
Extension	72°C for 30 s
Final extension step	72°C for 5 min

PCR amplified product was then observed on 1% agarose gel. DNA ladder of 100 bp (Fermentas) was used as a reference to compare the corresponding band sizes. After purification, sequencing of the amplified gene fragments was done using ABI PRISM 3730xl genetic analyzer (Applied Biosystems/Hitachi- USA). The obtained sequence was searched in Genbank using the NCBI BLAST program for similar homologs and deposited in the NCBI DNA repository. A phylogenetic tree was constructed using the program Clustal X to observe the evolutionary relatedness of the indigenous BLA 208 sequence with homologous counterparts.

Physicochemical characterization of B. licheniformis amylase 208

Expasy's ProtParam server was used to deduce the *insilico* physicochemical properties of the *B. licheniformis* amylase 208 such as molecular weight, theoretical isoelectric point (pI), the total number of positive and negative residues, extinction coefficient, instability index, aliphatic index, and grand average hydropathy (GRAVY) (Pramanik et al., 2021).

Secondary structure prediction

Self-Optimized Prediction Method of Alignment' (SOPMA) was employed for predicting the secondary structural features of the *B. licheniformis* amylase 208. Based on the homolog method, it calculates the percentage of α-helix, β-sheets, turns, random coils, and extended strands (Ji et al., 2020; Lugani and Sooch, 2017).

Tertiary structure prediction

The tertiary structure of *B. licheniformis* amylase 208 was elucidated using the ITASSER web server. The selection of the best predicted 3D model was based on a C score. A comparative account of *B. licheniformis* 208 amylase sequence was also established with the best-chosen template (Mukherjee, 2012).

Natural selection analysis

A total of 26 full-length fasta sequences of alpha amylase gene from different *B. licheniformis* variants were downloaded from the NCBI web server. Exclusion of non-coding regions of the nucleotide sequences along with premature stop codons was carefully done to ignore the possibility of frame shifting gapes. The programme "MEGA (version 06)" was used to create multiple sequence alignments. First, these sequences were translated, aligned, and then mapped back to constituent nucleotides. The signatures of selection operating on the *alpha amylase* genes were detected using "the single-likelihood ancestor counting (SLAC), fixed effects likelihood (FEL) and

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2204_33733388
© 2024, ALÖKI Kft., Budapest, Hungary

random effects likelihood (REL) methods” implemented in the ‘HyPhy package’ available at the ‘DataMonkey’ web server. The SLAC/FEL analysis was done using the TrN93 nucleotide substitution bias model at 0.1 level of significance whereas, for REL analysis Baye’s factor = 50 was used based on neighbor-joining tree (Tamura and Nei, 1993).

Statistical significance of the positive selection analysis was determined at a p-value of < 0.1. GARD recombination analysis was run using the HKY85 model (Argimón et al., 2013).

Results and discussion

Amylase gene sequencing and characterization

The amplification of the amylase gene using both pairs of primer set resulted in amplicons approximately 1 Kb in size (*Fig. 1*). The sequence of the amylase gene, consisting of 1313 base pairs, was determined by two overlapping fragments. After submitting the obtained amylase gene sequence to the NCBI BLAST server, it generated 49 hits, of which the first 7 hits showed a 99% similarity with 100% query coverage to the amylases of different strains of B. licheniformis. The amylase gene sequence was subsequently deposited to the GenBank repository and assigned the accession number KM 27230. Upon translation, 437 amino acid long protein sequence was deduced and the conserved domains were analyzed using NCBI CD server. The constructed evolutionary tree revealed the BLA 208 sequence similarity to the α- amylase gene sequence from *Bacillus licheniformis* ATCC 9945A and that of *Bacillus licheniformis* ATCC 14580 (*Fig. 2*).

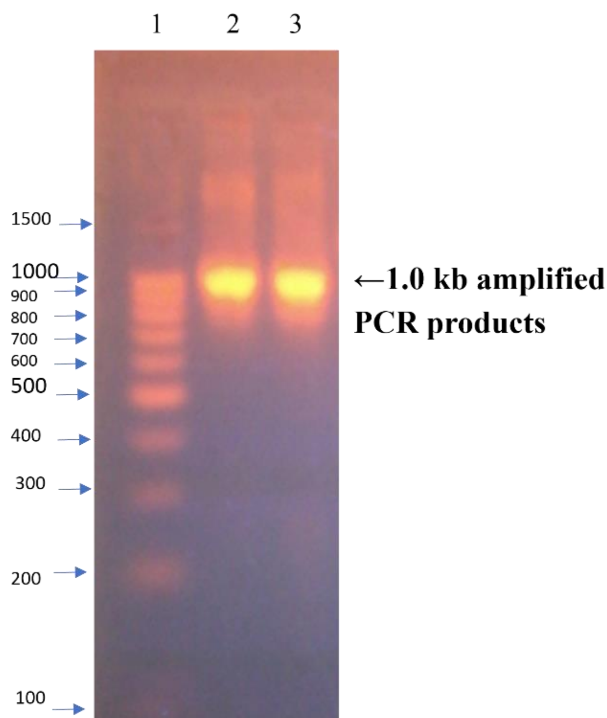

Figure 1. *Agarose gel electrophoresis of PCR amplicons of BLA 208 gene: (1) 100 bp DNA ladder, (2) BLA 208 gene fragment by primer set 1, (3) BLA 208 gene fragment by primer set 2*

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2204_33733388
© 2024, ALÖKI Kft., Budapest, Hungary

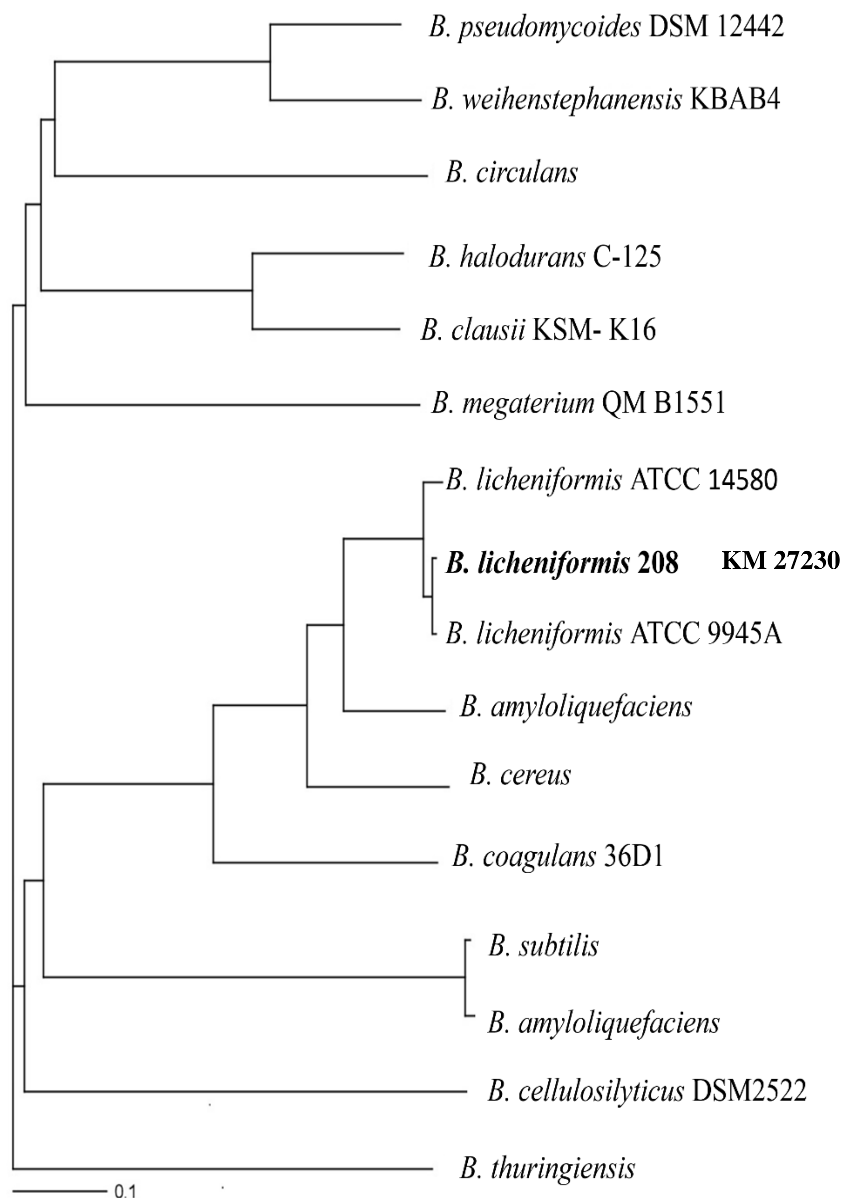

Figure 2. *Phylogram of amylase amino acid sequences from different members of Bacillus by Clustal X (Bootstrap value: 1000)*

In silico physicochemical characterization of BLA 208

The Expasy-ProtPram tool (http://www.expasy.org/tools/protparam.html) was used to compute the sequence-based physicochemical properties of alpha amylase 208 as shown in *Table 3*. Accordingly, the computed theoretical molecular weight was 50.23 kDa. In contrast, the estimated molecular weight of amylase 208 using SDS PAGE was 30 kDa (data not shown). It is very common to observe discrepancies in the molecular weights estimated using SDS-PAGE and other methods such as sequence analysis. The obvious reasons for these anomalies include post-translational modifications, poor binding to SDS micelles (which fasten the mobility in gel), protease-sensitive motifs in the sequence, proline-rich region (contributing rigidity in structure), and reduction of disulfide bonds, etc. (Rath et al., 2009).

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2204_33733388
© 2024, ALÖKI Kft., Budapest, Hungary

Table 3. *Physico-chemical characterization of BLA 208 (Accession Number KM 27230) computed via Expasy's ProtParam tool*

1	Mol. weight	50231.8
2	Atomic composition	$C_{2283}H_{3352}N_{614}O_{665}S_{6}$
3	No of amino acid	437
4	pI	5.95
5	Instability index	27.45
6	Aliphatic index	67.44
7	GRAVY	-0.627

Amino acid composition of α-amylase 208 revealed that a major proportion is comprised of alanine (8.2%), aspartic acid, and glycine (8.0%) followed by leucine (6.9%). No cysteine residue(s) was present in the sequence. This indicates the absence of disulfide bonds in the protein. However, the presence of 6 sulfur moieties can be accounted for by the presence of methionine (6 in number). It is somehow logical to maintain thermal stability at high temperatures as the amino acid cysteine is highly sensitive to degradation. Thus, they are rarely present in thermostable proteins (Dong et al., 1997). Presence of disulfide bond has been linked with accelerated protein evolution. In line with this hypothesis, Feyertag and Alvarez-Ponce (2001), found that membrane proteins with disulfide bonds evolved 88% faster than those without, and extracellular proteins with disulfide bonds evolved 49% faster than those without. Furthermore, genes encoding proteins with disulfide bonds are more likely to show positive selection signatures (Feyertag and Alvarez-Ponce, 2001). The absence of cysteine was also noticed by Shahhoseini et al. (2005) in the amino acid sequence of *B. licheniformis* alpha amylase (Shahhoseini et al., 2005). A study conducted by Dhar et al. (2012) further verifies this fact with particular reference to alpha amylase. They analyzed 31 full-length alpha amylase sequences belonging to different biological domains such as bacteria, plants, and fungi and observed a very high amino acid encoding frequency for glycine (9.42%) whereas, a very low codon frequency of 1.2% was noticed for cysteine. The isoelectric point (pI) for the protein sequences ranges from 5.12- 7.09 which describes the acidic nature of the considered α-amylase sequences (Arora et al., 2009). Isoelectric point (pI) of the currently investigated BLA 208 (KM27230) was found as 5.95. This may be due to a high number of acidic residues (56 in number) to the basic residues (44 in number) in its sequence and further explains its stability at the acidic pH range. This pI value is also in compliance with the findings of Shahhoseini et al. (2005) on alpha amylase of *B. licheniformis* isolated from flour mill wastewater (Shahhoseini et al., 2005). In vitro protein stability can be manifested by its instability index. A value < 40 is indicative of protein stability in the test tube. The BLA 208 was also found stable with a value of 27.45. Another stability pattern of a protein at higher temperatures can be witnessed by its elevated aliphatic index which is presented as the proportionate occupancy of aliphatic side chains in four amino acids including leucine, alanine, valine, and isoleucine. Alpha amylase of *B. licheniformis* 208 showed an aliphatic index of 67.44 (Ikai, 1980).

The grand average of hydropathicity (GRAVY) value of a protein can be calculated by adding the hydropathy values of all the amino acids divided by the length of the query sequence (Wayengera et al., 2009). GRAVY index is a measure of protein solubility, an increase towards the positive score describes the hydrophobic nature while a negative GRAVY value indicates hydrophilic nature. Negative GRAVY value for all the selected

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2204_33733388
© 2024, ALÖKI Kft., Budapest, Hungary

amylases including BLA 208(-0.627) revealed their hydrophilic nature (Hooda, 2011). The ability of a protein to absorb light at a specific wavelength can be expressed as its molar absorptivity or Extinction coefficient. It helps in the determination of protein concentration in a solution. The extinction coefficient of the α-amylase in the present study was calculated to be 124,220 M^{-1} cm^{-1} at 280 nm in water. This value is closer to the molar extinction coefficient of *B. licheniformis* amylase (139,690 M^{-1} cm^{-1}) reported by Nazmi et al. (2008) at the same wavelength in water (Nazmi et al., 2008).

Secondary structure prediction

The secondary structure features the location of α-helix, beta-strands, or random coils in the amino acid sequence. As shown in *Table 2,* alpha-helix (33.41%) and random coils (40.5%) constitute the majority of the secondary structure followed by extended strands (21.05%) and beta-strands (5.03%) as shown in *Table 4*.

Table 4. *Secondary structure prediction of BLA208 using SOPMA tools*

1	α–Helix (Hh) (%)	33.41
2	β–Turn (Tt) (%)	5.03
3	Extended strands (Ee) (%)	21.05
4	Random coils (%)	40.50

The predominance of the alpha helix contributes to the firmness of the secondary structures. Alpha amylase sequences have been reported to be rich in random coils varying from 51.14 to 33.54% with infrequent β-turns (Ashokan, et al., 2011; Sivakumar and Balaji, 2007). Liu et al. (2011) reported the 40% random coils in α-amylase of *Pseudomonas* sp. K6-28-040 (Liu et al., 2011). These results are also in consonance with the findings of Filiz (2014) that also followed the SOPMA for secondary structure prediction of β-amylase of *Brachypodium distachyon* (Filiz et al., 2014). They also noticed the abundance of random coils (44.49%) and α-helix (34.06%). This abundance of random coils can be explained as these are anticipated to induce the flexibility and conformational change(s) in protein structure like enzyme turnover (Buxbaum, 2007).

Tertiary structure analysis

The tertiary structure of *B. licheniformis* 208 α-amylase was elucidated with online server ITASSER i.e. rated as one of the best programs for high-resolution protein modelling and characterization. It uses the threading method which employs the alignment of the query sequence structure to search and identify the template protein with similar motifs. The unaligned regions are then aligned by *ab-initio* modeling (Roy and Zhang, 2012). Despite specifying a single template, multiple templates (10 in number) were chosen by I-TASSER. The 3D model predicted with highest C-score (0.67) with estimated accuracy of 0.8 ± 0.09 (TMscore) and 5.6 ± 3.6Å (RMSD) was opted for further evaluation (*Fig. 3*). Normalized B factor (NBF) characterizes the residue stability. Residues with normalized B factor greater than 2 generally exhibit less stability (Zhang, 2008). It was noticed that almost all the amino acids were predicted with NBF < 2.0 in *B. licheniformis* amylase 208. Out of the 10 best templates for threading of BLA 208 structure, normalized z-score was found highest for PDB hits 1vjsA and 3bh4a which refers to the crystal structure of precursor alpha amylase from *B. licheniformis* and *B.*

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2204_33733388
© 2024, ALÖKI Kft., Budapest, Hungary

amyloliquefaciencs with mean query coverage of 93 and 96% respectively (Alikhajeh et al.,2010; Yeon Hwang et al., 2022).

Figure 3. *Tertiary structure of B. licheniformis amylase 208*

Predicted E.C. number of BLA 208 sequence was 3.2.1.1 (the expected one) and the predicted active site residues were Histidine (122), Arginine (246), Aspartic acid (248), glutamic acid (278), tryptophan (280), histidine (344), aspartic acid (345). I-TASSER function prediction to enzymes is based on the assigned EC number of the templates closed to the query sequence (Roy and Zhang, 2012). There was a consensus for EC number (3.2.1.1) in all the structurally closed templates which predicts the alpha amylase activity of the target sequence of *B. licheniformis* 208. Comparative modelling with chosen templates revealed that all of the consistently predicted active site residues were conserved in the structure of *B. licheniformis* 208. In general, with precise reference to its catalytic role, location within the cell, and biological process involved, a protein may be tagged with miscellaneous gene oncology (GO) terms (*Table 5*). Again, a concurrence in these terms was noticed among all the selected templates. Based on the amino acid sequence, these Go terms further validated the extracellular secretion and role of BLA 208 in carbohydrate metabolism possessing the alpha amylase activity (*Table 5*). The quality of the GO prediction relies on the GOscore, A score >0.5 describes a reliable prediction and the results of the current study comply as such.

Alpha amylases are considered as metalloenzymes requiring a built-in calcium ion associated with the structure. Amino acid sequence-based I-TASSER search for these ion binding pockets in alpha amylase 208 elucidated the binding sites of sodium, potassium, and calcium with a high ligand binding C-score and BS > 1.0. The study emphasizes the first three predictions as the rest are with low ligand binding C-score, hence of less reliability. It was also inferred that except for a few residues, these ligand binding sites were shared for all the three ligands suggesting their occurrence as a metal triad (*Fig. 4*).

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2204_33733388
© 2024, ALÖKI Kft., Budapest, Hungary

Table 5. *Consistent GO terms showing function and location of the BLA 208*

Molecular function		**Biological process**		**Cellular function**
GO term	**GO-Score***	**GO term**	**GO-Score***	**GO term**
GO:0005509 (Ca ion binding)	1.00	GO:0005975 (Carbohydrate metabolic process)	1.00	GO:0005576 (Extracellular region)
GO:0004556 (alpha amylase activity)	0.98			
GO:0033927 (Glucan 1, 4-alpha-maltohexaosida**Hiba! A könyvjelző nem létezik.**se activity)	0.92			
GO:0005509 (Ca ion binding)	1.00	GO:0005975 (Carbohydrate metabolic process)	1.00	GO:0005576 (Extracellular region)

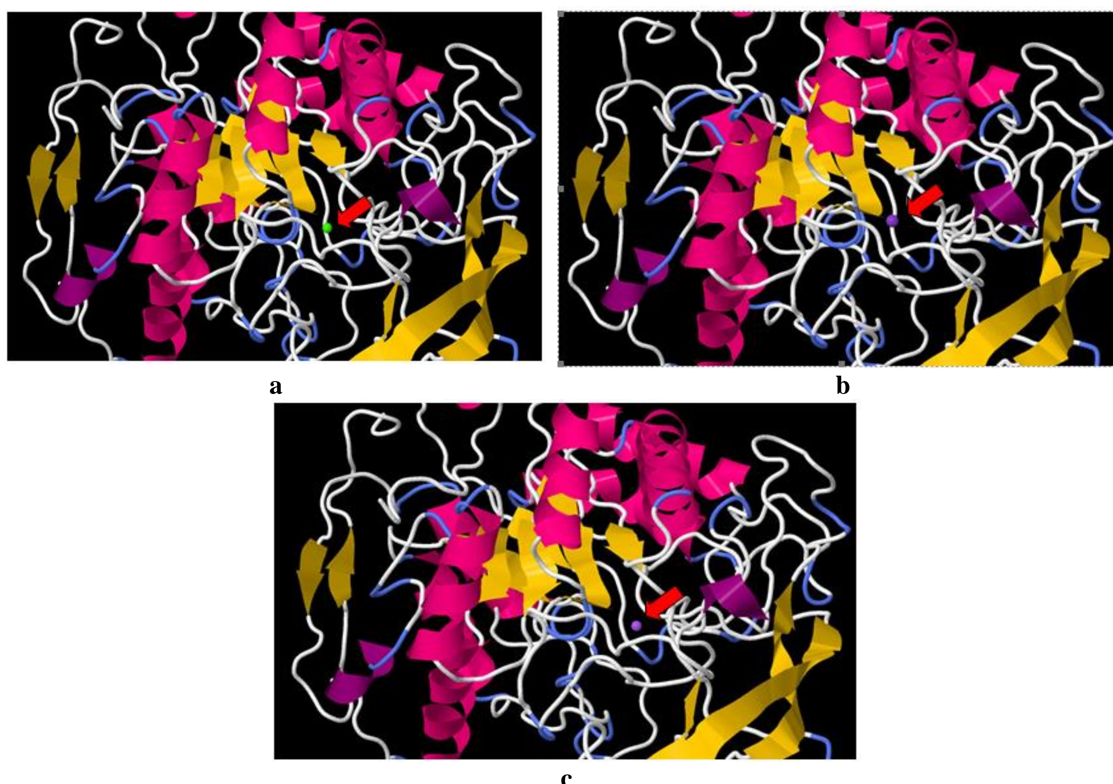

Figure 4. *Predicted ligand binding sites (a) calcium (b) sodium (c) potassium in BLA 208*

Natural selection analysis

Among the proposed methods for the detection of selection signatures in protein-coding data, the most emphasized approach is to emulsify evolutionary characteristics into codon-based models and to estimate the ratio (ω) of non-synonymous substitution (dN) to synonymous (dS). A ratio less than 1 explains the negative selection and vice versa (Vitti et al., 2013). The highest relative substitution rate was observed for A↔G

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2204_33733388
© 2024, ALÖKI Kft., Budapest, Hungary

while the lowest for C↔T (*Table 6a*). Substitution summary by residue (*Table 6b*) shows the maximum substitutions (47) between Asp↔Glu followed by Asn↔Lys (44), Tyr↔Asp (43), His↔Gln (42). It emerges that Gly and Phe are the most preferable amino acids for substitution in the amylase gene of *B. licheniformis*.

Evaluation of natural selective pressure operating on *B. licheniformis* amylase gene using SLAC, FEL, and REL analysis inferred that the gene is under purifying/negative selection.

SLAC analysis revealed 9 negatively selected sites and no positive selection sites at $p < 0.1$ with mean dN/dSratio = 0.322695.

Table 6a. *Relative substitution rates*

	A	C	G	T
A	*	0.940972	1	0.940972
C	-	*	0.940972	0.89283
G	-	-	*	0.940972
T	-	-	-	*

Table 6b. *Amino acid substitution summary in B. licheniformis alpha amylase variants*

	Leu	Ile	Met	Val	Ser	Pro	Thr	Ala	Tyr	His	Gln	Asn	Lys	Asp	Glu	Cys	Trp	Arg	Gly
Phe	9	8	4	10	7	2	6	6	18	4	3	14	2	6	5	9	3	3	5
Leu		19	14	32	10	11	4	11	3	23	19	7	9	15	5	1	15	14	11
Ile			8	19	15	1	15	10	3	3	1	29	11	7	2	2	2	3	4
Met				12	3		6	3	3		4	3	10	4	2	1	4	7	
Val					9	4	7	15	7	4	7	6	6	29	25	1	3	4	29
Ser						8	18	17	11	6	8	36	14	11	9	7	7	24	36
Pro							13	18	2	13	8	6	2	5	4		4	7	6
Thr								30	3	4	9	24	18	14	7	2	2	9	8
Ala									11	7	6	14	6	35	18	2	3	4	29
Tyr										29	6	27	8	43	8	6	8	1	6
His											42	29	3	38	3	1	2	9	11
Gln												9	23	8	10		2	11	3
Asn													44	40	11	2		7	15
Lys														10	27		5	20	5
Asp															40	3	8	7	47
Glu																2	5	8	20
Cys																	9	3	12
Trp																		16	14
Arg																			28
Gly																			

When the FEL analysis was run, 1 positively selected site and 89 negatively selected sites were found at $p < 0.1$. REL analysis indicates that no rates of non-synonymous > synonymous substitutions were deduced suggesting the influence of purifying selection at all the sites.

After running all three methods individually, integrative analysis at each codon site found that out of 551 codons, 86 were under negative selection using all three methods

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2204_33733388
© 2024, ALÖKI Kft., Budapest, Hungary

(*Table 7*). Codon 349 that encodes for glycine was the only identified site found under positive selection by FEL method only.

Table 7. *Integrative analysis using SLAC, FEL and REL methods*

Number of sites	SLAC (p < 0.1)	FEL (p < 0.1)	REL (BF > 50)	Integrated
Positively selected	0	1	0	1
Negatively selected	9	80	0	86

None of the sites was found to be influenced under positive selection by all three algorithms. Protein function includes a crucial component known as subcellular localization. There are conflicting data regarding the relationship between protein subcellular localization and evolutionary rate. The present results suggest that the alpha amylase gene in *B. licheniformis* natural variants is subjected to strong negative selective pressure. This illustrates the ongoing removal of deleterious mutation and sequence preservation pattern in alpha amylase gene of *B. licheniformis*. These results are in agreement with the findings of Liu et al. (2008), who observed that secreted proteins showed considerably higher dN/dSratio than non-secreted proteins and are certainly subject to less strict selection than cytoplasmic and nuclear proteins. The average A/S ratio of predicted extracellular proteins is 0.34, which is much greater than that of predicted nuclear and cytoplasmic proteins (Liu et al., 2008). Although the Genetic Algorithm for Recombination Detection (GARD) revealed the presence of two break points but the test for topological incongruence did not support these findings (*Fig. 5; Table 8*). Topological incongruence generally separates the recombination event from heterotachy—a process of varying site-specific gene substitution with time. It indicates the pseudo positive recombination events might be due to differences in the branch lengths or other processes such as heterotachy (Lopez et al., 2002; Delport et al., 2010).

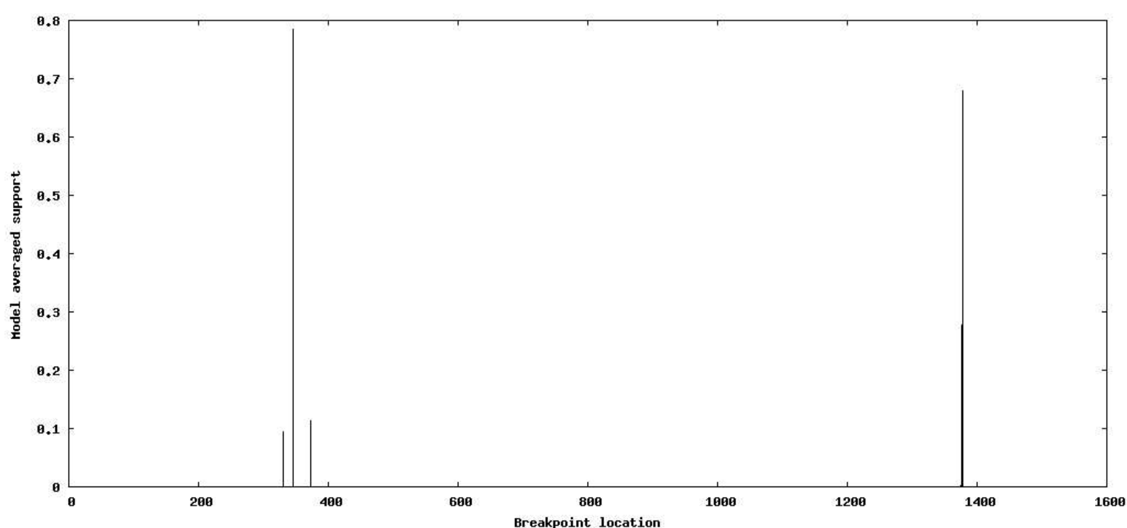

Figure 5. *GARD recombination plot*

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2204_33733388
© 2024, ALÖKI Kft., Budapest, Hungary

Table 8. *Kishino-Hasegawa test for topological incongruence*

Breakpoints	Significance
345	insignificant
1377	insignificant

Conclusions

The α-amylase gene (1313 bp; accession no, KM 27230) from *B. licheniformis* 208 was successfully amplified, sequenced, and characterized. Physico-chemical features of the studied enzyme witnessed a thermostable activity pattern with some indigenous differences. Analysis of selection signatures operating on the natural variants of *B. licheniformis* alpha amylase gene inferred that the gene is under purifying/negative selection stating its preservation during evolutionary changes.

Acknowledgments. This study was supported by a research grant from Dean faculty of Science, University of Karachi to the corresponding author.

Funding. The authors greatly acknowledge and express their gratitude to the Researchers Supporting Project number (RSP2024R335), King Saud University, Riyadh, Saudi Arabia.

Conflict of interests. All the authors declare no conflict of interest.

REFERENCES

[1] Alikhajeh, J., Khajeh, K., Ranjbar, B., Naderi-Manesh, H., Lin, Y.-H., Liu, E., Guan, H.-H., Hsieh, Y.-C., Chuankhayan, P., Huang, Y.-C., Jeyaraman, J., Liu, M.-Y., Chen, C.-J. (2010): Structure of Bacillus amyloliquefaciens α-amylase at high resolution: implications for thermal stability. – Acta Crystallogr Sect F Struct Biol Cryst Commun 66(2): 121-129.

[2] Andersson, S. G. E., Alsmark, C., Canbäck, B., Davids, W., Frank, C., Karlberg, O., Klasson, L., Antoine-Legault, B., Mira, A., Tamas, I. (2002): Comparative genomics of microbial pathogens and symbionts. – Bioinformatics. DOI: 10.1093/bioinformatics/18.suppl_2.s17.

[3] Argimón, S., Alekseyenko, A. V., DeSalle, R., Caufield, P. W. (2013): Phylogenetic analysis of glucosyltransferases and implications for the coevolution of mutans Streptococci with their mammalian hosts. – PLoS One 8. https://doi.org/10.1371/journal.pone.0056305.

[4] Arora, N., Banerjee, A. K., Mutyala, S., Murty, U. S. N. (2009): Comparative characterization of commercially important xylanase enzymes. – Bioinformation 3: 446-453. https://doi.org/10.6026/97320630003446.

[5] Asad, W., Saleem, F., Ajaz, M., Rasool, S. A. (2014): Optimization of the growth conditions for amylase production by *Bacillus licheniformis* 208 isolated from local hot springs of Karachi. – J. Chem. Soc. Pakistan 36: 366-370.

[6] Aziz T, Qadir R, Anwar F, Naz S, Nazir N, Nabi N, Cui H Lin L, Alharbi M, Alasmari AF. (2024). Optimal Enzyme-assisted Extraction of Phenolics from Leaves of Pangomia pinnata via Response Surface Methodology and Artificial Neural Networking. Applied Biochemistry and Biotechnology.. 2024, 1-8. https://doi.org/10.1007/s12010-024-04875-w

[7] Bustamante, C. D., Fledel-Alon, A., Williamson, S., Nielsen, R., Hubisz, M. T., Glanowski, S., Tanenbaum, D. M., White, T. J., Sninsky, J. J., Hernandez, R. D., Civello, D., Adams, M. D., Cargill, M., Clark, A. G. (2005): Natural selection on protein-coding genes in the human genome. – Nature 437: 1153-1157. https://doi.org/10.1038/nature04240.

© 2024, ALÖKI Kft., Budapest, Hungary

[8] Buxbaum, E. (2007): Fundamentals of Protein Structure and Function. – Springer, Cham. https://doi.org/10.1007/978-3-319-19920-7.

[9] Chakraborty, S., Khopade, A., Biao, R., Jian, W., Liu, X. Y., Mahadik, K., Chopade, B., Zhang, L., Kokare, C. (2011): Characterization and stability studies on surfactant, detergent and oxidant stable α-amylase from marine haloalkaliphilic Saccharopolyspora sp. A9. – J. Mol. Catal. B Enzym. 68: 52-58. https://doi.org/10.1016/j.molcatb.2010.09.009.

[10] 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. https://doi.org/10.1006/jmbi.2000.4025.

[11] Delport, W., Poon, A. et al. (2010): Datamonkey 2010: a suite of phylogenetic analysis tools for evolutionary biology. – Bioinformatics 26(19): 2455-2457.

[12] Dhar, D. V., Tanuj, S., Amit, P., Kumar, M. S. (2012): Insights to sequence. – Res. J. Biological Sci. 1(6): 38-42.

[13] Dong, G., Vieille, Claire, Savchenko, A., Zeikus, A. J. G. (1997): Cloning, sequencing, and expression of the gene encoding extracellular-amylase from Pyrococcus furiosus and biochemical characterization of the recombinant enzyme. – Appl. Environ. Microbiol. 63: 3569-3576.

[14] Feto, N. A. (2017): Bacillus spp. and Their Biotechnological Roles in Green Industry. – In: Islam, M. T. et al. (eds.) Bacilli and Agrobiotechnology. Springer International Publishing, Cham, pp. 143-162. https://doi.org/10.1007/978-3-319-44409-3_7.

[15] Feyertag, F., Alvarez-Ponce, D. (2001): Disulfide bonds enable accelerated protein evolution. – Mol. Biol. Evol 34: 1833-1837. https://doi.org/10.1093/molbev/msx135.

[16] Filiz, E., Koç, I. (2014): In silico sequence analysis and homology modeling of predicted beta-amylase 7-like protein in Brachypodium distachyon. – Journal of Bioscience and Biotechnology 2(1): 61-67.

[17] Hooda, V. (2011): Physiochemical, functional and structural characterization of wheat germin using in silico methods. – Curr. Res. J. Biol. Sci. 3: 35-41.

[18] Huang, Q., Liu, H., Zhang, J., Wang, S., Liu, F., Li, C., Wang, G. (2022): Production of extracellular amylase contributes to the colonization of Bacillus cereus 0-9 in wheat roots. – BMC Microbiol. 22: 205. https://doi.org/10.1186/s12866-022-02618-7.

[19] Ikai, A. (1980): Thermostability and aliphatic index of globular proteins. – Commun. J. Biochem 88: 1895-1898.

[20] Ji, Z. L., Peng, S., Chen, L. L., Liu, Y., Yan, C., Zhu, F. (2020): Identification and characterization of a serine protease from *Bacillus licheniformis* W10: A potential antifungal agent. – Int. J. Biol. Macromol. 145: 594-603. https://doi.org/10.1016/j.ijbiomac.2019.12.216.

[21] Khatoon N, Ullah N, Sarwar A, Aziz T, Alharbi M, Alshammari A. (2023). Isolation and Identification of Protease Producing Bacillus strain from Cold Climate Soil and Optimization of its Production by applying Different Fermentation Conditions. Appl Ecol Environ Res. 21(4):3391-3401

[22] King Jordan, I., Rogozin, I. B., Wolf, Y. I., Koonin, E. V. (2002): Microevolutionary Genomics of Bacteria. – Theor. Popul. Biol. 61: 435-447. https://doi.org/10.1006/tpbi.2002.1588.

[23] Kosiol, C., Anisimova, M. (2012): Selection on the Protein-Coding Genome. – In: Anisimova, M. (ed.) Evolutionary Genomics. Methods in Molecular Biology. Vol. 856. Humana Press, Totowa, NJ, pp. 113-140. https://doi.org/10.1007/978-1-61779-585-5_5.

[24] Ashokan, K. V., Mundaganur, D. S., Mundaganur, Y. D. (2011): Catalase: phylogenetic characterization to explore protein cluster – J. Res. Bioinforma. 1: 001-008.

[25] Lefébure, T., Stanhope, M. J. (2009): Pervasive, genome-wide positive selection leading to functional divergence in the bacterial genus Campylobacter. – Genome Res. 19: 1224-1232. https://doi.org/10.1101/gr.089250.108.

[26] Li, Z., Duan, X., Chen, S., Wu, J. (2017): Improving the reversibility of thermal denaturation and catalytic efficiency of *Bacillus licheniformis* α-amylase through

© 2024, ALÖKI Kft., Budapest, Hungary

stabilizing a long loop in domain B. – PLoS ONE 12. https://doi.org/10.1371/journal.pone.0173187.
[27] Liu, J., Zhang, Y., Lei, X., Zhang, Z. (2008): Natural selection of protein structural and functional properties: a single nucleotide polymorphism perspective. – Genome Biol. 9: R69. https://doi.org/10.1186/GB-2008-9-4-R69.
[28] Liu, J., Zhang, Z., Zhu, H., Dang, H., Lu, J., Cui, Z. (2011): Isolation and characterization of α-amylase from marine Pseudomonas sp. K6-28-040. – African J. Biotechnol. 10: 2733-2740. https://doi.org/10.5897/AJB10.2042.
[29] Lopez, P., Casane, D., Philippe, H (2002): Heterotachy, an important process of protein evolution. – Molecular Biology and Evolution 19(1): 1-7.
[30] Lugani, Y., Sooch, B. S. (2017): In silico characterization of cellulases from genus Bacillus. – Int J Cur Res Rev 09. https://doi.org/10.7324/IJCRR.2017.9136.
[31] Mukherjee, A. (2012): The putative synaptotagmin protein encoded by the SYT1 gene of the picoplanktonic alga Micromonas is a novel member of C2-domain containing proteins: evidence from in silico characterization and homology modeling International Journal of Biosciences (IJB). – Int. J. Biosci 2012: 36-52.
[32] Nazmi, A. R., Reinisch, T., Hinz, H.-J. (2008): Calorimetric studies on renaturation by CaCl2 addition of metal-free a-amylase from *Bacillus licheniformis* (BLA). – J. Therm. Anal. Calorim. 91: 141-149.
[33] Nguyen, Q. D., Rezessy-Szabó, J. M., Claeyssens, M., Stals, I., Hoschke, Á. (2002): Purification and characterisation of amylolytic enzymes from thermophilic fungus Thermomyces lanuginosus strain ATCC 34626. – Enzyme Microb. Technol. 31: 345-352. https://doi.org/10.1016/S0141-0229(02)00128-X.
[34] Novichkov, P. S., Wolf, Y. I., Dubchak, I., Koonin, E. V. (2009): Trends in prokaryotic evolution revealed by comparison of closely related bacterial and archaeal genomes. – J. Bacteriol. 91: 65-73. https://doi.org/10.1128/JB.01237-08.
[35] Orsi, R. H., Sun, Q., Wiedmann, M. (2008): Genome-wide analyses reveal lineage specific contributions of positive selection and recombination to the evolution of Listeria monocytogenes. – BMC Evol. Biol. 8: 233. https://doi.org/10.1186/1471-2148-8-233.
[36] Petersen, L., Bollback, J. P., Dimmic, M., Hubisz, M., Nielsen, R. (2007): Genes under positive selection in Escherichia coli. – Genome Res. 17: 1336-1343. https://doi.org/10.1101/gr.6254707.
[37] Pramanik, S. K., Mahmud, S., Paul, G. K., Jabin, T., Naher, K., Uddin, M. S., Zaman, S., Saleh, M. A. (2021): Fermentation optimization of cellulase production from sugarcane bagasse by Bacillus pseudomycoides and molecular modeling study of cellulase. – Curr. Res. Microb. Sci. 2: 100013. https://doi.org/10.1016/j.crmicr.2020.100013.
[38] Rana, N., Walia, A., Gaur, A. (2013): α-Amylases from microbial sources and its potential applications in various industries. – Natl. Acad. Sci. Lett. https://doi.org/10.1007/s40009-012-0104-0.
[39] Rath, A., Glibowicka, M., Nadeau, V. G., Chen, G., Deber, C. M. (2009): Detergent binding explains anomalous SDS-PAGE migration of membrane proteins. – Proc. Natl. Acad. Sci. U.S.A. 106: 1760-1765. https://doi.org/10.1073/pnas.0813167106.
[40] Roy, A., Zhang, Y. (2012): Protein structure prediction. – ELS. https://doi.org/10.1002/9780470015902.a0003031.pub2.
[41] Shahhoseini, M., Ziaee, A. A., Pourbabai, A. A., Ghaemi, N., Declerck, N. (2005): A natural variant of *Bacillus licheniformis* α-amylase isolated from flour mill wastewaters sheds light on the origin of high thermostability. – J. Appl. Microbiol. 98: 24-32. https://doi.org/10.1111/j.1365-2672.2004.02407.x.
[42] Sivakumar, K., Balaji, S. (2007): In silico characterization of antifreeze proteins using computational tools and servers. – J. Chem. Sci 119: 571-579.
[43] Suzuki, H., Lefébure, T., Hubisz, M. J., Bitar, P. P., Lang, P., Siepe, A., Stanhope, M. J. (2011): Comparative genomic analysis of the Streptococcus dysgalactiae species group:

© 2024, ALÖKI Kft., Budapest, Hungary

gene content, molecular adaptation, and promoter evolution. – Genome Biol. Evol. 3: 168-185. https://doi.org/10.1093/gbe/evr006.

[44] Tamura, K., Nei, M. (1993): Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. – Mol. Biol. Evol. 10(3): 512-526.

[45] Ullah N, Rehman MU, Sarwar A, Nadeem M, Nelofer R, Shakir HA, Irfan M, Idrees M, Naz S, Nabi G, et al. (2022). Purification, Characterization, and Application of Alkaline Protease Enzyme from a Locally Isolated Bacillus cereus Strain. Fermentation; 8(11):628. https://doi.org/10.3390/fermentation8110628

[46] Ullah, N., Mujaddad-ur-Rehman, M., Sarwar, A. et al. (2023). Effect of bioprocess parameters on alkaline protease production by locally isolated Bacillus cereus AUST-7 using tannery waste in submerged fermentation. Biomass Conv. Bioref. 2023, 1-9. https://doi.org/10.1007/s13399-023-04498-x

[47] Vitti, J. J., Grossman, S. R., Sabeti, P. C. (2013): Detecting natural selection in genomic data. – Annu. Rev. Genet. https://doi.org/10.1146/annurev-genet-111212-133526.

[48] Wayengera, M., Kajumbula, H., Kaddu-Mulindwa, D., Byarugaba, W., Olobo, J. (2009): Proteomics of Marburg and Ebola glycoproteins: insights into their physicochemical similarities and irregularities. – Afr J Biotechnol 8(17): 4025-31.

[49] Wu, X., Wang, Y., Tong, B., Chen, X., Chen, J. (2018): Purification and biochemical characterization of a thermostable and acid-stable alpha-amylase from *Bacillus licheniformis* B4-423. – Int. J. Biol. Macromol. 109: 329-337. https://doi.org/10.1016/j.ijbiomac.2017.12.004.

[50] Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S., Madden, T. L. (2012): Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. – BMC Bioinformatics 13: 134. https://doi.org/10.1186/1471-2105-13-134.

[51] Yeon Hwang, K., Kyu Song, H., Chang, C., Lee, J. (2022): Crystal structure of thermostable α-amylase from *Bacillus licheniformis* refined at 1.7 Å resolution. – Mol Cells 7(2): 251-258.

[52] Zhang, Y. (2008): I-TASSER server for protein 3D structure prediction. – BMC Bioinformatics 9. https://doi.org/10.1186/1471-2105-9-40.

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2204_33733388
© 2024, ALÖKI Kft., Budapest, Hungary