Protein Sequence Retrieval and Phylogenetic Analysis of Various *α*-amylase Producing *Bacillus* species

Sonali Priyadarshini[1*], Sukanta Kumar Pradhan[2], Pratima Ray[1]

[1]*Department of Microbiology, College of Basic Science and Humanities*
Orissa University of Agriculture and Technology, Surya Nagar
Bhubaneswar 751 003, Odisha, India
E-mails: sonalisony.priyadarshini@gmail.com, pratimary@yahoo.com

[2]*Department of Bioinformatics, Centre for Post Graduate Studies*
Orissa University of Agriculture and Technology, Surya Nagar
Bhubaneswar 751 003, Odisha, India
E-mail: ksukantapradhan@gmail.com

[*]*Corresponding author*

Received: December 13, 2018 **Accepted: January 25, 2020**

Published: September 30, 2020

Abstract: *Due to advancements made in biotechnology, genetic manipulation and medium optimization has been possible and that has led to high yield in enzyme especially α-amylase with improvised properties that find its application suitably in detergent, laundry and starch processing industries. Various species of the genus Bacillus produce α-amylases with varying properties. The amylases that find application miraculously in different fields can be easily obtained from plant, animal and microbial sources. Out of the three sources, microbial source especially the genus Bacillus is extensively studied for its ability to produce the enzyme. In the present study, protein sequences of primarily identified α-amylase producing Bacillus spp. were retrieved from protein database of National Center for Biotechnology Information. Multiple sequence alignment was performed and construction of Phylogenetic tree among these sequences was carried out using Neighbor-joining method; data sets based on 1000 re-samplings in molecular genetics evolutionary analysis.*

Keywords: *Alkaline α amylase, Bacillus, Phylogenetic analysis, Protein sequence.*

Introduction

Commercially *α*-amylases are applied in dairy, soft drinks, chocolates, pharmaceuticals, food processing, leather, textile, paper, wine, meat, fish processing industries and are mainly produced from the genus *Bacillus*. Among many species, *B. subtilis*, *B. stearothermophilus*, *B. licheniformis* and *B. amyloliquefaciens* are reported extensively as industrial amylase producers because of the limited space and short period of time required for their cultivation and their feasibility to genetic manipulation [3, 6, 7, 10-12]. *Bacillus* sp. is an industrial important microorganism because of its rapid growth rate, secretes enzyme into the extracellular medium and safe handling [14].

Phylogenetic analysis is a standard tool in any molecular biologist's bioinformatics toolkit that, in the context of protein sequence analysis, enables to study the evolutionary history and change of proteins and their function. Such analysis is essential to understanding major evolution such as the origins and history of macromolecules, developmental mechanisms, phenotypes, and life itself. Phylogenetic analysis of protein sequence data is integral to gene annotation, prediction of gene function, the identification and construction of gene families, and gene discovery. Phylogenetic trees are mathematical structures that depict the evolutionary history of a group of organisms or genes. A typical phylogenetic analysis of protein sequence data involves five

distinct steps [9]: (a) data collection, (b) inference of homology, (c) sequence alignment, (d) alignment trimming, and (e) phylogenetic analysis.

Phylogenetic methods for comparative analysis of protein sequences become vital with the rapid accumulation of molecular sequence detail. Phylogenetic analysis helps to find out the nature and extent of selective forces that shape the evolution of genes and species [4].

In the present study α-amylase producing *B. cereus* (60 spp.), *B. flexus* (12 spp.) and other different groups were selected. Comparative analysis of molecular sequence data was performed using MEGA 6.0 for understanding the evolutionary histories of species.

Materials and methods

Protein sequence retrieval

Microbial α-amylase protein sequence of bacterial isolates was retrieved from protein database of NCBI (http://www.ncbi.nlm.nih.gov).

Phylogenetic tree construction

Protein sequences of primarily identified α-amylase producing *Bacillus* spp. were retrieved from protein database of National Center for Biotechnology Information (NCBI) web server to perform sequence analysis. Multiple sequence alignment and construction of Phylogenetic tree among these sequences were performed using Neighbor-joining (NJ) method; data sets based on 1000 re-samplings in Molecular Genetics Evolutionary Analysis (MEGA) 6.0 package [13].

Physiochemical parameter analysis

ExPASy server was used to analyze the physiochemical properties of amylase protein in FASTA formed [1]. Different tools of the proteomic server such as ProtParam and Computer pI / MW were applied to figure out different physiochemical properties of bacterial amylase like amino acids number, molecular weight, pI, atomic compositions, instability index, aliphatic index and Grand average of hydropathicity (GRAVY).

Results and discussion

Sequence analysis

A total of sixty α-amylase protein sequences belonging to *B. cereus*, twelve protein sequences belonging to *B. flexus* and thirty one belonging to different spp. of genus *Bacillus* were taken for analysis. Multiple sequence alignment of α-amylase protein sequences of different group of bacteria showed the conserved pattern of the amino acid residues (Fig. 1). The conserved regions observed in amino acid positions ranging from 1-165, 185-240 and 275-350 in the alignment windows. The conserved pattern will be useful for structural and functional analysis of α-amylase protein and the resultant consensus sequence may be used to identify other members of the family.

Phylogenetic tree construction and divergence study

Three NJ based phylogenetic trees of α-amylase producing group i.e., *B. cereus* (60 spp.), *B. flexus* (12 spp.) and other different groups is depicted in Figs. 2, 3 and 4, respectively. These 60 numbers of α-amylase protein sequences of different strain of *B. cereus* were clustered into 3 groups: two major and a minor group while the twelve protein sequences of α-amylase producing *B. flexus* clustered into two groups. Thirty one α-amylase producing bacterial spp. including *Bacillus* spp. were studied to understand the divergence between different genera of bacteria involved in α-amylase production.

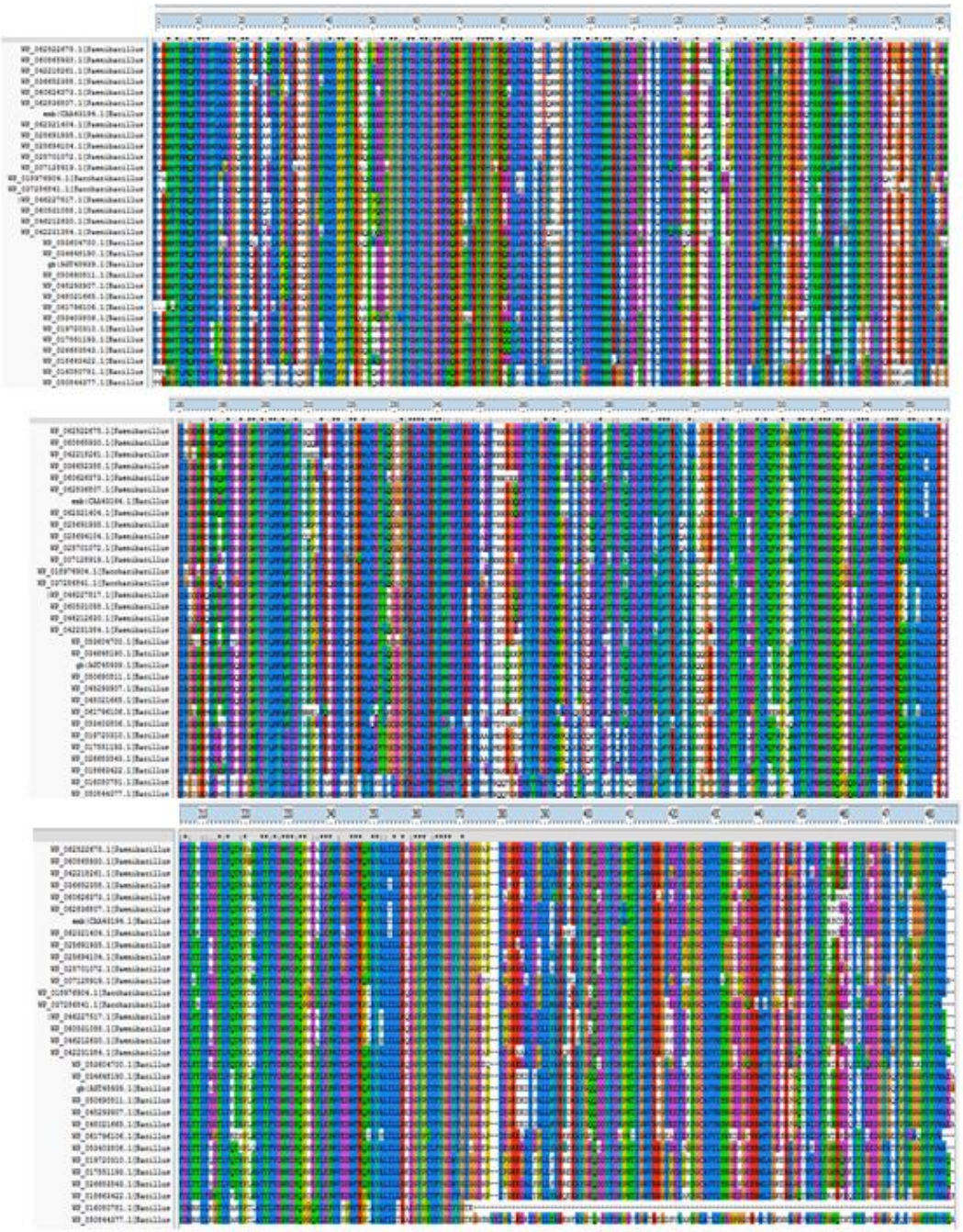

Fig. 1 Multiple sequence alignment showing conserved residues among bacterial *α* amylase

It was observed that most of the *Bacillus* were clustered into one group whereas *Paenibacillus* in other group. *B. flexus* closely related to *B. megaterium* while *B. cereus* and *B. thuringensis* were diverged from same node and closely related. The objective of studying divergence is to understand *α*-amylase producing bacterial diversity at protein level. These clusters indicated the functional as well as the structural similarities among the various *Bacillus* spp. with respect to *α*-amylase protein sequence though the strains may be distantly related. Distance matrix for different *α*-amylase producing bacterial spp. is presented in Fig. 5.

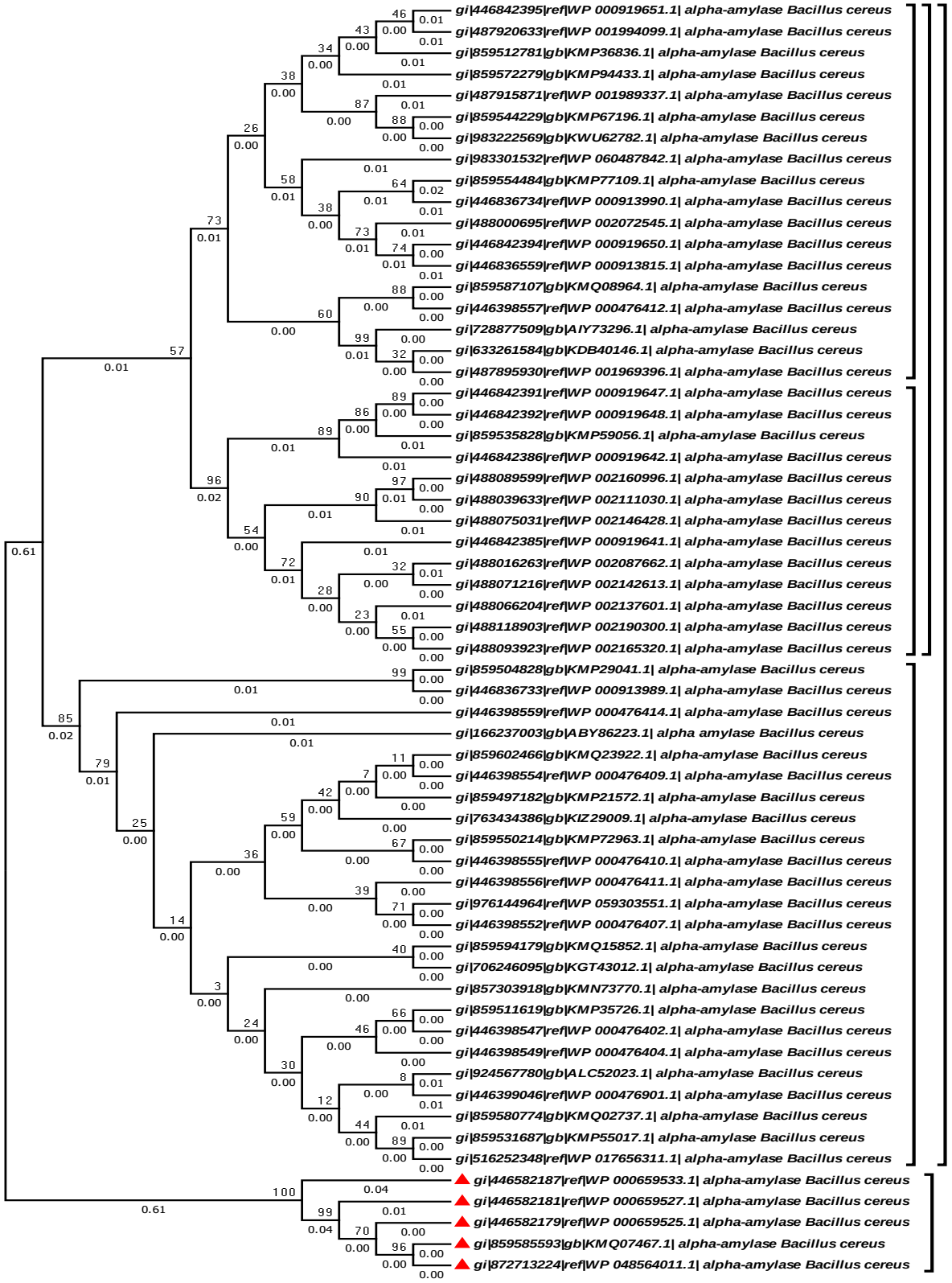

Fig. 2 Phylogenetic tree of α-amylase protein sequences from different strains of *B. cereus* constructed by NJ method

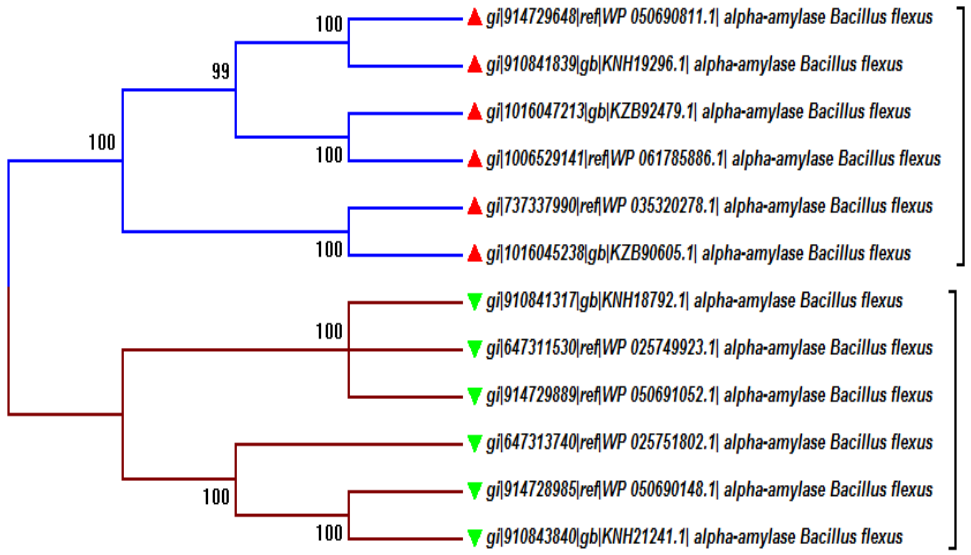

Fig. 3 Phylogenetic tree of α-amylase protein sequences from different strains of *B. flexus* constructed by NJ method

Physiochemical parameter analysis

The physiochemical features of the bacterial (*Bacillus* spp.) α-amylase protein sequences were listed in Table 1. The total no of amino acid residues ranged from 482 to 688 with variable molecular weight. The pI values showed small scale variability which ranged between 4.39-5.83.

The other physiochemical properties such as number of negatively charged amino acid residues (Aspartic acid and Glutamic acid), positively charged amino acid residues (Arginine and Lysine) and GRAVY of the amylase protein sequences showed a wide range of variability. The measurement of the relative volume occupied by aliphatic amino acid residues viz., alanine, valine, leucine and isoleucine was defined as the aliphatic index and the thermo stability of a globular protein was directly proportional to that of aliphatic index [5].

The analyzed α-amylase protein sequences were thermostable based on high value of aliphatic index. *In vivo*, half-life of a protein was calculated in the form of instability index [2]. Previous reports claimed that the proteins with instability index more than 40, showed the half-life period of less than 5 h whereas proteins with less than 40 instability index showed half-life period of 16 h [8]. In our study, the instability index values of studied amylase protein sequences were found lower than 40, which indicated that they have *in vivo* half-life as high as 16 h except six bacteria (*Paenibacillus jiluni, P. wulumuqiensis, P. dauci, P. bovis, B. koreensis* and *Pseudomonas stutzeri*).

Conclusion

A total of sixty α-amylase protein sequences belonging to *B. cereus*, twelve protein sequences belonging to *B. flexus* and thirty one belonging to different spp. of genus *Bacillus* were analysed which were retrieved from protein database of NCBI. Multiple sequence alignments were carried out for three different groups of α-amylase producing bacteria using Clustal W to find out the conserved pattern of the amino acid residues. Phylogenetic tree of α-amylase protein sequences from *B. cereus*, *B. flexus* and different *Bacillus* spp. showed clusters which indicated

the functional as well as the structural similarities among the various *Bacillus* spp. with respect to *α*-amylase protein sequence though the strains may be distantly related.

The physiochemical features of the bacterial (*Bacillus* spp.) *α*-amylase protein sequences were studied. The total number of amino acid residues ranged from 482 to 688 with variable molecular weight. The pI values showed small scale variability which ranged between 4.39-5.83. The analyzed *α*-amylase protein sequences were thermostable based on high value of aliphatic index. In our study, the instability index values of studied amylase protein sequences were found lower than 40, which indicated that they have *in vivo* half-life as high as 16 h except six bacteria (*Paenibacillus jiluni*, *P. wulumuqiensis*, *P. dauci*, *P. bovis*, *B. koreensis* and *Pseudomonas stutzeri*).

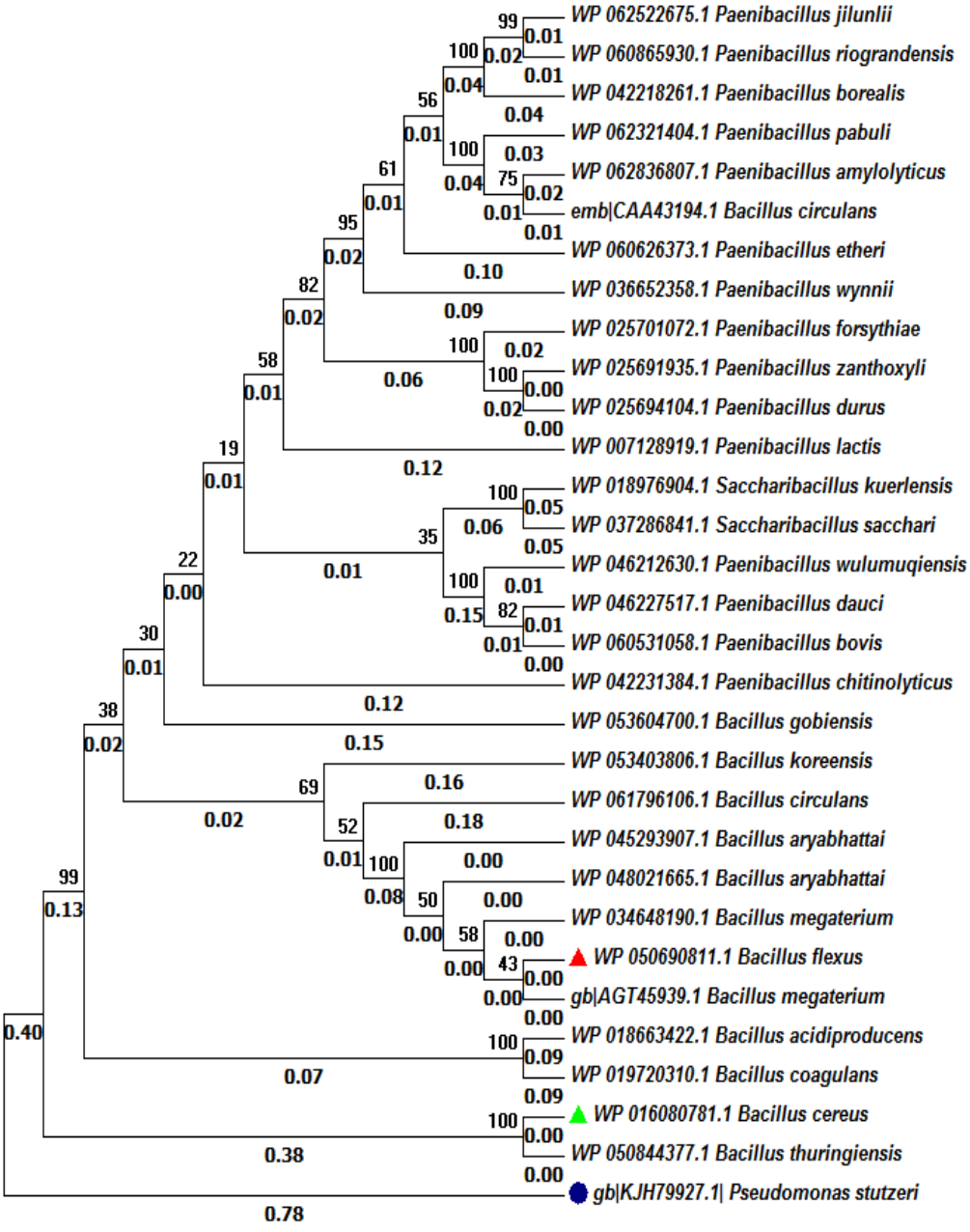

Fig. 4 Phylogenetic tree of *α*-amylase protein sequences from different *Bacillus* spp. constructed by NJ method

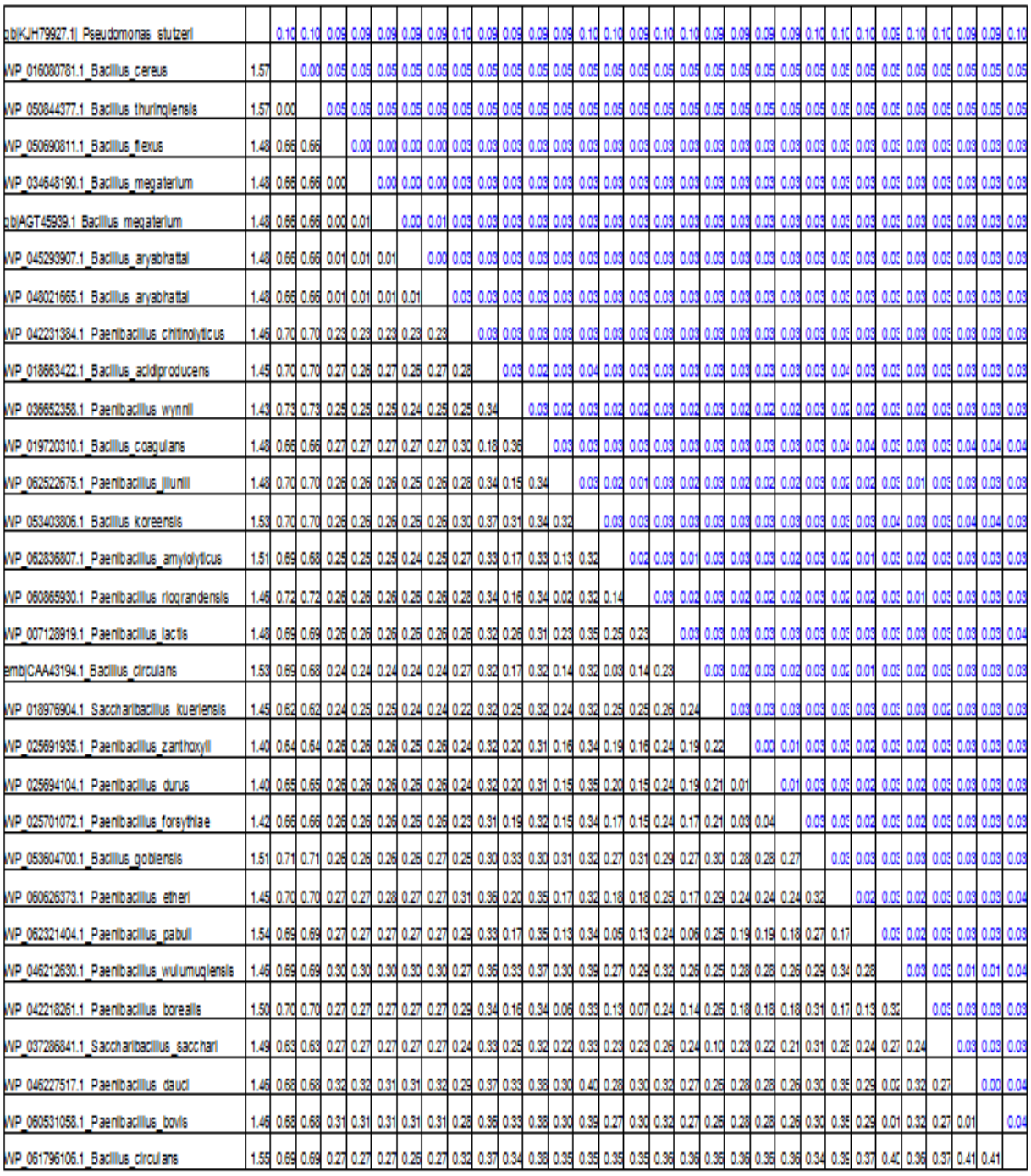

Fig. 5 Distance matrix for different α-amylase producing bacterial spp.

Acknowledgements

The authors are thankful to the Heads of the Department of Microbiology and Bioinformatics, College of Basic Science and Humanities and Centre for Post Graduate Studies respectively, Odisha University of Agriculture and Technology, Bhubaneswar, Odisha, India, for their constant support and allowing this research work to be accomplished.

Table 1. Physiochemical features of α-amylase protein sequences observed in *Bacillus* spp.

Sl. No.	Organism	Accession No.	Residues	MW	Theoritical pI	Negative residue	Positive residue	Instability index	Aliphatic index	GRAVY
1	*Bacillus flexus*	WP050690811.1	483	56116.9	4.58	93	41	36.72	66.34	0.684
2	*Paenibacillus jiluni*	WP062522675.1	514	58276.4	4.64	90	47	41.34	67.92	0.557
3	*Paenibacillus riograndensis*	WP060865930.1	511	58163.8	4.85	86	53	39.59	69.08	0.567
4	*Paenibacillus borealis*	WP042218261.1	494	56107.1	4.59	88	47	35.94	68.08	0.558
5	*Paenibacillus pabuli*	WP062321404.1	493	56444.7	4.75	90	50	37.09	71.60	0.532
6	*Paenibacillus amylolyticus*	WP062836807.1	493	56341.4	4.64	93	47	36.7	70.20	0.531
7	*Bacillus circulans*	emb\CAA43194.1	493	56536.8	4.64	94	51	37.21	70.41	0.554
8	*Bacillus megaterium*	WP034648190.1	483	56109.7	4.60	93	42	37.48	65.13	0.696
9	*Paenibacillus etheri*	WP060626373.1	490	56177.7	5.34	76	54	28.06	70.02	0.529
10	*Paenibacillus wynnii*	WP036652358.1	504	57641.9	4.84	86	51	29.86	68.85	0.565
11	*Paenibacillus forsythiae*	WP025701072.1	511	57991.6	4.93	82	53	30.12	71.92	0.484
12	*Paenibacillus zanthoxyli*	WP025691935.1	511	58017.1	4.77	86	50	31.41	71.35	0.478
13	*Paenibacillus durus*	WP025694104.1	511	57994.6	4.89	81	53	29.78	71.55	0.464
14	*Paenibacillus lactis*	WP007128919.1	488	56056.7	5.83	73	57	33.83	66.74	0.600
15	*Saccharibacillus kuerlensis*	WP018976904.1	499	57316.3	4.52	93	46	37.88	66.21	0.571
16	*Saccharibacillus sacchari*	WP037286841.1	499	57172.0	4.6	92	47	29.77	65.07	0.639
17	*Paenibacillus wulumuqiensis*	WP046212630.1	588	63703.1	4.41	103	43	41.64	67.44	0.606
18	*Paenibacillus dauci*	WP046227517.1	560	64123.4	4.40	106	42	42.51	66.14	0.647
19	*Paenibacillus bovis*	WP060531058.1	560	64069.3	4.39	107	42	45.92	66.14	0.646
20	*Paenibacillus chitinolyticus*	WP042231384.1	486	55627.7	4.67	89	50	34.17	67.22	0.593
21	*Bacillus gobiensis*	WP053604700.1	485	56275.8	5.21	77	53	31.27	69.90	0.573
22	*Bacillus koreensis*	WP053403806.1	486	56492.3	4.56	94	46	42.24	67.41	0.703
23	*Bacillus circulans*	WP061796106.1	482	55819.0	4.66	89	51	31.46	69.17	0.643

24	*Bacillus aryabhattai*	WP045293907.1	483	56104.6	4.62	93	42	38.29	64.93	0.696
25	*Bacillus aryabhattai*	WP048021665.1	483	56162.8	4.65	93	43	37.52	64.93	0.701
26	*Bacillus megaterium*	gb/AGT45939.1	483	56094.6	4.60	92	41	37.24	64.93	0.699
27	*Bacillus acidiproducens*	WP018663422.1	484	55838.0	5.10	81	51	28.58	67.67	0.665
28	*Bacillus coagulans*	WP019720310.1	487	56064.2	5.08	81	50	26.45	63.84	0.702
29	*Bacillus cereus*	WP016080781.1	513	58259.5	5.48	63	49	19.67	67.45	0.617
30	*Bacillus thuringiensis*	WP050844377.1	513	58350.6	5.48	64	50	20.49	66.69	0.624
31	*Pseudomonas stutzeri*	gb/KJH79927.1	688	76370.8	5.43	86	62	41.45	73.43	0.505

References

1. Gough J., K. Karplus, R. Hughey, C. Chothia (2001). Assignment of Homology to Genome Sequences Using a Library of Hidden Markov Models that Represent All Proteins of Known Structures, Journal of Molecular Biology, 313, 903-919.
2. Guruprasad K., B. V. Reddy, M. W. Pandit (1990). Correlation between Stability of a Protein and Its Dipeptide Composition: A Novel Approach for Predicting *in vivo* Stability of a Protein from Its Primary Sequence, Protein Engineering, 4(2), 155-161.
3. Hayashida S., Y. Teramoto, T. Inoue (1988). Production and Characteristics of Raw-potato-starch-digesting *α*-amylase from *Bacillus subtilis* 65, Applied Environmental Microbiology, 54(6), 1516-1522.
4. Huma T., A. Maryam, S. Rehman, M. Tahir, T. Shaheen, A. Haque, B. Shaheen (2014). Phylogenetic and Comparative Sequence Analysis of Thermostable Alpha Amylases of Kingdom Archea, Prokaryotes and Eukaryotes, Bioinformation, 10(7), 443-448.
5. Ikai A. (1980). Thermostability and Aliphatic Index of Globular Proteins, Journal of Biochemistry, 88(6), 1895-1998.
6. Khire J., A. Pant (1992). Thermostable, Salt-tolerant Amylase from *Bacillus* sp. 64, World Journal of Microbiology and Biotechnology, 8(2), 167-170.
7. Prakasham R., C. S. Rao, R. S. Rao, S. Rajesham, P. Sarma (2005). Optimization of Alkaline Protease Production by *Bacillus* sp. Using Taguchi Technology, Applied Biochemistry and Biotechnology, 120(2), 133-144.
8. Rogers S., R. Wells, M. Rechsteiner (1986). Amino Acid Sequences Common to Rapidly Degraded Proteins: The Pest Hypothesis, Science, 234(4774), 364-368.
9. Rokas A. (2011). Phylogenetic Analysis of Protein Sequence Data Using the Randomized Axelerated Maximum Likelihood (RAXML) Program, Current Protocols in Molecular Biology, Chapter 19, Unit19.11, doi: 10.1002/0471142727.mb1911s96.
10. Saxena S. (2015). Microbial Enzymes and Their Industrial Applications, Applied Microbiology, Springer, India, New Delhi.
11. Selvamohan T., S. Sherin (2010). Optimization of Protease Production from *Bacillus cereus* Using Different Substrates, Plant Archives, 10(2), 651-656.
12. Souza P. M., P. de O. Magalhaes (2010) Application of Microbial *α*-amylase in Industry – A Review, Brazilian Journal of Microbiology, 41(4), 850-861.
13. Tamura K., G. Stecher, D. Peterson, A. Filipski, S. Kumar (2013). MEGA 6: Molecular Evolutionary Genetics Analysis Version 6.0, Molecular Biology and Evolution, 30(12), 2725-2729.

14. Vijayalakshmi K. S., S. Abha, P. Chander (2012). Isolation and Characterization of *Bacillus subtilis* KC3 for Amylolytic Activity, International Journal of Bioscience, Biochemistry and Bioinformatics, 2(5), 336-341.

Dr. Sonali Priyadarshini
E-mail: sonalisony.priyadarshini@gmail.com

Dr. Sonali Priyadarshini currently works as a part time Lecturer in Buxi Jagabandhu Bidyadhar (B.J.B) Autonomous College, Bhubaneswar. Prior to that she worked as teaching assistant in the Department of Biotechnology, Rama Devi Women's University (RDWU), Bhubaneswar, Odisha, India. She has published six papers in national and international journals. Her research area of interest is environmental microbiology and microbial enzyme production.

Assist. Prof. Sukanta Kumar Pradhan
E-mail: ksukantapradhan@gmail.com

Sukanta Kumar Pradhan is an Assistant Professor and PI of DBT-BIF Centre in the Department of Bioinformatics, Centre for Post Graduate Studies, Orissa University of Agriculture and Technology, Bhubaneswar, Odisha, India. He has published fifteen papers in journals of international repute. His area of interest is in comparative genomics of different field crops, microbial bioinformatics, metagenomics study, environmental bioinformatics and bioremediation, and NGS analysis at genome level. He has guided twenty-five P.G. students of M.Sc. Bioinformatics.

Prof. Pratima Ray, Ph.D.
E-mail: pratimary@yahoo.com

Dr (Mrs) Pratima Ray is a Professor in the Department of Microbiology, College of Basic Science and Humanities, Orissa University of Agriculture and Technology, Bhubaneswar, Odisha, India. Her area of interest is in industrial microbiology and extremozymes (amylase; alkaline protease; cold active cellulase and lipase; heat stable and alkaline amylase) from micro-organisms. She has guided twenty five P.G. and five Ph.D. students. She has published more than forty papers in national and international journals.