Structure and Function Predictions of Hypothetical Proteins in Vibrio Phages

Swapnil G. Sanmukh swamukh1985in@rediffmail.com
Applied Aquatic Ecosystem Division,
National Environmental Engineering Research Institute,
Nehru Marg, Nagpur-440020, India

Waman N. Paunikar wn_paunikar@neeri.res.in
Applied Aquatic Ecosystem Division,
National Environmental Engineering Research Institute,
Nehru Marg, Nagpur-440020, India

Tarun K. Ghosh tk_ghosh@neeri.res.in
Applied Aquatic Ecosystem Division,
National Environmental Engineering Research Institute,
Nehru Marg, Nagpur-440020, India

Tapan Chakrabarti t_chakrabarti@neeri.res.in
National Environmental Engineering Research Institute,
Nehru Marg, Nagpur-440020, India

Abstract

The Vibriophages are the potential agents for the transfer of the virulence factor to their host through lateral gene transfer. The complete genome sequencing of various known vibriophages has been done which deciphered the presence of various gene sequences for hypothetical proteins whose function is not yet understood. We analyzed complete genome of 21such Vibriophages for hypothetical proteins from which 13 phages were sorted for our studies. 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 of the hypothetical protein was done by using Bioinformatics web tools like CDD-BLAST, INTERPROSCAN, PFAM and COGs by searching sequence databases for the presence of orthologous enzymatic conserved domains in the hypothetical sequences. While tertiary structures were constructed using PS^2 Server (Protein Structure Prediction server). These study revealed presences of enzymatic functional domain in 92 uncharacterized proteins; their roles are yet to be discovered in Vibriophages. These deciphered enzymatic data for hypothetical proteins can be used for the understanding of functional, structural, evolutionary and metabolic development of Vibriophages and its life cycle along with their role in host evolution and pathogenicity.

Keywords: Bioinformatics Web Tools, Conserved Domains, Protein Structure Prediction, Uncharacterized Proteins, Life Cycle and Pathogenicity.

1. INTRODUCTION

The etiologic agent of cholera, *Vibrio cholerae* is a gram negative bacterium which has been reported to be infected by various specific filamentous phages (Campos, et al., 2003, Faruque, et al., 2005, Waldor, et al., 1997, Ikema, et al., 1998, Jouravleva, et al., 1998, Kar, et al., 1996, Honma, et al., 1996). CTXΦ phage has been the most studied due to its role in pathogenicity and horizontal gene transfer (Davis, et al., 2003). The phage is potentially responsible for transducing the cholera toxin genes into nonpathogenic environmental strains along with replicating directly from the bacterial chromosome for producing infective phage particles (Davis, et al., 2003, Waldor, et al., 2003). The VGJΦ is able to recombine with the CTXΦ genome to originate a hybrid phage with the full potential for virulence conversion. The hybrid phage shows an increased infectivity due to its specificity for the receptor mannose-sensitive hemagglutinin (receptor mannose- sensitive hemagglutinin pilus), which is ubiquitous among environmental strains (Campos, et al., 2003a, Campos, et al., 2003b). The vibriophages KVP40 differs from many described vibriophages in having a broad host range and is reported to infect eight *Vibrio* species, including *Vibrio cholerae* and *Vibrio parahaemolyticus*, the nonpathogenic species *Vibrio natriegens*, and *Photobacterium leiognathi* (Matsuzaki, et al., 1992).

Vibriophages (family Vibrionaceae) contains the greatest number of reported phage-host systems for the marine environment (Moebus 1987), with the genus *Vibrio* comprising most of the hosts (Moebus & Nattkemper 1981). The phage VpVs phage infect only *V. parahaemolyticus* strains (Koga et al., 1982; Kellogg et al., 1995), phage P4 (Baross et al., 1974) and KVP20 (Matsuzaki et al., 1998) infect other *Vibrio* spp. (as the VpVs in this study), whereas phage V14 (Nakanishi et al., 1966) and KVP40 (Matsuzaki et al., 1992) have been reported to infect other genera. Vibriophage has also proved to be useful in studying the host chromosomes (Guidolin and Manning, 1987).

Vibrio cholera-specific filamentous bacteriophages CTXf was first identified in 1996 (Waldor and Mekalanos, 1996). Its genome includes the genes encoding cholera toxin, an AB 5- subunit type toxin secreted by *V. cholera* during its growth in the small intestine which causes secretory diarrhoea (Lencer and Tsai, 2003). The acquisition of CTXf is an important factor for *V. cholera* virulence. Virulence factors are frequently encoded within mobile genetic elements such as phages and plasmids (Davis and Waldor, 2002). The first reported filamentous phage horizontally transmitting a virulence factor that results in lysogenic conversion of a host to become virulent was CTXf (Waldor and Mekalanos, 1996; Ochman et al., 2000). Most of the characterized phages that integrate into their respective host chromosomes also undergo a reverse reaction wherein the phage genome excises from the chromosome (Azaro and Landy, 2002). However, excision of the CTXf prophage from the *V. cholera* chromosome has never been observed (Davis and Waldor, 2000). Instead, the chromosomally integrated CTXf prophage acts as a template for synthesis of viral DNA (Davis and Waldor, 2000; Moyer et al., 2001).

The study of Vibriophages is limited to the expressed genetic characteristics which are observed through experimental studies, but to get some insight of the Vibriophages and how its acquisition imparts host to gain various new characteristics leading to virulence and evolution of both phage-host systems, the study of phage genome is essential. The *in-silico* studies of hypothetical proteins (Uncharacterized proteins) for identifying their structure and function is an attempt to understand Vibriophages and their genomes with some possible implications.

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 enzymatic family.

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 (Edward et al., 2000; Dilip and Alankar,

2009). This study will helps us to understand the probable functions of hypothetical proteins in Vibriophages.

Several 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 and based on that, constructs the 3Dstructure for target protein using best scored template of orthologous family member. Here, we have predicted 3-D structure using Protein Structure Prediction Server (PS^2 server) (Dilip and Alankar, 2009; Zafer et al., 2006; Chih-Chieh et al., 2006).

2. MATERIALS AND METHODS

2.1 Sequence Retrieval

The Complete protein sequences for 21 different Vibrio phages *were* downloaded from the Database of KEGG (http://www.genome.jp/kegg/). The phages under study includes Vibrio phage kappa (Ehara, et. al., unpublished), Vibrio phage VP93,Vibrio phage VEJphi (Campos, 2010), Vibrio phage N4 (Das, et. al., Unpublished),Vibrio phage fs1 (Honma, et. al., 1997),Vibrio phage K139 (Kapfhammer, et. al., 2002),Vibrio phage KVP40 (Miller, et. al., 2003),Vibrio phage fs2 (Ikema, et. al., 1992),Vibrio phage VfO3K6,Vibrio phage VfO4K68,Vibrio phage Vf33,Vibrio phage Vf12,Vibrio phage VSK (Basu, Unpublished), Vibrio phage VpV262 (Hardies, et. al., 2003),Vibrio phage VHML, Vibrio phage VGJphi (Campos, et. al., Unpublished), Vibrio phage VP2 (Wang, Unpublished),Vibrio phage VP5,Vibrio phage VP882,Vibrio phage KSF-1phi (Faruque, et. al., 2005) and Vibrio phage VP4.

2.2 Functional Annotations

Hypothetical proteins were screened for the presence of enzymatic conserved domains using sequence similarity search with close orthologous family members available in various protein databases using the web-tools. Four bioinformatics web tools like CDD-BLAST (http://www.ncbi.nlm.nih.gov/BLAST/) (Altschul et al., 1997; Schaffer et al., 2001; Aron et al., 2006), INTERPROSCAN (http://www.abi.ac.uk/interpro) (Zdobnov and Rolf, 2001), Pfam (http://www.pfam.sanger.ac.uk/) (Alex et al., 2004) and COGs (http://www.ncbi.nih, gov/cog) (Roman et al., 2000) 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.

2.3 Functional Categorization

Hypothetical proteins analyzed by the function prediction web tools such as CDD-BLAST, INTERPROSCAN, PFAM and COGs have shown the variable results when searched for the conserved domains in hypothetical sequences.

2.4 Protein Structure Prediction

Several online protein structure prediction servers are available. Out of that, online PS^2 (PS Squared) Protein Structure Prediction Server was used (http://www.ps2.life.nctu.edu.tw/) (Chih-Chieh et al., 2006; Altschul et al.,1997; Schaffer et al., 2001; Cédric et al., 2000; Wendy et al.,2000), which accepts the protein (query) sequences in FASTA format and uses the strategies of Pair-wise and multiple alignment by combining powers of the programs PSI-BLAST, IMPALA and T-COFFEE in both target – template selection and target–template alignment and resultant target proteins 3D structures were 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. RESULTS AND DISCUSSION

The *in silico* structure and function of the Vibriophages was worked out for 21 phages. Out of 21 Vibriophages, conserved domain prediction in hypothetical proteins was possible in 13 phages. The hypothetical proteins were screened for the presence of enzymatic conserved domains using

sequence similarity search with close orthologous family members available in various protein databases using the web tools. The 3-D structure prediction of protein (query) sequences in FASTA format and uses the strategies of Pair-wise and multiple alignment by combining powers of the programs PSI-BLAST, IMPALA and T-COFFEE in both target – template selection and target– template alignment and resultant target proteins 3D structures were 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.

3.1 Functional Annotations and Protein Structure Prediction

The analysis of hypothetical proteins of Vibriophages was accomplished by using web tools for their classification into particular enzymatic family based on enzymatic conserved domain available in the sequence which are represented in respective Table 1 through 13. In 13 different Vibriophages, 215 hypothetical proteins resulted in 205 functional annotations out of which 92 are showing enzymatic conserved domains.

The $(PS)^2$ Server built the three dimensional structures for hypothetical proteins. Where in 17 different Vibriophage genome analyzed, $(PS)^2$ satisfactorily predicted structures of 54 hypothetical proteins using best scored orthologous template. The resulted 10 structures out of 54 showed no functional conserved domains may be due to lack of due to the lack of defined 3D structures for the aligned templates. The 3-D structures built are represented sequentially in respective Vibriophage 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 Vibriophage gene analyzed. The structure and functional data for Vibrio phage VfO3K6 (Table 1), Vibrio phage Vf33 (Table 2), Vibrio phage KSF-1phi Table 3), Vibrio phage VP4 (Table 4), Vibrio phage kappa (Table 5), Vibrio phage fs1 (Table 6), Vibrio phage K139 (Table 7), Vibrio phage KVP40 (Table 8), Vibrio phage VP93 (Table 9) , Vibrio phage N4 (Table 10) , Vibrio phage VP2 (Table 11) , Vibrio phage VP5 (Table 12) and Vibrio phage VP882 (Table 13) are given in their respective tables.

4. CONCLUSION

This study sorted some functional hypothetical proteins of Vibriophages 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 Vibriophages. Development in sequence analysis programming and ever growing genome sequence databases enhanced this methodology to draw conclusive functional relationships in the hypothetical proteins under study. Bioinformatics Web Tools like CDD-BLAST, INTERPROSCAN, PFAM and COGs have shown the ability to predict structure and functions in 215 hypothetical proteins of Vibriophages, in that sense assisted in predicting functional activity in 205 hypothetical proteins, out of which 10 showed only structural results and no functional activity was found in them. In all 54 3-D structures for hypothetical proteins was constructed using $(PS)^2$ serves as fast automated homology modeling web server. This predicted three dimensional structures may assist in establishing their role in life cycle of Vibriophages whose exact role in phage-host lifecycle is still unclear and can be used in future for the study of virulence and evolution of both phage-host systems.

5. DISCUSSION

The in-silico analysis of the hypothetical proteins is proved only on expression of the selective gene through cloning. The results obtained are concluded on the bases of available information in different databases and are valid till date.

Table 1 :Vibrio phage VfO3K6

NCBI gene ID	CDD BLAST	INTERPROSCAN	PFAM	COGS	Structures
1262767	No	No	Phage related protein & MraW methylase family	No	3cecA -34-35 -0.005

Table 2 Vibrio phage Vf33

NCBI gene ID	CDD BLAST	INTERPROSCAN	PFAM	COGS	Structures
2853318	No	No	Chromate transporter	No	No

Table 3 Vibrio phage KSF-1phi

NCBI gene ID	CDD BLAST	INTERPROSCAN	PFAM	COGS	Structures
3031573	No	No	Retrograde transport protein Dsl1 N & CRISPR-associated protein	No	No
3031575	No	No	Baculovirus 11 kDa family	No	No
3031578	No	No	Archaeal ATPase	ABC-type multidrug/protein/lipid transport system, ATPase component	No

Table 4 Vibrio phage VP4

NCBI gene ID	CDD BLAST	INTERPROSCAN	PFAM	COGS	Structures
3800005	Nucleoside/nucleotide kinase (NK)	No	No	No	No
3800011	No	No	No	No	1dekA-17- 132-6e-32

Table 5 Vibrio phage kappa

NCBI gene ID	CDD BLAST	INTERPROSCAN	PFAM	COGS	Structures
5850542	No	No	Thiamin pyrophosphokinase, catalytic domain & Sporulation related domain	No	No
5850551	S-adenosylmethionine-dependent methyltransferases	DNA methylase, N-6 adenine-specific, conserved site & N6 adenine-specific DNA methyltransferase	Methyltransferase small domain & Ribosomal L32p protein	No	2okcB -13- 64- 1e-11
5850544	Helix-turn-helix	Helix-turn-helix & Lambda repressor-like, DNA-binding	Helix-turn-helix	Predicted transcriptional regulators	1b0nA-20- 51- 1e-07
5850553	phage zinc-binding transcriptional activators	Phage transcriptional activator, Ogr/Delta	Ogr/Delta-like zinc finger,Insertion element protein & Dam-replacing	No	No
5850575	P2_Phage_GpR super family[cl06104	P2 phage tail completion R	P2 phage tail completion protein R (GpR)	No	No
5850569	No	No	MerC mercury resistance protein ,Diacylglycerol acyltransferase	No	No
5850560	No	No	Anti-sigma-K factor rskA , Bacteriophage lysis protein , Hepatic lectin, N-terminal domain	No	1i84S- 22-33- 0.010
5850548	Baseplate_J super family[cl01294	Baseplate assembly protein J-like, predicted	Baseplate J-like protein	No	No
5850543	No	No	Phage tail protein (Tail_P2_I)	No	No
5850540	No	No	Baculovirus polyhedron envelope protein, PEP, C terminus , FlgN protein	Methyl-accepting chemotaxis protein	No
5850550	No	No	BRO family, N-terminal domain , NTF2-like N-terminal transpeptidase domain	No	No
5850584	No	No	No	No	3cddD-14- 35 -3e-04

Table 6 Vibrio phage fs1

NCBI gene ID	CDD BLAST	INTERPROSCAN	PFAM	COGS	Structures
955575	No	No	Procyclic acidic repetitive protein (PARP) & Potato leaf roll virus readthrough protein	No	No
955576	No	No	Exonuclease VII,Bacillus transposase protein ,Reovirus sigma C capsid protein,Allexivirus 40kDa protein ,Baculovirus polyhedron envelope protein,Filoviridae VP35,Biogenesis of lysosome & Nucleopolyhedrovirus P10 protein	No	1m1jC-17- 38-0.002

955584	DNA replication initiation protein	No	No	Putative phage replication protein RstA	2gtqA- 24- 37- 0.002
955585	Rep_trans super family, Plasmid replication is initiated by the replication initiation factor (REP).	Replication initiation factor	Replication initiation factor	Putative phage replication protein RstA	No

Table 7 Vibrio phage K139

NCBI gene ID	CDD BLAST	INTERPROSCAN	PFAM	COGS	Structures
929070	No	No	Thiamin pyrophosphokinase, catalytic domain & Sporulation related domain	No	No
929074	No	No	Indoleamine 2,3-dioxygenase	No	No
929077	Helix-turn-helix	Helix-turn-helix	Helix-turn-helix	Predicted transcriptional regulators	1b0nA- 20- 51- 1e-07
929087	P2_Phage_GpR super family	P2 phage tail completion R	P2_Phage_GpR super family	No	No
929093	No	No	MerC mercury resistance protein & Diacylglycerol acyltransferase	No	No
929096	No	No	No	No	1i84S- 22- 33- 0.010
929101	Baseplate_J super family[cl01294], The P2 bacteriophage J protein lies at the edge of the baseplate.	Baseplate assembly protein J-like, predicted	Baseplate J-like protein	No	No
929102	No	No	Phage tail protein (Tail_P2_I)	No	No
929106	No	No	FlgN protein & Baculovirus polyhedron envelope protein	Methyl-accepting chemotaxis protein	No
929107	No	No	NTF2-like N-terminal transpeptidase domain & BRO family, N-terminal domain BRO-A and BRO-C are DNA binding proteins that influence host DNA replication and/or transcription	No	No
929108	No	No	No	No	3cddD- 14- 35- 3e-04
929110	Breast Cancer Suppressor Protein (BRCA1) & NAD-dependent DNA ligase	BRCT	BRCA1 C Terminus (BRCT) domain	NAD-dependent DNA ligase	No

Table 8 Vibrio phage KVP40

NCBI gene ID	CDD blast	Interproscan	Pfam	Cogs	Structure
2545647	F420_ligase super family[cl00644]	No	DUF218 domain	No	No
2545650	No	No	Caf1 Capsule antigen	No	No
2545653	No	No	Transglycosylase associated protein	No	No
2545654	Phage_head_chap super family[cl12668]	Bacteriophage T4, Gp40, head assembly	Head assembly gene product	Predicted ATP-dependent protease	No
2545674	No	No	Gap junction channel protein cysteine-rich domain	No	No
2545675	No	No	Plasmid conjugative transfer entry exclusion protein TraS	No	No
2545680	No	No	Fibronectin type III domain	No	No
2545681	DUF3307 super family[cl13235]	No	Protein of unknown function (DUF3307)	No	No
2545684	No	No	Glycosyltransferase family 52,Monogalactosyldiacylglycerol (MGDG) synthase	No	No
2545686	PRTase_typeII super family[cl12019], Phosphoribosyltransferase (PRTase) type II	Nicotinate phosphoribosyltransferase-like	Nicotinate phosphoribosyltransferase (NAPRTase) family	Nicotinic acid phosphoribosyltransferase	2g95B- 18- 320- 2e-88
2545687	30.2 super family[cl14359], hypothetical protein	No	No	No	2jarA- 14- 63- 5e-11
2545688	Radical_SAM super family[cl14056],NrdG[COG0602], Organic radical activating enzymes	No	Radical SAM superfamily	Organic radical activating enzymes	1tv8A- 16- 41- 3e-04
2545689	No	No	Poly(ADP-ribose) polymerase catalytic domain, RNA 2'-phosphotransferase, Tpt1 / KptA family	No	3c4hA- 15- 37- 0.008
2545691	No	No	Frag1/DRAM/Sfk1 family,(DUF2976), (DUF1625),(DUF2569), (DUF373)	No	1ii7A- 12- 40- 6e-04
2545694	MPP_superfamily super family[cl13995], Metallophosphatases	Metallo-dependent phosphatase	Calcineurin-like phosphoesterase	Predicted phosphohydrolases	No

	(MPPs),Metallophos[pfam00149]				
2545704	MPP_superfamily super family[cl13995], Metallophosphatases (MPPs)	No	No	Predicted phosphohydrolases	1xm7A- 24- 103- 2e-23
2545705	No	Prepilin-type cleavage/methylation, N-terminal	Prokaryotic N-terminal methylation motif	No	2hi2A- 63- 39- 3e-04
2545706	No	No	(DUF1469), (DUF973),(HAP),(DUF2614),(DUF2062),Vpu protein,UNC-50 family,ABC-2 type transporter,Secretion system effector C (SseC) like family	No	No
2545718	No	No	Post-segregation antitoxin CcdA,Ribosomal L29 protein, Leucine permease transcriptional regulator helical domain, Region found in RelA / SpoT proteins	No	No
2545725	No	No	Biofilm regulator BssS	No	No
2545728	No	No	PKC-activated protein phosphatase-1 inhibitor	No	No
2545733	DUF458 super family[cl00861]	Protein of unknown function DUF458, RNase H-like	Protein of unknown function (DUF458)	No	No
2545735	DUF2828[pfam11443]	No	Domain of unknown function (DUF2828)	No	1yvrA- 12- 94- 1e-19
2545736	No	No	Special lobe-specific silk protein SSP160	No	No
2545743	No	No	Acyl-ACP thioesterase	No	No
2545744	No	No	GatB domain,Septum formation initiator ,TATA element modulatory factor 1 DNA binding ,PspA/IM30 family,,She9 / Mdm33 family, Flagellar protein FliT	No	No
2545748	No	No	(DUF1014),ZF-HD protein dimerisation region	No	No
2545749	No	No	M protein trans-acting positive regulator (MGA) HTH domain	No	2hbtA- 13- 38- 0.002
2545750	No	No	CYTH domain	No	2fblB- 22- 79- 1e-15
2545751	No	No	Bacterial virulence protein (VirJ)	No	No
2545752	No	No	ORF6C domain, Gal4-like dimerisation domain,Bacillus transposase protein ,Acetyl co-enzyme A carboxylase carboxyltransferase alpha subunit, Tetrahydromethanopterin S-methyltransferase subunit B ,Toxic anion resistance protein (TelA),Baculovirus polyhedron envelope protein	Chromosome segregation ATPases	No
2545753	Adenine nucleotide alpha hydrolases superfamily including N type ATP PPases	No	No	Predicted ATPase (PP-loop superfamily), confers aluminum resistance	2pg3A- 14- 202- 3e-53
2545759	No	No	Bacterial alpha-L-rhamnosidase , ARP2/3 complex 16 kDa subunit (p16-Arc)	No	2i6gB- 15- 36- 0.006
2545760	No	No	Histidine kinase ,DUF576	No	No
2545761	30.2 super family[cl14359], hypothetical protein, COG1011[COG1011], Predicted hydrolase (HAD superfamily)	No	haloacid dehalogenase-like hydrolase	No	1q92A- 18- 52- 1e-07
2545763	No	No	Ubiquitin-fold modifier-conjugating enzyme 1	No	No
2545764	No	No	KRAB box ,Cytochrome C biogenesis protein,Septum formation topological specificity factor MinE	No	No
2545767	23 super family[cl14344], major capsid protein	No	Sucrose-6F-phosphate phosphohydrolase, Major capsid protein Gp23	No	No
2545768	No	No	XisH protein	No	No
2545771	No	Thioredoxin-like fold	Glutaredoxin	No	No
2545773	No	No	No	1,4-alpha-glucan branching enzyme	No
2545779	No	No	No	No	1potA -27- 36 -0.004
2545780	No	No	Iron dependent repressor, N-terminal DNA binding domain	No	No
2545782	GatB_Yqey super family[cl11497]	Aspartyl/glutamyl-tRNA amidotransferase subunit B-related,Protein of unknown function YOR215C, mitochondrial	Yqey-like protein	Uncharacterized ACR	1ng6A- 25- 125- 2e-30
2545788	No	No	RIO1 family , Beta-trefoil ,(DUF2972)	No	No
2545795	No	No	Herpes virus protein UL24	No	No
2545796	57B super family[cl14352]	RNA ligase/cyclic nucleotide phosphodiesterase	2',5' RNA ligase family	No	1jh6A- 18- 44- 2e-05

2545798	No	No	No	ATP-dependent protease Clp, ATPase subunit	No
2545799	No	No	Flagellar motor switch protein FliM	No	No
2545801	No	No	Nickel-containing superoxide dismutase	No	1qgrA- 10- 32- 0.004
2545805	No	No	Serpentine type 7TM GPCR chemoreceptor Srbc	No	No
2545809	No	No	PAAR motif, phosphotransferase system, EIIB	No	
2545814	Band_7 super family[cl02525]	Band 7 protein	SPFH domain / Band 7 family	membrane protease subunits, somatin/prohition homologs	3bk6A- 13- 128- 9e-31
2545825	NT5C super family[cl01869]	5'(3')-deoxyribonucleotidase,	5' nucleotidase, deoxy (Pyrimidine), cytosolic type C protein (NT5C)	No	3bwvB- 25- 159- 2e-40
2545826	DUF1768 super family[cl01271],	Bacteriophage GP30.3	Bacteriophage protein GP30.3	No	2b3wA- 20- 89 -5e-19
2545830	No	No	Thymidine kinase	No	No
2545834	alt[PHA02566]	No	No	No	1r45A- 16- 41- 4e-04
2545837	No	No	Integral membrane protein	No	No
2545840	No	No	Enterobacterial protein of unknown function	No	No
2545841	No	No	Zinc-binding domain of primase-helicase , PHD-finger	No	No
2545843	No	No	Ion transport protein ,(DUF2530) , (DUF1119)	No	No
2545844	No		Saposin-like type B, region 1	No	1oygA- 23- 34- 0.008
2545848	No	No	Flagellar protein FliT , MerR, DNA binding, Potyviridae polyprotein, Proteasome complex subunit Rpn13 ubiquitin receptor	No	No
2545849	No	Zinc finger, C2H2-like	Ribosomal protein L33	No	No
2545851	No	No	Enhancer of rudimentary, Lysophospholipase catalytic domain,Pyrrolo-quinoline quinone coenzyme N-terminus ,(DUF3228)	No	No
2545854	No	No	Predicted permease, (DUF2899)	No	
2545861	No	No	Ca2+ regulator and membrane fusion protein	No	No
2545863	No	No	Predicted membrane protein (DUF2324)	No	No
2545865	No	No	Epstein-Barr virus nuclear antigen 3 (EBNA-3)	No	No
2545866	No	No	M61 glycyl aminopeptidase	No	No
2545871	No	No	Protein of unknown function (DUF2682)	No	No
2545876	No	No	Periplasmic binding protein	No	No
2545877	No	DNA methylase, C-5 cytosine-specific, active site	No	No	No
2545878	No	No	Baculoviral E56 protein, specific to ODV envelope, Orbivirus NS3 ,Calcium-activated chloride channel	No	No
2545879	Nuc-transf super family[cl01417	Nucleotidyltransferase, predicted	Predicted nucleotidyltransferase	No	2v3cC- 14 -36- 0.007
2545884	No	No	WW domain	No	No
2545885	Lysine_decarbox super family[cl00695]	No	No	No	No
2545886	No	Neuraxin/MAP1B repeat	No	No	No
2545887	No	No	Agenet domain	No	No
2545889	No	No	CobW/HypB/UreG, nucleotide-binding domain , Borrelia burgdorferi BBR25 lipoprotein	ATPases with chaperone activity, ATP-binding subunit	1qvrA- 27- 35- 0.004
2545897	TFold super family[cl00263],	6-pyruvoyl tetrahydropterin synthase-related	6-pyruvoyl tetrahydropterin synthase	6-pyruvoyl-tetrahydropterin synthase	1y13A- 17- 49- 1e-06
2545899	No	Glutamine amidotransferase, type II	No	No	1ao0A- 17- 113- 2e-26
2545900	No	No	Prokaryotic membrane lipoprotein lipid attachment site	No	No
2545901	No	No	Queuine tRNA-ribosyltransferase	No	No
2545905	No	No	Glycosyl hydrolase family 98 ,(DUF1795)	No	No
2546061	NO	NO	DUF1219,Orthopoxvirus A49R protein,	NO	No
			DUF1967, Terpene synthase family, metal binding domain		
2546059	NO	No	Prokaryotic membrane lipoprotein lipid attachment site ,MLTD_N	NO	No
2546055	UvsW super family[cl13141]	DNA helicase, ATP-	ATP-dependant DNA helicase UvsW	NO	No

		dependent, UvsW			
	This family of proteins represents the DNA helicase UvsW from bacteriophage T4				
2546054	NO	NO	Phage portal protein, lambda family	NO	2jpnA- 28- 41- 2e-04
2546050	NO	No	Type IV leader peptidase family	NO	No
2546049	NO	No	Type IV leader peptidase family	NO	No
2546033	NO	NO	Adenoviral DNA terminal protein,FFD and TFG box motifs	NO	No
2546032	NO	NO	Quinohemoprotein amine dehydrogenase, gamma subunit	NO	No
2546029	NO	No	DNA binding domain of tn916 integrase	NO	No
2546021	NO	NO	Fibritin C-terminal region	NO	1nayA- 70- 39- 0.003
2546017	NO	NO	DNA gyrase C-terminal domain, beta-propeller	NO	No
2546012	NO	NO	Restriction endonuclease EcoRII, N-terminal Opioid growth factor receptor (OGFr) conserved region	NO	No
2546007	NO	No	Predicted membrane protein	NO	No
2546006	NO	No	General secretion pathway, M protein ,Bacterial protein of unknown function (DUF948) Cytomegalovirus TRL10 protein ,Sodium ion transport-associated	NO	No
2546005	NO	No	Colicin V production protein ,Srg family chemoreceptor SNARE associated Golgi protein ,Protein of unknown function (DUF3590)	NO	No
2546004	Macro_Poa1p_like[cd02901], Macro domain, Poa1p_like family	Appr-1-p processing	Macro domain	NO	1vhuA- 18- 40- 5e-04
2546002	SprT super family[cl01182], Predicted to have roles in transcription elongation	No	SprT-like family	Uncharacterized BCR	No
2546001	NO	NO	Uncharacterized protein conserved in archaea, NAF domain	NO	No
2545995	NO	NO	Cancer susceptibility candidate 1	NO	No
2545993	NO	NO	Iron/manganese superoxide dismutases, C-terminal domain	NO	No
2545991	NO	NO	Aldehyde dehydrogenase family ,Phage GP30.8 protein	NO	No
2545990	NO	No	Glycosyl transferases group 1	NO	No
2545987	NO	ATPase, AAA+ type, core,	NO	NO	No
2545983	DUF2829 super family[cl12744],This proteins found in bacteria and bacteriphages.	No	No	NO	No
2545981	NO	Hedgehog/DD-peptidase, zinc-binding motif Peptidase M15A, C-terminal	Peptidase M15	NO	1lbuA- 25- 42- 4e-05
2545979	NO	NO	NO	ATPase involved in DNA repair	No
2545978	NO	No	Heat shock factor binding protein 1 , Bacterial flagellin N-terminal helical region	No	No
2545977	NO	NO	Ribonuclease R winged-helix domain ,DUF1514	No	No
2545976	NO	NO	G10 protein	Predicted GTPase	No
2545970	NO	NO	AP2 domain	No	No
2545957	AdoMet_MTases[cd02440],S-adenosylmethionine-dependent methyltransferases	No	Methyltransferase small domain,DNA N-6-adenine-methyltransferase (Dam)	No	No
2545954	NO	NO	Sporulation lipoprotein YhcN/YlaJ (Spore_YhcN_YlaJ)	No	No
2545950	NO	NO	NO	Molecular chaperone	No
2545949	NO	No	Invasion associated locus B (IalB) protein	No	No
2545946	NO	No	Sugar (and other) transporter	No	No
2545940	NO	NO	Peptidyl-tRNA hydrolase PTH2	No	1rzwA - 23- 40- 4e-04
2545939	NO	NO	DUF2536,Uncharacterized protein conserved in bacteria (DUF2312)	No	No
2545936	NO	NO	NO	No	2hbtA- 14- 39- 0.001

2545929	NO	NO	MSV199 domain	No	No
2545926	NO	Armadillo-type fold	HEAT repeat	No	1lrvA- 21 - 49- 2e-06
2545924	NO	NO	Mor transcription activator family	No	No
2545923	NO	NO	Merozoite surface protein (SPAM) ,DUF2392, CTP synthase N-terminus	No	No
2545922	NO	NO	MULE transposase domain	No	No
2545917	NO	No	SUR7/PalI family ,DUF2499	No	No
2545914	NO	No	Colicin V production protein,Transmembrane amino acid transporter protein Wnt-binding factor required for Wnt secretion,DUF3021	No	No
2545913	NO	NO	Rad4 beta-hairpin domain 3 ,DUF2585	No	No
2545907	NO	NO	Arabidopsis thaliana protein of unknown function (DUF821)	No	No

Table 9 Vibrio phage VP93

NCBI Gene ID	CDD BLAST	INTERPROSCAN	PFAM	COGS	Structures
7853570	NO	NO	POPLD (NUC188) domain	NO	No
7853571	NO	NO	Ribosomal protein S9/S16 CENP-B N-terminal DNA-binding domain	NO	No
7853573	NO	NO	tRNA synthetases class II (A)	NO	No
7853580	NO	NO	SOCS box	NO	No
7853581	No	No	No	No	1u3eM - 16- 41- 4e-05
7853583	NO	NO	Prolyl 4-Hydroxylase alpha-subunit, N-terminal region	NO	No
7853584	NT_Pol-beta-like super family[cl11966]	NO	Poly A polymerase head domain	tRNA nucleotidyltransferase/poly(A) polymerase	1ou5A- 41- 42- 1e-04
7853585	PHA02030[PHA02030], hypothetical protein	NO	NO	NO	No
7853587	NO	Phosphoribosyl-ATP pyrophosphohydrolase-like	Phosphoribosyl-ATP pyrophosphohydrolase	Predicted pyrophosphatase	1vmgA- 27- 47- 3e-06
7853592	cyt_kin_arch[TIGR02173]	NO	AAA domain (dynein-related subfamily)	ATPase involved in DNA replication	1dekA- 15 -42- 5e-05
7853594	NO	NO	6-phosphofructo-2-kinase	NO	No
7853600	NO	NO	Rickettsia 17 kDa surface antigen Bacteriocin class II with double-glycine leader peptide LMBR1-like membrane protein	NO	2i1kA- 14- 37- 0.004
7853605	PHA02046 super family[cl10354]	NO	EspA-like secreted protein	DNA-directed RNA polymerase sigma subunits (sigma70/sigma32)	No
7853607	NO	NO	IRSp53/MIM homology domain ,Histidine kinase Phi29 scaffolding protein,Centromere protein H (CENP-H)	NO	No
7853608	NO	NO	Ribosomal protein S30	NO	No
7853609	Peptidase_M15_3 super family[cl01194], Peptidase M15	Hedgehog/DD-peptidase, zinc-binding motif Peptidase M15A, C-terminal	Peptidase M15	TPR-repeat-containing proteins	1lbuA- 21- 50- 2e-07
7853613	NO	NO	Cyclin, N-terminal domain	NO	No

Table 10 Vibrio phage N4

NCBI Gene ID	CDD BLAST	INTERPROSCAN	PFAM	COGS	Structures
8676422	NO	NO	Villin headpiece domain	NO	No
8676425	Wzz super family[cl01623]	NO	DUF848, STAT protein, all-alpha domain Autophagy protein 16 (ATG16),DUF641 Tumour-suppressor protein CtIP N-terminal domain TATA element modulatory factor 1 DNA binding Afadin- and alpha -actinin-Binding Spc7 kinetochore protein,Kinesin-related Cobalamin adenosyltransferase,TMPIT-like protein CorA-like Mg2+ transporter protein Erp protein C-terminus	ATPase involved in DNA repair	No
8676426	No	No	No	No	1dekA-

					17- 125- 8e-30
8676431	NO	Beta tubulin, autoregulation binding site	Endoplasmic reticulum-based factor for assembly of V-ATPase	NO	No
8676439	NO	NO	DUF2675,Brucella outer membrane protein 2	NO	No
8676447	NO	Bacteriophage T7-like, gene 6.7	NO	NO	No
8676464	NO	NO	DUF2133,GHMP kinases N terminal domain	NO	No
8676465	NO	NO	Lysis protein,Prokaryotic membrane lipoprotein lipid attachment site	NO	No

Table 11 Vibrio phage VP2

NCBI Gene ID	CDD BLAST	INTERPROSCAN	PFAM	COGS	Structures
2948097	NO	NO	Outer membrane lipoprotein LolB	NO	No
2948099	NO	NO	VRR-NUC domain	NO	No
2948114	NO	NO	D-Ala-teichoic acid biosynthesis protein,Rap-phr extracellular signalling	NO	No
2948115	NO	Peptidase S26A, signal peptidase I, serine active site	NO	NO	No
2948116	NO	NO	Minor capsid	NO	No
2948120	NO	NO	Tim17/Tim22/Tim23 family,Mitochondrial ribosomal protein L28	NO	No
2948127	NO	NO	(DUF2459), short chain dehydrogenase	NO	No
2948137	NO	NO	Mediator complex subunit 3 fungal	NO	No
2948139	NO	NO	Sulfolobus plasmid regulatory protein	NO	No

Table 12 Vibrio phage VP5

NCBI Gene ID	CDD BLAST	INTERPROSCAN	PFAM	COGS	Structures
5741329	NO	No	Glycosyl hydrolases family 6	NO	No
5741330	NO	NO	SYF2 splicing factor	NO	No
5741333	HDc super family[cl00076] Metal dependent phosphohydrolases with conserved 'HD' motif	Metal-dependent phosphohydrolase, HD domain	HDOD domain	Predicted hydrolases of HD superfamily	2gz4A- 22- 52- 7e-08
5741338	NO	NO	Sulfolobus plasmid regulatory protein	NO	No
5741340	NO	NO	PaaX-like protein,Glycosyl transferase family, helical bundle domain	NO	No
5741347	NO	Hedgehog/DD-peptidase, zinc-binding motif Peptidase M15A, C-terminal	Peptidase M15	NO	1lbuA- 27-43- 2e-05
5741353	NO	NO	DUF2459, short chain dehydrogenase	NO	No
5741361	NO	NO	VRR-NUC domain	NO	No
5741365	NO	NO	Mediator complex subunit 3 fungal	NO	No
5741366	NO	NO	Minor capsid	NO	No

Table 13 Vibrio phage VP882

NCBI Gene ID	CDD BLAST	INTERPROSCAN	PFAM	COGS	Structures
5076227	NO	NO	Probable metal-binding protein (DUF2387) Zn-ribbon-containing, possibly nucleic-acid-binding protein (DUF2310)	NO	No
5076229	NO	No	Organic Anion Transporter Polypeptide (OATP) family	NO	No
5076232	NO	NO	Caenorhabditis protein of unknown function, DUF268	NO	No
5076244	NO	Phage DNA packaging Nu1 Winged helix-turn-helix transcription repressor DNA-binding	Phage DNA packaging protein Nu1	NO	1j9iA- 40-63- 5e-11
5076267	NO	NO	Vps23 core domain	NO	No
5076268	NO	NO	Asp/Glu/Hydantoin racemase , PsbP	NO	No

6. ACKNOWLEDGEMENT

We are thankful to Miss. Kimi Patel and Miss. Lekha Patel for their help and assistance in this work.

7. REFERENCES

1. A. Guidolin and P.A.Manning: Genetics of *Vibrio cholerae* and its bacteriophages. *Microbiol Rev* (1987), 51:285-298.

2. 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., (2004). The Pfam families' database. Nucleic Acids Research, Vol. 32, D138-D141.

3. Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., Lipman, D. J., (1997). Gapped BLAST and PSI-BLAST: "*a new generation of protein database search programs"*. Nucleic Acids Res. 25 (17), 3389-402.

4. 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., (2006). CDD: "*a conserved domain database for interactive domain family analysis.* "Nucleic Acids Research, Vol. 35, D237–D240.

5. Azaro, M.A., and Landy, A. (2002) l integrase and the l Int family. In: *Mobile DNA II*. Craig, N.L. (ed.). Washington, DC: American Society for Microbiology Press, pp. 118–148

6. B. M. Davis and M. K. Waldor. "*Filamentous phages linked to virulence of Vibrio cholera"*. Curr. Opin. Microbiol. (2003), 6:35–42.

7. Basu,N., Kar,S. and Ghosh,R.K. "*Molecular analysis of filamentous phage VSK of Vibrio cholerae 0139*: "A possible clue to genetic transmission Unpublished

8. C. A. Kellogg, J.B. Rose, S.C. Jiang, J.M. Thurmond and J.H. Paul," *Genetic diversity of vibriophages isolated from marine environments around Florida and Hawaii,"* USA. *Mar Ecol Prog Ser* (1995), 120: 89–98.

9. Campos, J., Martinez, E., Izquierdo, Y. and Fando, R. VEJ {phi*}, "A novel filamentous phage of Vibrio cholerae able to transduce the cholera toxin genes*." Microbiology (Reading, Engl.) 156 (PT 1), 108-115 (2010)

10. Campos,J., Martinez,E., Suzarte,E., Rodriguez,B.L., Marrero,K., Silva,Y.K., Ledon,T.Y., Del Sol,R.E. and Fando,R.A. VGJphi: "*A Novel Lysogenic Filamentous Phage of Vibrio cholerae which Shares the Same Integration Site with CTXphi* " Unpublished

11. Cédric, N., Desmond, G. H., Jaap, H., (2000). T-coffee: "*a novel method for fast and accurate multiple sequence alignment*." J. Mol. Biol. 302, 205-217.
12. Chih-Chieh, C., Jenn-Kang, H., Jinn-Moon, Y., (2006). $(PS)^2$: "*protein structure prediction server Nucl*." Acids Res. 34, W152-W157.

13. Das,M., Bhowmick,T.S., Sarkar,B.L., Nair,G.B., Yamasaki,S. and Nandy,R.K. "*Complete genome sequence of lytic vibriophage N4 indicates close relativeness of T7 viral supergroup Unpublished*

14. Davis, B.M., and Waldor, M.K. "*(2000) CTXf contains a hybrid genome derived from tandemly integrated elements." Proc Natl Acad Sci USA* 97: 8572–8577.

15. Davis, B.M., and Waldor, M.K. (2002) '"*Mobile genetic elements and bacterial pathogenesis".* In: *Mobile DNA II*. Craig, N.L. (ed.). Washington, DC: American Society for Microbiology Press, pp. 1040–1059.

16. Dilip, G., Alankar, R., (2009). "*Computational Function and Structural Annotations for Hypothetical Proteins Bacillus anthracis.* Biofrontiers", 1, 27-36.

17. E. A. Jouravleva, G. A. McDonald, C. F. Garon, M. B. Finkelstein, and R. A. Finkelstein. "*Characterization and possible function of a new filamentous bacteriophage from Vibrio cholera".* Microbiology (1998), 144:315–324.

18. Edward, E., Gary, L. G., Osnat, H., John, M., John, O., Roberto, J. P., Linda, B., Delwood, R., Andrew, J. H., (2000). "*Biological function made crystal clear- annotation of hypothetical proteins via structural genomics."* Current Opinion in Biotechnology 11, 25-30.

19. Ehara, M., Nguyen, M.B., Nguyen,T.D., Ngo,C.T., Le,H.T., Nguyen,T.H. and Iwami,M. Integrated kappa phage genome Unpublished

20. Faruque,S.M., Bin Naser., Fujihara., Diraphat., Chowdhury., Kamruzzaman., Qadri., Yamasaki,S., Ghosh,R.K. and Mekalanos,J.J. Genomic sequence and receptor for the Vibrio cholerae phage KSF-1phi: "*evolutionary divergence among filamentous vibriophages mediating lateral gene transfer."* J. Bacteriol. 187 (12), 4095-4103 (2005)

21. H. Nakanishi, Y. Iida, K. Maeshima, T. Teramoto, Y. Hosaka and M. Ozaki. "*Isolation and properties of bacteriophages of Vibrio parahaemolyticus*. "*Biken J* (1966), 9: 149– 157.

22. Hardies, S.C., Comeau, A.M., Serwer, P. and Suttle, C.A. "*The complete sequence of marine bacteriophage VpV262 infecting vibrio parahaemolyticus indicates that an ancestral component of a T7 viral supergroup is widespread in the marine environment";* Virology 310 (2), 359-371 (2003)

23. Honma, Y., Ikema, M., Toma, C., Ehara, M. and Iwanaga, M." *Molecular analysis of a filamentous phage (fsl) of Vibrio cholerae O139*;" Biochim. Biophys. Acta 1362 (2-3), 109-115 (1997)

24. J. Campos, E. Martinez, E. Suzarte, B. L. Rodriguez, K. Marrero, Y. Silva, T. Ledo´n, R. del Sol, and R. Fando. VGJ_, "*a novel filamentous phage of Vibrio cholerae, integrates into the same chromosomal site as CTX"*_. J. Bacteriol. (2003), 185:5685–5696.

25. J. Campos, E. Martinez, K. Marrero, Y. Silva, B. L. Rodriguez, E. Suzarte, T. Ledon, and R. Fando*.' Novel type of specialized transduction for CTX_ or its satellite phage RS1 mediated by filamentous phage VGJ_ in Vibrio cholera"*. J. Bacteriol. (2003), 185:7231–7240.

26. J.A. Baross, J. Liston, and R.Y. Morita. "*Some implications of genetic exchange among marine vibrios, including Vibrio parahaemolyticus, naturally occurring in the Pacific oyster".* In *International Symposium on* Vibrio parahaemolyticus. Fujino, T., Sakaguchi, G., Sakazaki, R., and Takeda, Y. (eds). Tokyo, Japan: Saikon Publishing, (1974), pp. 129–137.

27. K. Moebus and H. Nattkemper. '*Bacteriophage sensitivity patterns among bacteria isolated from marine waters. Helgolander Meeresunters.*"(1981), 34: 375-385

28. K. Moebus. Ecology of manne bacteriophages. In: Goyal, S. M., Gerba. C. P, Bitton, G. (eds.) Phage ecology. John Wiley & Sons. New York, (1987), p. 137-156

29. Kapfhammer, D., Blass, J., Evers, S. and Reidl, J. Vibrio cholerae phage K139: "*complete genome sequence and comparative genomics of related phages";* J. Bacteriol. 184 (23), 6592-6601 (2002)

30. M. Ikema and Y. Honma*. "A novel filamentous phage, fs2, of Vibriocholerae O139*." Microbiology (1998), 144:1901–1906.

31. M. K. Waldor and J. J. Mekalanos"*. Lysogenic conversion by a filamentous phage encoding cholera toxin".* Science (1996), 272:1910–1914.

32. Miller,E.S., Heidelberg,J.F., Eisen,J.A., Nelson,W.C., Durkin,A.S., Ciecko,A., Feldblyum,T.V., White,O., Paulsen,I.T., Nierman,W.C., Lee,J., Szczypinski,B. and Fraser,C.M. "*Complete genome sequence of the broad-host-range vibriophage KVP40: comparative genomics of a T4-related bacteriophage"* J. Bacteriol. 185 (17), 5220-5233 (2003)

33. Moyer, K.E., Kimsey, H.H., and Waldor, M.K. (2001) "*Evidence for a rolling-circle mechanism of phage DNA synthesis from both replicative and integrated forms of CTXf. Mol Microbiol 41*:" 311–323.

34. Ochman, H., Lawrence, J.G., and Groisman, E.A. (2000)" *Lateral gene transfer and the nature of bacterial innovation*." *Nature* 405: 299–304.

35. Roman, L. T., Michael, Y., Galperin, Darren A. Natale, Eugene V. Koonin (2000). "*The COG database: a tool for genome –scale analysis of protein functions and evolution. Nucleic Acid Research*." 28, 33-36.

36. S. Kar, R. K. Ghosh, A. N. Ghosh, and A. Ghosh." *Integration of the DNA of a novel filamentous bacteriophage VSK from Vibrio cholerae O139 into the host chromosomal DNA. FEMS Microbiol*." Lett. (1996), 145:17–22.

37. S. M. Faruque, I. Bin Naser, K. Fujihara, P. Diraphat, N. Chowdhury, M. Kamruzzaman, F. Qadri, S. Yamasaki, A. N. Ghosh, and J. J. Mekalanos. "*Genomic sequence and receptor for the Vibrio cholerae phage KSF-1_: evolutionary divergence among filamentous vibriophages mediating lateral gene transfer"*. J. Bacteriol. (2005), 187:4095–4103.

38. S. Matsuzaki, S. Tanaka, T. Koga, and T. Kawata. "*A broad-host-range vibriophage, KVP40, isolated from sea wate*r". Microbiol. Immunol. (1992), 36:93–97.

39. S. Matsuzaki, T. Inoue, M. Kuroda, S. Kimura and S. Tanaka. "*Cloning and sequencing of major capsid protein (mcp) gene of a vibriophage, KVP20, possibly related to Teven coliphages" Gene* (1998), 222: 25–30.

40. Schaffer, A. A., Aravind, L., Madden, T. L., Shavirin, S. Spouge, J. L., Wolf, Y. I., Koonin, E. V., Altschul, S. F., (2001*)." Improving the accuracy of PSI-BLAST protein database searches with composition-based statistics and other refinements".* Nucleic Acids Res. 29(14), 2994-3005.

41. T. Koga, S. Toyoshima and T. Kawata. "*Morphological varieties and host ranges of Vibrio parahaemolyticus bacteriophages isolated from seawater". Appl Environ Microbiol* (1982), 44: 466–470.

42. W.I. Lencer and B. Tsai. "*The intracellular voyage of cholera toxin: going retro.* "Trends Biochem Sci (2003), 28: 639– 645.

43. Wang,D., Kan,B., Li,Y., Liu,Z., Gao,S., Liu,Y., Liang,W., Zhang,L., Yan,M., Li,W., Liu,G., Liu,Y., Li,J., Diao,B., Zhu,Z. and Qiu,H. Vibrio cholerae phage VP2 complete genome Unpublished

44. Wendy, B. et al., (2000). "*The EMBL Nucleotide Sequence Database".* Nucleic Acid Research. 28, 19-23

45. Zafer, A., Yucel, A., Mark, B., (2006). "*Protein secondary structure prediction for a single-sequence using hidden semi-Markov models, BMC Bioinformatics*, "7, 178.

46. Zdobnov, E. M., Rolf, A., (2001*)." Interproscan- an integration platform for the signatures recognition methods in InterPro. Bioinformatics* "17, 847-848