EasyChair Preprint

№ 2243

Inferring Staurosporine Gene Evolution In Bacteria Using Metagenomics Approach

Priya Kumari

December 25, 2019

Inferring Staurosporine Gene Evolution In Bacteria Using Metagenomics Approach

Priya Kumari
Birla Institute of Technology, Mesra
Ranchi, Jharkhand (835215)

ABSTRACT

Soil is predicted to contain thousands of unique bacterial species per gram. A large diversity of biosynthetic genes dwells in soil DNA from which diverse secondary metabolite gene clusters can be recovered and studied. The screening and heterologous expressions of an archived soil DNA library functionally identified biosynthetic gene cluster producing a potent anticancer/antibiotic compound i.e. Staurosporine. Simultaneously, a recently discovered obligate marine actinomycetes genus produces a diverse range of secondary metabolites including the compound Staurosporine. The amazing fact is that this genus resides in marine sediments and display fundamental physiological differences from those that occur on land. This short study predicted the diversity and evolution of the regulatory gene ranging from the primitive archean group to bacteria. It may help to screen the population from different geological niche for their potential role in synthesizing anticancer/antibiotic compounds.

Metagenomics is the study of genetic materials of microbial DNA that are extracted directly from the environment. Culture independent studies of genomes unlock the hidden diversities of microscopic life and suggest that there is genes large group of microorganism exists in many environments that cannot be cultured. Also, it offers a powerful lens for exploring the microbial world that has the potential to revolutionize the living world. To measure phylogenetic diversity in microbial communities we have used the small subunit of the rRNA gene to quantify phylogenetic relationships among microbial taxa.

Keywords: biosynthetic gene, secondary metabolite genes, Staurosporine, marine actinomycetes

INTRODUCTION

A natural compound Staurosporine first reported in the year 1977 was obtained from the bacterium *Streptomyces staurosporine* which is a representative member of indolocarbazole that is under current study due to their potential as anticancer drugs. These are frequently found with several analogs used in cancer clinical trials [1]. The main biological activity of Staurosporine is the inhibition of protein kinases through the prevention of ATP binding to the kinase through the stronger affinity of Staurosporine to the ATP-binding site on the kinase [1].

The Staurosporine biosynthesis regulatory gene cluster consists of 14 ORFs lying in the same operon from which three genes play vital role in the production of the compound are StaO, StaD and StaP. The biosynthesis of Staurosporine starts with the dimerization of amino acid L-tryptophan in its zwitterionic form. The Staurosporine is coded by StaO, StaD and StaP enzymes involved in biosynthesis, StaG and StaN responsible for the bond formation between aglycone and deoxy-sugar, StaA, StaB, StaE, StaJ, StaI, StaK, StaMA and StaMB direct the deoxysugar biosynthesis and StaR is a transcriptional regulator [2].

Recently discovered Actinobacteria genera, found from temperate marine sediments typically distribute secondary metabolic genes. This actinobacteria genus is Salinispora whose three known species have been reported, namely *Salinispora arenicola*, *Salinispora tropica* and *Salinispora pacifica* sharing 99% 16S sequence identity. The genus is broadly distributed in tropical and subtropical marine sediments, with *S.arenicola* displaying the widest geographical range [3]. The main geographical location of *S. Arenicola* found in Sea of Cortez/Gulf of California, Red Sea, Island of Guam and Bahamas in about 1 meter of

depth at the temperature range of 10–30^{o}C with an optimum at 20–28^{o}C [4]. The most significant aspect of studying this Actinobacterial genus is the production of the diverse range of secondary metabolites [5] which can be explored further to unlock its specific functions.

Early cloning and sequencing study of genomes unlock the fact that there is probably a large group of microorganisms in many environments that cannot be cultured and thus cannot be sequenced [6,7]. The metagenomics approach reveals hidden diversity of microscopic life and also offers a powerful lens for exploring the microbial world that has the potential to revolutionize the entire living world.

To measure phylogenetic diversity in microbial communities we have used the small subunit of the rRNA gene to quantify phylogenetic relationships among microbial taxa [8].

METHODOLOGY

1. Staurosporine sequence extraction and Multiple Sequence Alignment (MSA) analysis

The cluster of genes that regulate Staurosporine biosynthesis was extracted from the NCBI database (https://www.ncbi.nlm.nih.gov/) and subjected for further analysis. The sequence is used for the MSA using the BLAST tool [9] against the database with default parameters. The homology study of that staurosporine gene indicated that S. arenicola and S. pacifica show maximum similarity with these regulatory genes.

2. Staurosporine biosynthesis gene

From the literature, three important genes localized in the Staurosporine regulatory gene cluster namely StaO, StaD and StaP code for the enzymes involved in the Staurosporine biosynthesis. The details of amino acid gene sequences of all those three genes were explored from the Uniprot database (https://www.uniprot.org/) and briefly described in the table given below.

Table no.1: List of genes mainly responsible for Staurosporine production.

Gene No.	Gene Name	Orthology ID	Uniprot ID	Name
504	Sta O	K20075	Q83WG4	L- Amino Acid Oxidase
1096	Sta D	K20076	Q8GRF3	Chromo Pyrrolic Acid Synthase
417	Sta P	K20078	Q83WG3	Cytochrome P450

3. Metabolic pathway analysis

The metabolic pathway for the staurosporine biosynthesis (map00404) is reported in the KEGG database [10]. Here we observed that the key element is Tryptophan for the synthesis of staurosporine in the presence of different genes expressing corresponding enzymes.

4. Analysis using Metagenomic samples

On the basis of geographical distribution of *S. arenicola* reported in the literature, there were 12 metagenomic samples consisting of 16s rRNA that were collected from the EBI metagenomics database (https://www.ebi.ac.uk/metagenomics/). The taxonomic distribution of species present in metagenomic samples was done using Phyloshop software (http://omics.informatics.indiana.edu/mg/phyloshop/). The

evidence found from the result of phyloshop software indicates that marine sample consists of homolog of Streptomyces species which is widely distributed in the marine habitat. Actinobacteria species are present in the entire sample.

5. Orthologous analysis

Each gene that is under study i.e. staO, staP and staD is identified in the COG database (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC102395/) to check its diversity and prepared orthology phylum table from obtained list of phyla that contain the same gene.

6. Evolutionary analysis

To explore more about the evolution of the respective genes responsible for the biosynthesis of the staurosporine phylogenetic tree was drawn using the MEGA X [11]. Randomly sequences were extracted from each phylum consisting orthologous gene separately for each gene which was subjected to the phylogenetic tree using Maximum likelihood Algorithm in MEGA X software.

RESULTS AND DISCUSSION

The staurosporine biosynthetic regulatory gene cluster extracted from the NCBI database consisting of accession number AB088119.1 is 22999nt bp which was sequenced in the year 2002. The multiple sequence alignment of the staurosporine regulatory gene suggests that *S.arenicola* and *S. pacifica* shows the maximum similarity reported in the literature of 99% similarity. The alignment is shown in the form of phylogenetic tree in the figure below (figure 1).

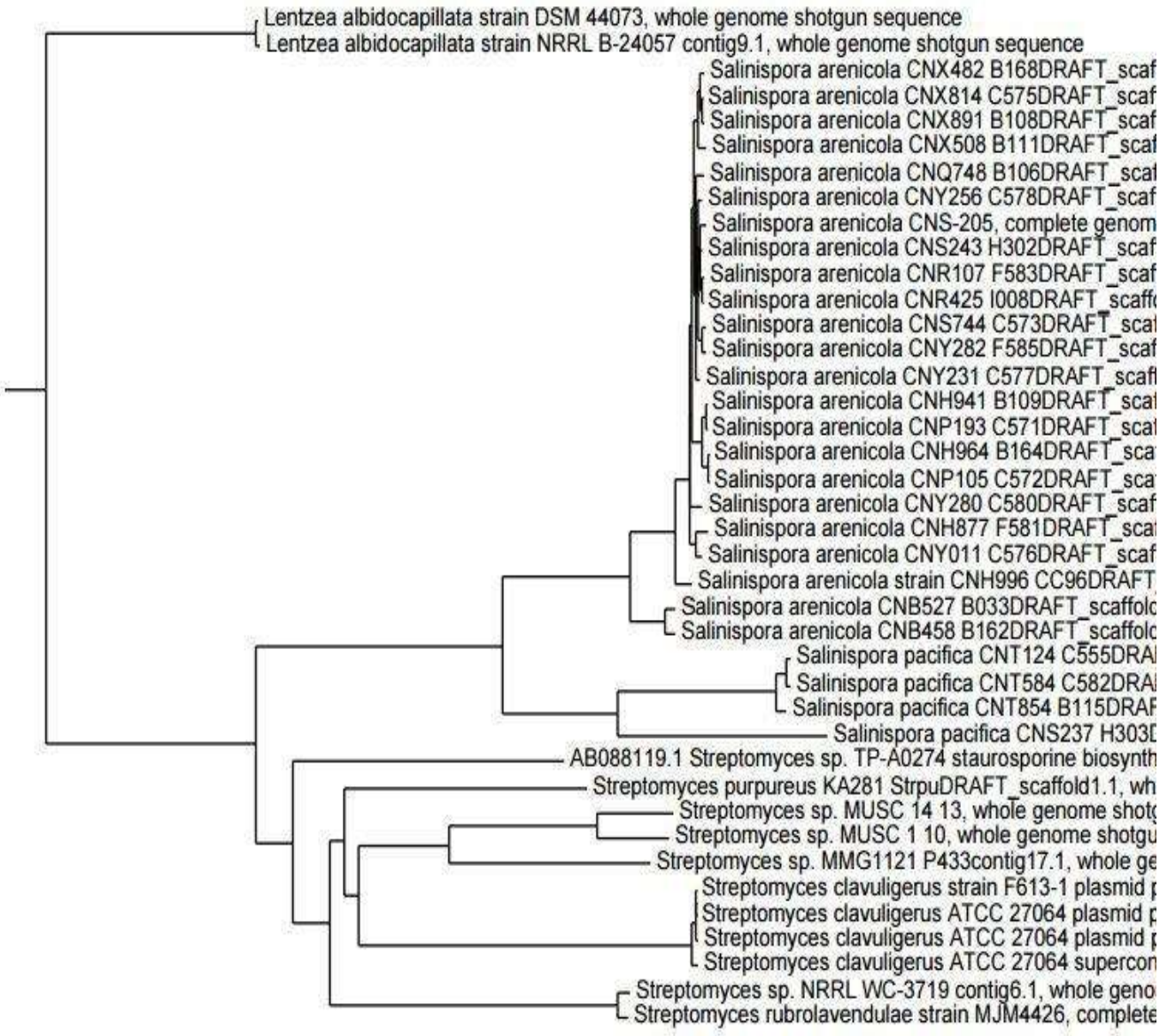


Figure 1: Phylogenetic tree for evolutionary study of staurosporine gene retrieved from NCBI database where each node showing a single species. Scale bar represents frequency of expected number of amino acid substitutions.

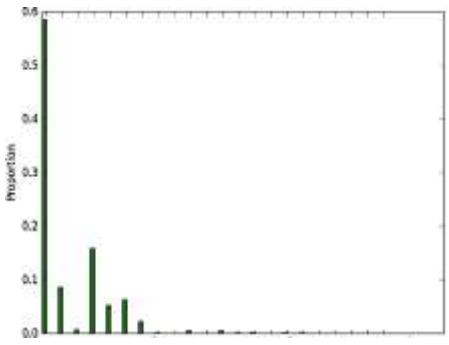
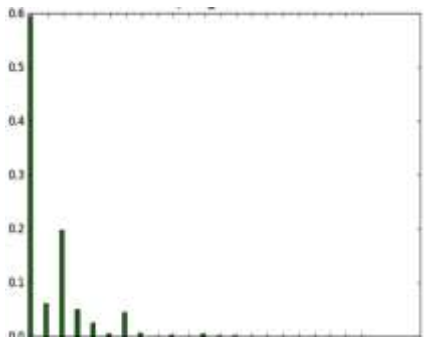
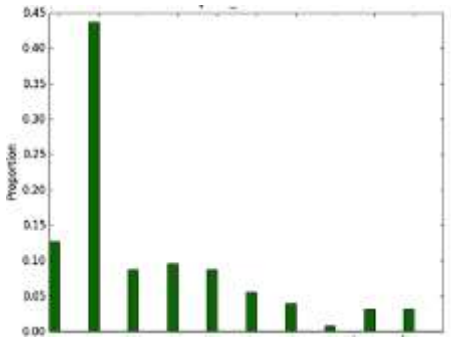
Now, the work proceeded by using those collected metagenomic samples as input for phyloshop software to classify the microorganism existing in the sample. The software predicted very positive results in the form of bar graph that each sample consists of Actinobacteria species which is listed in the table given below (refer table 2).

Table 2: List of metagenomic samples taken out from the database having approximately the same geographical distribution

Sample no.	Metagenomic Sample ID	Total No. of	Actinobacteria Position in the graph

		Species	
1	ERS667608	1696	5th bar
2	ERS667609	3808	7th bar
3	ERS667622	148	5th bar
4	ERS667623	1292	7th bar
5	ERS667625	1232	8th bar
6	ERS667629	1314	-----
7	ERS667630	1859	9th bar
8	ERS667633	992	6th bar
9	ERS667634	1046	7th bar
10	ERS667640	1046	6th bar
11	ERS667526	1148	4th bar
12	ERS667544	1688	4th bar

The software also generate histogram graphs shown in the figure below (refer figure 2), against each sample showing per cent concentration of existence of the genus actinobacteria (shown in bar representation) in that particular sample. Bars in the histogram are representing individual genus present in the sample. Actinobacteria present in the sample in the respective bar number which is listed in table number 2. In sample 1 (A) the actinobacteria is present in the 5th bar of the histogram, 7th bar is representing the actinobacteria in sample 2 (B), again 5th bar on the histogram of sample 3 (C) is showing the actinobacteria sample likewise in sample C, D, E, F, G, H, I, J and K the actinobacteria genus is present in the 7th, 8th, not defined, 9th, 6th, 7th, 6th, 4th and 4th bar of the given histogram which is shown in figure 2 below.

A	B	C
		
D	E	F

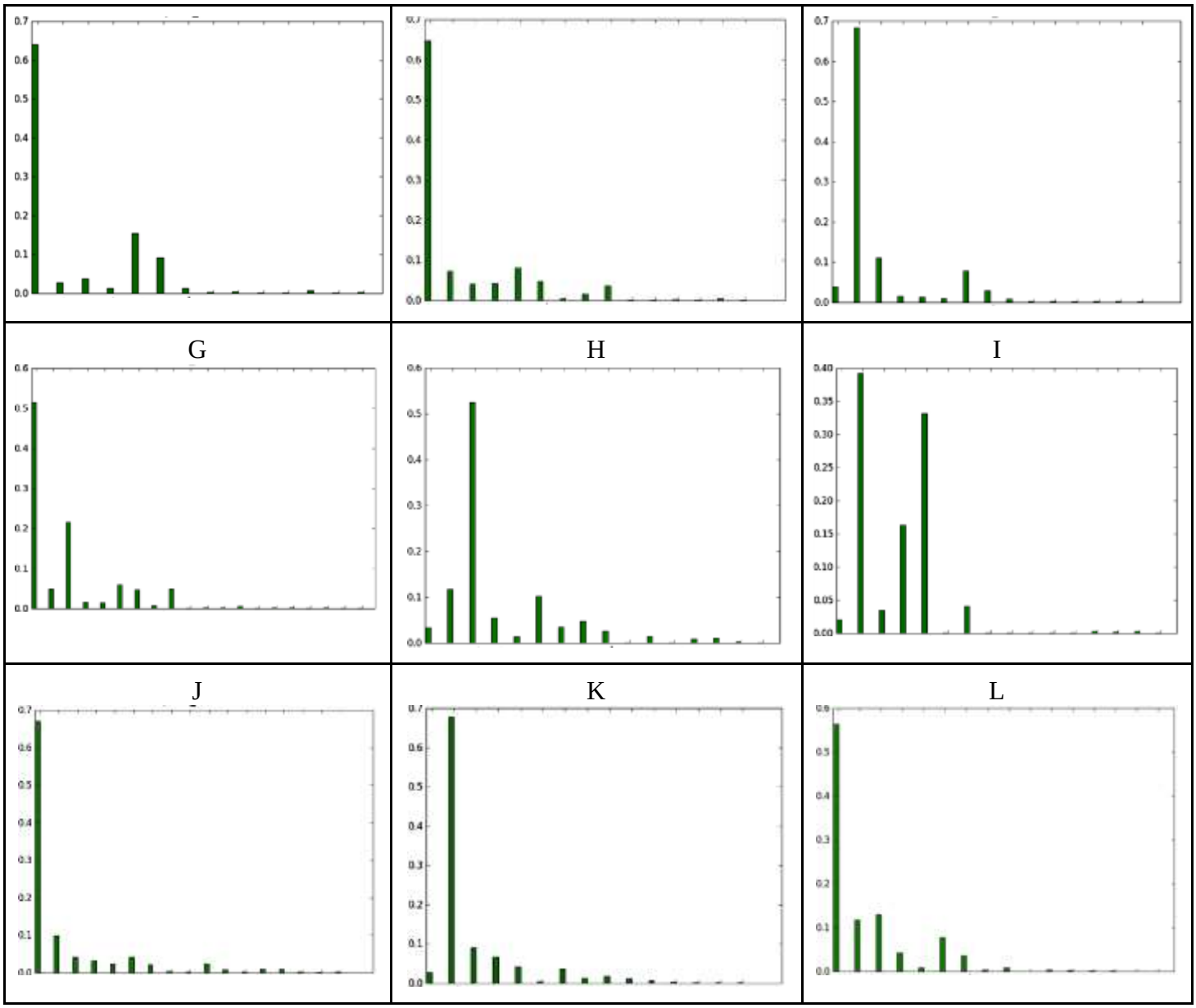


Figure 2: Graph showing phylum distribution of genus in all the samples. The X-axis represents the genus present in the sample and Y-axis represents the proportion of the existence.

The orthologous gene that duplicated and evolved through the process of evolution in prokaryotes is explored using the NCBI-COG database. Since StaD was not reported in the database so only StaO (COG0665) and StaP (COG2124) were taken into consideration. The list of genes was found in different phylum exhibiting orthologous genes ultimately showing the evolution of the gene which is listed in table numbers 3 and 4 for StaO and StaP gene respectively. There were a total of 25 phyla reported that contain StaO gene and 16 phyla containing StaP gene. The gene is evolved from very primitive microorganism i.e. Archaea to Firmicutes and also majorly the genes are found from those phyla that belong to marine niche.

Table no.3: List of genes that are orthology of StaO

S.I. no.	Phylum	Total species	Species consisting StaO gene
1	Crenarchaeota	21	18
2	Euryarchaeota	56	30
3	Thaumarchaeota	4	1
4	Other Archaea	2	1
5	Acidobacteria	6	6
6	Actinobacteria	74	63
7	Aquificae	8	5
8	Bacteroidetes	55	46
9	Chlorobi	5	1
10	Chlamydiae	6	6
11	Chloroflexi	9	7
12	Cyanobacteria	31	31
13	Deinococcus-Thermus	6	6
14	Bacilli	10	29
15	Clostridia	49	20
16	Mollicutes	10	0
17	Fusobacteria	5	1
18	Planctomycetes	6	6
19	Alphaproteobacteria	75	67
20	Betaproteobacteria	52	47
21	Deltaproteobacteria	28	13
22	Epsilonproteobacteria	11	9
23	Gammaproteobacteria	103	90
24	Spirochaetes	7	4
25	Synergistetes	5	3
26	Thermotogae	7	4

Table no.4: List of genes that are orthology of StaP

S.I. no.	Phylum	Total Species	Species consisting StaP gene
1	Crenarchaeota	21	1
2	Euryarchaeota	56	15
3	Thaumarchaeota	4	0
4	Other Archaea	2	0
5	Acidobacteria	6	6
6	Actinobacteria	74	54

7	Aquificae	8	0
8	Bacteroidetes	55	21
9	Chlorobi	5	0
10	Chlamydiae	6	1
11	Chloroflexi	9	5
12	Cyanobacteria	31	28
13	Deinococcus-Thermus	6	4
14	Bacilli	10	12
15	Clostridia	49	3
16	Mollicutes	10	0
17	Fusobacteria	5	0
18	Planctomycetes	6	6
19	Alphaproteobacteria	75	48
20	Betaproteobacteria	52	28
21	Deltaproteobacteria	28	6
22	Epsilonproteobacteria	11	1
23	Gammaproteobacteria	103	25
24	Spirochaetes	7	1
25	Synergistetes	5	0
26	Thermotogae	7	0

Randomly amino acid sequences from each phylum were chosen and extracted out and used to built-up a phylogenetic distance tree (figure no. 3 and 4) using Maximum Likelihood Algorithm with the help of MEGA X software. The evolutionary history was inferred using the JTT based matrix model [12]. The branch lengths of drawn trees are measured in the number of substitutions per site. The evolutionary analysis involved 25 amino acid sequences and 18 amino acid sequences for the StaP and StaO respectively.

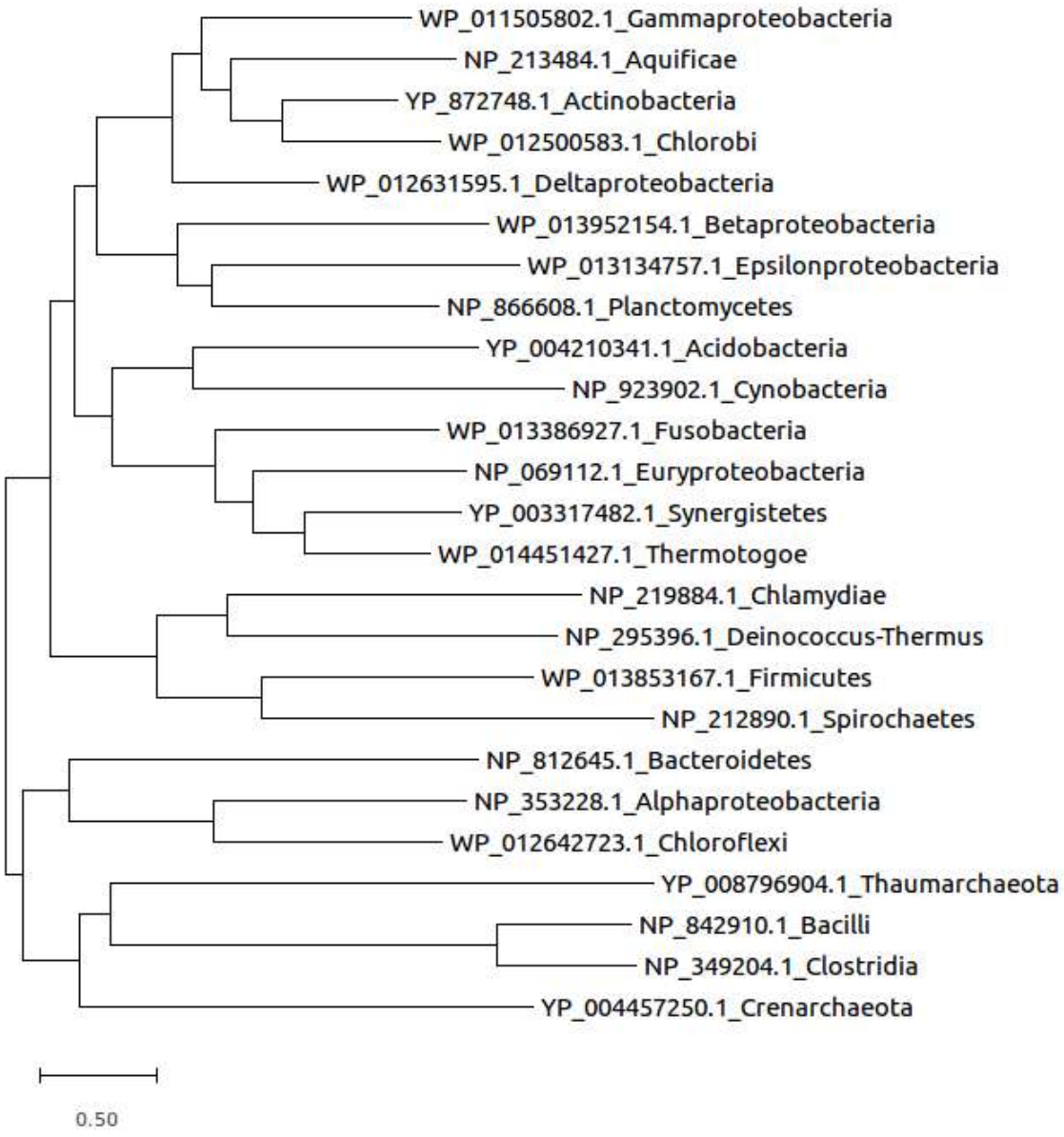


Figure 3: Phylogenetic tree of phylum consisting StaO orthologous genes based on Maximum Likelihood.

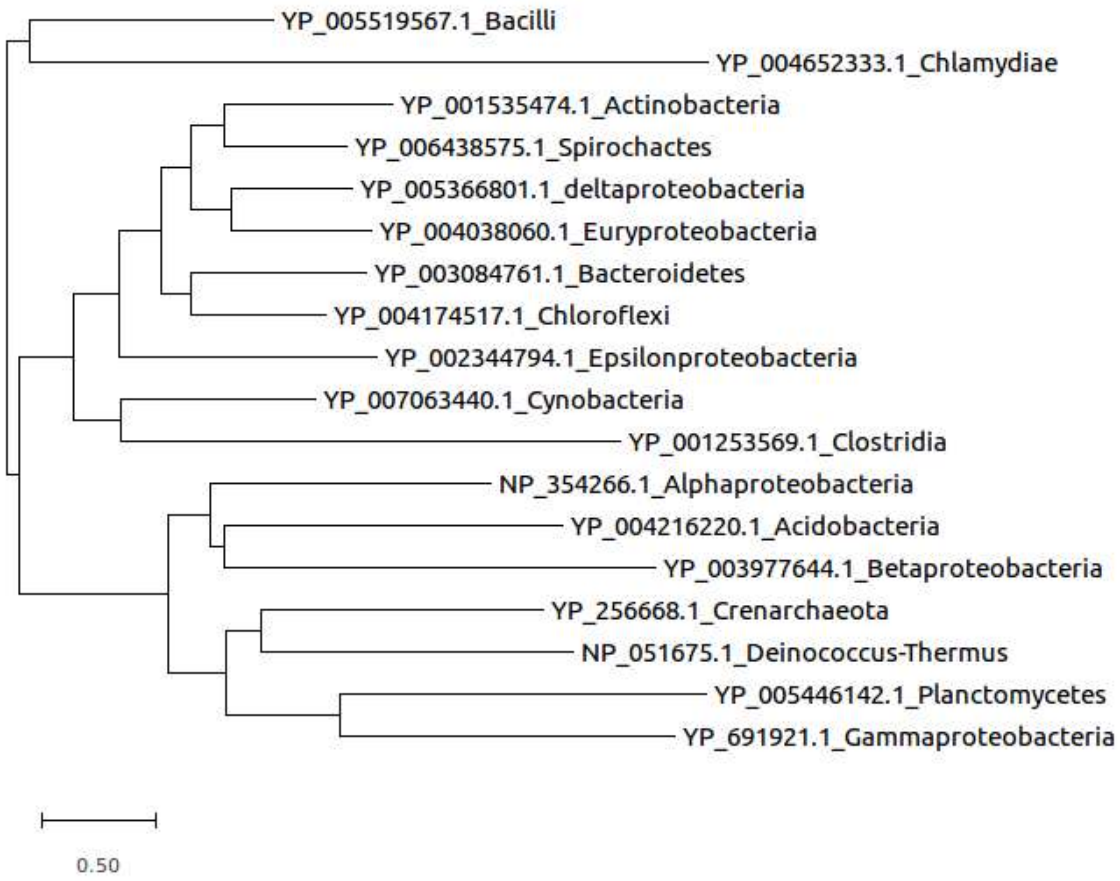


Figure 4: Phylogenetic tree of phylum consisting StaP orthologous genes based on Maximum Likelihood.

CONCLUSION AND FUTURE ASPECTS

The local alignment of the entire operon of the Staurosporine regulatory gene cluster, as well as individual genes, show the same result that marine residing Actinobacteria *S.arenicola* and *S. pacifica* has the maximum similarity with the Staurosporine regulatory gene. This indicates that regulatory gene has evolved whether from soil residing bacteria to marine residing bacteria and vice versa over a period of time.

The gene is still evolving with respect to time as suggested by metagenomics part of our study which is existing from Archaean period of geologic time scale. The Orthology of Staurosporine regulatory gene basically shows the diversity of genes belonging from Archaeal to bacterial domain.

The results of this study provide evidence of much wider geographical distribution and secondary metabolism diversity in this genus. Various anticancer compounds originate from natural products or from natural product-derived products. The variety of structures in products is key for new therapeutics.From the perspective of natural product discovery, archaeal screening of populations from distant locations can enhance the discovery of new natural products.

ACKNOWLEDGMENT

I am grateful to the Central University of South Bihar for providing me platform to carry forward this study. I sincerely give thanks to my mentor Dr. Krishna Kumar Ojha for his guidance and my family and friends for supporting me.

REFERENCES

1. Nakano, H., & Ōmura, S. (2009). Chemical biology of natural indolocarbazole products: 30 years since the discovery of staurosporine. *The Journal of antibiotics*, *62*(1), 17.
2. Chang, F. Y., & Brady, S. F. (2011). Cloning and characterization of an environmental DNA-derived gene cluster that encodes the biosynthesis of the antitumor substance BE-54017. *Journal of the American Chemical Society*, *133*(26), 9996-9999.
3. Edlund, A., Loesgen, S., Fenical, W., & Jensen, P. R. (2011). Geographic distribution of secondary metabolite genes in the marine actinomycete Salinispora arenicola. *Appl. Environ. Microbiol.*, *77*(17), 5916-5925.
4. Maldonado, L. A., Fenical, W., Jensen, P. R., Kauffman, C. A., Mincer, T. J., Ward, A. C., ... & Goodfellow, M. (2005). Salinispora arenicola gen. nov., sp. nov. and Salinispora tropica sp. nov., obligate marine actinomycetes belonging to the family Micromonosporaceae. *International Journal of Systematic and Evolutionary Microbiology*, *55*(5), 1759-1766.
5. Goo, K. S., Tsuda, M., & Ulanova, D. (2014). Salinispora arenicola from temperate marine sediments: new intra-species variations and atypical distribution of secondary metabolic genes. *Antonie Van Leeuwenhoek*, *105*(1), 207-219.
6. Bashir, Y., Singh, S. P., & Konwar, B. K. (2014). Metagenomics: An Application Based Perspective. *Chinese Journal of Biology, 2014*, 1-7.
7. Handelsman, J. (2004). Metagenomics: application of genomics to uncultured microorganisms. *Microbiol. Mol. Biol. Rev.*, *68*(4), 669-685.
8. Shah, N., Tang, H., Doak, T. G., & Ye, Y. (2011). Comparing bacterial communities inferred from 16S rRNA gene sequencing and shotgun metagenomics. In *Biocomputing 2011* (pp. 165-176).
9. Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of molecular biology*, *215*(3), 403-410.
10. Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M., & Tanabe, M. (2015). KEGG as a reference resource for gene and protein annotation. *Nucleic acids research*, *44*(D1), D457-D462.
11. Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). MEGA X: molecular evolutionary genetics analysis across computing platforms. *Molecular biology and evolution*, *35*(6), 1547-1549
12. Jones, D. T., Taylor, W. R., & Thornton, J. M. (1992). The rapid generation of mutation data matrices from protein sequences. *Bioinformatics*, *8*(3), 275-282.