ISSN: 0976-3104
REGULAR ISSUE
Amara

THE IIOAB JOURNAL

RESEARCH ARTICLE OPEN ACCESS

PROTEASES AND THEIR ADAPTATION

Amro Abd Al Fattah Amara[1]*

[1]*Prof. Department of Protein Research, Genetic Engineering and Biotechnology Research Institute, City for Scientific Research and Technology Applications, Universities and Research Centre District, New Borg El-Arab, P.O. Box 21934, Alexandria, EGYPT*

ABSTRACT

The amino acids constituents of the various proteins play a crucial role in their structures/functions. The proteases variations are well explained due to their differences in their of amino acids constituents. However, such differences are not explained at DNA level. But, it is assumed that, at DNA level, much variation could be seen and many factors could be responsible for such differences. However, which factor(s) could affect more than the other are not yet elucidated. One factor could be the differences in environmental conditions, because the thermophilic proteases are existed in thermophilic environment, but mesophilic ones existed in their favorite habitat. Such genetic changes and subsequent changes in polypeptide chains are due to the environmental adaptation. Here, proteases coding twelve DNAs (nucleotides) and their translated products (amino acids) from different microbes adopted in different environmental conditions were analyzed using a combined bioinformatics approach. The results show that, there is a clear correlation between the nucleotides and amino acid constituents of different proteases therefore, proving a link between the microbes' types and environments they are adopted in. This study opens a new window and warrens to give equal attention to DNA in understanding the adaptation of microbes in various environmental conditions

Received on: 02nd-April-2012
Revised on: 07th-May-2012
Accepted on: 25th– Apr-2012
Published on: 8th –May-2012

KEY WORDS

Nucleotides, Amino acids, Protease, Protein Model, Comparative analysis

***Corresponding author:** Email: amroamara@web.de Tel: +00201010777905; Fax: +2034593423

[I] INTRODUCTION

The diversity is a common criterion within microbes [1]. Such diversity gives each microbe a fingerprint, which makes it different somehow from the other ones [2]. Within the same species there is a big level of diversity can be found particularly within strains existed in different habitats [3]. This is mainly because of unicellular microbes reflect clearly any genetic variation, while they follow mostly the role of "Single cell with single copy of each gene; gives a single protein". That protein gives (mostly) single function [1, 4]. Unicellular microbes have different routes of gaining new genes including, transformation, transduction, hyperdization, transposones elements, phage infections and mutations [5]. As an example, which could explain the dynamics of gene transfer, β-lactamase is given [1]. Another example is the enterohemorrhagic *E. coli* isolates containing virulence plasmids and pathogenicity islands similar to those found in *Shigella spp* [6]. In higher genera; chromosomal recombination, mitosis and sexual reproduction are other sources of variation. Lower, microbes could also do DNA recombination when big DNA part (might be transferred from dead microbes) or homologues DNA fragment inter to the cytoplasm and find a place (with special sequence) to do recombination. The elevation of the antibiotic resistance in sensitive microbes is a clear example. [1, 6]. Microbial resistance teaches us that microbes have uncommon ability to survive depending on one of the mechanisms of the different types of gene(s) transformation. The DNA is the code for the protein, which mediated through the synthesis of intermediate macromolecules, the RNA. Changes in the DNA will be reflected on the protein and will be affecting on its original structure/functions. Mutagenesis, (particularly the site directed mutagenesis), the protein engineering and the enzyme activity (and many others) prove that [7, 8]. The different DNA and protein databases could add some points to such prove further, such as changes in the DNA and protein sequences. Particularly the differences within the same type (class) of enzyme on the strain level within a certain species will be more interesting. Because, such differences will be able to reflect the role of both of the amino acids and the nucleotides. The protein primary structure, could explain the structures/functions, because it is the main macromolecules, which do the function. However, the DNA can be also an interesting subject for investigating such variation within the same class or within different classes [9]. That because, two similar proteins did not mean two similar genes. Therefore, regarding the diversity, the DNA will be the best choice for detecting minor variation. However, study the DNA variation due to different functions or adaptation was neglected particularly if one relating such variation in protein functions to their environmental adaptation. Alternatively, this study investigates if there is a relation between the DNA protection (from any unusual conditions such as the environmental

conditions) and the protein behavior. In this study I try to put a spot on the DNA as an element, could side by side with its protein explain the variation within similar proteins and different function as a result for different environmental and adaptation factors.

[II] MATERIALS AND METHODS

2.1. The used protein and DNA sequences

Twelve protein and DNA sequences have been collected from BLAST (NIH) database and represent: (1) *Bacillus sp*. L010 serine alkaline mesophilic protease precursor (sprD) KC153302.1 [10]; (2) Uncultured bacterium protease for serine mesophilic protease (Antarctic costal sediment) FM163400.1 [11]; (3) *Pseudoalteromonas sp*. SM9913 cold-adapted halophilic subtilase serine protease IFO 3455 (AP) (deep-sea psychrotolerant bacterium) MCP-03 [12]; (4) *Pseudomonas aeruginosa* mesophilic alkaline metalloprotein proteinase-D87921.1 [13]; (5) *Flavobacterium indicum* GPTSA100.9 (hot spring water) [14]; (6) Antarctic psychrotroph *B. subtilis* (TA41) sub gene for subtilisin-X63533.1 [15]; (7) AY028615. 1 *Bacillus* (*Geobacillus*) *stearothermophilus* thermophilic alkaline protease [16]; (8) *B. sp*. Subtilisin-like thermophilic serine SG-1-1101501000595 [17]; (9) U31759. 1 *Thermoactinomyces sp*. thermostable alkaline protease [18]; (10) HM192828. 1 *Laceyella sacchari* strain DSM 43353 thermitase [19]; (11) *Bacillus thuringiensis str*. Thermitase Al Hakam chromosome [20]; (12) *Thermosipho africanus* TCF52B-NC_011653.1 serine protease MucD [21]. The GeneBank or the NCBI Reference sequence for each protease and its producing microbes are: GenBank: AY028615.1 *Bacillus stearothermophilus*| GenBank: KC153302.1 *Bacillus sp*. L010| GenBank: FM163400.1 Uncultured bacterium| GenBank: DQ422814.1 *Pseudoalteromonas sp*.| GenBank: D87921.1 *Pseudomonas aeruginosa* |GenBank: X63533.1 *B. subtilis* | GenBank: ABCF01000018.1 *Bacillus sp* | GenBank: U31759.1 *Thermoactinomyces sp*.| GenBank: HM192828.1 *Laceyella sacchari*| NCBI Reference Sequence: YP_005356102.1 *Flavobacterium indicum* | NCBI Reference Sequence: NC_008600.1 *Bacillus thuringiensis* str. Al Hakam| NCBI Reference Sequence: NC_011653.1 *Thermosipho africanus*.

The amino acids and nucleotides sequences of those strains can be found in the amino acids and the nucleotides alignment in this study. The amino acids and the DNA sequences are adjusted to FASTA format to enable various types of analysis using the different software used in this study [22-25].

2.2. The software used in this study

Several software were used in this study to do various proteins' and DNAs' amino acids and nucleotides sequences analysis. Clustal W v. 1.7 was used to alignment both of the amino acids and the nucleotides used in this study **[Figures– 1,4]** and to generate BOOTSTRAP N-J tree **[Figures–3a,3c]**. FigTree v.1.4 was used to visualize the trees obtained from the amino acids and the nucleotides alignments **[Figure–3a, 3c]**. PAST statistical package was used to do clustering for the different numeric data **[Figures – 3b,3d]**. MEGA v. 5.1 was used to generate comparative analysis for the twelve amino acids sequences as in **Table–1** and **Figure–2**. BioEdit v.7.1.11 was used for the analysis of the nucleotides compositions as in Table 2. MODELLER 9v8 was used in protein models generation for the twelve amino acids sequences used in this study **[Figure – 5]** against six published protease models. And for calculating the % of similarity of each protein sequence with the six used models as in **Table 3 [26-33]**.

2.3. Generating amino acids profiles

For each of the twelve different proteins of the proteases enzymes, an amino acids profile was generated **[Table – 1]**. For each profile, each amino acid has been given as % and the overall data has been summarized in **Table– 3**. For that the software OMGA 5.1 was used to analyze the sequences collected for each protein individually and for all of the twelve used sequences collectively. An average for each of the twenty amino acids for the twelve sequences have been also calculated and given as an average %. OMGA 5.1 enables calculating the % of each amino acid in each protein. The average of each amino acid % for each the twelve proteins was summarized in **Table– 1**.

Commented [D1]: No Result section, No caption for Fig 2

2.4. Allignment and phylogenic trees

Alignments and Phylogenic trees for the proteins primer sequences of amino acids and nucleotides have been generated **(Figure – 1, 4)**. The sequences alignment and the phylogenic trees have been generated using Clustal W version 1.7. The software does alignment for both of the amino acids and the nucleotides used in this study and generate a BOOTSTRAP N-J tree for each. FigTree v1.4 has been used to visualize the trees obtained from both of the alignment of the different proteins' amino acids and nucleotides **(Figure –3a, 3c)**.

2.5. Generating proteases protein models

A model for each of the twelve proteases has been generated using the software MODELLER v 9.8 **[Figure– 5]**. Six published protease models have been used to build the hypothetical model for each of the twelve proteases using MODELLER v 9.8. against six published protease models represent 2GKO (S41 *Psychrophilic subtilisin* (x-ray) [34], 2IXT S41 *Psychrophilic subtilisin* (x ray) [35], 1O0T Cold-adapted alkaline psychrophilic metalloprotease (x-ray) [36], 2PEF Mesophilic protease A (x-ray) [37] 1DBI Thermostable serine protease (x-ray) [38] , and 1SNG Thermophilic serpin in native state (x-ray) [39] models. An overall alignment for the twelve generated models has been generated as in **Figure–6**.

Because the proteases are related to microbes from different environmental conditions they are different from each other. The proteases investigated in this study represent cold-adapted, mesophilic, thermophilic, thermitase, metalloprotease, alkaline, serine, endopeptidases, subtilisine and subtilisine like. For any protein, the 3D structure plus the chemical and physical conditions (where it existed) determined and control the protein structures/functions. However, the 3D structure is depending on the primary structure of the protein. In other word, the protein 3D structure depend on its constituents of amino acids. In addition, the constituent of the amino acids is depending on the DNA nucleotides sequences. For that, both of the proteins' amino acids and nucleotides sequences were investigated in this study to enable better understanding for any variation could be linked to the amino acids and the nucleotides primary structure; the protein structures/functions as well as the protein physical adaptation. The amino acids alignment and the phylogenetic tree show that there are clustering for the thermophilic proteases as well as for the mesophilic proteases **[Figure– 1, 3]**. The same results could be obtained from the nucleotides analysis as in **Figure–3, 4**. The amino acids distributions within the twelve tested protein which were

IIOAB JOURNAL
ISSN: 0976-3104

determined using MEGA 5.1 software were summarized in **Table–1** and **Figure– 2**. Cys is represented in the lesser percentage. That might be explained; that proteases did not need any sort of rigidity or within molecules sulfur bonds, which is a characteristic future for the Cys. The amino acids constituents of the twelve investigated proteases could be ranked from the lower to the higher % as in **Table– 1**. Cys was the lowest one according to its % and followed by Trp, His, Met, Arg, Phe, Glu, Gln, Pro, Tyr, Ile, Leu, Asp, Lys, Thr, Asn, Val, Ser, Gly and Ala. Serine which is an important residue in the proteases active site was ranked a number eighteen. That may explain the presence of protease activity even after the fragmentation of the enzyme as suggested by Amara et al 2013 [41]. The analysis also spot the homogeneity regarding to the amino acids constituent between the investigated proteases. Only *Pseudomonas aeruginosa* has remarkable number of amino acids different from the other compared strains as in **Figure– 2**. That might explained the role of the biofilm formation by *Pseudomonas aeruginosa*, its survive, under certain conditions enable the death of other microbe(s) and the possibility of its acquiring DNA fragments from such microbes [5]. Some strains have minor differences (e.g. *Laceyella sacchari*). Proteases for such unique constituents of amino acids are fast folded and fast interact with their substrate. They could reach their optimum activities in seconds. The nucleotides sequences analysis, which was conducted by BioEdit v. 7.1.11 to calculate each of GC percentage, AT percentage, Molecular weight of single and duple strand and used the number of each of A, T, G and C nucleotides. The cluster analysis for the amino acids percentage data in **Table–1** and which obtained from the PAST statistical package software could better summarize the data in **Table–1**. The cluster put amino acids with similar percentage in groups near each other based on the similarity. As well as it put microbes similar in their amino acids profile in groups. The same was done for the nucleotides sequences in **Table– 2** as well as the GC percentage and AT percentage in **table– 2** and **Figure– 3.** There are minor differences could be found in the three clusters mediated by PAST software in **Figure – 3** and generated from the data in **Table– 1, 2**. This increase the chance of the DNA to be more observed in the future research. From the **Figure– 3** thermophilic proteases as well as the nonthermophilic ones are clustered to similar groups. This might not be clearly distinguished from the protein 3D alignment for the proteases models. For the twelve amino acids sequence, twelve protein models were generated. The models were generated against six resolved and published proteases models [34-39]. The template models represent three thrmophilic, one mesophilic, one psychrophilic and one cold-adapted alkaline protease. For more details, refer to references numbers [34-39]. The three dimensional structure of the different models apparently are similar. Most of them give similar total configuration but with differences in some of their chains, mainly represented in shorter or longer chain if compared with other ones **[Figure– 4]**. The alignment for the all proves that they could be fitted somehow to one alignment model as in **Figure–5**. Interestingly the % of similarity between them and between the six template models are ranging from 11% to 19%. Only one odd protease gives 47.6% of similarity [*B. subtilis* as in **Table– 3**]. The protease of Pseudoalteromonas model gives % of similarity bigger than the four other models **[Table–3]**. Pseudoalteromonas protease model expected to give the smallest % of similarity because its molecular weight is nearly 1.5 bigger than the other proteases. In general, the 3D structures of the proteases even are similar but still able to give some valuable information. Enzymes are used in different kind of medicinal and technical applications [42]. They are protein in nature. Their function can be determined against different substrates as unit of activity [42]. They are sequences of amino acids. And have a DNA codes. The variation between them for that is mainly due to the variation in their nucleotides. In the age of the protein engineering there is a real need for doing more comparative analysis to understand what going on, and what are the player in determining the protein structures/functions [42]. In the EMBO conference for Comparative genomic of Eukaryotic microorganisms (2009) Spain Amara et al introduce the importance of the collective amino acids constituents in the PhaC synthases different classes functions and that similar classes have some similarity in their amino acids constituents, particularly the conserved amino acids residue [43-45]. In fact, such observation has come out from the random mutagenesis study published in 2002 and conducted during 1999 to 2002 for three years on the PhaC synthases classes I and II aiming to change their substrate specificity from one to another [46]. This study leads to determining enhanced PhaC enzyme activity mutants with minor change in the produced polymers' monomeric constituents with minor change in the substrate specificity and enhancement in the enzymes activities [47]. Amino acids in places other than the catalytic residue are responsible for such enzyme activity enhancement. The amino acids collectively for the different enzymes related to classes of PhaC were also analyzed [43-45]. This differences have somehow significant roles in the minor changes within the same type of enzyme in a particular class and lead to the clear variation within the different classes [46]. In fact there are many scientific work discuss the similarity between the amino acids within related enzymes. However, lesser who have been attracted to study differences within nucleotides and related such differences to function or adaptation. This study furthermore attract the attention that, differences within the same protein from a DNA point of view could be also a good guide for understanding the protein structure function. Particularly, if the protein and the DNA differences are due to environmental conditions. DNA phylogenic tree show more variation than the amino acids tree does. Not only, such differences might be due to strain diversity, but also it might due environmental adaptation. In simpler words, I suggest that there might be some built in tools and mechanisms in each microbe induce changes in the DNA to adapt certain environmental conditions firstly and to adjust the produced protein itself. Epigenetic teach us that such possibility could

be happened [48, 49]. Such changes might have a controlled mechanisms and its machinery. As an example the spore formation is, considers a great change happened in the microbe structure and controlled by genes and proteins inside its genome [50]. This will lead us to a conclusion that, we have a very little information about the mechanisms govern the changes happened due to external inducer. This might be due to internal mechanism enable such changes and not due to what is commonly known as mutation. It might be certain similar structures and elements govern the microbes' adaptation, which we still did not know.

[IV] CONCLUTION

This study simply targeted twelve proteins and their DNA sequences to map, similarity, differences, conduct various molecular and bioinformatics analysis and do a logical analysis for both of the similarities and differences and to correlate that to the environmental conditions where the protein is work based on its properties (alkaline, thermophilic, serine, cold-adapted, psychlotrophic, thermophilic, mesophilic etc). For that different software could do different analysis have been used. Mainly to do alignment, phylogenic tree and similarity analysis, building protein models, calculating the nucleotides and the amino acids constituents and do comparative analysis for both of them as well as calculating the GC, and AT%. For starting such work, both of the amino acids and nucleotides sequences of the twelve protein (which have various properties and represent different bacterial strains live in different ecological systems and habitat) and DNA sequences have been collected from the NIH BLAST database. Clustal W v. 1.7, have been used to do the different analysis as described in details in the MMs section. The sequences have adjusted to a FASTA format for analysis. Six well identified and published proteases models have been used for building the various proteases models. The protein models could not explain clearly the differences in the protein functions alone. However, one point could be highlighted here that some of the dissimilarity in the different protein 3D structure are due to the length of the amino acids sequences, of course and the differences in the amino acid constituents. Each microbe has been created to adapt certain conditions. The different phylogenetic trees and cluster analysis for the data in this study prove that proteases with similar environmental conditions cluster together in similar chains. The phylogenic trees prove that DNA sequences could be used in the understanding of the protein type of adaptation while similar proteins clustered near to each other in most case. The data from the protein models in Figure 1 and the alignment between all proteins prove the similarity. DNA analysis if related to the protein physical conditions or functions could also explain the type of changes happened due to different habitat, activity, physical and chemical conditions..

CONFLICT OF INTEREST

The authors have no conflict of interests to declare.

FINANCIAL DISCLOSURE

Institutional support

ACKNOWLEDGEMENT

The author acknowledges the City of Scientific Research and Technological Applications for supporting this work.

REFERENCES

[1] Amara AA, Shibl AM, Tawfik AF. [2012] Deeper in the antibiotics paradigm: Microbes react individually and collectively for spreading the resistance," *African Journal of Microbiology Research* vol. 6(9): 2010v2019,

[2] Blattner FR, Plunkett G, Bloch CA. et al. [1997] The complete genome sequence of *Escherichia coli* K-12, *Science* 277: 1453–1474.

[3] G. De La Cueva-Méndez and B. [2007]bPimentel "Gene and cell survival: lessons from prokaryotic plasmid R1," *EMBO. Report*, 8(5): 458–464

[4] Mendoza Mdel C, Herrero A, Rodicio MR [2009] [Evolutionary engineering in Salmonella: mergence of hybrid virulence-resistance plasmids in non-typhoid serotypes]. Enferm. Infect. *Microbiol Clin* 7(1): 37–43.

[5] Amara, AA. [2011] Opportunistic pathogens and their biofilm "food for thought" In Science against microbial pathogens: Communicating current research and technology advances: Edit by Mendez-Vilas. FORMATEX Microbiology Series No 3 Vol 2: 813–824.

[6] Makino K, Ishii K, Yasunaga T, et al. [1998] Complete nucleotide sequences of 93-kb and 3.3-kb plasmids of an enterohemorrhagic *Escherichia coli* O157:H7 derived from Sakai outbreak. *DNA Res* 5: 1–9

[7] Hutchison CA, Phillips S, Edgell MH. Et al. [1978]. Mutagenesis at a specific position in a DNA sequence. *J Biol Chem* 253(18): 6551– 6560.

[8] Winter G, Fersht AR, Wilkinson AJ, et al. [1982] Redesigning enzyme structure by site directed mutagenesis: Tyrosyl tRNA synthetase and ATP binding. *Nature* 299(5885): 756–758.

[9] Satyanarayana T, Raghukumar C, Shivaji S. [2005] Extremophilic microbes: Diversity and perspectives. *Current Science* 89(1): 78–90

[10] Li D, Huang F, Jiang Y. et al. [2012] Molecular cloning and expression of a novel mesophilic protease from Bacillus sp. L010 in *Escherichia coli* Unpublished Direct Submission (11-NOV-2012) *College of Life Sciences*, Sichuan

[11] Zhang Y, Zhao J, Zeng,R. [2011] Expression and characterization of a novel mesophilic protease from metagenomic library derived from Antarctic coastal sediment *JOURNAL Extremophiles* 15 (1): 23–29

[12] Yan,BQ, Chen XL, Hou XY. [2009] Molecular analysis of the gene encoding a cold-adapted halophilic subtilase from deep-sea psychrotolerant bacterium Pseudoalteromonas sp. SM9913: cloning, expression, characterization and function analysis of the C-terminal PPC domains *JOURNAL Extremophiles* 13 (4):725–733

[13] Okuda K., Morihara K., Atsumi Y. et al. [1990] Complete nucleotide sequence of the structural gene for alkaline proteinase from Pseudomonas aeruginosa IFO 3455 *JOURNAL Infect. Immun* 58 (12): 4083–4088

[14] Barbier P, Houel A, Loux V, et al. [2012] Complete genome sequence of Flavobacterium indicum GPSTA100-9T, isolated from warm spring water *J Bacteriol* 194 (11):3024–3025
[15] Davail S, Feller G, Narinx E. et al. [1992] TITLE Sequence of the subtilisin-encoding gene from an Antarctic psychrotroph Bacillus TA41 JOURNAL Gene 119 (1), 143-144 (1992) Bacillus sp. SG-1 1101501000595, whole genome shotgun sequence
[16] Fu Z, Hamid SB, Razak CN. [2003] Basri,M., Salleh,A.B. and Rahman,R.N.Secretory expression in Escherichia coli and single-step purification of a heat-stable alkaline protease *Protein Expr. Purif* 28 (1):63–68
[17] Tebo B, Ferriera S, Johnson J. et al. [2007] Direct Submission (08-JUN-2007) J Craig Venter Institute, 9704 Medical Center Drive, Rockville, MD 20850, USA
[18] Lee JK, KimYO, Kim HK., et al. [1996] Purification and characterization of a thermostable alkaline protease from Thermoactinomyces sp. E79 and the DNA sequence of the encoding gene *Biosci. Biotechnol. Biochem* 60 (5): 840–846
[19] Jorgensen CM, Madsen SM, Hansen OC et al. [2010] TITLE Original-like thermitase from Laceyella and its safe heterologous expression Direct Submission (06-MAY-2010) Bioneer a/s, Kogle Alle 2, Horsholm DK-2970, Denmark
[20] Challacombe JF, Altherr MR, Xie G, et al.[2007] The complete genome sequence of Bacillus thuringiensis *Al Hakam JOURNAL J Bacteriol* 189 (9):3680–3681
[21] Nesbo,C.L., Bapteste,E., Curtis,B., et al. [2009] The genome of Thermosipho africanus TCF52B: lateral genetic connections to the Firmicutes and *Archaea J Bacteriol* 191 (6): 1974–1978
[22] Altschul SF, Gish W, Miller W. et al. [1990] Basic Local Alignment Search Tool. *J Mol Biol* 215:403–410.
[23] Altschul SF, Madden TL, Schaffer AA, et al. [1997] Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25:3389–3402.
[24] Madden T. [2002] The BLAST Sequence Analysis Tool. 2002 Oct 9 [Updated 2003 Aug 13]. In: McEntyre J, Ostell J, editors. The NCBI Handbook [Internet]. Bethesda (MD): National Center for Biotechnology Information (US); 2002-. Chapter 16. Available from: http://www.ncbi.nlm.nih.gov/books/NBK21097/
[25] Tao Tao. [2011] "Single Letter Codes for Nucleotides". NCBI Learning Center. National Center for Biotechnology Information. Retrieved 2012-03-15. "IUPAC code table". NIAS DNA Bank.
[26] Tamura K, Peterson D, Peterson N, et al. [2011] MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. *Mol Biol Evol* 28: 2731–2739
[27] Thompson JD, Higgins DG, Gibson TJ. [1994] CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-speci?c gap penalties and weight matrix choice. *Nucleic Acids Research* 22: 4673–4680.
[28] Rambaut A. [2012] Figtree v 131, http:// tree.bio.ed.ac.uk/software/figtree 12.05-v 1.4
[29] Hall TA. [1999] BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl Acids Symp Ser* 41: 95-98.
[30] Felsenstein J. [1992] Estimating effective population size from samples of sequences: a bootstrap Monte Carlo integration method. *Genet Res* 60(3):209–220.
[31] Allerberger F [2003] Listeria: growth, phenotypic differentiation and molecular microbiology. FEMS. *Immunol. Med. Microbiol* 35(3): 183–189.
[32] MA Marti-Renom, A Stuart, A Fiser, R Sanchez, et al.[2000] Comparative protein structure modeling of genes and genomes. *Annu. Rev. Biophys. Biomol. Struct*. 29: 291–325
[33] Hammer Ø, Harper DAT, Ryan PD. [2001] PAST: Palaeontological statistics software package for education and data analysis. *Palaeontologia Electronica* 4(1):http://palaeoelectronica.org/2001_1/past/issue1_01.htm
[34] Walter RL, Mekel MJ, Grayling RA, et al. [2009] The crystal structures of the psychrophilic subtilisin S41 and the mesophilic subtilisin Sph reveal the same calcium-loaded state. *Proteins* 74: 489–496
[35] Almog O, Gonzalez A, Godin N.[2009] The Crystal Structures of the Psychrophilic Subtilisin S41 and the Mesophilic Subtilisin Sph Reveal the Same Calcium-Loaded State. *Proteins* 74: 489
[36] Ravaud S, Gouet P, Haser R et al. [2003] Probing the role of divalent metal ions in a bacterial psychrophilic metalloprotease: binding studies of an enzyme in the crystalline state by x-ray crystallography. *J Bacteriol* 185: 4195–4203
[37] Kelch BA, Agard DA. [2007] Mesophile versus Thermophile: Insights Into the Structural Mechanisms of Kinetic Stability *J Mol.Biol* 370: 784–795
[38] Smith CA, Toogood HS, Baker HM, et al. [1999] Calcium-mediated thermostability in the subtilisin superfamily: the crystal atructure of Bacillus AK.1 protease at 1.8 a resolution. *J MOL BIO* 294:1027–1040
[39] Fulton KF, Buckle AM, Cabrita LD, et al. [2005] The high resolution crystal structure of a native thermostable serpin reveals the complex mechanism underpinning the stressed to relaxed transition. *J Biol Chemb*280 : 8435–8442
[40] Feng DF, Doolittle RF. [1987] Progressive sequence alignment as a prerequisite to correct phylogenetic trees. *J Mol Evol* 25: 351–360.
[41] Amara AA, Younis AMM, Salem SR Shabeb MSA. [2013] Mixed enzymes show enhanced activities. *International Science and Investigation Journal* 1–11
[42] Amara AA. [2013] Pharmaceutical and Industrial protein engineering: Where we are? PJPS. 217–232
[43] AA Amara, SM Matar, E Serour. [2012] comparative amino acids studies on PhaC synthases and proteases as well as establishing a new trend in experimental design. IIUM *Engineering Journal* Vol. 13 No. 1, 35–47
[44] E Serour, AA Amara, MS Matar. [2009]"Comparative study between thermophilic and mesophilic lipases and proteases". EMBO –Conference- "Comparative genomics of Eukaryotic microorganisms" 77
[45] S Mater, AA Amara, E Serour. [2009] "Study the possibility of lipases overproduction using media constituents designed based on genomic comparative

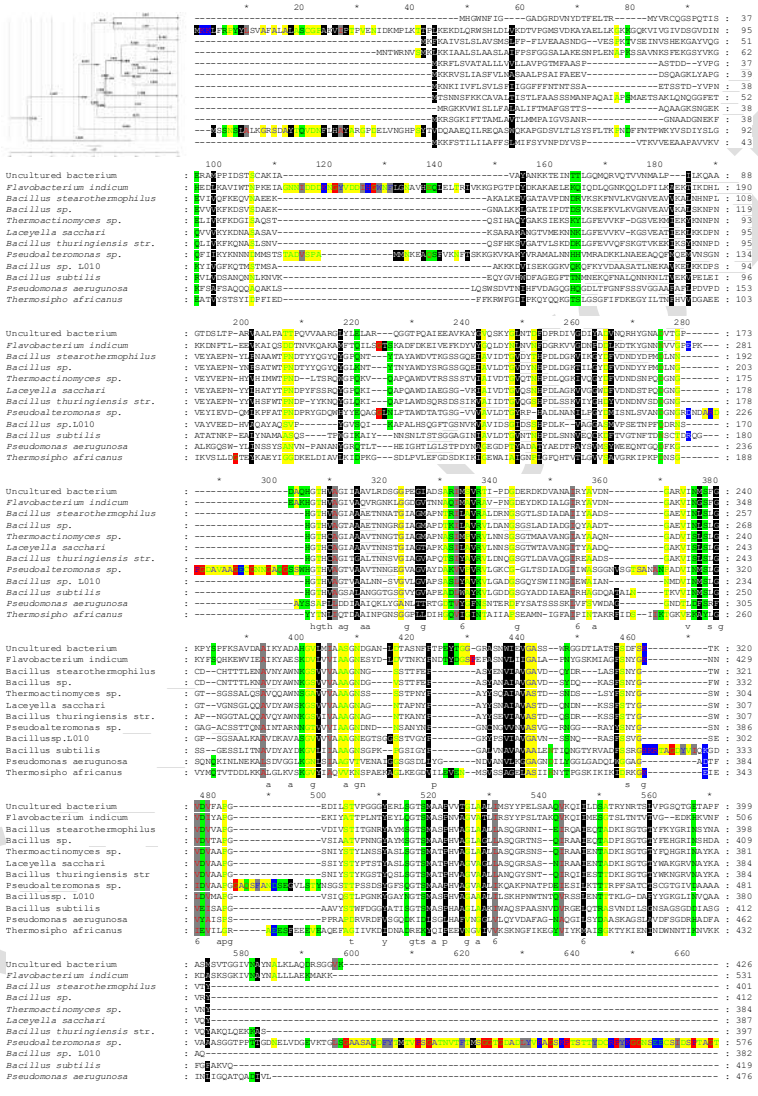
study". EMBO –Conference "Comparative genomics of Eukaryotic
[46] AA Amara, MS Matar E Serour. [2009]"Comparative study on the amino acids constituents of PhaC synthases Class I, II, III and IV" EMBO –Conference- "Comparative genomics of Eukaryotic microorganisms" 59: 2009
[47] A A, Steinbüchel A, Rehm BHA. [2002] In vivo evolution of the Aeromonas punctata polyhydroxyalkanoate (PHA) synthase: Isolation and characterization of modified PHA synthases with enhanced activity. *Appl. Microbiol. Biotechnol* 59:477–482
[48] Proceedings of the First Russian-European Workshop on DNA Repair and Epigenetic Regulation of Genome Stability. St. Petersburg, Russia. Dedicated to the Memory of Nikolay Tomilin. *Mutat. Res.* [2008] 685: 1–102.
[49] Egger G, Liang G, Aparicio A et al. [2004] Epigenetics in human disease and prospects for epigenetic therapy. *Nature* 429(6990), 457–463.
[50] Errington J. [2003] Regulation of endospore formation in Bacillus subtilis. *Nature reviews* 1:117–126

ABOUT AUTHOR

*Prof. **AMRO ABD-AL-FATTAH AMARA** is working in the field of molecular microbiology and, biotechnology. He is the Head of the protein research department in (Mubarak) City for scientific research and, technological applications. He is interested in linking science to the real life to find simple solutions for problems could be costly for many.*

IIOAB JOURNAL
ISSN: 0976-3104

Supplementary Materials (As supplied by author)



Supplementary fig. 1: Multiple alignment of the primary sequences of twelve proteases' amino acids.

IIOAB JOURNAL
ISSN: 0976-3104

Supplementary table: 1. Proteases different amino acids % and an average for each amino acid of the twelve tested proteases.

Protease bacterial host	Amino acids %																				
	Cys	Trp	His	Met	Arg	Phe	Glu	Gln	Pro	Tyr	Ile	Leu	Asp	Lys	Thr	Asn	Val	Ser	Gly	Ala	Total
B. subtilis Antarctic psychrotroph	0.5	1.2	1.7	1.9	2.4	2.6	3.6	3.1	2.9	3.1	5	5.5	5.7	4.5	6.9	8.6	7.4	8.1	11.5	13.8	419
Pseudoalteromonas sp. Serine/cold adapted/halophilic subtilase	1.6	0.8	1.4	2.7	2.1	2.5	2.7	3.4	4.4	3.8	3	3.9	5.4	3.7	9	8.2	7.3	9.6	12.1	12.4	709
Uncultured bacterium Serine mesophilic	0.5	0.7	1.2	2.6	5.2	2.3	3.3	3.8	5.2	4	5.6	5.4	6.8	3.5	7.3	4.2	7.7	7.3	10.3	13.1	426
Bacillus sp. L010 Serine/alkaline/mesophilic	0	1	1.8	2.4	1	2.4	2.9	3.9	3.7	3.4	4.2	5.5	3.4	7.3	5	5.2	10.5	11.8	10.7	13.9	382
Pseudomonas aeruginosa Alkaline/mesophilic	0	1.5	1.7	0.4	2.5	4.8	2.3	5.3	3.2	4.8	3.6	8.8	8.4	3.4	5.3	6.1	5.7	10.3	10.5	11.6	476
Flavobacterium indicum Hot spring subtilisin/thermophilic	0.2	0.8	1.7	1.7	1.5	2.6	5.1	3.4	4	4.7	6.2	7.3	8.1	11.1	5.1	7.2	8.5	4.5	7.9	8.5	531
Bacillus sp. Serine/subtilisine-like/thermophilic	0.5	1.2	1.5	1.5	1.9	2.7	3.9	1.9	3.6	5.6	3.9	5.8	6.6	5.8	6.6	7.5	8.3	8.5	9.5	13.3	412
Thermoactinomyces sp. Alkaline/thermophilic	0.3	1.3	2.1	2.1	1.8	2.1	2.1	4.9	4.2	4.9	5.2	5.2	4.4	3.6	6	7	8.6	12.2	8.9	13	384
Laceyella sacchari Thermitase	0.3	1.6	1.3	0.8	1.8	1.6	2.6	3.6	4.1	5.2	4.4	4.1	4.4	6.2	6.5	6.7	9.3	10.3	9.8	15.5	387
Bacillus thuringiensis Thermitase	0.3	1.3	1.8	0.5	1.5	2.8	2.3	5.8	3	5.8	5.8	4.8	5.3	6.8	6.5	6.8	8.6	11.6	8.8	10.1	397
Thermosipho africanus Thermitase	0	0.7	0.9	1.3	1.3	4.9	7.5	2.2	3.8	5.1	11.5	5.3	5.7	9.9	4.4	4.6	8.6	6	9.3	7.1	453
Bacillus stearothermophilus Alkaline/thermophilic	0.5	1	1.7	1.5	2.2	2.5	5	2.7	3.2	5.7	5.2	5.2	6	4.5	6.7	8	10.2	7	8.2	12.7	401
Average	0.4	1.1	1.5	1.7	2.1	2.8	3.6	3.6	3.8	4.6	5.2	5.6	5.9	5.9	6.4	6.8	8.3	8.8	9.9	12	448.1

Supplementary table: 2. Proteases different nucleotides analysis

Protease bacterial host	G+C content	A+T content	Length/base pairs	Molecular Weight Daltons, single [illegible]	Molecular Weight Daltons, double stranded	Number A	Mol% A	Number T	Mol% T	Number C	Mol% C	Number G	Mol% G
B. subtilis Antarctic psychrotroph	48.10%	51.90%	1260	382866	765582	370	29.37	284	22.54	278	22.06	328	26.03
Pseudoalteromonas sp. Serine/cold adapted/halophilic subtilase	47.19%	52.81%	1689	511701	1025972	453	26.82	439	25.99	362	21.43	435	25.75
Uncultured bacterium SERINE MESOPHILIC	61.36%	38.64%	1281	389902	781229	275	21.47	220	17.17	358	27.95	428	33.41
Bacillus sp. L010 Serine/alkaline/mesophilic	51.21%	48.79%	828	251994	503548	206	24.88	198	23.91	215	25.97	209	25.24
Pseudomonas aeruginosa Alkaline/mesophilic	63.45%	36.55%	1431	438054	873213	281	19.64	242	16.91	478	33.4	430	30.05
Flavobacterium indicum Hot spring subtilisin/thermophilic	34.27%	65.73%	1596	482786	965977	442	27.69	607	38.03	320	20.05	227	14.22
Bacillus sp. Serine/subtilisine-like/thermophilic	44.87%	55.13%	1239	375180	752145	293	23.65	390	31.48	286	23.08	270	21.79
Thermoactinomyces sp. Alkaline/thermophilic	52.29%	47.71%	1155	352988	702611	306	26.49	245	21.21	325	28.14	279	24.16
Laceyella sacchari Thermitase	55.24%	44.76%	1164	356707	708669	289	24.83	232	19.93	363	31.19	280	24.05
Bacillus thuringiensis Thermitase	36.77%	63.23%	1194	361759	723183	361	30.23	394	33	234	19.6	205	17.17
Thermosipho africanus Thermitase	32.16%	67.84%	1362	413362	823864	412	30.25	512	37.59	286	21	152	11.16
Bacillus stearothermophilus Alkaline/thermophilic	41.21%	58.79%	1206	365508	731361	382	31.67	327	27.11	227	18.82	270	22.39

IIOAB JOURNAL
ISSN: 0976-3104

Supplementary table: 3. Different Proteins similarity % to the 1D0T, 1SNG, 2GKO, 2IXT and 2PEF models

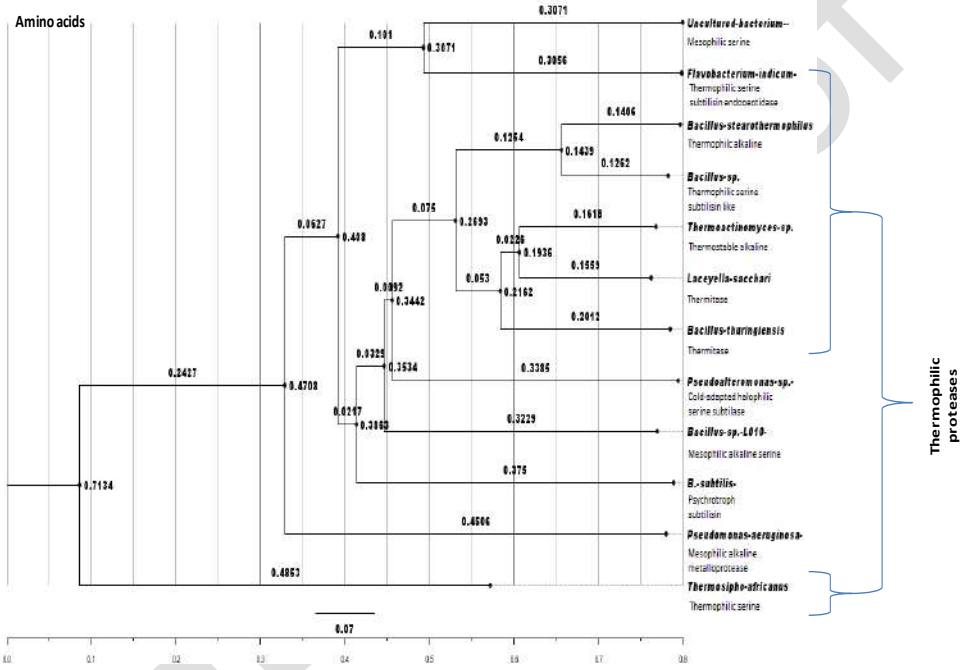
Protease bacterial host names	Similarity % to the used cited six models [10-21]	Rank
B. subtilis	47.573	1
Laceyella sacchari	19.188	2
Thermoactinomyces sp	18.450	3
Uncultured bacterium	18.122	4
Bacillus stearothermophilus	18.080	5
Bacillus sp	17.712	6
Bacillus thuringiensis	15.858	7
Pseudoalteromonas sp	13.916	8
Flavobacterium indicum	13.226	9
Bacillus sp. L010	12.903	10
Thermosipho africanus	12.258	11
Pseudomonas aeruginosa	11.327	12

Supplementary figure: 2. Proteases different amino acids %

IIOAB JOURNAL
ISSN: 0976-3104

Supplementary figure: 3. Phylogenic trees and cluster for each of. 3a) amino acids sequences; 3b) nucleotides sequences (The branching order and distance score were calculated by the program tree as described by Feng and Doolittle (1987) [40]; The trees have been visualized using FigTree v. 1.4.); 3b) Cluster analysis for the amino acids 3d) nucleotides constitents; 3e) GC% and AT% nucleotides constitents. Generated using PAST software using Paird group algorithm (two way with constraind) in the cluster analysis option.

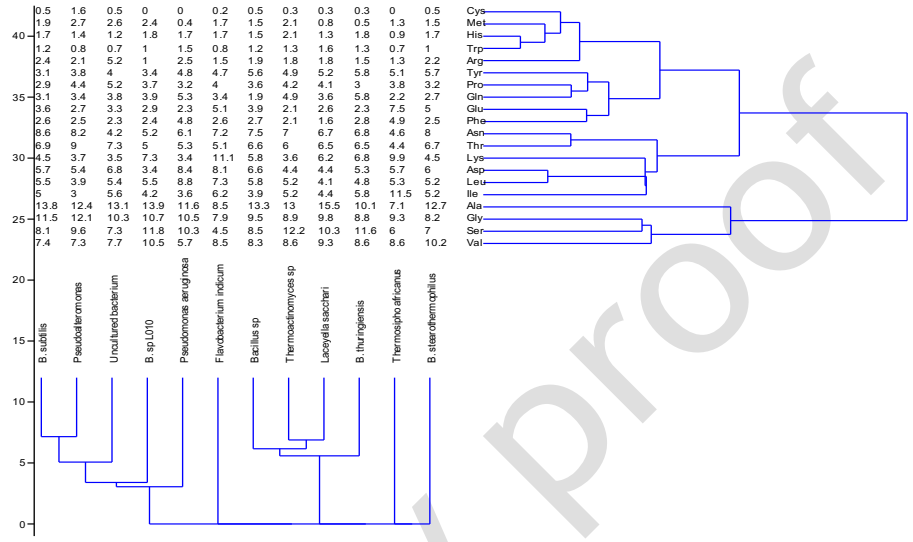
Supplementary figure: 3a



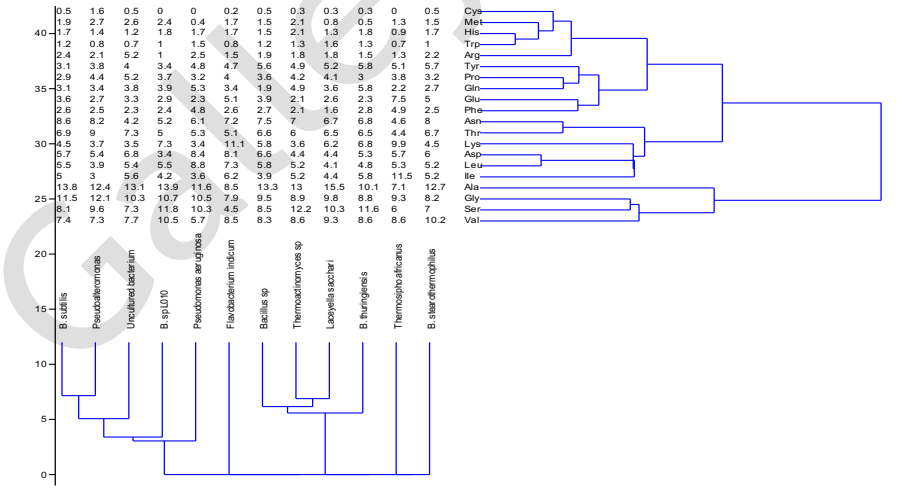
CHEMISTRY

IIOAB JOURNAL
ISSN: 0976-3104

Supplementary figure: 3b.

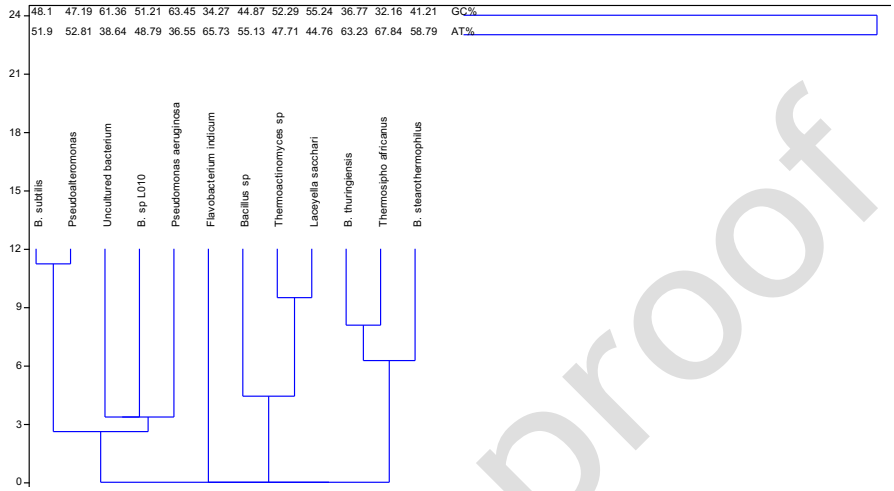


Supplementary figure: 3c.

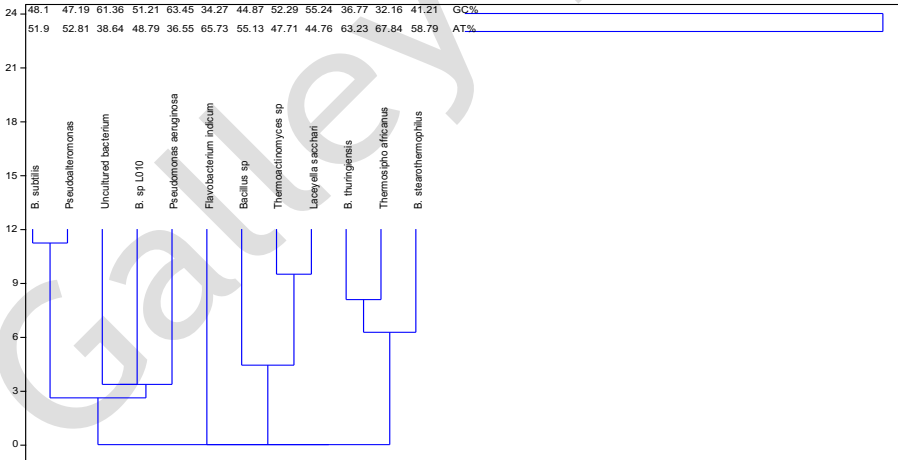


IIOAB JOURNAL
ISSN: 0976-3104

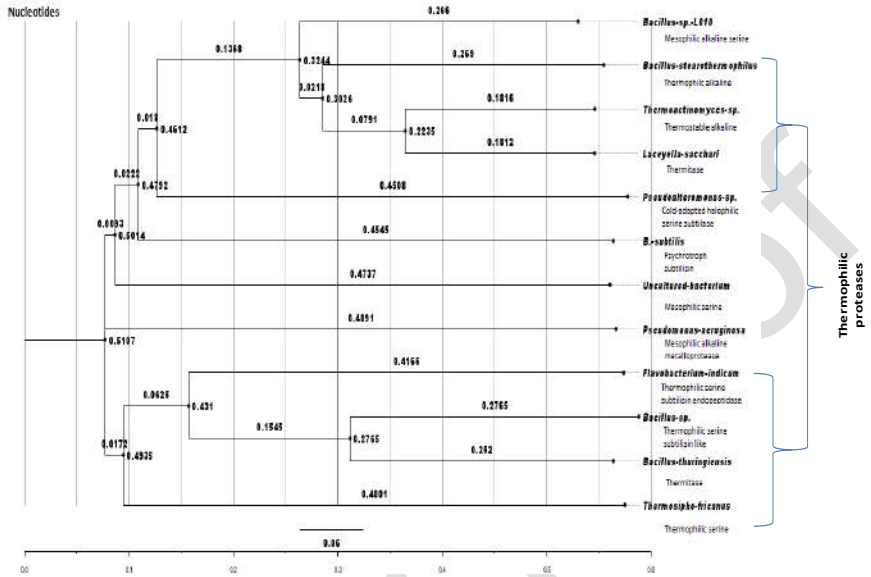
Supplementary figure: 3d.



Supplementary figure: 3e



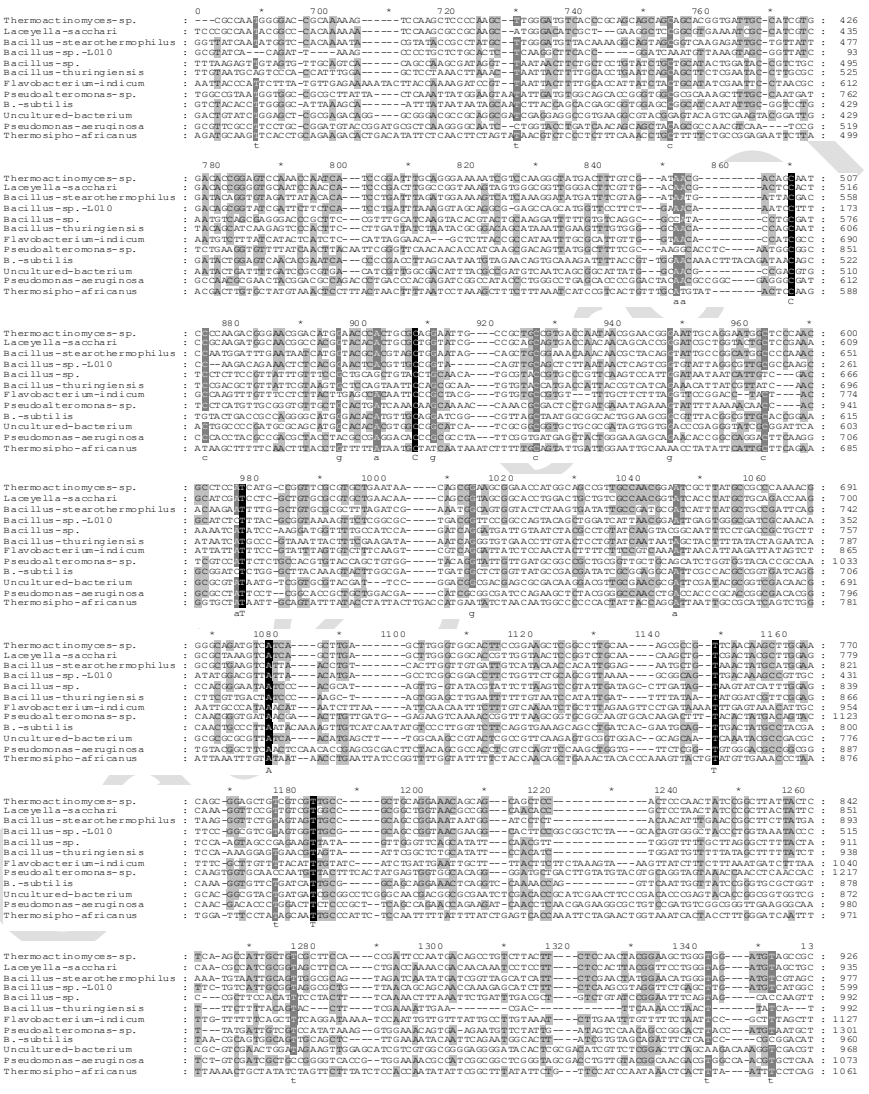
IIOAB JOURNAL
ISSN: 0976-3104



Nucleotides
0.266
Bacillus-sp.-LH9
Mesophilic alkaline serine
0.1368
0.3344
0.269
Bacillus-stearothermophilus
Thermophilic alkaline
0.3219
0.3026
0.1816
Thermoactinomyces-sp.
Thermostable alkaline
0.018
0.4612
0.0791
0.3235
0.1812
Laceyella-sacchari
Thermitase
0.0223
0.4752
0.4508
Pseudoalteromonas-sp.
Cold-adapted halophilic serine subtilase
0.0093
0.5014
0.4545
B.-subtilis
Psychrotroph subtilisin
0.4737
Uncultured-bacterium
Mesophilic serine
0.4851
Pseudomonas-aeruginosa
Mesophilic alkaline metalloprotease
0.5107
0.4166
Flavobacterium-indicum
Thermophilic serine subtilisin endopeptidase
0.0628
0.431
0.2765
Bacillus-sp.
Thermophilic serine subtilisin like
0.1548
0.0172
0.4935
0.2765
0.282
Bacillus-thuringiensis
Thermitase
0.4891
Thermosipho-fricanus
Thermophilic serine
Thermophilic proteases
0.06
0.0
0.1
0.2
0.3
0.4
0.5
0.6

IIOAB JOURNAL
ISSN: 0976-3104

Supplementary figure: 4. Multiple alignment of primary structure of twelve proteases' nucleotides. Shaded and less shaded represent conserved and highly conserved regions respectively

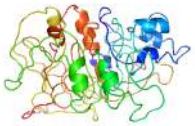


IIOAB JOURNAL
ISSN: 0976-3104

Supplementary figure: 5.Twelve protein models represent the twelve used proteases in this study. The source microbes are included in the Figure.

Bacillus stearothermophilus protease model

Thermosipho africanus protease model



Bacillus sp. L010 protease model

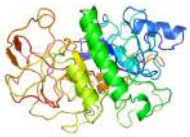
Flavobacterium indicum protease model

Pseudoalteromonas sp. protease model

Pseudomonas aeruginosa protease model

B. subtilis protease model

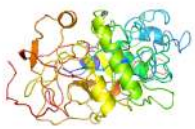
Uncultured bacterium protease model



Laceyella sacchari protease model

Bacillus thuringiensis protease model

***Bacillus sp.* protease model**



***Thermoactinomyces sp.* protease model**

IIOAB JOURNAL
ISSN: 0976-3104

Supplementary figure: 6. Alignment of the twelve models in Fig: 5