ejpmr, 2020,7(4), 614-621

SJIF Impact Factor 6.222

EUROPEAN JOURNAL OF PHARMACEUTICAL AND MEDICAL RESEARCH

www.ejpmr.com

Research Article

ISSN 2394-3211

EJPMR

INSILICO STUDY OF 'DNA POLYMERASE I' IN *CAUDOVIRALES'S* FAMILY WITH THEIR HOST CELL TO CONCLUDE THE INVOLVEMENT OF HORIZONTAL GENE TRANSFER

[1]*Subodh Choukidar, [1]Dr. Sanjay Harke and [2]Dr. Talib Y.A.

[1]Dept of Bioinformatics, MGM Institute of Biosciences and Technology, MGM Campus, Aurangabad.(M.S) 431001.
[2]Dept of Biotechnology and Bioinformatics, Dr. Rafiq Zakaria Campus, Maulana Azad College, Aurangabad. (M.S) 431001.

***Corresponding Author: Subodh Choukidar**

Dept of Bioinformatics, MGM Institute of Biosciences and Technology, MGM Campus, Aurangabad.(M.S) 431001.

Article Received on 10/02/2020 **Article Revised on 02/03/2020** **Article Accepted on 23/03/2020**

ABSTRACT

Viruses which infects bacterium are known as bacteriophages. They utilize host's machinery to encode their own proteins needed to develop the new progeny. They are the vectors of horizontal gene transfer and drivers of bacterial evolution. A phage can exhibits two types of life cycles one is lytic and another is lysogenic. In lytic life cycle phage infects and rapidly kill the bacterial host via the process of generalized transduction. In lysogenic life cycle phage instead of killing the host integrates in its genome or exists as plasmid in the host cell. bacteriophages and bacteria are co-existed and evolved together in evolutionary time. *Caudovirales* is an order of tailed bacteriophages also known as class I bacteriophages or DNA phages. They exists in three families namely *myoviridae*, *siphoviridae*, and *podoviridae*. bioinformatics is playing significant role in all the areas of life sciences specially in evolutionary studies and tertiary structure prediction of macromolecules like protein, With the help of various algorithms and methods. Here in this paper an attempt is made to identify the horizontal gene transfer events in these three families of *caudovirales* and their hosts using 'DNA polymerase I' as a molecular marker. The said study is to implement the correlation between bacteria & *caudovirales* with the aid of advanced computational Biology named Bioinformatics which will help to pull the evolutionary relationship, thier structure and able to predict the functional elements in the said samples. SWISS model(Automated protein structure homology-modelling server),MEGA (Molecular Evolutionary Genetics Analysis), DALI (Network service for comparing protein structures in 3D) above said server and program is used to analyse the activity.

KEYWORDS: Bacteriophage, Bacteria, *Caudovirales*, HGT, Structure prediction, Structure alignment, Phylogenetic tree construction.

INTRODUCTION

bacteriophages's and bacterial relationship

'Phages' were firstly discovered by Williom twort and felix d'Herelle in 1915 and 1917 respectively, realized that they had a potential capacity to kill bacteria. Phages are the most abundantly available on biosphere. The name given bacteriophage because, they infect and kill bacteria. They are very species specific, usually infect a single bacterial species or specific strains within a species .The infection is done with help of their tail fibres and bacterial receptors. After the successful penetration of DNA into host's cell they generally exhibits lytic and lysogenic types of life cycles. In lytic life cycle phages infects and rapidly kill the bacterial host via the process of generalized transduction. In lysogenic life cycle phages instead of killing the host integrates in its genome or exists as plasmid in the host cell. Phages are the vectors of horizontal gene transfer and drivers of bacterial evolution. Evidences shows that the phages and bacteria are co-existed and evolved together in evolutionary time. Bacteria develop different mechanisms to prevent the infection like bacterial receptor modification, degradation of invading phage DNA. But, phages circumvent the resistance and evolve mechanisms to target such resistance of bacteria. Studies shows that the 20% of the bacterial genetic content is acquired. Genetic homology of nucleotide and protein sequences of phages and their bacterial hosts can be used to identify the horizontal gene transfer events because, these sequences may represent the sequences were acquired by phage during past infection event.

Caudovirales

'Caudovirales' is an order of bacteriophages also known as 'tailed bacteriophages'. They are also known as DNA phages due to presence of double stranded DNA as a genetic material. They have tail with which they can easily attach to surface receptors of bacteria to inject the

DNA into host cell. They have three different families namely *myoviridae*, *siphoviridae*, *podoviridae*. Each family have different morphological structures of tail. *myoviridae* have long contractile tail, *siphoviridae* have long non-contractile tail, *podoviridae* have short non-contractile tail. Their genomic size can be 18 to 500 kb in length. *Caudovirales* account for the 95% of the phages reported in scientific literature and 90% of approximately 6,200 phages examined under electron microscopy(EM).

Bioinformatics and Phylogenetics

The term 'bioinformatics' was coined by Paulien Hogeweg and Ben Hesper. It is new science of studying, retrieving and analysing large amount of biological data/information. It is highly interdisciplinary field of science, and playing key role in all the areas of life sciences specially in the evolutionary studies. The study of evolutionary relationship among individuals or group of organisms is called 'phylogenetics'. Various anatomical methods used by taxonomists take too much time. By using bioinformatics, phylogenetic tree is constructed based on the alignment of the nucleotide or protein sequences using various algorithms and methods. There are various algorithms and methods developed for the construction of phylogenetic tree. Although, which algorithm is to be used for the study depends on the evolutionary lineages. To study the evolutionary relationship between phages and their hosts we need to take a set of phages with their specific known hosts.

What is horizontal gene transfer?

Horizontal gene transfer is the process in which the genetic information is transferred between organisms, rather than parents to offspring. In the case of bacteriophage and bacteria, the genes can be transfer horizontally by the process of transduction in which bacteriophage is a vector. It is said that, the genes other than 'drug resistance genes' can be transfer horizontally and multiply by natural selection.

DNA polymerase I as molecular marker

For construction of phylogenetic tree and to study horizontal gene transfer events with the help of molecular markers we need to either choose nucleotide or protein sequence. Selected markers can make a major difference in obtaining a correct tree. For studying closely related organisms we should use nucleotide sequence because, nucleotide sequences evolve rapidly than protein sequence. In this study of evolutionary relationship between phages (*caudovirales*) and their specific hosts, which are widely divergent group of microorganisms we selected protein sequence as molecular marker. Because, protein sequences are more conserved and can be used to for distantly related organisms. It is said that, the amino acid sequence alignments of 'DNA polymerases' can show frequent events of horizontal gene transfer in *caudovirales*. And also their are no genes which are conserved within all phages like *caudovirales*. There fore we have selected 'DNA polymerase I' as molecular marker for this study. Because, 'DNA polymerase I' plays role in processing RNA primers during synthesis of lagging-strand and fills small gaps during 'DNA repair' reaction mechanisms. The protein sequences and pairs of three phages from three different families of *caudovirales* (*myoviridae*, *siphoviridae*, *podoviridae*) with their known hosts are taken from the NCBI's (National Center for Biotechnology Information) taxonomy browser and *Virus-Host* database (https://www.genome.jp/virushostdb/) ,which organizes data about the relationships between viruses and their hosts, represented in the form of pairs of NCBI taxonomy IDs for viruses and their hosts [Table no. 01].

Table no. 01: Table showing the pairs of *caudovirales's* three different family phages with their specific known hosts with their accession id's of protein sequence taken from Virus-host database and NCBI's taxonomy browser.

Sr no.	Bacteriophage	Order/family	Accession ID	Bacterial host	Order/family	Accession ID
1	*Bacillus phage BalMu-1*	Viruses; Caudovirales; Myoviridae.	NCBI Reference Sequence: YP_009276820.1	*Bacillus alcalophilus*	Bacteria; Firmicutes; Bacilli; Bacillales; Bacillaceae; Bacillus.	NCBI Reference Sequence: WP_040323772.1
2	*Arthrobacter phage Amigo*	Viruses; Caudovirales; Siphoviridae; Amigovirus.	GenBank: ALY08405.1	*Arthrobacter sp. ATCC 21022*	Bacteria; Actinobacteria; Micrococcales; Micrococcaceae; Arthrobacter.	GenBank: AMB40537.1
3	*Erwinia phage Ea9-2*	Viruses; Caudovirales; Podoviridae; Johnsonvirus.	NCBI Reference Sequence: YP_009007430.1	*Erwinia amylovora*	Bacteria; Proteobacteria; Gamma proteobacteria; Enterobacterales; Erwiniaceae; Erwinia.	NCBI Reference Sequence: WP_004167759.1

Structure prediction

Protein tertiary structure prediction provides basis of understanding its function. Experimental techniques like X-ray crystallography and NMR (Nuclear Magnetic Resonance), are expensive and time consuming compared to computational methods. Computational methods are relatively cheaper and faster. There are three methods for prediction of tertiary structure of protein Homology modelling, AB initio and Threading. Homology modelling builds an atomic model from amino acid sequence known as 'query sequence' of interest, based on sequence homology with known experimentally determined structures known as 'templates' which are closely related at sequence level. The principle behind homology modelling is that 'if two proteins share a high enough sequence similarity they are likely to have very similar three dimensional structures'. If one protein has known structure, then the structure can be copied to the unknown protein structure with high degree of confidence. The structures of 'DNA polymerase I' are predicted with the help of 'SWISS model', which is online tool for homology modelling developed by Torsten Schwede's structural bioinformatics group. In this process three steps are involved template recognition, target-template alignment, model building. Three dimensional structures of phage's and their host's 'DNA polymerase I' are predicted [Table no. 02.1 and 02.2].QMEAN (Qualitative Model Energy ANalysis) Z-scores provides 'degree of nativeness' of the structural features observed in the model on global scale. QMEAN Z-scores around zero (0) indicate good agreement between the model structure and experimental structures of similar size[Graph no. 01]. 'Ramachandran favoured scores' generated by the Ramachandran plots of predicted structures ranging from '80% - 90%' are also showing good quality of predicted structures [Ramachandran plots are given with species name and Ramachandran favoured scores in Graph A,B,C,D,E & F].

Table no. 02.1: Table Showing species name with predicted structures and Qmean score.

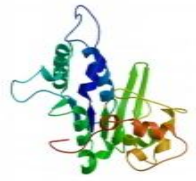
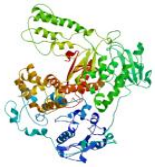
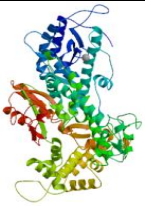
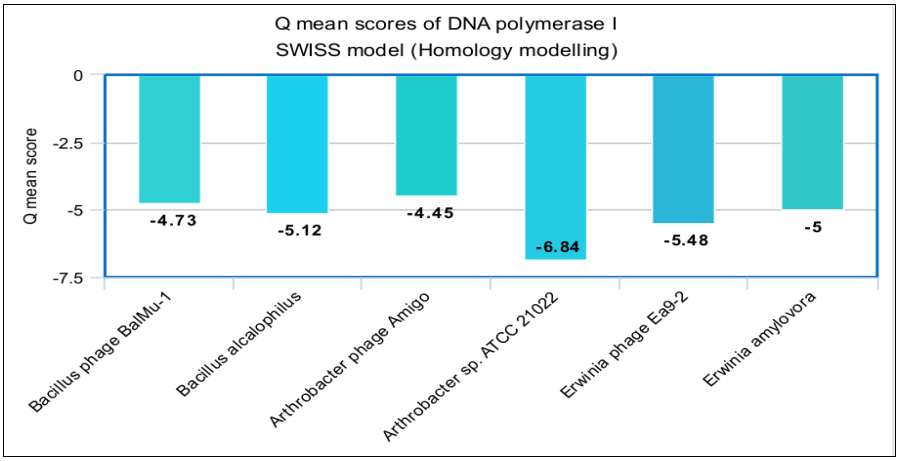
Species name	Q mean score	Predicted Structure
Bacillus phage BalMu-1 *(Viruses; Caudovirales; Myoviridae.)*	**-4.73**	
Bacillus alcalophilus *(Bacteria; Firmicutes; Bacilli; Bacillales; Bacillaceae; Bacillus.)*	**-5.12**	
Arthrobacter phage Amigo *(Viruses; Caudovirales; Siphoviridae; Amigovirus.)*	**-4.45**	

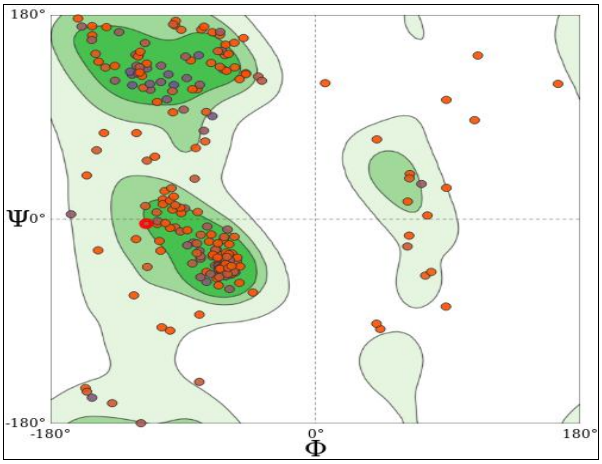
Table no. 02.2: Table Showing species name with predicted structures and Qmean scores.

Species name	Q mean score	Predicted Structure
Arthrobacter sp. ATCC 21022 *(Bacteria; Actinobacteria; Micrococcales; Micrococcaceae; Arthrobacter.)*	**-6.84**	
Erwinia phage Ea9-2 *(Viruses; Caudovirales; Podoviridae; Johnsonvirus.)*	**-5.48**	

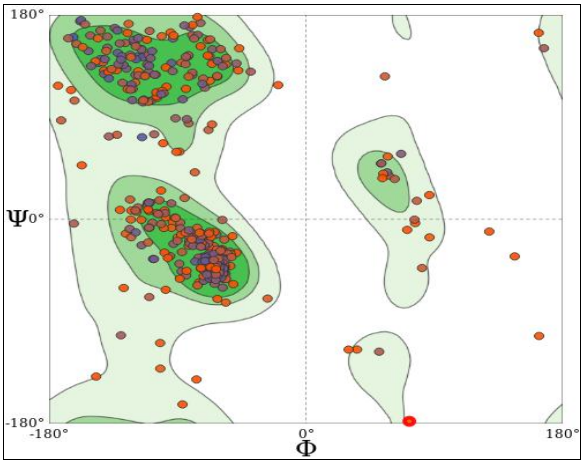
Erwinia amylovora *(Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacterales; Erwiniaceae; Erwinia.)*	**-5**	



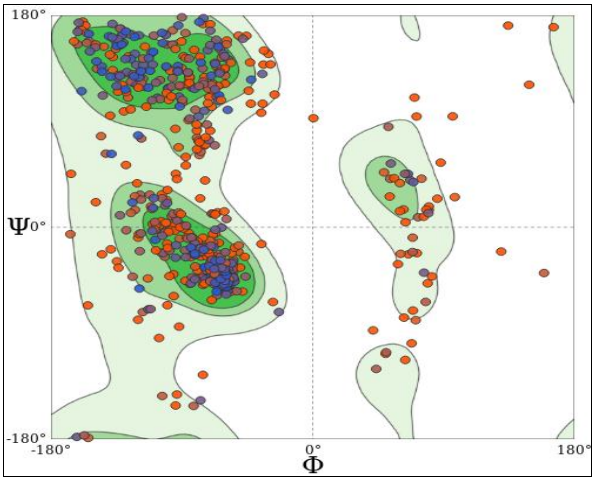
Graph no. 01: Graph Showing Q-mean scores of predicted structures of 'DNA polymerase I'.



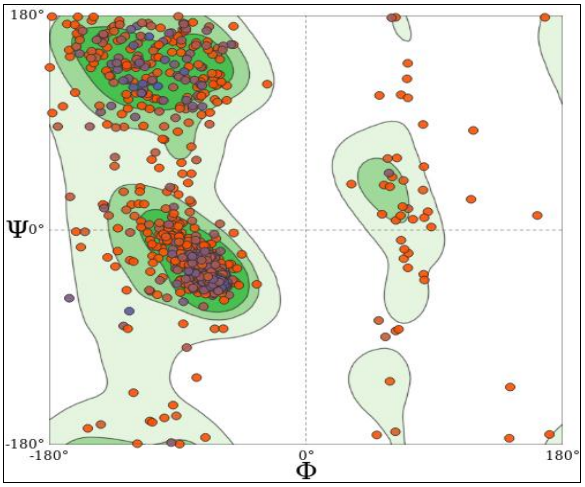
Graph (A): Showing general Ramachandran plot for *Bacillus phage BalMu-1*



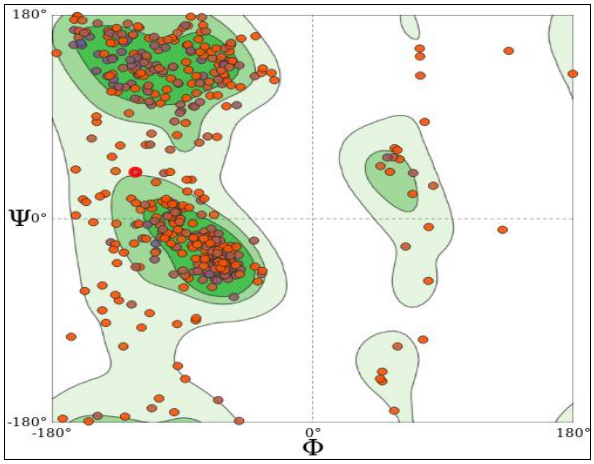
Graph (C): Showing general Ramachandran plot for *Arthrobacter phage Amigo*



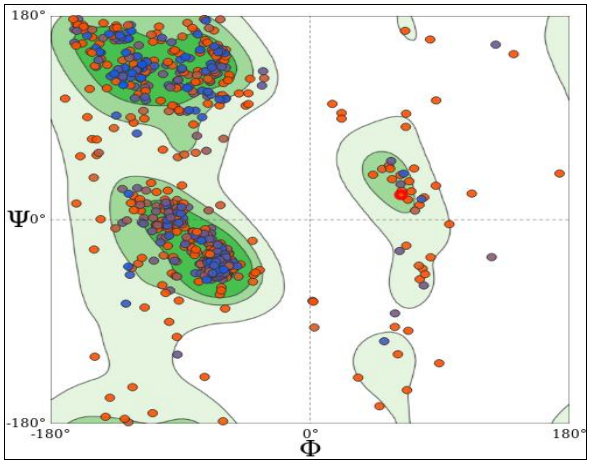
Graph (B): Showing general Ramachandran plot for *Bacillus alcalophilus.*



Graph (D): Showing general Ramachandran plot for *Arthrobacter sp. ATCC 21022.*



Graph (E): Showing general Ramachandran plot for *Erwinia phage Ea9-2.*



Graph (F): Showing general Ramachandran plot for *Erwinia amylovora.*

Graph (A): Ramachandran favoured score of *Bacillus phage BalMu-1's* predicted structure of 'DNA Polymerase I' is 81.82 %, showing good quality of structure.

Graph (B): Ramachandran favoured score of *Bacillus alcalophilus's* predicted structure of 'DNA Polymerase I' is 85.37 %, showing good quality of structure.

Graph (C): Ramachandran favoured score of *Arthrobacter phage Amigo's* predicted structure of 'DNA Polymerase I' is 90.03 %, showing good quality of structure.

Graph (D): Ramachandran favoured score of *Arthrobacter sp. ATCC 21022's* predicted structure of 'DNA Polymerase I' is 85.6%, showing good quality of structure.

Graph (E): Ramachandran favoured score of *Erwinia phage Ea9-2's* predicted structure of 'DNA Polymerase I' is 84.58 %, showing good quality of structure.

Graph (F): Ramachandran favoured score of *Erwinia amylovora's* predicted structure of 'DNA Polymerase I' is 87.64%, showing good quality of structure.

Structure alignment

Protein structural alignment provides a functional basis for deriving principles for functional and evolutionary relationships. Structure alignment of 'DNA polymerase-I' of phages with their specific hosts taken for the study are done with the help of 'DALI server' (Network service for comparing protein structures in 3D). And their alignment scores are given in the forms of Z-score and identity score. The Z-scores which are more than two (2) clearly indicates the significant similarities found in the predicted structures of phage's and their bacterial host's 'DNA polymerase I' [Table no. 03].

Table no. 03: Table Showing structure alignment of predicted 'DNA polymerase I' structures of phages's with their host's.

Bacteriophage	Bacterial host	Z score	Identity score
Bacillus phage BalMu-1	*Bacillus alcalophilus*	19.9	37 %
Arthrobacter phage Amigo	*Arthrobacter sp. ATCC 21022*	27.6	19 %
Erwinia phage Ea9-2	*Erwinia amylovora*	21.9	19 %

Phylogenetic tree construction

Phylogenetics is the study of the evolutionary history of living organisms. Pedigrees of these organisms are represented in tree like diagrams. The tree can be rooted or unrooted. Rooted trees are more informative than unrooted trees, because unrooted trees do not have the direction of an evolutionary path. There are two types of tree construction methods distance based and character based. Distance based methods generates tree based on amount of dissimilarities between pairs of sequences which are computed based on sequence alignment. UPGMA and NJ are the algorithms of distance based methods. In other hand character based methods are based on sequence characters. Character based method algorithms are of two types : Maximum Likelihood and Maximum Parsimony [figure no. 01].

Maximum likelihood : ML finds a tree that most likely to reflects the actual evolutionary process. It is exhaustive method which searches every tree topology. It also considers all the positions in an alignment, not just informative sites.

So, when we are trying to study the evolutionary relationship between two different species we must use maximum likelihood algorithm to construct phylogenetic tree.

Maximum parsimony: Parsimony method of tree construction chooses a tree that has fewest evolutionary changes or shortest overall branch length.

Here in this study we used the offline MEGA software (Molecular Evolutionary Genetics Analysis) for the

multiple alignment and phylogenetic tree construction. Firstly sequences are stored in FASTA file format (.fasta) and then aligned by using clustalW algorithm of MEGA. And then the alignment is exported in the mega format (.meg file). Then using 'phylogeny tool bar' with the help of exported (.meg) file, Maximum likelihood and maximum parsimony trees are generated [Figure no. 02 and 03]. The representation used in the tree, for three different families of *caudovirales* and their respective hosts are given [Table no. 04].

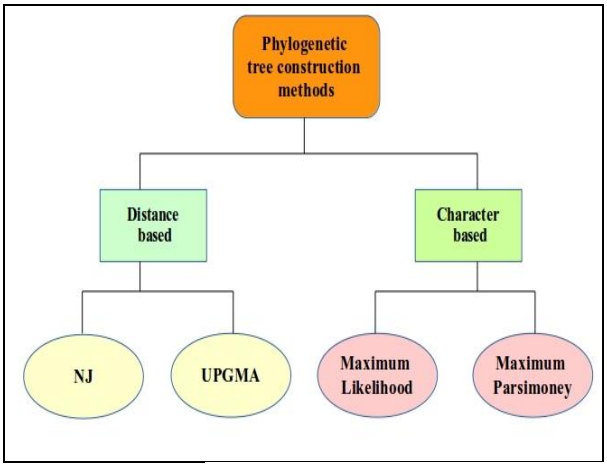


Figure no. 01: Figure showing phylogenetic tree construction methods.

Table no. 04: Table Showing the representation used in this study for tree construction.

Bacteriophage	Representation used for tree construction	Bacterial host	Representation used for tree construction
Bacillus phage BalMu-1	Phage 1	Bacillus alcalophilus	Host 1
Arthrobacter phage Amigo	Phage 2	Arthrobacter sp. ATCC 21022	Host 2
Erwinia phage Ea9-2	Phage 3	Erwinia amylovora	Host 3

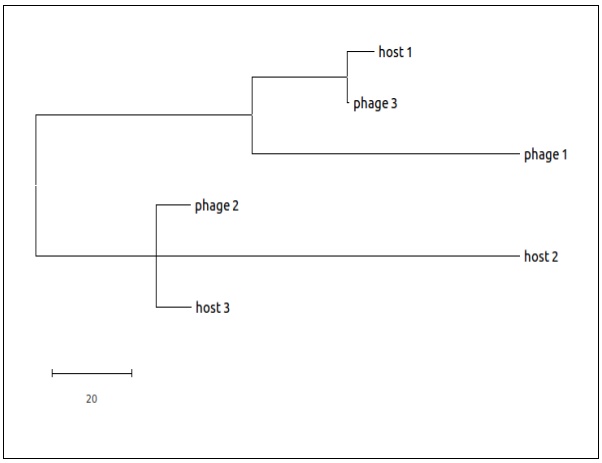


Figure no 02: Phylogenetic tree by using Maximum likelihood.

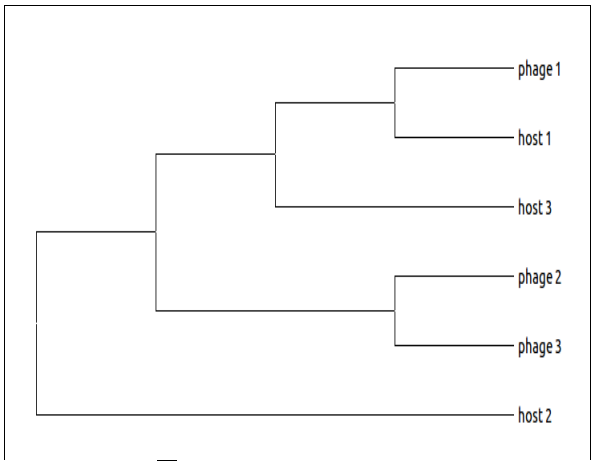


Figure no 03: Phylogenetic tree by using Maximum parsimony.

CONCLUSION

Study is based on the secondary and tertiary structure prediction of 'DNA polymerase I' taken as a molecular marker from caudovirales's family phages and their specific hosts. QMEAN scores and Ramachandran favoured scores showing good quality of predicted structures. Further, structures are aligned in the pairs of phage and it's specific host, in order to get the similarities between the structures. The Z-scores indicating that, significant similarities found in the phage's and its specific host's 'DNA polymerase I'. After that the sequence alignment is done by using offline MEGA software and phylogenetic trees are constructed by using character based methods.

In maximum likelihood tree 'phage-1' which is representing *myoviridae* family i.e. *Bacillus phage BalMu-1* and 'phage-2' which representing *siphoviridae* family i.e. *Arthrobacter phage Amigo* are showing clear relationship (as they are in single clade) with their hosts represented as 'host-1' and 'host-2' i.e. *Bacillus alcalophilus* and *Arthrobacter sp. ATCC 21022* respectively. But 'phage-3' which is representing *podoviridae* family i.e. *Erwinia phage Ea9-2* not showing relationship with its host represented as 'host-3' i.e. *Erwinia amylovora*. It can be due to the 'host switching' may occurred by horizontal gene transfer in closely related phages (i.e. phage 2 and phage 3 – their relation is shown in parsimony).

In Maximum parsimony tree 'phage-1' is showing clear relationship with its host. But 'phage 3' and 'Phage-2' are not showing relationship with their hosts 'host-2'and 'host-3' respectively. Because, parsimony allows fewest evolutionary changes So that 'phage-2' and 'phage-3' which are closely related are shown in one clade.

From the all results of structure prediction, structure alignment and phylogenetic tree we have reached to conclude that, the horizontal gene transfer events found in 'DNA polymerase I' of *caudovirales's* family phages and their specific hosts. The genes which are coding for 'DNA polymerase I' may transferred horizontally during co-evolution and life cycles of phage. after interpreting the evolutionary phylogenetic tree relationship we can comment, the respective phages of three different families of *caudovirales* and their specific hosts belongs to class of 'xenologs' between their studied species.

REFERENCES

1. Martha RJ Clokie, Andrew D Millard, Andrey V Letarov, ShaunHeaphy.Phages in nature. Bacteriophage, Jan-Feb, 2011; 1(1): 31–45. [PMC3109452].
2. Robert A. Edwards, Katelyn McNair, Karoline Faust, Jeroen Raes, Bas E. Dutilh. Computational approaches to predict bacteriophage–host relationships.FEMS Microbiol Rev., Mar 1, 2016; 40(2): 258–272. [PMC5831537].
3. Jingfen Zhang, Zhiquan He, Qingguo Wang, Bogdan Barz, Ioan Kosztin, Yi Shang, Dong Xu.Prediction of protein tertiary structures using MUFOLD. Published in final edited form as: Methods Mol Biol., 2012; 815: 3–13. [PMC4365503].
4. Haiyou Deng, Ya Jia, Yang Zhang. Int J Mod Phys B. Protein structure prediction, Jul 20, 2018; 32(18): 1840009. [PMC6407873].
5. Walter Pirovano, K Anton Feenstra, Jaap Heringa.The meaning of alignment: lessons from structural diversity.BMC Bioinformatics, 2008; 9: 556. [PMC2630330].
6. L. Holm, S. Kääriäinen, P. Rosenström, A. Schenkel. Searching protein structure databases with DaliLite v.3.Bioinformatics, Dec 1, 2008; 24(23): 2780–2781. [PMC2639270].
7. Soraya Chaturongakul, Puey Ounjai.Phage–host interplay: examples from tailed phages and Gram-negative bacterial pathogens. Front Microbiol, 2014; 5: 442. [PMC4138488].
8. Alita R. Burmeister. Horizontal Gene Transfer. Evol Med Public Health, 2015; 2015(1): 193–194. [PMC4536854].
9. Laura M. Kasman; La Donna Porter. Bacteriophages. StatPearls. [NBK493185].
10. Ackermann HW. Tailed bacteriophages: the order caudovirales. Adv Virus Res., 1998; 51: 135-201. [Pubmed: 9891587].
11. Elena V Orlova. How viruses infect bacteria?. EMBO J., Apr 8, 2009; 28(7): 797–798. [PMC2670874].
12. Kumar A, Chordia N. Role of Bioinformatics in Biotechnology. Res Rev Biosci, 2017; 12(1): 116.
13. Jennifer M. Allen, David M. Simcha, Nolan G. Ericson, David L. Alexander, Jacob T. Marquette, Benjamin P. Van Biber, Chris J. Troll, Rachel Karchin, Jason H. Bielas, Lawrence A. Roles of DNA polymerase I in leading and lagging-strand replication defined by a high-resolution mutation footprint of ColE1 plasmid replication .Nucleic Acids Res., Sep, 2011; 39(16): 7020–7033.[PMC3167613].
14. Edwards YJ, Cottage A. Bioinformatics methods to predict protein structure and function. A practical approach. Mol Biotechnol, Feb, 2003; 23(2): 139-66. [PMID: 12632698].
15. Andrew Waterhouse, Martino Bertoni, Stefan Bienert, Gabriel Studer, Gerardo Tauriello, Rafal Gumienny, Florian T Heer, Tjaart A P de Beer, Christine Rempfer, Lorenza Bordoli, Rosalba Lepore, Torsten Schwede. SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Res., Jul 2, 2018; 46(Web Server issue): W296–W303. [PMC6030848].
16. Vincent Daubin, Gergely J. Szöllősi. Horizontal Gene Transfer and the History of Life. Cold Spring Harb Perspect Biol., Apr, 2016; 8(4): a018036. [PMC4817804].
17. Koonin EV, Makarova KS, Aravind L. Horizontal gene transfer in prokaryotes: quantification and classification. Annu Rev Microbiol, 2001; 55: 709-42. [PMID: 11544372].
18. Tom Burr, " Phylogenetic Trees in Bioinformatics", Current Bioinformatics, 2010; 5: 40. https://doi.org/10.2174/157489310790596367.
19. J Felsenstein. Maximum-likelihood estimation of evolutionary trees from continuous characters. Am J Hum Genet, Sep, 1973; 25(5): 471–492. [PMC1762641].
20. Mount DW. Maximum parsimony method for phylogenetic prediction. CSH Protoc, Apr 1, 2008; 2008:pdb.top32. [PMID: 21356815].
21. Anne Lopes, Paulo Tavares, Marie-Agnès Petit, Raphaël Guérois, Sophie Zinn-Justin. Automated classification of tailed bacteriophages according to their neck organization. BMC Genomics, 2014; 15(1): 1027. [PMC4362835].
22. Kasman LM, Porter LD. Bacteriophages .StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing, 2019-. 2019 Aug 12. [PMID: 29630237].
23. Robin G. Allaby, Mathew Woodwark. Phylogenetics in the Bioinformatics Culture of Understanding. Comp Funct Genomics, Mar, 2004; 5(2): 128–146. [PMC2447345].
24. Kumar S, Tamura K, Nei M. MEGA: Molecular Evolutionary Genetics Analysis software for

microcomputers. Comput Appl Biosci, Apr, 1994; 10(2): 189-91. [PMID: 8019868].
25. Benkert P, Tosatto SC, Schomburg D. QMEAN: A comprehensive scoring function for model quality assessment. Proteins, Apr, 2008; 71(1): 261-77. [PMID: 17932912].
26. Fiona S.L. Brinkman. "Phylogenetic Analysis".In: Andreas D. Baxevanis, B.F. Francis Ouellette . Bioinformatics : A practical guide to the analysis of genes and proteins .3rd ed. Wiley-India edition.S.P. Printers: Delhi., 2011; 465-92.
27. Jin Xiong. "Phylogenetics Basics". Essential Bioinformatics.1st ed. First South Asian edition. Replika Press Pvt. Ltd.: India., 2007; 127-41.
28. Jin Xiong. "Protein Tertiary Structure Prediction". Essential Bioinformatics.1st ed. First South Asian edition. Replika Press Pvt. Ltd.: India, 2007; 200-213.
29. Charlotte A Darby, Maureen Stolzer, Patrick J Ropp, Daniel Barker, Dannie Durand. Xenolog classification. Bioinformatics, Mar 1, 2017; 33(5): 640–649.[PMC5860392].
30. Mihara, T., Nishimura, Y., Shimizu, Y., Nishiyama, H., Yoshikawa, G., Uehara, H., Hingamp, P., Goto, S., and Ogata, H.; Linking virus genomes with host taxonomy. Viruses, 2016; 8: 66. doi:10.3390/v8030066 [PMID: 26938550]
31. Phylogenetic and molecular evolutionary analyses were conducted using MEGA version X (Kumar, Stecher, Li, Knyaz, and Tamura 2018).
32. Sudhir Kumar, Masatoshi Nei, Joel Dudley,Koichiro Tamura. MEGA: A biologist-centric software for evolutionary analysis of DNA and protein sequences.Brief Bioinform, Jul, 2008; 9(4): 299–306. [PMC2562624].
33. Jin Xiong. "Phylogenetic Tree Construction Methods and Programs". Essential Bioinformatics.1st ed. First South Asian edition. Replika Press Pvt. Ltd.: India, 2007; 142-69.