BIOINFORMATION
Discovery at the interface of physical and biological sciences

open access www.bioinformation.net Hypothesis

Volume 10(7)

Phylogenetic and Comparative Sequence Analysis of Thermostable Alpha Amylases of kingdom Archea, Prokaryotes and Eukaryotes

Tayyaba Huma[1]*, Arooma Maryam[1], Shahid ur Rehman[2], Muhammad Tahir ul qamar[1], Tayyaba Shaheen[1], Asma Haque[1] &Bushra Shaheen[1]

[1]Department of Bioinformatics and Biotechnology, Government College University (GCUF), 38000, Faisalabad, Punjab, Pakistan; [2]Department of Poultry Sciences, University of Agriculture (UAF), 38000, Faisalabad, Punjab, Pakistan; Tayyaba Huma – Email: tayyabashahbaz@gmail.com; Phone: +92300-6606933; *Corresponding author

Received June 26, 2014; Accepted June 27, 2014; Published July 22, 2014

Abstract:
Alpha amylase family is generally defined as a group of enzymes that can hydrolyse and transglycosylase α-(1, 4) or α-(1, 6) glycosidic bonds along with the preservation of anomeric configuration. For the comparative analysis of alpha amylase family, nucleotide sequences of seven thermo stable organisms of Kingdom Archea i.e. *Pyrococcus furiosus* (100-105ºC), Kingdom Prokaryotes i.e. *Bacillus licheniformis* (90-95ºC), *Geobacillus stearothermophilus* (75ºC), *Bacillus amyloliquefaciens* (72ºC), *Bacillus subtilis* (70ºC) and *Bacillus KSM K38* (55ºC) and Eukaryotes *i.e. Aspergillus oryzae* (60ºC) were selected from NCBI. Primary structure composition analysis and Conserved sequence analysis were conducted through Bio Edit tools. Results from BioEdit shown only three conserved regions of base pairs and least similarity in MSA of the above mentioned alpha amylases. In Mega 5.1 Phylogeny of thermo stable alpha amylases of Kingdom Archea, Prokaryotes and Eukaryote was handled by Neighbor-Joining (NJ) algorithm. Mega 5.1 phylogenetic results suggested that alpha amylases of thermo stable organisms i.e. *Pyrococcus furiosus* (100-105ºC), *Bacillus licheniformis* (90-95ºC), *Geobacillus stearothermophilus* (75ºC) and *Bacillus amyloliquefaciens* (72ºC) are more distantly related as compared to less thermo stable organisms. By keeping in mind the characteristics of most thermo stable alpha amylases novel and improved features can be introduced in less thermo stable alpha amylases so that they become more thermo tolerant and productive for industry.

Key words: Alpha amylase, Kingdom archea, Thermo stable, Phylogenetic analysis

Background:
Alpha amylase family is generally defined as a group of enzymes that can hydrolyse and transglycosylase α-(1, 4) or α-(1, 6) glycosidic bonds along with the retention of anomeric configuration. They have a conserved $(\alpha/\beta)_8$TIM barrel fold, catalytic residues and substrate binding residues in three dimensional structures of all family members **[1, 2, 3].** Three dimensional structures of alpha amylases consist of three domains namely Domain A, Domain B and Domain C. Domain A is composed of eight parallel β-strands that are surrounded by eight helices. Unique arrangement of these eight parallel beta strands and alpha helices form a $(\alpha/\beta)_8$TIM barrel fold in three dimensional structure of alpha amylases. It originates from third β-strand and third α-helix of catalytic domain. It present between A and C domain and composed of many β-strands and few α-helices **[2].** Domain B is important for enzyme specificity and substrate binding. Domain C of alpha amylases also composed of β sheets and act as an independent domain. It stabilizes the catalytic domain by protecting the hydrophobic residues of catalytic domain from solvent and also

plays role in substrate binding. Alpha amylases also supplemented with additional domains i.e. domain D and E in some members of alpha amylases **[4].** Four short conserved sequence regions cluster at Domain A and play significant role in structure conformation of particular enzyme. These regions have seven fully conserved residues in their amino acid sequence while their positions in the sequence may vary among different alpha amylases. Amylases find potential application in a number of industrial processes such as in the food, fermentation, textiles and paper industries. Microbial amylases have successfully replaced the chemical hydrolysis of starch in starch-processing industries. These amylases also reduce viscosity and haze formation in sugar syrups and ice juices respectively **[5, 6].** Detergent industry is the primary consumers of enzymes, in terms of both volume and value. The utilization of amylases in 90% of liquid detergents formulations enhances ability of detergents to remove tough starch stains and making the detergent environmentally safe **[7].** They would be potentially useful in the pharmaceutical and fine chemicals industries if enzymes with suitable properties could be prepared **[8].** The spectrum of amylase application has widened in many other fields, such as clinical, medical, and analytical chemistry **[9].**

Biological systems are natural producers of alpha amylases. Even though alpha amylases are widely produced from animals and plants sources but scientists normally prefer to isolate them from microbial sources. Microbial derived alpha amylases are easy to modify and optimize **[10].** Fungal and bacterial alpha amylases especially *Bacillus species* alpha amylases are of special concern because of their significant thermo stability. Among *Bacillus specie, Bacillus licheniforms, Bacillus amyloliquefaciens, Bacillus subtilis and Bacillus stearothermophilus* are well-known producers of commercial thermo stable alpha amylases **[3].** Alpha amylase of *Pyrococcus furiosus* from kingdom Archea is also hyperthermophilic alpha amylase **[11].** Phylogenetic methods for comparative analysis of DNA and protein sequences become ever more important with the rapid accumulation of molecular sequence detail. Phylogenetic analysis is helpful to find out the nature and extent of selective forces that shape the evolution of genes and species. Since no such α-amylase was characterized and no evolutionary relationship between all kingdom was reported to date, thus in the present study alpha amylase from seven microbial species were selected and comparative analysis of molecular sequence data was performed for reconstructing the evolutionary histories of species.

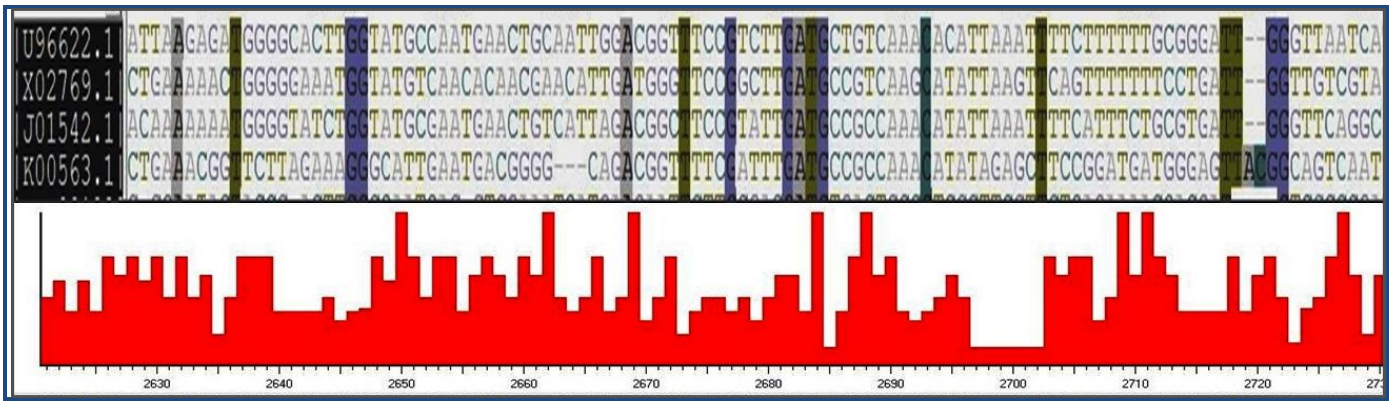

Figure 1: Example of Multiple Sequence Alignment and Information Scan of Nucleotide Sequences of Archea: Prokaryote: Eukaryotes α-amylases.gb File generated by Bio Edit (Archea i.e. *Pyrococcus furiosus (100-105°C); Prokaryotes i.e. Bacillus licheniformis (90-95°C), Geobacillus stearothermophilus (75°C), Bacillus amyloliquefaciens (72°C), Bacillus subtilis (70°C), Bacillus KSM K38 (55°C)* and *Eukaryotes i.e. Aspergillus oryzae (60°C).*

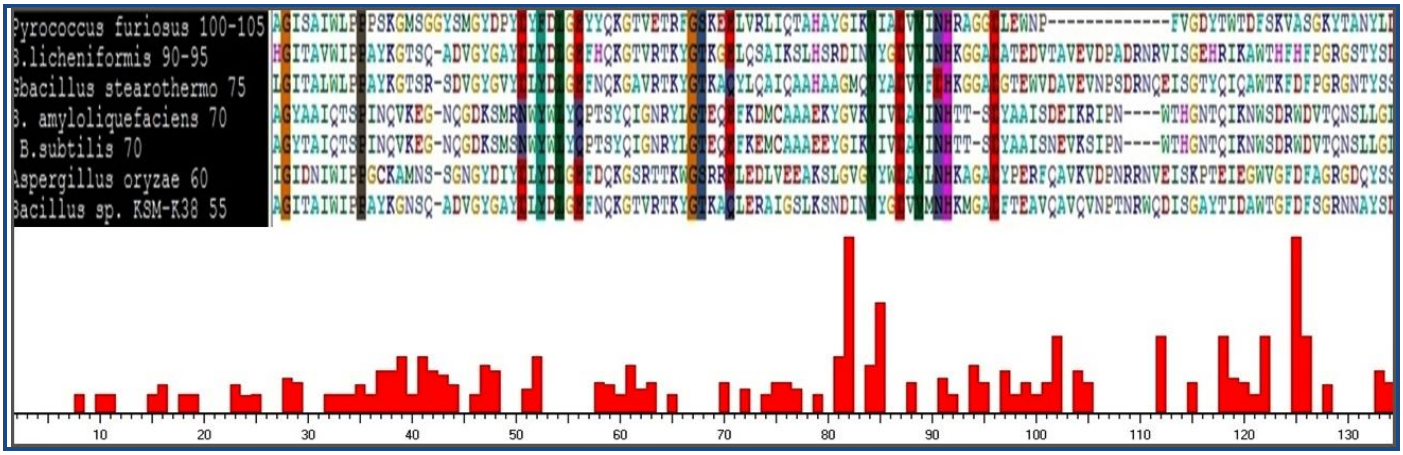

Figure 2: Example of Multiple Sequence Alignment and Information Scan of Protein Sequences of Archea: Prokaryote: Eukaryotes α-amylases.gb File. generated by Bio Edit. (Archea i.e. *Pyrococcus furiosus (100-105°C); Prokaryotes i.e. Bacillus licheniformis (90-95°C), Geobacillus stearothermophilus (75°C), Bacillus amyloliquefaciens (72°C), Bacillus subtilis (70°C), Bacillus KSM K38 (55°C)* and *Eukaryotes i.e. Aspergillus oryzae (60°C).*

ISSN 0973-2063 (online) 0973-8894 (print)
Bioinformation 10(7):443-448 (2014)

Methodology:

Primary Structure Information

Sequence Retrieval

Nucleotide and protein fasta sequences of alpha amylase 3.2.1.1 were retrieved from NCBI and Uniprot databases **Table 1 (see supplementary material).** Out of these seven sequences, four sequences belong to *Bacillus* specie and one sequence was from *Geobacillus* specie. On the whole five sequences were from kingdom prokaryotes. Alpha amylase of *Aspergillus oryzae* was from kingdom eukaryotes and *Pyrococcus furiosus* was from kingdom archea.

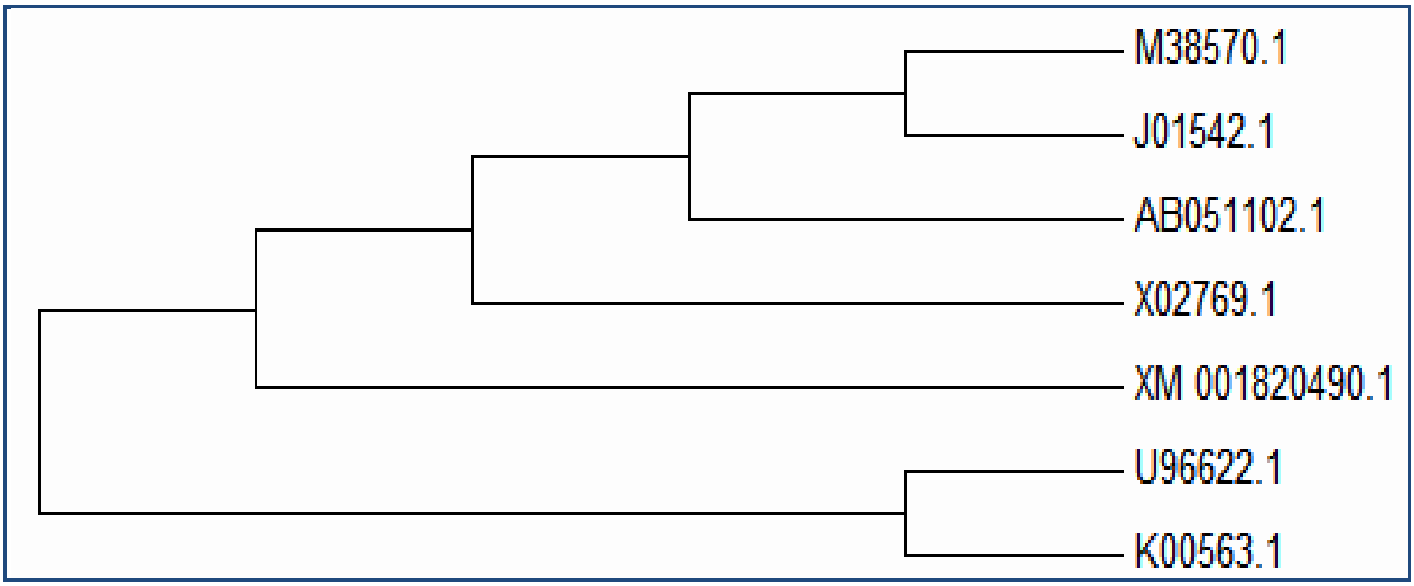

Figure 3: Phylogenetic Tree generated by Mega 5.1 results with branch lengths of Kingdom Archea (Pyrococcus furiosus), Kingdom Prokaryotes (*Bacillus licheniformis, Bacillus amyloliquefaciens, Bacillus subtilis, Bacillus KSM-K38, Geobacillus stearothermophilus*) and Eukaryotes (*Aspergillus oryzae*) alpha amylase.

Sequence Analysis in BIOEDIT Tools

Multiple sequence alignment was carried out for both DNA and protein sequences of above mentioned alpha amylases through ClustalW. Multiple sequence analysis was done to assess sequence conservation of each individual nucleotide and amino acids. Temperature based alignment was carried out and important residues and regions were defined in multiple sequence alignments.

Primary structure composition analysis and Conserved Sequence Analysis

Nucleotide and Protein Sequence alignments were manipulated to find out the conserved regions of the sequence alignments. Conserved regions were found by through BioEdit. Consensus sequence was obtained at the end of multiple sequence alignment and it was comprised of the all identical residues of the organisms given in multiple sequence alignment. Through Information Scan a graphical plot was obtained which was a clear description of multiple sequence alignments. Multiple sequence alignment results were further studied in later *in silico* analysis of two dimensional and three dimensional alpha amylases structures.

Phylogenetic Analysis of Alpha Amylase

For this purpose phylogenetic analysis was performed in Mega 5.1 tool. This tool was preferred because it uses heuristic approach and progressive alignment method, aligns the most similar sequences having best alignment score first then progressively add more distant groups of sequences to the alignment until an optimal global alignment is obtained. ClustalW algorithm computes a rough distance matrix between each pair of sequences based on pairwise sequence alignment scores and Tree Construction was done through Neighbor-Joining method. Neighbour-joining method is a distance based method that defines the distance on the basis of mismatches. It generally provides a close-to-optimal result by applying greedy algorithms. As data set for the study was small so Neighbour Joining method was preferred for generation of initial evolutionary hypotheses with maximum accuracy. Phylogenetic tree reconstruction was done for verification through Maximum likelihood method as well. As most of our sequences belong to distantly related sequences and only seven sequences were in our data set so it was computationally inexpensive to perform Maximum Likelihood method fastDNAml tool. Same dataset was also verified through BioEdit software by applying Neighbor-Joining/UPGMA method.

Results & Discussion:

Primary Structure Analysis of Archea, Prokaryotes and Eukaryotes Alpha Amylases

Archea: Prokaryote: Eukaryote.gb files with nucleotide and protein sequences of alpha amylases from *Pyrococcus furiosus* (100-105^{o}C), *Bacillus licheniformis* (90-95^{o}C), *Geobacillus stearothermophilus* (75^{o}C), *Bacillus amyloliquefaciens* (72^{o}C), *Bacillus subtilis* (70^{o}C), *Bacillus KSM K38* (55^{o}C) and *Aspergillus oryzae* (60^{o}C) were analysed through multiple sequence alignment in BioEdit software. Least similarity was observed in MSA results of above mentioned alpha amylases as given in **Figure 1 & Figure 2**. Only three conserved regions of base pairs were observed while no protein based conserved region was present in the multiple sequence alignment analysis of Prokaryote and Eukaryote alpha amylases **Table 2 (see supplementary material).** While single nucleotide and amino

acid identities were evident, this shows that some common base pairs and amino residues are present among them. Consensus protein sequence of Archea, Prokaryotes and Eukaryotes alpha amylases is Trp59, Glyc81, Pro90, Leu111, Gly125, Glu32I, Val144, Asp147, Asp157, Ile418 and Gly424.

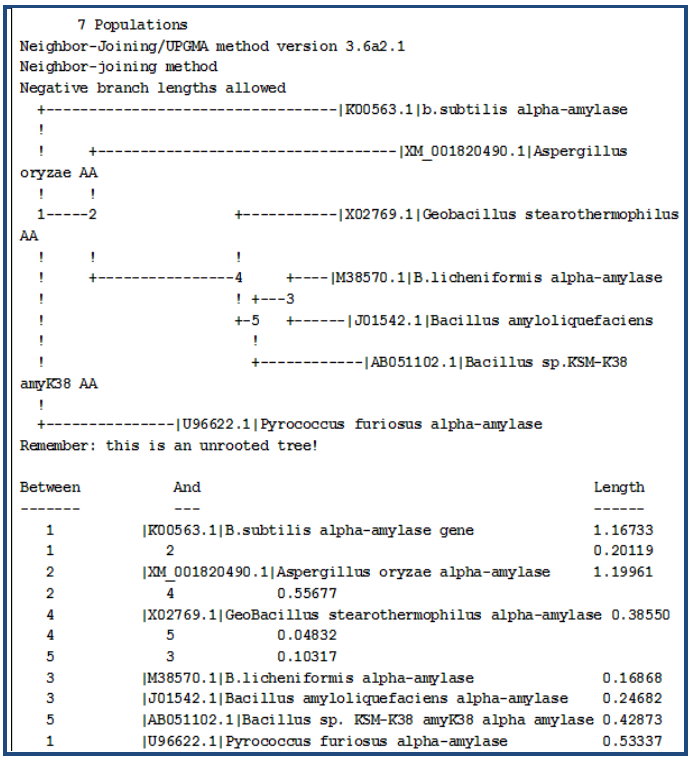

Figure 4: Phylogenetic Tree generated by BioEdit results with branch lengths of Kingdom Archea (*Pyrococcus furiosus*), *Kingdom Prokaryotes (Bacillus licheniformis, Bacillus amyloliquefaciens, Bacillus subtilis, Bacillus KSM-K38, Geobacillus stearothermophilus*) and Eukaryotes (*Aspergillus oryzae*) alpha amylase .

Phylogenetic Relationship Analysis

Distance based phylogenetic analysis was performed through Mega 5.1 software. Data set for phylogenetic analysis includes nucleotide sequences of alpha amylase gene from seven species i.e. *Pyrococcus furiosus* (U96622.1), *Bacillus licheniformis* (M38570.1), *Bacillus amyloliquefaciens* (J01542.1), *Bacillus stearothermophilus* (X02769.1), *Aspergillus oryzae* (XM_001820490.1) and *Bacillus subtilis* (K00563.1). In this analysis Neighbor-Joining (NJ) algorithm was used to predict phylogeny of these organisms. This algorithm employs a matrix of pairwise distances estimated under the Tamura and Nei model for nucleotide sequences **[12, 13, 14].** In this model, maximum composite likelihood approach is used that compares two sequences at a time and calculates the number of base substitutions i.e. single nucleotide polymorphism and insertion deletion events within the two aligned sequences. An estimate of these calculated transitions and transversions generated by all the possible sequence pairs comparison is known as pairwise distance **[15].** With the help of the calculated pairwise distances a matrix is generated which is helpful in construction of an un-rooted tree. The optimal tree with the sum of branch length = 3.90657519 is shown in **Figure 3.** Total 1547 positions in the final dataset were analysed for tree construction.

In the present study phylogenetic results shown an unrooted tree which takes alpha amylase gene from *Pyrococcus furiosus* as a base of analysis and as an outgroup taxa shown in **Figure 3**. Out group taxa is actually most distant related specie from which other species are evolving. Here *Pyrococcus furiosus* alpha-amylase gene is behaving as a reference group for the determination of evolutionary relationship among other species. Alpha amylase from *Bacillus subtilis* has 41% sequence similarity and closely related with *Pyrococcus furiosus* alpha-amylase gene as compared alpha amylases of other species. From the out group node (1) a bifurcating branch of length 0.20119 originates whose one leaf shares evolutionary distance of 1.1996 and represents *Aspergillus oryzae* alpha amylase gene as recent descendent of *Pyrococcus furiosus* and *Bacillus subtilis* alpha amylase genes as it shares 39.9% similarity with both the sequences. While the second leaf is sharing evolutionary distance of 0.55677 with formation of a clade (node 4) which shows that Geo*bacillus stearothermophilus* alpha-amylase evolves from its ancestor *Aspergllius oryzae* alpha amylase and has distant relationship with the genes of *Pyrococcus furiosus* and *Bacillus subtilis* alpha amylase. Sequence similarity results, phylogenetic tree speciation events and branch lengths show that *Geobacillus stearothermophilus* alpha-amylase gene is a common ancestor of other *Bacillus* species i.e. *Bacillus sp. KSM-K38*, *Bacillus licheniformis and Bacillus amyloliquefaciens* alpha amylase sequences. They share more similarity among each other as compared to *Aspergllius oryzae* alpha amylase and *Pyrococcus furiosus* alpha amylase. Alpha amylases of *Bacillus licheniformis* and *Bacillus amyloliquefaciens* are sister clades having 64.5% sequence similarity between them and observed as most closely related operational taxonomic units (OTUs). *Bacillus licheniformis* alpha amylase has 52% and 40% nucleotide sequence identity with *Bacillus KSM K38* alpha amylase and *Bacillus subtilis* alpha amylase respectively.

Alikhajeh *et al.,* (2010) also reported in their study that *Bacillus licheniformis* alpha amylase and *Bacillus amyloliquefaciens* alpha amylase are more closely related to each other than *Bacillus subtilis* alpha amylase and *Bacillus KSM K38* alpha amylase. While reasonable homology of *Bacillus KSM K38* alpha amylase (<63% identity) to alpha amylases from other *Bacillus species* e.g. *Bacillus amyloliquefaciens, Bacillus licheniformis* was reported by Suvd et al., (2001) and Hagihara et al., (2001) which is also supported by the phylogenetic tree in **Figure 4** . Jorgensen et al., 1997 also reported moderate homology between *Pyrococcus furiosus* alpha amylase and *Bacillus licheniformis* alpha amylases that means both are not closely related ones. Distant relationship of both is clearly depicted in the phylogenetic tree. This phylogenetic tree is according to the sequence identities given above in primary structure analysis. Same dataset was also verified through BioEdit software by applying Neighbor-Joining/UPGMA method. Results of BioEdit verify the phylogenetic tree given by Mega 5.1 software.

Conclusion:

From the phylogenetic analysis it is concluded that alpha amylases from Bacillus subtilis and Aspergillus oryzae having optimum activity at 70ºC and 60ºC respectively are closest descendants of hyperthermophilic Pyrococcus furiosus alpha amylase (Opt100-105 ºC), as compared to other Prokaryotes i.e. Geobacillus stearothermophilus (Opt 75ºC), Bacillus licheniformis (Opt 90-95ºC) and Bacillus amyloliquefaciens

(Opt 72°C). As alpha amylases from Bacillus subtilis and Aspergillus oryzae having optimum activity up to 70°C are sharing more sequence similarity with thermo-tolerant Pyrococcus furiosus alpha amylase, their primary structures can be exploited on the basis of thermo-tolerant Pyrococcus furiosus alpha amylase to alter their intrinsic stability. For designing hyper thermo stable alpha amylases from less thermo stable alpha amylases further comparative sequence and structure based study would be required to elucidate mechanism of thermostability.

References:

[1] Van der Maarel MJ *et al. J Biotechnol* 2002 **94**: 137 [PMID: 11796168]

[2] De Souza PM & de Oliveira Magalhaes F, *Braz J Microbiol*. 2010 **41**: 850 [PM1D: 24031565]

[3] Aiyer PV, *African J Biotechnology*. 2005 **4**: 1525

[4] Ohdan K *et al. Appl Environ Microbiol*. 2002 **66**: 3058 [PM1D: 10877806]

[5] Gomes I *et al. Bioresour Technol*. 2003 **90**: 207 [PMID: 12895565]

[6] Kirk O *et al. Curr Opin Biotechnol*. 2000 **13**: 345 [PM1D: 12323357]

[7] Mitidieri S *et al. Bioresours Technol*. 2006 **97**: 1217 [PMID: 16112858]

[8] Prakash O & Jaiswal N, *Appl Biochem Biotechnol*. 2010 **160**: 2401 [PMID: 19763902]

[9] Mukherjee AK, *J Biochem Engineering* 2009 **43**: 149

[10] Arikan B, *Bioresour Technol*. 2008 **99:** 3071 [PMID: 17689242]

[11] Reddy NS *et al. African J Biotechnol*. 2003 **2**: 645

[12] Saitou N & Nei M, *Mol Biol Evol*. 1987 **4**: 406 [PMID: 3447015]

[13] Jones DF & Foulds LM, *Mathematical Bioscience* 1992 **53**: 131

[14] Tamura K *et al. Proc Natl Acad Sci U S A*. 2004 **101**: 11030 [PMID: 15258291]

[15] Kumar S *et al. Brief Bioinform*. 2004 **5**: 150 [PMID: 15260895]

Edited by P Kangueane

Citation: **Huma *et al.*** Bioinformation 10(7): 443-448 (2014)

BIOINFORMATION

open access

Supplementary material:

Table 1: Sequential information of alpha amylase (3.2.1.1) from *Pyrococcus furiosus, Bacillus licheniformis, Bacillus amyloliquefaciens, Geobacillus stearothermophilus, Bacillus subtilis, Bacillus KSM*-K38 and *Aspergillus oryzae* along with their sequence lengths.

Source Of Alpha Amylase 3.2.1.1	Nucleotide Sequence Id	Protein Sequence Id	References
Archea*: Pyrococcus furiosus*	U96622.1 1740bp	O08452 460aa	Park *et al.*, 2013
Prokaryotes*: Bacillus licheniformis*	M38570.1 1948bp	C0LZX2 512aa P06278 512aa	Yuuki *et al.*, 1985
Prokaryotes:*Geobacillus stearothermophilus*	X02769.1 2169bp	PO6279 549aa	Nakajima *et al.*, 1985
Prokaryotes:*Bacillus amyloliquefaciens*	J01542.1 2084bp	K2IAF9 659aa	Lee *et al.*, 2012
Prokaryotes:*Bacillus subtilis*	K00563.1 2294bp	A8W7JI 659aa P00691 659aa	Xin *et al.*, 2007
Prokaryotes:*Bacillus KSM K38*	AB051102.1 1759bp	Q93I48 501aa	Hagihara *et al.*, 2001
Eukaryotes*: Aspergillus oryzae*	XM_001820490.1 1650bp	XP_001820542.1 549aa	Zhao *et al.*, 2012

Table 2: Nucleotide Based Conserved Regions in Pairwise Sequence Alignment of Archea: Prokaryotes: Eukaryotes α-amylases.gb File. (Archea i.e. *Pyrococcus furiosus* (100-105ºC); Prokaryotes i.e. *Bacillus licheniformis* (90-95ºC), *Geobacillus stearothermophilus* (75ºC), *Bacillus amyloliquefaciens* (72ºC), *Bacillus subtilis* (70ºC), *Bacillus KSM K38* (55ºC) and Eukaryotes i.e. *Aspergillus oryzae* (6Oº C).

Conserved Regions	Position	Sequence	Segment Length
Region1	2265 to 2267	CAA	3
Region2	2780 to 2783	GATG	4
Region3	3004 to 3006	ATA	3