

Comparative Genomic Studies of hypothetical proteins in Cyanophages

S. G. Sanmukh
Applied Aquatic Eco-system
Division,
National Environmental
Engineering Research Institute
(NEERI)
Nehru Marg, Nagpur-440020,
Maharashtra (India)

M. Rahman
Applied Aquatic Eco-system
Division,
National Environmental
Engineering Research Institute
(NEERI)
Nehru Marg, Nagpur-440020,
Maharashtra (India)

W. N. Paunikar
Applied Aquatic Eco-system
Division,
National Environmental
Engineering Research Institute
(NEERI)
Nehru Marg, Nagpur-440020,
Maharashtra (India)

ABSTRACT

The Cyanophages are the phages infecting cyanobacteria. They are potential agents for the horizontal gene transfer. The complete genome of 10 known Cyanophages deciphered the presence of various gene sequences for hypothetical proteins whose functions are not yet understood. Our attempt is to predict the structure and function of these hypothetical proteins by the application of computational methods and Bioinformatics. The probable function prediction for the hypothetical proteins was done by using Bioinformatics web tools like CDD-BLAST, INTERPROSCAN, PFAM and COGs by searching protein databases for the presence of conserved domains. While tertiary structures were constructed using PS² Server- Protein Structure Prediction server. This study revealed presences of functional domain in 258 uncharacterized proteins. These deciphered enzymatic data for hypothetical proteins can be used for the understanding of functional, structural and evolutionary development of cyanophages and its life cycle along with their role in host evolution.

Keywords

Bioinformatics web tools, conserved domains, Protein structure prediction, uncharacterized proteins, life cycle.

1. INTRODUCTION

The cyanophages plays an important role in both horizontal gene transfer and host mortality of the microbial populations that are responsible for the biogeochemical processes that run the planet [5] [6] [13] [25] [27]. They shape the ecology and evolution of both [the host and the phage] over evolutionary time. However, we have only barely begun to understand the genomic repertoire of these important genetic vectors [6] [16]. More than 70% of sequenced bacterial genomes contain prophages [4] [7] [8]. They often represent the primary constituent of strain-to-strain variability [4] [20] [21] and their genes are among the most highly expressed genes in genome-wide expression studies [22] [35]. The in-silico studies of hypothetical proteins (Uncharacterized proteins) for identifying their structure and function is an attempt to understand cyanophages and their role in cyanobacterial evolution [20] [33].

Computational biology assists us to predict the functionality in the uncharacterized sequences using the different strategies of comparative proteomics. The program's ability of homology searching using defined databases and by choosing standard parameters, the presence of the enzymatic conserved domain/s in the sequences could be searched out

and it may assist in the categorizing protein into specific family [29] [32]. Bioinformatics web tools like CDD-BLAST,INTERPROSCAN, PFAM and COGs can search the orthologous sequence in biological sequence databases for the target sequence, while assist in classification of target sequence in particular family [12] [26] [27] [28]. This study will help us to understand the probable functions of hypothetical proteins in cyanophages. The online automated servers are available which can predict the three dimensional structures for protein sequences by using the strategy of aligning target sequences with orthologous sequences by virtue of sequence homology using best scored template of orthologous family member [30] [31]. Here, we have predicted 3-D structure using Protein Structure Prediction Server (PS² server) [11] [37].

2. MATERIALS AND METHODS

2.1 Sequence Retrieval

The Complete protein sequences for 10 different cyanophages phages were downloaded from the Database of KEGG (<http://www.genome.jp/kegg/>). The phages under study includes Cyanophage PSS2 [24], Microcystis phage Ma-LMM01 [36], Prochlorococcus phage P-SSM2 [23], Prochlorococcus phage P-SSM4 [23], Prochlorococcus phage P-SSP7 [23], Synechococcus phage S-RSM4 [15], Synechococcus phage P60 [10], Synechococcus phage S-PM2 [14], Synechococcus phage syn9 [34] and Synechococcus phage Syn5 [17].

2.2 Functional Annotations

The hypothetical proteins from all 10 cyanophages were screened for the presence of conserved domains using the web-tools. The four bioinformatics web tools like CDD-BLAST (<http://www.ncbi.nlm.nih.gov/BLAST/>) [2] [19] [3], INTERPROSCAN (<http://www.abi.ac.uk/interpro>) [38], Pfam (<http://www.pfam.sanger.ac.uk/>) [1] and COGs (<http://www.ncbi.nih.gov/cog>) [18] were used, which shows the ability to search the defined conserved domains in the sequences and assist in the classification of proteins in appropriate family. Hypothetical proteins analyzed by the function prediction web tools have shown variable results depending upon the information available in databases, when searched for the conserved domains in hypothetical sequences.

2.3 Protein Structure Prediction

Online PS² Protein Structure Prediction Server was used to generate 3D-structures of hypothetical proteins

(<http://www.ps2.life.nctu.edu.tw/>) [2] [9] [11] [19]. The server accepts the protein (query) sequences in FASTA format and uses the strategies of Pair-wise and multiple alignments to generate resultant proteins 3D structures, Which are constructed using structural positioning information of atomic coordinates for known template in PDB format using best scored alignment data. Where the selection of template was based on the same conserved domain detected in the functional annotations and which must be available in the structure alignment for modeling purpose.

3. RESULT AND DISCUSSION

The in silico structure and function of the cyanophages was done for 10 phages. The hypothetical proteins were classified by using sequence similarity search with close orthologous family members available in various protein databases using the web tools. The predictions of 3-D structures of protein were done by using online automated PS2 server. The analysis of hypothetical proteins of cyanophages for function annotations was done by using web tools for their classification into particular enzymatic family based on

conserved domain available in the sequence which are represented in respective Table 1 through 10. From 10 different cyanophages, 258 hypothetical proteins were classified depending upon the presence of conserved domains. The (PS)² Server built the three dimensional structures for 28 hypothetical proteins satisfactorily using best scored orthologous template. The 3-D structures built are represented sequentially in respective cyanophage specific gene. The templates with best scoring with hypothetical protein sequences are represented in the order as Template ID, Identity, Score and E-value which represented in structure column of each cyanophage analyzed. The structure and functional data for Cyanophage PSS2 (Table 1), Microcystis phage Ma-LMM01 (Table 2), Prochlorococcus phage P-SSM2 (Table 3), Prochlorococcus phage P-SSM4 (Table 4), Prochlorococcus phage P-SSP7 (Table 5), Synechococcus phage S-RSM4 (Table 6), Synechococcus phage P60 (Table 7), Synechococcus phage S-PM2 (Table 8), Synechococcus phage syn9 (Table 9) and Synechococcus phage Syn5 (Table 10) are given in their respective tables.

Table 1 Structure and Functional Data for CYANOPHAGE PSS2

NCBI Gene	CDD-Blast	Interproscan	pfam	COGs	Structures
8207392	No	No	mRNA_cap_C mRNA capping enzyme, C-terminal domain	No	No
8207393	No	No	Predicted transmembrane and coiled-coil 2 protein	No	No
8207398	No	No	Prefoldin subunit	No	No
8207408	No	Endodeoxyribonuclease, RusA-like	RusA Endodeoxyribonuclease RusA	No	2h8eA- 21- 59- 4e-10
8207411	No	No	Mec-17 Touch receptor neuron protein Mec-17	No	No
8207412	No	Endodeoxyribonuclease, RusA-like	RusA Endodeoxyribonuclease RusA	No	2h8eA- 21- 59- 4e-10
8207413	HemN_C super family[cl06150], Members of this family are all oxygen-independent coproporphyrinogen-III oxidases (HemN)	No	No	No	No
8207422	No	No	SDH_alpha Serine dehydratase alpha chain,Decorin_bind Decorin binding protein	No	No
8207426	No	No	SprT-like SprT-like family	No	No
8207428	No	No	OEP Outer membrane efflux protein	No	No
8207436	No	No	No	COG0050 GTPases - translation elongation factors	2jesA -14 -42- 3e-06
8207438	No	No	RF-1 domain	No	No

8207450	FliI_clade3[TIGR03498], flagellar protein export ATPase FliI; Members of this protein family are the FliI protein of bacterial flagellum systems	No	Lac_bphage_repr Lactococcus bacteriophage repressor	No	No
8207451	No	No	CBM_17_28 Carbohydrate binding domain	No	No
8207453	No	No	Activator_LAG-3 Transcriptional activator LAG-3	No	No
8207461	No	No	SBP_bac_1 Bacterial extracellular solute-binding protein	No	No
8207472	No	No	Phage_tail_S Phage virion morphogenesis family	No	No
8207478	No	Fibrin/6-phosphogluconate dehydrogenase, C-terminal extension	No	No	1v1hA- 33- 36- 0.009
8207485	No	Phage major tail protein TP901-1	No	No	No
8207493	HOOK protein; This family consists of several HOOK1, 2 and 3 proteins from different eukaryotic organisms.	Endothelin-like toxin, conserved site, Legume lectin, beta chain, Mn/Ca-binding site	No	No	No
8207500	No	No	Tail_tube Phage tail tube protein	No	No
8207506	No	NUMOD4	NUMOD4 motif	No	No
8207507	No	No	HTH_8 Bacterial regulatory protein, Fis family , Putative excisionase	No	No
8207510	No	Homeodomain-related	HTH_7 Helix-turn-helix domain of resolvase	No	

Table 2 Structure and Functional Data for MICROCYSTIS PHAGE MA-LMM01

NCBI Gene	CDD-Blast	Interproscan	pfam	COGs	Structures
4484313	No	LamB porin	No	No	
4484317	FGE-sulfatase super family[cl00556], Formylglycine-generating sulfatase enzyme	Sulphatase-modifying factor,C-type lectin fold	FGE-sulfatase Formylglycine-generating sulfatase enzyme	No	1z70X- 26- 219- 3e-58
4484320	No	No	LIM_bind LIM-domain binding protein	No	No
4484323	Peptidase M15, PG_binding_1 super family, Putative peptidoglycan binding domain	Peptidoglycan binding-like,Hedgehog/DD-peptidase, zinc-binding motif, Peptidase M15A, C-terminal	PG_binding_1 Putative peptidoglycan binding domain ,Peptidase M15	No	1lbuA- 23- 145- 7e-36
4484326	No	No	Curlin_rpt Curlin associated repeat	No	No
4484330	No	No	PAAR motif	No	No

4484331	FGE-sulfatase super family, Formylglycine-generating sulfatase enzyme	Sulphatase-modifying factor, C-type lectin fold	FGE-sulfatase Formylglycine-generating sulfatase enzyme	No	2q17E- 25- 144- 3e-35
4484335	No	Concanavalin A-like lectin/glucanase	No	No	1a8dA- 12- 41- 4e-04
4484337	No	No	Predicted RNA-binding protein ,Probable molybdopterin binding domain,Angiomotin C terminal ,RIM-binding protein of the cytomatrix active zone , Poly(hydroxyalcanoate) granule associated protein (phasin) ,Autophagy protein 16 (ATG16),Tropomyosin ,Biogenesis of lysosome-related organelles complex-1 subunit 2 ,Phage minor structural protein GP20 , Seryl-tRNA synthetase N-terminal domain , TATA element modulatory factor 1 DNA binding ,bZIP transcription factor , TSC-22/dip/bun family ,Axonemal dynein light chain ,SNARE domain	No	No
4484338	No	No	RyR domain , Oestrogen-responsive protein Fam102A-B	No	No
4484339	No	No	Sigma-70, region 4	No	No
4484345	No	HNH nuclease	HNH endonuclease	No	No
4484350	No	Profilin/allergen	ER protein Pkr1 ,PRA1 family protein	NADH:ubiquinone oxidoreductase subunit 5 (chain L)/Multisubunit Na+/H+ antiporter, MnHA subunit	No
4484351	No	No	TfoX N-terminal domain	Archaeal/vacuolar-type H+-ATPase subunit A	No
4484356	Baseplate_J super family, Baseplate J-like protein	Baseplate assembly protein J-like, predicted	Baseplate_J Baseplate J-like protein	No	No
4484376	putative RNA ligase	No	No	No	No
4484378	No	No	RVT_2 Reverse transcriptase (RNA-dependent DNA polymerase)	No	No
4484380	No	No	KilA-N domain	No	No
4484383	No	No	No	Ribosomal protein S12	No
4484386	No	No	NfeD-like	No	No
4484393	No	No	RINT-1 / TIP-1 family	No	No
4484397	No	No	NOGCT domain	No	No

4484398	Cas_Cas4 super family,ATP-dependent nuclease, subunit B [DNA replication, recombination, and repair]	Restriction endonuclease, type II-like	No	No	1w36B- 18-45- 2e-05
4484406	No	No	KGG Stress-induced bacterial acidophilic repeat motif	General stress protein	No
4484413	No	No	Ferric reductase NAD binding domain	No	No
4484417	No	No	Proteobacterial lipase chaperone protein	No	No
4484419	No	No	Guanylate kinase	No	No
4484421	No	No	LTXQ motif	No	No
4484424	No	No	CTP synthase N-terminus	CTP synthase (UTP-ammonia lyase)	No
4484470	No	No	DNA polymerase III, theta subunit	No	No
4484471	No	No	Synapsin N-terminal	Ribonuclease HII	No
4484477	No	PLD	No	Methyl-accepting chemotaxis protein	NO
4484479	No	No	N-glycosylation protein ,Predicted membrane protein (DUF2232)	No	NO
4484481	No	Zinc finger, SWIM-type	SWIM zinc finger	No	NO

Table 3 Structure and Functional Data for PROCHLOROCOCCUS PHAGE P-SSM2

NCBI Gene	CDD-Blast	Interproscan	pfam	COGs	Structures
3294165	NO	NO	Pestivirus NS3 polyprotein peptidase S31	NO	NO
3294168	NO	NO	GSPII_E N-terminal domain	NO	NO
3294169	NO	NO	GSPII_E N-terminal domain	NO	NO
3294175	NO	NO	Serine dehydratase alpha chain Fimbrial, major and minor subunit	NO	NO
3294185	NO	NO	Microtubule associated ,Baculovirus LEF-11 protein ,HECT-domain (ubiquitin-transferase) ,TATA-binding protein interacting (TIP20)	NO	NO
<u>3294200</u>	NO	Collagen triple helix repeat	Collagen triple helix repeat (20 copies) ,Collagen triple helix repeat (20 copies)	NO	NO
<u>3294204</u>	NO	NO	DNA circulation protein N-terminus	NO	NO

<u>3294217</u>	NO	NO	Rab5 binding , Fibrinogen alpha/beta chain family ,Flagella accessory protein C (FlaC) , TATA element modulatory factor 1 TATA binding , Spc24 subunit of Ndc80 ,Septum formation initiator , bZIP transcription factor,Herpesvirus BLRF2 protein ,Hr1 repeat	NO	NO
<u>3294225</u>	NO	Collagen triple helix repeat	Collagen triple helix repeat (20 copies) ,	NO	NO
<u>3294229</u>	NO	NO	BMC domain	NO	NO
<u>3294232</u>	NO	NO	Ribosomal protein S27 ,Desulfoferrodoxin, N-terminal domain	NO	NO
<u>3294245</u>	NO	NO	S-adenosyl-L-homocysteine hydrolase, NAD binding domain ,EF hand associated	NO	NO
<u>3294251</u>	NO	NO	Selenium binding protein , Fumarate reductase subunit D	Ribosome recycling factor	NO
<u>3294254</u>	NO	NO	FerA (NUC095) domain	NO	NO
<u>3294255</u>	NO	NO	Acetyl co-enzyme A carboxylase carboxyltransferase alpha subunit ,emp24/gp25L/p24 family/GOLD ,Flagella accessory protein C (FlaC), Trimeric coiled-coil oligomerisation domain of matrilin , Tetrahydromethanopterin S-methyltransferase, subunit G ,Bacillus transposase protein Mating-type protein MAT alpha 1	ATPase involved in DNA repair	NO
<u>3294256</u>	NO	NO	Borrelia burgdorferi BBR25 lipoprotein	NO	NO
<u>3294268</u>	NO	NO	NO	Acyl-CoA synthetases (AMP-forming)/AMP-acid ligases II	NO
<u>3294276</u>	NO	NO	Chitin binding Peritrophin-A domain , Glycine rich protein family	NO	NO
<u>3294280</u>	NO	NO	C3 binding domain 4 of IgG-bind protein SBI	NO	NO
<u>3294282</u>	NO	NO	Sigma-54 factor, core binding domain	NO	NO
<u>3294285</u>	NO	NO	Apolipoporphin-III precursor (apoLp-III)	NO	IdlcA- 18- 31- 0.007

<u>3294288</u>	NO	NO	Predicted coiled-coil domain-containing protein (DUF2360)	NO	1y0eA- 24 -38- 0.003
<u>3294297</u>	NO	NO	Phage antitermination protein Q	NO	NO
<u>3294305</u>	NO	NO	IRSp53/MIM homology domain ,Intermediate filament protein ,NUDE protein, C-terminal conserved region ,Fibrinogen alpha/beta chain family ,fzo-like conserved region ,IncA protein ,Mitochondrial ATPase inhibitor, IATP , Outer membrane protein (OmpH-like)	NO	2dfsA -36 -34- 0.005
<u>3294306</u>	NO	NO	Collagen triple helix repeat (20 copies)	NO	NO
<u>3294325</u>	NO	NO	FERM central domain	NO	NO
<u>3294333</u>	NO	Parallel beta-helix repeat	Collagen triple helix repeat (20 copies)	NO	NO
<u>3294334</u>	NO	NO	R67 dihydrofolate reductase	NO	1cygA -13 -30 -0.009
<u>3294345</u>	NO	NO	HEPN domain	NO	NO
<u>3294350</u>	NO	HNH nuclease	YvbH-like oligomerisation region, Putative GTPase activating protein for Arf ,HNH endonuclease , GrpE	NO	NO
<u>3294355</u>	NO	NO	Zinc ribbon domain	NO	NO
<u>3294365</u>	NO	NO	Conserved mid region of cactin	NO	NO
<u>3294370</u>	NO	NO	mbt repeat	NO	NO
<u>3294374</u>	NO	NO	PDZ domain (Also known as DHR or GLGF)	NO	NO
<u>3294379</u>	NO	NO	RHS Repeat	NO	NO
<u>3294386</u>	NO	DNA methylase, N-6 adenine-specific, conserved site	Eco57I restriction endonuclease	NO	NO
<u>3294394</u>	NO	NO	yrrolo-quinoline quinone coenzyme C-terminus	NO	NO
<u>3294396</u>	NO	NO	Motilin/ghrelin-associated peptide, Phage Mu protein F like protein	NO	NO
<u>3294410</u>	NO	NO	Multicopper oxidase , AraC-like ligand binding domain	NO	2hbTA- 17- 38- 0.002
<u>3294422</u>	NO	NO	VPS28 protein	NO	NO
<u>3294425</u>	NO	NO	PPP4R2	NO	NO
<u>3294426</u>	NO	NO	Thrombin inhibitor from mosquito	NO	NO
<u>3294446</u>	NO	NO	Spherical virus-type peptidase	NO	NO
<u>3294461</u>	NO	NO	SprT-like family	NO	NO

<u>3294467</u>	NO	NO	Herpesvirus UL4 family	NO	NO
<u>3294469</u>	NO	Photosystem II PsbN	Photosystem II reaction centre N protein (psbN)	NO	NO
<u>3294473</u>	NO	NO	Invariant surface glycoprotein	NO	NO
<u>3294478</u>	NO	NO	Mediator complex subunit 28 ,SlyX	NO	NO
<u>3294494</u>	NO	NO	OMS28 porin , Transcription factor/nuclear export subunit protein 2 ,Microtubule associated , Golgi transport complex subunit 5 ,Erp protein C-terminus ,Septum formation initiator ,Vesicle transport v-SNARE protein N-terminus ,Heat shock factor binding protein 1	NO	NO

Table 4 Structure and Functional Data for PROCHLOROCOCCUS PHAGE P-SSM4

NCBI Gene	CDD-Blast	Interproscan	pfam	COGs	Structures
<u>10021827</u>	NO	NO	Lipase (class 2)	NO	NO
<u>10021830</u>	NO	Photosystem II PsbN	Photosystem II reaction centre N protein (psbN) ,NADH dehydrogenase subunit 5 C-terminus ,	NO	NO
<u>10021841</u>	NO	NO	Syntaxin 6, N-terminal	NO	NO
<u>10021846</u>	NO	NO	Lecithin:cholesterol acyltransferase	NO	NO
<u>10021847</u>	NO	NO	Zinc finger C-x8-C-x5-C-x3-H type (and similar)	NO	NO
<u>3294505</u>	NO	NO	Nuclear fragile X mental retardation-interacting protein 1 (NUFIP1)	NO	NO
<u>3294509</u>	NO	NO	RST domain of plant C-terminal , Hr1 repeat	NO	NO
<u>3294513</u>	NO	NO	Mitochondrial glycoprotein ,Staphylococcal complement inhibitor SCIN ,Bacteriophage FRD3 protein	NO	NO
<u>3294514</u>	NO	NO	Macrophage scavenger receptor	NO	NO
<u>3294516</u>	NO	NO	Respiratory-chain NADH dehydrogenase 24 Kd subunit	NO	NO
<u>3294520</u>	NO	NO	N-acetylmuramoyl-L-alanine amidase	NO	NO
<u>3294525</u>	NO	NO	Spc7 kinetochore protein	NO	NO
<u>3294527</u>	NO	NO	Carbohydrate binding domain	NO	1guiA -16 -39 -0.002

<u>3294542</u>	NO	NO	Protein phosphatase 2A regulatory B subunit (B56 family)	NO	NO
<u>3294543</u>	NO	NO	Ezrin/radixin/moesin family , Clusterin-associated protein-1	NO	NO
<u>3294564</u>	NO	NO	Type I restriction and modification enzyme - subunit R C terminal	NO	NO
<u>3294567</u>	NO	NO	Preprotein translocase SecG subunit	NO	NO
<u>3294590</u>	NO	NO	Mitochondrial ribosomal protein L28	NO	NO
<u>3294613</u>	NO	NO	Orbivirus outer capsid protein VP2 ,RNA polymerase recycling family C-terminal	NO	NO
<u>3294614</u>	NO	NO	Chordopoxvirus fusion protein ,Resolvase, N terminal domain	NO	NO
<u>3294634</u>	NO	Fibrinogen, alpha/beta/gamma chain, C-terminal globular	FadR C-terminal domain	NO	1fzcB -23- 44- 4e-05
<u>3294636</u>	NO	NO	bZIP transcription factor ,WW domain binding protein 11	NO	NO
<u>3294655</u>	NO	NO	NO	Translation initiation factor 2 (GTPase)	NO
<u>3294670</u>	NO	NO	BOPINT (NUC169) domain	NO	NO
<u>3294673</u>	NO	NO	STAT1 TAZ2 binding domain	ATPase involved in DNA repair	NO
<u>3294682</u>	NO	NO	T4-like virus tail tube protein gp19	NO	NO
<u>3294691</u>	NO	NO	Phage regulatory protein Rha (Phage_pRha)	NO	NO

Table 5 Structure and Functional Data for PROCHLOROCOCCUS PHAGE P-SSP7

NCBI Gene	CDD-Blast	Interproscan	pfam	COGs	Structures
<u>3294704</u>	NO	NO	MutS domain III	NO	NO
<u>3294712</u>	NO	NO	NO	S-adenosylmethionine synthetase	NO
<u>3294713</u>	NO	NO	YvrJ protein family		NO
<u>3294721</u>	NO	NO	HicB family	NO	NO
<u>3294723</u>	NO	NO	Polyketide synthesis cyclase ,Ubiquitin-specific protease 7	NO	NO
<u>3294725</u>	NO	NO	Chitin synthesis regulation, resistance to Congo red	NO	NO

<u>3294726</u>	NO	Winged helix-turn-helix transcription repressor DNA-binding	NO	NO	NO
<u>3294728</u>	NO	NO	Phycobilisome degradation protein nblA	NO	NO
<u>3294735</u>	NO	NO	ERF superfamily	NO	NO

Table 6 Structure and Functional Data for SYNECHOCOCCUS PHAGE S-RSM4

NCBI Gene	CDD-Blast	Interproscan	pfam	COGs	Structures
<u>8303263</u>	NO	NO	Coatomer (COPI) alpha subunit C-terminus	NO	NO
<u>8303264</u>	NO	NO	Restriction endonuclease EcoRV	NO	NO
<u>8303265</u>	NO	NO	Herpesvirus UL87 family	NO	NO
<u>8303268</u>	NO	Thioredoxin fold	NO	NO	NO
<u>8303284</u>	NO	NO	CARDB	NO	NO
<u>8303289</u>	NO	Prevent-host-death protein	Phd_YefM	NO	2a6qb- 26 -34 -0.008
<u>8303293</u>	NO	NO	GIY-YIG catalytic domain	NO	NO
<u>8303304</u>	NO	NO	Prolyl 4-Hydroxylase alpha-subunit, N-terminal region	NO	NO
<u>8303326</u>	NO	NO	Formiminotransferase domain, N-terminal subdomain	NO	NO
<u>8303337</u>	NO	NO	BPS (Between PH and SH2)	NO	NO
<u>8303341</u>	NO	NO	Glycine rich protein	NO	NO
<u>8303342</u>	NO	NO	Phytanoyl-CoA dioxygenase (PhyH)	NO	2fcuA -11- 38- 0.002
<u>8303353</u>	NO	NO	Bacterial-like globin ,Major Outer Sheath Protein C-terminal region	NO	NO
<u>8303371</u>	NO	NO	NusA N-terminal domain	NO	NO
<u>8303376</u>	NO	NO	Transposase family tnp2 ,Associated with zinc fingers	NO	NO
<u>8303378</u>	NO	NO	T4 bacteriophage base plate protein	NO	NO
<u>8303382</u>	NO	Transcription regulator AbrB/SpoV, predicted	SpoVT / AbrB like domain	NO	NO

<u>8303383</u>	NO	NO	SF-assemblin/beta giardin , Endoribonuclease XendoU , Colicin immunity protein / pyocin immunity protein , Phage capsid scaffolding protein (GPO) serine peptidase ,Serpin (serine protease inhibitor) ,Biofilm formation regulator YbaJ , UDP-glucose/GDP-mannose dehydrogenase family, central domain	NO	NO
<u>8303389</u>	NO	NO	Inorganic pyrophosphatase	NO	NO
<u>8303391</u>	NO	NO	YebG protein	NO	NO
<u>8303393</u>	NO	NO	Ribulose bisphosphate carboxylase large chain, N-terminal domain	NO	NO
<u>8303394</u>	NO	NO	Peptidase propeptide and YPEB domain	NO	NO
<u>8303395</u>	NO	NO	RED-like protein N-terminal region	NO	NO
<u>8303397</u>	NO	NO	RmlD substrate binding domain	NO	NO
<u>8303401</u>	NO	NO	Eco57I restriction-modification methylase	NO	
<u>8303402</u>	NO	SP1917 domain	C-5 cytosine-specific DNA methylase ,Hydrolytic ATP binding site of dynein motor region D1	NO	NO
<u>8303406</u>	NO	NO	Alpha and gamma adaptin binding protein p34 ,Autophagy protein Apg6 ,MbeD/MobD like , MerR, DNA binding ,Haemolysin XhIA, FlgN protein , WW domain binding protein 11	NO	NO
<u>8303408</u>	NO	NO	WASP-binding domain of Sorting nexin protein	NO	NO
<u>8303425</u>	NO	NO	WSK motif	NO	NO
<u>8303428</u>	NO	NO	3D domain	NO	NO
<u>8303442</u>	NO	NO	Rubredoxin	NO	3ct5A -19- 81- 2e-16
<u>8303451</u>	NO	NO	GnsA/GnsB family ,Calcium binding and coiled-coil domain (CALCOCO1) like , Mitochondrial ATPase inhibitor, IATP ,Herpesvirus UL14-like protein , IncA protein	NO	NO
<u>8303463</u>	NO	NO	Gryzun, putative Golgi trafficking	NO	NO

<u>8303464</u>	NO	NO	T4-like virus tail tube protein gp19	NO	NO
<u>8303480</u>	NO	NO	Glycosylphosphatidyl inositol-anchored merozoite surface protein, Secretion system effector C (SseC) like family, FAD binding domain in molybdopterin dehydrogenase	NO	NO
<u>8303493</u>	NO	NO	Iron only hydrogenase large subunit, C-terminal domain	NO	NO
<u>8303499</u>	NO	NO	Heat-labile enterotoxin beta chain	NO	NO
<u>8303501</u>	NO	NO	Calponin homology (CH) domain	NO	
<u>8303503</u>	NO	NO	Cell division inhibitor SulA	NO	NO

Table 7 Structure and Functional Data for SYNECHOCOCCUS PHAGE P60

NCBI Gene	CDD-Blast	Interproscan	pfam	COGs	Structures
<u>932709</u>	NO	NO	GspL periplasmic domain	NO	NO
<u>932720</u>	NO	NO	Glycosylphosphatidyl inositol-anchored merozoite surface protein , Peptidase inhibitor I78 family	NO	NO
<u>932725</u>	NO	NO	GspL periplasmic domain	NO	NO
<u>932740</u>	NO	NO	TNF(Tumour Necrosis Factor) family	NO	NO
<u>932761</u>	NO	NO	Occludin homology domain	NO	NO
<u>932783</u>	NO	NO	Type III secretion protein (HpaP) ,Down syndrome cell adhesion molecule C terminal , Major surface antigen from hepadnavirus	NO	NO

Table 8 Structure and Functional Data for SYNECHOCOCCUS PHAGE S-PM2

NCBI Gene	CDD-Blast	Interproscan	pfam	COGs	Structures
<u>3260259</u>	NO	NO	CBP4	NO	NO
<u>3260265</u>	NO	NO	Phage antitermination protein Q	NO	NO
<u>3260269</u>	NO	NO	Tetrahydromethanopterin S-methyltransferase, subunit G ,Spc24 subunit of Ndc80 , Proteins of 100 residues with WXG , Flagella accessory protein C (FlaC)	NO	1f6wA- 35- 35 -0.003

<u>3260271</u>	NO	NO	Thermoplasma acidophilum protein TA0956	NO	NO
<u>3260279</u>	NO	NO	Prominin , DASH complex subunit Dad2 , Predicted SPOUT methyltransferase , DNA polymerase processivity factor (UL42)	NO	NO
<u>3260284</u>	NO	NO	Anti-sigma-28 factor, FlgM	NO	NO
<u>3260290</u>	MazG nucleotide pyrophosphohydrolase domain	NTP pyrophosphohydrolase MazG, putative catalytic core; NTP Pyrophosphohydrolase MazG-related, GP37	MazG nucleotide pyrophosphohydrolase domain	Predicted pyrophosphatase	2gtaA -11- 58- 8e-10
<u>3260294</u>	clamp loader small subunit	NO	NO	NO	NO
<u>3260299</u>	NO	NO	Clavanin	NO	NO
<u>3260300</u>	MoxR-like ATPases	ATPase, AAA-5	AAA domain (dynein-related subfamily) , CbbQ/NirQ/NorQ C-terminal	MoxR-like ATPases	2r44A -16- 205- 7e-54
<u>3260301</u>	NO	NO	Neurotransmitter-gated ion-channel transmembrane region	NO	NO
<u>3260309</u>	NO	NO	Virion protein N terminal domain , Fibrillar collagen C-terminal domain	NO	NO
<u>3260312</u>	NO	NO	GnsA/GnsB family , IncA protein ,Prefoldin subunit ,Calcium binding and coiled-coil domain (CALCOCO1) like , Rho Binding ,Microtubule-binding stalk of dynein motor	NO	NO
<u>3260316</u>	NO	Pectin lyase fold/virulence factor	NO	NO	NO
<u>3260319</u>	NO	NO	EAP30/Vps36 family ,Vps51/Vps67 , Prefoldin subunit	NO	NO
<u>3260320</u>	NO	NO	BRE1 E3 ubiquitin ligase	NO	NO
<u>3260327</u>	NO	NO	Transcriptional activator HlyU	NO	NO
<u>3260330</u>	NO	NO	Firmicute eSAT-6 protein secretion system EssA	NO	NO
<u>3260331</u>	Peptidase family M23	Duplicated hybrid motif,Peptidase M23	Peptidase family M23	Membrane proteins related to metalloendopeptidases	NO
<u>3260332</u>	Protein Disulfide Oxidoreductases and Other Proteins with a Thioredoxin fold	NAD:arginine ADP-ribosyltransferase, ART,Arsenate reductase-like	NAD:arginine ADP-ribosyltransferase , ArsC family	Methyl-accepting chemotaxis protein	NO

<u>3260338</u>	NO	NO	TM1410 hypothetical-related protein	NO	NO
<u>3260341</u>	NO	NO	Second Mitochondria-derived Activator of Caspases	NO	NO
<u>3260347</u>	NO	NO	FtsX-like permease family	NO	NO
<u>3260350</u>	NO	NO	FIP domain,Paired amphipathic helix repeat	NO	NO
<u>3260355</u>	NO	NO	Hypothetical protein (DUF2410) ,Transcription-initiator DNA-binding domain IBD	NO	NO
<u>3260357</u>	NO	NO	Glucitol operon activator protein (GutM)	NO	NO
<u>3260359</u>	2OG-Fe(II) oxygenase superfamily	Oxoglutarate/iron-dependent oxygenase,Prolyl 4-hydroxylase, alpha subunit	2OG-Fe(II) oxygenase superfamily,Poxvirus C4/C10 protein	NO	2hbTA- 17- 73 -5e-14
<u>3260360</u>	NO	NO	Vanadium/alternative nitrogenase delta subunit	NO	NO
<u>3260362</u>	NO	NO	Survival motor neuron protein (SMN)	NO	NO
<u>3260370</u>	NO	NO	Mitochondrial inner membrane protein , Chlamydia-phage Chp2 scaffold (Chlamy_scaf) , Ribosomal L29 protein	NO	NO
<u>3260373</u>	NO	NO	Pedibin/Hym-346 family	NO	NO
<u>3260380</u>	NO	NO	Signal transducing histidine kinase, homodimeric domain; Bacillus haemolytic enterotoxin (HBL);Coat F domain;HOOK protein;Autophagy protein 16 (ATG16);Outer membrane protein (OmpH-like); Plasmid replication region DNA-binding N-term;GDP/GTP exchange factor Sec2p ; Reovirus sigma C capsid protein; PspA/IM30 family; Laminin Domain II; Microtubule associated; Rab5 binding; Intermediate filament protein; Apolipoporphin-III precursor (apoLp-III) ; ATP synthase subunit D ; Uso1 / p115 like vesicle tethering protein, C terminal region	NO	NO

			:Seryl-tRNA synthetase N-terminal domain ; Phage capsid scaffolding protein (GPO) serine peptidase ; SlyX ; Baculovirus polyhedron envelope protein, PEP, C terminus ; FlgN protein ; Prominin ; Tat binding protein 1(TBP-1)-interacting protein (TBPIP) ;Axonemal dynein light chain ; Subunit 21 of Mediator complex ;bZIP transcription factor ;Tektin family ;Homeobox associated leucine zipper ;JNK_SAPK-associated protein-1 ; Intra-flagellar transport protein 57 ;Transposase protein ;Spc7 kinetochore protein ;Cortexillin I, coiled coil ;UV radiation resistance protein and autophagy-related subunit 14 ;MbeD/MobD like ;DASH complex subunit Spc34 F;Septum formation initiator ; USP8 interacting ; Nucleopolyhedrovirus P10 protein ; IncA protein ;TATA element modulatory factor 1 DNA binding ; Transcription factor Opi1 ; Cartilage oligomeric matrix protein ;Ribosomal L29 protein		
<u>3260383</u>	NO	NO	NO	NO	NO
<u>3260385</u>	NO	NO	NO	NO	NO
<u>3260386</u>	NO	NO	Alphamannosidase, middle domain	NO	NO
<u>3260388</u>	2OG-Fe(II) oxygenase superfamily	NO	NO	NO	2hbta -14 -62 -6e-11
<u>3260389</u>	2OG-Fe(II) oxygenase superfamily	Oxoglutarate/iron-dependent oxygenase, Prolyl 4-hydroxylase, alpha subunit	2OG-Fe(II) oxygenase superfamily	NO	2jigA -16 -76- 4e-15
<u>3260392</u>	NO	NO	UvrB/uvc motif	NO	NO
<u>3260393</u>	NO	NO	SpoVT / AbrB like domain	NO	NO
<u>3260394</u>	NO	NO	Fork head domain , Transcription factor Vhr1	NO	NO
<u>3260396</u>	NO	NO	Mis12-Mtw1 protein family	NO	NO

<u>3260399</u>	NO	Globin-like	Phycobilisome protein	NO	NO
<u>3260405</u>	NO	NO	Low affinity iron permease ,6-pyruvoyl tetrahydropterin synthase	NO	NO
<u>3260420</u>	NO	NO	ZPR1 zinc-finger domain ,OMS28 porin ,HAND	NO	NO
<u>3260421</u>	NO	NO	Putative papain-like cysteine peptidase (DUF1796)	NO	NO
<u>3260422</u>	NO	NO	2-hydroxyglutaryl-CoA dehydratase, D-component ,Mediator complex subunit 2	NO	NO
<u>3260423</u>	NO	NO	NADH-ubiquinone/plastoquinone oxidoreductase chain 4L	NO	NO
<u>3260427</u>	NO	NO	Pheromone A receptor	NO	NO
<u>3260428</u>	NO	NO	NO	Membrane protease subunits, stomatin/prohibitin homologs	NO
<u>3260429</u>	SpoVT / AbrB like domain	NO	NO	ATPase components of ABC transporters with duplicated ATPase domains	NO
<u>3260432</u>	NO	Transmembrane Fragile-X-F-associated protein	Transmembrane Fragile-X-F protein , Bacterial signalling protein N terminal repeat	ABC-type dipeptide/oligopeptide/nickel transport systems, permease components	NO
<u>3260433</u>	NO	NO	KNOX1 domain	NO	NO
<u>3260435</u>	NO	NO	Zinc ribbon domain ,NAD-dependent DNA ligase C4 zinc finger domain ,MIZ/SP-RING zinc finger ,Rubredoxin	NO	NO
<u>3260437</u>	NO	NO	ATP cone domain , GTPase-activator protein for Ras-like GTPase	NO	NO
<u>3260442</u>	NO	NO	Acetate kinase	NO	NO
<u>3260451</u>	NO	NO	Galactose-3-O-sulfotransferase	NO	NO
<u>3260487</u>	2OG-Fe(II) oxygenase superfamily	Oxoglutarate/iron-dependent oxygenase,Prolyl 4-hydroxylase, alpha subunit	2OG-Fe(II) oxygenase superfamily	NO	2hbTA -15- 88 -2e-18
<u>3260493</u>	NO	NO		DNA-directed RNA polymerase beta subunit/140 kD subunit (split gene in Mjan, Mthe, Aful)	2eutA- 18- 34- 0.010
<u>3260501</u>	2OG-Fe(II) oxygenase superfamily	Prolyl 4-hydroxylase, alpha subunit	Taurine catabolism dioxygenase TauD, TfdA family	NO	2jigA -19-150 -2e-37
<u>3260507</u>	NO	NO	stress-induced bacterial acidophilic repeat motif	NO	NO
<u>3260515</u>	NO	NO	Antirestriction protein (ArdA)	NO	NO

Table 9 Structure and Functional Data for SYNECHOCOCCUS PHAGE SYN9

NCBI Gene	CDD-Blast	Interproscan	pfam	COGs	Structures
4239015	No	No	VP_N-CPKC Virion protein N terminal domain	No	No
4239049	No	No	Poxvirus C7/F8A protein	No	No

Table 10 Structure and Functional Data for SYNECHOCOCCUS PHAGE SYN5

NCBI Gene	CDD-Blast	Interproscan	pfam	COGs	Structures
5220168	No	Acyl-CoA N-acetyltransferase	No	No	No
5220169	No	No	GHMP kinases C terminal	No	No

4. CONCLUSIONS

This in-silico study have sorted some functionally important hypothetical proteins of cyanophages applying the parameters of pair-wise and multiple sequence alignment tools along with structure prediction tools, which suggests that many probable functional uncharacterized proteins are available in the cyanophages. Bioinformatics Web Tools have shown the ability to predict structure and functions in 258 hypothetical proteins of cyanophages. In all 28 3-D structures for

5. ACKNOWLEDGMENTS

We thanks to all the researchers who have contributed for preparation and reviewing of this manuscript.

6. REFERENCES

- [1] Alex, B., Lachlan, C., Richard, D., Robert, D. F., Volker, H., Sam, G.J., Ajay, K., Mhairi, M., Simon, M., Erik, L. L. S., David, J. S., Corin Y., Sean, R. E. The Pfam families' database. Nucleic Acids Research, Vol. 32, D138-D141, 2004.
- [2] Altschul, S., F., Madden, T., L., Schaffer, A., A., Zhang, J., Zhang, Z., Miller, W., Lipman, D., J. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25 (17), 3389-402, 1997.
- [3] Aron, M. Bauer., John, B. A., Myra, K. D., Carol, D. S., Noreen, R. G., Marc, G., Luning, H., Siqian, H., David, I. H., John, D. J., Zhaoxi, K., Dmitri, K., Christopher, J. L., Cynthia A. L., Chunlei, L., Fu, L., Shennan, L., Gabriele, H. M., Mikhail, M., James, S. S., Narmada, T., Roxanne, A. Y., Jodie, J. Y., Dachuan, Z., Stephen, H. B. CDD: a conserved domain database for interactive domain family analysis. Nucleic Acids Research, Vol. 35, D237-D240, 2006.
- [4] Baba, T., Takeuchi, F., Kuroda, M., Yuzawa, H., Aoki, K., Oguchi, A., et al. Genome and virulence determinants of high virulence community-acquired MRSA. Lancet 359: 1819–1827, 2002.
- [5] Barre, F.X., and Sherratt, D.J. Xer site-specific recombination: promoting chromosome segregation. In Mobile DNA II, 2002.
- [6] Breitbart, M., Thompson, L.R., Suttle, C.A., Sullivan, M. B. Exploring the Vast Diversity of Marine Viruses. Oceanogr 20: 135–139, 2007.
- [7] Canchaya, C., Proux, C., Fournous, G., Bruttin, A., and Brussow, H. Prophage genomics. Microbiol Mol Biol Rev 67: 238–276, 2003.
- [8] Casjens, S. Prophages and bacterial genomics: what have we learned so far? Mol Microbiol 49: 277–300, 2003.
- [9] Cédric, N., Desmond, G. H., Jaap, H. T-coffee: a novel method for fast and accurate multiple sequence alignment. J. Mol. Biol. 302, 205-217, 2000.
- [10] Chen, F. and Lu, J. Genomic sequence and evolution of marine cyanophage P60: a new insight on lytic and lysogenic phages. Appl. Environ. Microbiol. 68 (5), 2589-2594, 2002.
- [11] Chih-Chieh, C., Jenn-Kang, H., Jinn-Moon, Y. (PS)2: protein structure prediction server Nucl. Acids Res. 34, W152-W157, 2006.
- [12] Edward, E., Gary, L. G., Osnat, H., John, M., John, O., Roberto, J. P., Linda, B., Delwood, R., Andrew, J. H. Biological function made crystal clear- annotation of hypothetical proteins via structural genomics. Current Opinion in Biotechnology 11, 25-30, 2000.
- [13] Fuhrman, J. A. Marine viruses and their biogeochemical and ecological effects. Nature 399: 541–8, 1999.
- [14] Mann, N.H., Clokie, M.R., Millard, A., Cook, A., Wilson, W.H., Wheatley, P.J., Letarov, A. and Krisch, H. M. The genome of S-PM2, a 'photosynthetic' T4-type bacteriophage that infects marine Synechococcus strains. J. Bacteriol. 187 (9), 3188-3200, 2005.
- [15] Millard, A. D., Zwirglmaier, K., Downey, M.J., Mann, N.H. H. and Scanlan, D. J. Comparative genomics of marine cyanomyoviruses reveals the widespread occurrence of Synechococcus host genes localized to a hyperplastic region: implications for mechanisms of cyanophage evolution. Environ. Microbiol. 11 (9), 2370-2387, 2009.

hypothetical proteins was constructed using (PS) ² server, which serves as a fast automated 3D- structure generating web server. This predicted three dimensional structures may assist in establishing their importance in life cycle of cyanophages whose exact role in phage-host lifecycle is still unclear and can be used in future for the understanding of functional, structural and evolutionary development of cyanophages and its life cycle along with their role in host evolution.

- [16] Paul, J. H. and Sullivan, M. B. Marine phage genomics: what have we learned? *Curr Opin Biotechnol* 16: 299–307, 2005.
- [17] Pope,W.H., Weigele,P.R., Chang,J., Pedulla,M.L., Ford,M.E., Houtz,J.M., Jiang,W., Chiu,W., Hatfull,G.F., Hendrix,R.W. and King, J. Genome sequence, structural proteins, and capsid organization of the cyanophage Syn5: a 'horned' bacteriophage of marine Synechococcus. *J. Mol. Biol.* 368 (4), 966-981, 2007.
- [18] Roman, L. T., Michael, Y., Galperin, Darren A. Natale, Eugene V. Koonin. The COG database: a tool for genome –scale analysis of protein functions and evolution. *Nucleic Acid Research*. 28, 33-36, 2000.
- [19] Schaffer, A. A., Aravind, L., Madden, T. L., Shavirin, S. Spouge, J. L., Wolf, Y. I., Koonin, E. V., Altschul, S. F. Improving the accuracy of PSI-BLAST protein database searches with composition-based statistics and other refinements. *Nucleic Acids Res.* 29(14), 2994-3005, 2001.
- [20] Simpson, A.J., Reinach, F.C., Arruda, P., Abreu, F.A., Aencio, M., Alvarenga, R., et al. The genome sequence of the plant pathogen *Xylella fastidiosa*. The *Xylella fastidiosa* Consortium of the Organization for Nucleotide Sequencing and Analysis. *Nature* 406: 151–157, 2002.
- [21] Smoot, J.C., Barbian, K.D., Van Gompel, J.J., Smoot, L.M., Chaussee, M.S., Sylva, G.L., et al. Genome sequence and comparative microarray analysis of serotype M18 group A *Streptococcus* strains associated with acute rheumatic fever outbreaks. *Proc Natl Acad Sci USA* 99: 4668–4673, (2002).
- [22] Smoot, L.M., Smoot, J.C., Graham, M.R., Somerville, G.A., Sturdevant, D.E., Migliaccio, C.A., et al. Global differential gene expression in response to growth temperature alteration in group A *Streptococcus*. *Proc Natl Acad Sci USA* 98: 10416–10421, 2001.
- [23] Sullivan, M. B., Huang, K. H., Ignacio-Espinoza, J. C., Berlin, A. M., Kelly, L., Weigele, P. R., Defrancesco, A. S., Kern, S. E., Thompson, L. R., Young, S., Yandava, C., Fu, R., Krastins, B., Chase, M., Sarracino, D., Osburne, M. S., Henn, M. R. and Chisholm, S. W. Genomic analysis of oceanic cyanobacterial myoviruses compared with T4-like myoviruses from diverse hosts and environments. *Environ. Microbiol.* 12 (11), 3035–3056, 2010.
- [24] Sullivan, M. B., Krastins, B., Hughes, J. L., Kelly, L., Chase, M., Sarracino, D. and Chisholm, S.W. The genome and structural proteome of an ocean siphovirus: a new window into the cyanobacterial 'mobilome'. *Environ. Microbiol.* 11 (11), 2935-2951, 2009.
- [25] Suttle, C. A. Viruses in the sea. *Nature* 437: 356–61, 2005.
- [26] Swapnil, G. S., Dilip, B. M., Waman, N. P., Tarun, K. G. Computational characterizations for structure and function of unclassified proteins in *Ictalurus punctatus*. *CiiT International Journal of Artificial Intelligent Systems and Machine Learning DOI: AIML052011001*, 2011.
- [27] Swapnil, G. S., Waman, N. P. Study of hypothetical proteins in *Shigella* phages. *CiiT International Journal of fuzzy Systems* 2011.
- [28] Swapnil, G. S., Waman, N. P., Dilip, B. M., Tarun, K. G. Functionality search in hypothetical proteins of *Halobacterium salinarum* *CiiT International Journal of fuzzy Systems* 2011.
- [29] Swapnil, G. S., Waman, N. P., Dilip, B. M., Tarun, K. G. Insilico function prediction for hypothetical proteins in *Vibrio parahaemolyticus* Chromosome II. *CiiT International Journal of Data Mining and Knowledge Engineering*. DOI: DMKE052011003, 2011.
- [30] Swapnil, G. S., Waman, N. P., Tarun, K. G. Computational approach for structure and functionality search for hypothetical proteins in *Mycobacterium leprae* *CiiT International Journal of Data Mining and Knowledge Engineering* DOI: DMKE032011014, 2011.
- [31] Swapnil, G. S., Waman, N. P., Tarun, K. G. Study of Hypothetical Proteins in *Salmonella* Phages and Predicting their Structural and Functional Relationship *CiiT International Journal of Biometrics and Bioinformatics*. DOI: BB022011001, 2011.
- [32] Swapnil, G. S., Waman, N. P., Tarun, K. G. & Tapan, C. Structure and Function Predictions of Hypothetical Proteins in *Vibrio* Phages. *International Journal of Biometrics and Bioinformatics*. 4, 161-175, 2010.
- [33] Waman, N. P., Swapnil, G. S. and Tarun, K. G.. Exploring the hypothetical proteins in Rhizophages and their role in influencing *Rhizobium* species in soil. *CiiT International Journal of Artificial Intelligent systems and Machine Learning* DOI: AIML042011002, 2011.
- [34] Weigele,P.R., Pope,W.H., Pedulla,M.L., Houtz,J.M., Smith, A. L., Conway, J. F., King,J., Hatfull,G.F., Lawrence, J. G. and Hendrix,R.W. Genomic and structural analysis of Syn9, a cyanophage infecting marine *Prochlorococcus* and *Synechococcus*. *Environ. Microbiol.* 9 (7), 1675-1695, 2007.
- [35] Whiteley, M., Bangera, M.G., Bumgarner, R.E., Parsek, M.R., Teitzel, G.M., Lory, S., and Greenberg, E.P. Gene expression in *Pseudomonas aeruginosa* biofilms. *Nature* 413: 860–864, 2001.
- [36] Yoshida,T., Nagasaki,K., Takashima,Y., Shirai,Y., Tomaru,Y., Takao,Y., Sakamoto,S., Hiroishi,S. and Ogata,H. Ma-LMM01 infecting toxic *Microcystis aeruginosa* illuminates diverse cyanophage genome strategies. *J. Bacteriol.* 190 (5), 1762-1772, 2008.
- [37] Zafer, A., Yucl, A., Mark, B. Protein secondary structure prediction for a single-sequence using hidden semi-Markov models, *BMC Bioinformatics* ,7, 178, 2006.
- [38] Zdobnov, E. M., Rolf, A. Interproscan- an integration platform for the signatures recognition methods in InterPro. *Bioinformatics* 17,847-848, 2001.