Vibrio cholerae Typing Phage N4: Genome Sequence and Its Relatedness to T7 Viral Supergroup

Mayukh Das^a R.K. Nandy^a Tushar Suvra Bhowmick^a S. Yamasaki^b A. Ghosh^a
G.B. Nair^a B.L. Sarkar^a

^aDivision of Bacteriology, National Institute of Cholera and Enteric Diseases, Kolkata, India;

^bLaboratory of International Prevention of Epidemics, Department of Veterinary Science, Graduate School of Life and Environmental Sciences, Osaka Prefecture University, Osaka, Japan

Key Words

Cholera • Vibriophage • Lytic phage • Hybridization • Genome • Comparative genomics

Abstract

Objectives: In countries where cholera is endemic, *Vibrio cholerae* O1 bacteriophages have been detected in sewage water. These have been used to serve not only as strain markers, but also for the typing of *V. cholerae* strains. Vibriophage N4 (ATCC 51352-B1) occupies a unique position in the new phage-typing scheme and can infect a larger number of *V. cholerae* O1 biotype El Tor strains. Here we characterized the complete genome sequence of this typing vibriophage.

Methods: The complete DNA sequence of the N4 genome was determined by using a shotgun sequencing approach.

Results: Complete genome sequence explored that phage N4 is comprised of one circular, double-stranded chromosome of 38,497 bp with an overall GC content of 42.8%. A total of 47 open reading frames were identified and functions could be assigned to 30 of them. Further, a close relationship with another vibriophage, VP4, and the enterobacteriophage T7 could be established. DNA-DNA hybridization among *V. cholerae* O1 and O139 phages revealed homology among O1 vibriophages at their genomic level. **Conclusions:**

This study indicates two evolutionary distinctive branches of the possible phylogenetic origin of O1 and O139 vibriophages and provides an unveiled collection of information on viral gene products of typing vibriophages.

Copyright © 2011 S. Karger AG, Basel

Introduction

Cholera is an acute food and waterborne communicable disease caused by Gram-negative, γ -proteobacteria *Vibrio cholerae* under the family Vibrionaceae. It has been a common cause of infant and adult deaths in developing countries since the early 19th century. *V. cholerae* is divided into more than 206 serogroups based on LPS structure [1]. Primarily, O1 and O139 serogroups have been associated with epidemic and pandemic cholera in humans. The O1 serogroup of *V. cholerae* can further be subdivided into two biotypes: classical and El Tor [1]. Of the seven pandemics that have taken place, the first six were due to the classical strain of *V. cholerae* while the seventh was due to the El Tor [1].

During a cholera outbreak, it is often important to know quickly the serotype and biotype of the causative strain, and bacteriophages have proven invaluable in as-

certaining this. A phage-typing scheme was developed to differentiate *V. cholerae* O1 El Tor by Basu and Mukerjee [2] (1968) in Kolkata (previously called Calcutta). NICED is a WHO collaborating center for research and training in diarrheal diseases and is recognized as a vibriophage reference laboratory. Since 1968, it has been receiving strains of *V. cholerae* from all parts of India and abroad for biotyping, serotyping and phage typing. With the passage of time, identification of the limitations of this scheme led to the development of a new phage-typing scheme for the *V. cholerae* O1 biotype El Tor [3]. The new typing scheme is used to categorize strains of *V. cholerae* O1 and is an eloquent testimony of its importance in the area of *V. cholerae* epidemiology [4]. After the emergence of *V. cholerae* O139 strains, a new phage-typing scheme specific to *V. cholerae* O139 was developed [5]. Vibriophage N4 is one of the key phages of the new phage-typing scheme [3]. It has the ability to lyse a broader array of *V. cholerae* O1 biotype El Tor strains compared to the other phages belonging to the new typing scheme. Though quite a large number of *V. cholerae* phages have been described, the complete genome sequence of only a few of them are known. Moreover, to the best of our knowledge, with one exception, all of them are lysogenic phages. There is a dearth of sequence information on the genome sequence of lytic vibriophages, many of which occupy key positions in the typing schemes. In this paper, we report the complete genome sequence of vibriophage N4 and compare it with other known bacteriophages available in the database. We show it shares maximum similarity with vibriophage VP4 and *Escherichia coli* phage T7, both at the nucleotide and the amino acid levels.

Materials and Methods

Bacteria and Bacteriophages

The bacterial strains used for the present study are *V. cholerae* O1 biotype El Tor strain MAK 757 (ATCC 51352) and *V. cholerae* O139 strain NPR-4 as the propagating strains for *V. cholerae* O1 and O139 phages, respectively. All strains of *V. cholerae* were identified and confirmed by using standard techniques [6]. Vibriophage N4 (ATCC 51352-B1) used in this study for molecular characterization was obtained from the phage repository bank of the Vibrio Phage Reference Laboratory, Division of Bacteriology, NICED, Kolkata. Other phages of the new phage-typing scheme, i.e. S5, S20, M4, D10 of O1 and MAD-5 of the O139 phage-typing scheme, and one phage for the *V. cholerae* O1 biotype El Tor isolated from aquatic environs in Kolkata are incorporated in this study.

Isolation and Purification of Bacteriophage and Phage DNA

Phage lysate was prepared using culture of *V. cholerae* O1 El Tor MAK 757 and O139 NPR-4 mixed with the O1 and O139 phages, respectively, at a multiplicity of infection of 0.01 by standard methods adopted previously in our laboratory [3, 4]. The phage particles were purified by the method of Yamamoto et al. [7] (1970) with certain modifications, and genomic DNA from purified phage particles free from host genomic DNA was prepared following the protocol as described [8].

Construction of the Phage Genomic DNA Library

Purified vibriophage N4 DNA was confirmed to be sensitive to cleavage with *Hind*III as reported previously [3]. Agarose gel electrophoresis of the digested genome gave an estimated genome size of circa 38.02 kb. Phage DNA fragments between 500 and 600 bp long generated by random physical shearing using an ultrasonic processor (Vibra-Cell; Sonics and Materials Inc., USA) were used for the genomic DNA library. The phage DNA fragments were blunt-ended by using Mung bean nuclease according to the procedure supplied by the manufacturer (New England Biolabs, Ipswich, Mass., USA). Blunt-ended phage DNA fragments were cloned by using Zero Background/Kan Cloning Kit (Invitrogen, USA) and were ligated into this cloning vector by using T4 DNA ligase (Invitrogen). Ligated products were electroporated into *E. coli* DH10B cells in a gene pulser apparatus and pulse controller (Bio-Rad, USA). Transformed electrocompetent *E. coli* DH10B cells were incubated overnight at 37° on Luria agar (HiMedia, India) plates containing kanamycin (50 µg/ml; Sigma, USA) and 1 mM IPTG (isopropyl-β-D-thiogalactoside; Sigma, USA). The colony obtained from each individual clone was mixed with 20% glycerol (Merck, Germany) and was stored at -70° with proper nomenclature to complete the preparation of the phage genomic DNA library.

Recombinant plasmid DNA-carrying a phage DNA insert was isolated from the *E. coli* cells by alkaline lysis/PEG treatment method. Digested plasmids with *Not*I restriction endonuclease (Takara, Japan) along with uncut pZErO-2 vector and *Eco*RV digested pZErO-2 vector were separated on a 1.5% agarose (Sigma, USA) gel to examine isolated plasmids with fragmented phage DNA inserts.

Complete Nucleotide Sequence of Genome of Vibriophage N4 and Prediction of Open Reading Frames

The nucleotide sequences were assembled using the Sequencher 4.1 software package (Gene Codes). Trimmed and edited clone sequences were assembled into three contigs. Three gaps were formed between these contigs, which were connected by primer walking. Six primers were required to cover up these gaps and designated as walking primers. Primer walking using six primers facilitated to complete the whole genome and a draft sequence of the phage N4 genome was made.

Nucleotide Sequencing and Sequence Assembly

The isolated and purified plasmid DNA inserts of randomly selected clones from the phage genomic DNA library were sequenced from both ends with primers M13F and M13R using the BigDye Terminator v.3.1 Cycle Sequencing Kit with the ABI 3100 Genetic Analyzer Sequencer (Applied Biosystems, USA). DNA sequence data were assembled into contigs using the Sequencher 4.1 software package (Gene Codes) and finally rechecked with the

Lasergen software suite (DNASTAR Inc., USA). Sequencing reads were trimmed of vector using Sequencher and further trimmed manually after chromatograms were inspected for ambiguities. The initial assembly was done under stringent assembly conditions (90% sequence match with a minimum overlap of 50 bases), which yielded contigs consisting of sequences meeting the assembly criteria. Contigs were edited by removing nucleotide gaps and replacing miscalled bases by using a majority rule from the base-quality chromatograms. Additional ambiguities were corrected by determining a consensus for base calling. The sequence is expected to be 99.999% accurate, or have a statistical error of less than 1 in 10,000 bp. Several rounds of assembly and editing were performed, and more relaxed parameters were used at each successive assembly procedure (minimum 75% sequence match, and a 30-base overlap region) until no more assemblies were found by the software. After performing several rounds of assembly of clone sequence data, many gaps were formed between several pairs of contigs which were connected by primer walking. Pools of primers were used in a simplex PCR to amplify regions of genomic DNA corresponding to DNA sequences between contigs. PCR was performed with a PTC 200, Peltier Thermal Cycler (M. J. Research, USA), with a 200- μ l PCR tube containing a reaction mixture volume of 25 μ l. Each reaction contained 2.5 μ l of 10 $\times$ PCR buffer [500 mM KCl, 100 mM Tris HCl (pH 8.0), 15 mM MgCl₂ and 0.01% gelatin] (Takara, Japan), 2.5 μ l of 25 mM dNTP, 1 μ l each of the forward and reverse primer (10 pmol/ μ l) of a single primer pair, 0.2 μ l of AmpliTaq DNA polymerase (Takara, Japan) and 1 μ l of phage DNA template and Mili-Q water to a final volume of 25 μ l. Amplification conditions were 5 min at 94 $^{\circ}$ for initial denaturation followed by 30 cycles consisting of 30 s at 94 $^{\circ}$, 40 s at 50 $^{\circ}$ and 3.5 min at 72 $^{\circ}$, with a final round of extension for 7 min at 72 $^{\circ}$ at the end, followed by maintenance at 4 $^{\circ}$. After amplification, 5 μ l of each reaction mixture was separated on a 1.5% agarose (Sigma, USA) gel. After completion of whole genome sequence data assembly, it was reconfirmed by two methods. In one method, comparison was performed between the experimental restriction digestion profile of phage DNA with *AccI*, *EcoRV*, *HaeIII* and *HindIII* and the theoretical restriction digestion virtual profile with the same enzymes created by pDRAW32 (AcaClone Software, USA). Another method was comprised of comparison between experimental restriction digestion profile of PCR-amplified products using custom primers and the theoretical restriction digestion virtual profile of amplified regions of the genome. PCR-amplified products using custom primers were 3–4 kb in size and were distributed throughout the genome with 500-bp overlaps with each other.

Eighty-two custom primers were designed to fill in single-stranded gaps and were used to resolve the ambiguities of the assembled phage genome. These primers were distributed throughout the phage genome and had 500-bp overlaps with each other. Direct genome sequencing of the phage DNA as a template with these custom-made primers was carried out to confirm the draft sequence, thereby making it the final one. Each nucleotide position was sequenced a minimum of two times, and both cloned DNA and direct genome sequencing were used. The complete assembled nucleotide sequence data was rechecked with the Lasergen software suite (DNASTAR Inc.).

Open Reading Frames Prediction and Genome Annotation

Nucleotide sequence data of whole genome of phage was compared by BLAST search (blastn) against the National Center for Biotechnology Information (NCBI) GenBank database to find the sequence homology. Open reading frame (ORF) coordinates were predicted by using Heuristic GeneMark (http://opal.biology.gatech.edu/GeneMark/heuristic_hmm2.cgi) and also by ORF Finder software from the NCBI (<http://www.ncbi.nlm.nih.gov/gorf/gorf.html>) using the bacterial start and stop codon option. Predicted ORFs were compared using a BLAST search (blastp and psi-blast) against the NCBI GenBank nonredundant protein database (nr), Prosite, GOanna, Pfam and Fasta3 to investigate protein domains and families.

Southern Hybridization

A probe of 1 μ g of N4 vibriophage genomic DNA, double-digested with *HaeIII* and *HindIII* endonucleases (Takara, Japan), was labeled with the AlkPhos direct labeling and detection system (Amersham Biosciences Corp., USA) according to the manufacturer's instructions as described below. The concentration of the phage DNA for the preparation of labeled probe was selected according to the protocol used previously [9]. Purified phage DNAs (2 μ g from each phage) from vibriophage N4, S5, D10, MAD-5 and A1 were digested with restriction endonucleases *EcoRV*, *HaeIII*, *HindIII* (Takara, Japan) and electrophoresed on a 1% agarose (Sigma, USA) gel. The resulting fragments were transferred onto Hybond N+ nylon membrane (Amersham Biosciences Corp., USA) by a modification of the Southern transfer procedure [10].

Results

After completion of whole genome draft sequence, reconfirmation of sequence data assembly was carried out by comparison between the experimental restriction digestion profile of phage DNA with *AccI*, *EcoRV*, *HaeIII* and *HindIII* and the theoretical restriction digestion virtual profile with the same enzymes created by pDRAW32 (AcaClone Software, USA). An experimental finding (fig. 1) exhibited that each fragment of experimental digestion profile matched with the theoretical profile by number and by molecular weight and confirmed the authenticity of whole genome draft sequence data assembly of vibriophage N4. This draft sequence exposed a fact that the vibriophage N4 has a circular genome which is documented in figure 2 as it was created on this draft sequence.

The entire phage genome was amplified by 12 PCR product fragments using 24 custom-made primers for final validation. The fragments were between 2,137 and 5,064 bp in size and were distributed throughout the genome with 500 bp overlaps with each other. The finding showed that the expected (theoretical) size of each PCR product and digested fragment matched with the experimentally obtained size (data not shown). Successful com-

Fig. 1. Comparison between experimental and theoretical restriction digestion profile of phage N4. **a** Experimental restriction digestion profile. Lanes 1 and 6: λ -HindIII-digested marker; lanes 2–5: digested with *EcoRV*, *HindIII*, *HaeIII* and *AccI*, respectively. **b** Theoretical restriction digestion virtual profile. Lanes 1 and 6: λ -HindIII-digested marker; lanes 2–5: digested with *EcoRV*, *HindIII*, *HaeIII* and *AccI*, respectively.

Fig. 2. Overview of the vibriophage N4 genome showing predicted ORFs with provisional functional gene clusters.

Color version available online

Fig. 3. Comparison at nucleotide level between the predicted ORFs of vibriophage N4, VP4 and enterobacteriophage T7.

pletion of whole genome sequencing revealed that vibriophage N4 comprised one circular, double stranded chromosome of 38,497 bp in size. The overall GC content of this genome was 42.8%.

ORFs within the N4 genome were located using Heuristic GeneMark (http://opal.biology.gatech.edu/GeneMark/heuristic_hmm2.cgi) and also with ORF Finder software from NCBI (<http://www.ncbi.nlm.nih.gov/gorf/gorf.html>) using the bacterial start and stop codon options. Forty-seven ORFs were identified, distributed throughout the phage genome and predicted as coding regions (table 1). A DNA region was considered if START and STOP codons were present, at least one of the base composition methods was positive and the length of the ORF was at least 100 bp. These 47 ORFs represented 90.7% of the total genome. The average ORF length was 743 bp and intergenic regions 109 bp. Genes varied in length from 126 bp (ORF7) to 3,384 bp (ORF39).

Vibriophage N4 contained 9.3% of the noncoding sequence; in many cases there was little or no space between genes. Five gene pairs overlapped with the effect of overlap between 3 and 16 nucleotides (ORF11-ORF12, ORF14-ORF15, ORF16-ORF17, ORF19-ORF20 and ORF22-ORF23). One overlap was found involving three genes (ORF35-ORF36-ORF37) which suggested that translation of some ORFs could be coupled or there is possibility for a programmed translational frameshift [11].

Comparison of Genome Sequences of Vibriophage N4 to Vibriophage VP4

When the genome sequence of N4 was compared with those of the other phages for which the data are available in the gene bank, it was found that N4 shares maximum homology with the vibriophage VP4, whose complete genome sequence has been submitted to the gene bank (Accession No. AAP33945.1). Seven regions were found which had 89–96% identity at the nucleotide level. Based on this, functions of the predicted ORFs of N4 phage were assigned. A comparison at the nucleotide level of the predicted ORFs of the two phages is presented in figure 3. It can be seen that phage VP4 contains 31 ORFs, whereas N4 contains 47 predicted ORFs. Twenty-seven ORFs were found in these two phages with 73–100% homology. Three ORFs with predicted functions (ORF16, ORF28 and ORF29; table 1) were found only in the vibriophage N4 while 2 ORFs (ORF11 and ORF19) of vibriophage VP4 were found to be absent in N4.

Close Relatedness between Vibriophage N4 and Enterobacteriophage T7

A search for homology in the database also revealed that besides VP4, the N4 phage genome is also closely related to *E. coli* phage T7. The predicted ORFs of vibriophage N4, whose functions have been assigned based on homology search, found significant similarity with the ORFs of enterobacteriophage T7 as well (fig. 3). From

Fig. 4. Genome homologies among different vibriophages by Southern hybridization. **a** Agarose gel electrophoresis of digested vibriophage DNAs. **b** Autoradiogram obtained after Southern transfer of DNAs from the agarose gel in **a**, followed by hybridization with labeled N4 DNA digested with *Hind*III and *Hae*III. Lanes 1 and 17: λ -*Hind*III-digested marker; lanes 2–4: digested vibriophage N4 DNA with *Eco*RV, *Hae*III and *Hind*III, respec-

tively; lanes 5–7: digested vibriophage S5 DNA with *Eco*RV, *Hae*III and *Hind*III, respectively; lanes 8–10: digested vibriophage D10 DNA with *Eco*RV, *Hae*III and *Hind*III, respectively; lanes 11–13: digested vibriophage MAD-5 DNA with *Eco*RV, *Hae*III and *Hind*III, respectively; lanes 14–16: digested vibriophage A1 DNA with *Eco*RV, *Hae*III and *Hind*III, respectively.

these results, it appears that N4 and VP4 could be related to the phage T7 and, hence, could be a member of the T7 viral supergroup.

Prediction of Gene Function

For the prediction of gene function homology, searches of 47 ORFs and their gene products were performed using blastp and psi-blast algorithms [12], with the non-redundant databases against protein sequences in the protein database at NCBI (<http://www.ncbi.nlm.nih.gov>). Association of the gene products with the protein families and other similarities were established using the secondary database search tools PrositeSearch [13], Hmmpfam [14], Goanna [15], and Fasta3 [16]. Descriptions of the predicted gene functions are summarized in table 1.

As can be seen from table 1, putative functions could be assigned to a number of ORFs. It was also seen that the vibriophage N4 genome could provisionally be grouped into four clusters consisting of genes with related functions. One cluster was related to DNA replication while the others were associated with structural, regulatory and lytic functions (table 1). A schematic diagram of vibrio-

phage N4 genome with predicted ORFs is given in figure 2. Out of 47 ORFs of the N4 genome functions, 30 could be predicted by homology. No significant homology in the database for the remaining 17 could be found. Based on this information, 7 ORFs could be assigned to replication, 6 to regulatory and 2 to the lytic function cluster. Thirteen ORFs belonged to the cluster associated with the synthesis of the structural components of N4.

Overview of the Putative Functional Gene Clusters Structural Gene Cluster

The cluster that could be defined best was the one containing the genes that encode the structural parts of the phage particle and the proteins needed for the assembly process. It embraced 13 predicted ORFs (genes 30–40, 43, 46). ORF30 (VN4_30) showed 99% identity to the head-to-tail joining protein of vibriophage VP4 and around 60% identity to the same protein from *Yersinia* phage phiYeO3–12, enterobacteria phage T7, *Yersinia* phage Berlin and enterobacteria phage T3. Homology searches on the remaining 12 ORFs revealed the highest match with the vibriophage VP4 (73–99%) and significant

Table 1. Predicted ORFs with their functional annotation of vibriophage N4

ORF No.	Gene name	Strand	Start	End	Nucleotide length	Amino acid length	Predicted function	Functional gene clusters
ORF1	VN4_01	+	509	940	432	143	putative S-adenosyl-L-methionine hydrolase	regulatory
ORF2	VN4_02	+	940	1,122	183	60	unknown function	unclassified
ORF3	VN4_03	+	1,169	3,820	2,652	883	RNA polymerase	regulatory
ORF4	VN4_04	+	4,003	5,013	1,011	336	DNA ligase	regulatory
ORF5	VN4_05	+	5,309	5,635	327	108	unknown function	unclassified
ORF6	VN4_06	+	5,647	6,330	684	227	unknown function	unclassified
ORF7	VN4_07	+	6,420	6,545	126	41	putative host RNA polymerase inhibitory protein	regulatory
ORF8	VN4_08	+	6,715	6,867	153	50	unknown function	unclassified
ORF9	VN4_09	+	6,906	7,187	282	93	unknown function	unclassified
ORF10	VN4_10	+	7,203	7,598	396	131	unknown function	unclassified
ORF11	VN4_11	+	7,703	7,900	198	65	unknown function	unclassified
ORF12	VN4_12	+	7,897	9,603	1,707	568	DNA primase/helicase	replication
ORF13	VN4_13	+	9,664	9,978	315	104	unknown function	unclassified
ORF14	VN4_14	+	10,064	10,528	465	154	unknown function	unclassified
ORF15	VN4_15	+	10,515	10,772	258	85	unknown function	unclassified
ORF16	VN4_16	+	10,779	11,090	312	103	putative gene 13.5 protein	replication
ORF17	VN4_17	+	11,074	13,110	2,037	678	DNA polymerase	replication
ORF18	VN4_18	+	13,121	13,285	165	54	unknown function	unclassified
ORF19	VN4_19	+	13,415	13,807	393	130	unknown function	unclassified
ORF20	VN4_20	+	13,794	13,985	192	63	unknown function	unclassified
ORF21	VN4_21	+	14,121	14,318	198	65	putative gene 5.7 homolog	replication
ORF22	VN4_22	+	14,318	15,181	864	287	exonuclease	replication
ORF23	VN4_23	+	15,169	15,420	252	83	unknown function	unclassified
ORF24	VN4_24	+	15,513	16,184	672	223	ssDNA binding protein	replication
ORF25	VN4_25	+	16,184	16,636	453	150	endonuclease	replication
ORF26	VN4_26	+	16,636	17,070	435	144	lysozyme	lysis
ORF27	VN4_27	+	17,504	17,755	252	83	unknown function	unclassified
ORF28	VN4_28	+	17,791	18,198	408	135	host specificity protein A	regulatory
ORF29	VN4_29	+	18,198	18,488	291	96	host specificity protein B	regulatory
ORF30	VN4_30	+	18,506	20,104	1,599	532	head to tail joining protein	structural
ORF31	VN4_31	+	20,184	21,062	879	292	capsid assembly protein	structural
ORF32	VN4_32	+	21,182	22,306	1,125	374	major capsid protein	structural
ORF33	VN4_33	+	22,372	22,965	594	197	tail tubular protein A	structural
ORF34	VN4_34	+	22,977	25,361	2,385	794	tail tubular protein B	structural
ORF35	VN4_35	+	25,449	26,081	633	210	putative deoxyribonucleoside kinase	structural
ORF36	VN4_36	+	26,074	26,592	519	172	internal virion protein A	structural
ORF37	VN4_37	+	26,579	27,169	591	196	internal virion protein B	structural
ORF38	VN4_38	+	27,201	29,414	2,214	737	internal virion protein C	structural
ORF39	VN4_39	+	29,427	32,810	3,384	1,127	internal virion protein D	structural
ORF40	VN4_40	+	32,890	34,674	1,785	594	tail fiber protein	structural
ORF41	VN4_41	+	34,689	34,883	195	64	putative lysis protein	lysis
ORF42	VN4_42	+	34,883	35,353	471	156	unknown function	unclassified
ORF43	VN4_43	+	35,365	35,598	234	77	DNA packaging/maturation protein A	structural
ORF44	VN4_44	+	35,617	35,760	144	47	unknown function	unclassified
ORF45	VN4_45	+	35,760	35,957	198	65	unknown function	unclassified
ORF46	VN4_46	+	36,066	37,784	1,719	572	DNA packaging/maturation protein B	structural
ORF47	VN4_47	+	38,031	38,162	132	43	unknown function	unclassified

matches with other bacteriophages like enterobacteria phage T7 and enterobacteria phage T3. Other predicted protein functions of this cluster were thought to be capsid assembly protein, major capsid protein, tail tubular protein A, tail tubular protein B, internal virion protein A, internal virion protein B, internal virion protein C, tail fiber protein, DNA packaging/maturation protein A and DNA packaging/maturation protein B.

Lytic Gene Cluster

In this study, ORF41 shared complete (100%) identity with the putative lysis protein of vibriophage VP4. It has 52% homology with the lysis protein of enterobacteria phage T7 and enterobacteria phage T3. A homology search revealed that the other ORF of this cluster, ORF26, was closely related to the lysozyme of vibriophage VP4 with 97% similarity, and suggested that this protein (144 aa) is most likely a general lysozyme involved in general cell lysis.

Regulatory Gene Cluster

Six ORFs, namely ORF1, ORF3, ORF4, ORF7, ORF28 and ORF29, were involved in this cluster and were assigned as putative S-adenosyl-L-methionine hydrolase, RNA polymerase, DNA ligase, putative host RNA polymerase inhibitory protein, host specificity protein A and host specificity protein B, respectively. The largest ORF of this cluster was ORF3 (883 aa), which shared 98% homology with RNA polymerase of vibriophage VP4 and 61% similarity with that of enterobacteria phage T7.

Replicatory Gene Cluster

The transition zone between the regulatory and the replicatory clusters was not sharply defined. A functional prediction of ORFs based on significant hits with database material revealed that the replicatory cluster contained at least 7 ORFs, namely ORF12, ORF16, ORF17, ORF21, ORF22, ORF24 and ORF25, which appeared to encode DNA primase/helicase, putative gene 13.5 protein, DNA polymerase, putative gene 5.7 homolog, exonuclease, ssDNA binding protein and endonuclease, respectively.

Homologies among the Different Vibriophages

In light of the homology of phage N4 with VP4, it was of interest to see if N4 phage is related to those other phages as well. In the absence of sequence information, the extent of homology of N4 with some of the other well-characterized vibriophages mentioned was determined by Southern hybridization. The hybridization data in figure 4 shows that the probes made of the whole-phage N4

genome hybridized with phages S5 and D10, but not with MAD5, indicating that at the genomic level, these vibriophages showed no homology with the O139-specific phage. To corroborate this finding, two genes, VN4_28 and VN4_29, which code for the putative host specificity genes A and B, were chosen and examined for their relatedness to the similar putative genes that could be present on the phages S5 and D10. The forward and reverse primer designated as HSPf and HSPr flanking these two genes were designed and then used for PCR amplification. PCR amplification of the putative genes coding for host specificity protein A and B was performed in S5 and D10. Amplified products of 802 bp obtained in both the phages were sequenced. Results revealed 100% similarity at the nucleotide level of these two genes with that of vibriophages S5 and D10.

Discussion

Vibriophage N4 is the first lytic typing vibriophage whose complete genomic sequence has been determined. The genomic map demonstrated that the coding regions are tightly packed, with very few intergenic spaces between them (fig. 2). The phage proteomes, in particular, are rich sources of untapped protein sequence diversity. For example, recent studies of *Staphylococcus aureus* and *Mycobacterium tuberculosis* bacteriophages indicate that these could be rich reservoirs of novel protein information, which have yet to be mined and analyzed [17, 18]. Our results show that a significant proportion of the predicted ORFs (17 genes; 36.17% of the proteome) which are unique to the vibriophage may also belong to this group.

A homology search using the publicly available entries in the genome database revealed that N4 had maximum similarity with vibriophage VP4 (GenBank Accession No. FJ409640). These two phages showed 89–96% sequence relatedness. Further, these two phage genomes shared a common set of essential genes that are distributed throughout the chromosome. Nonetheless, some of the VP4 sequences, such as nucleotide reductase subunit C, putative gene 6.5 homolog and probable N-terminus of major capsid protein genes have changed in the vibriophage N4. Except the shared ORFs of these two phages, some ORFs of unknown function are in sites in the N4 genome. These ORFs may have been transferred horizontally into the N4 genome. It has been reported previously that modular exchange of segments of the viral genome could generate enormous diversity within the context of a fairly standard T4-type genome [19]. Besides, VP4, as

the results reported here, showed that N4 is also closely related to enterobacteriophage T7. It has been previously reported that by definition the supergroup is not intended to supersede the T7-like genera, which remains confined to phages with a T7-like RNA polymerase. This supergroup is intended to be larger, and is apparently ancestral to the more strictly defined T7-like phages and exchanges modules with them [20].

These three vibriophages, N4, VP4 and T-7, belong to the family Podoviridae and genus T7-like viruses based on their having a hexagonal head, a short noncontractile tail, and a double-stranded DNA genome [3]. Based on this and in view of the results described in this paper, it appears that vibriophage N4 could belong to the T7 supergroup. Further analyses of both DNA and the described protein sequences support the conjecture that phage genomes are mosaic structures which make phylogenetic relations of phages more nonlinear compared to other organisms.

Hybridization experiments showed that O1 vibriophages are more closely related to one another than to vibriophages active on *V. cholerae* O139. Limited homology at the DNA level also indicated that the O1 and O139 vibriophages could represent two evolutionarily distinctive branches of their possible phylogenetic origin. In this context, it may be mentioned here that recent comparative studies have indicated that different phage families can have major differences in the characteristics of the genomic plasticity that they manifest [21–23]. Also the similarity at the nucleotide level of the predicted genes in the vibriophages infecting the same host might provide significant insight into the genetic background of typing vibriophages and their evolution of typing vibriophages.

Database Submission

The nucleotide sequence of the whole genome of vibriophage N4 was submitted to GenBank under Accession No. FJ409640.

References

- Kaper JB, Morris JG Jr, Levine MM: Cholera. Clin Microbiol Rev 1995;8:48–86.
- Basu S, Mukerjee S: Bacteriophage typing of *Vibrio* eltor. Experientia 1968;24:299–300.
- Chattopadhyay DJ, Sarkar BL, Ansari MQ, Chakrabarty BK, Roy MK, Ghosh AN, Pal SC: New phage typing scheme for *Vibrio cholerae* O1 biotype El Tor strains. J Clin Microbiol 1993;31:1579–1585.
- Sarkar BL, De SP, Saha MR, Niyogi SK, Roy MK: Validity of new phage typing scheme against *Vibrio cholerae* O1 biotype El Tor strains. Indian J Med Res 1994;99:159–161.
- Chakraborty AK, Ghosh AN, Nair GB, Niyogi SK, Bhattacharya SK, Sarkar BL: Development and evaluation of phage typing scheme for *Vibrio cholerae* O139. J Clin Microbiol 2000;38:44–49.
- World Health Organisation: Guidelines for cholera control. Geneva, World Health Organisation, 1993.
- Yamamoto KR, Alberts BM, Benzinger R, Lawhorne L, Treiber G: Rapid bacteriophage sedimentation in the presence of polyethylene glycol and its application to large scale virus purification. Virology 1970;40:734.
- Kurazono H, Sasakawa C, Yoshikawa M, Takeda Y: Cloning of a Vero toxin (VT1, Shiga-like toxin I) gene from a VT1-converting phage isolated from *Escherichia coli* O157:H7. FEMS Microbiol Lett 1987;44:23–26.
- Werquin M, Ackermann HW, Levesque RC: A study of 33 bacteriophages of *Rhizobium meliloti*. Appl Environ Microbiol 1988;54:188–196.
- Southern EM: Detection of specific sequences among DNA fragments separated by gel electrophoresis. J Mol Biol 1975;98:503–517.
- Miller ES, Heidelberg JF, Eisen JA, Nelson WC, Durkin AS, Ciecko A, Feldblyum TV, White O, Paulsen IT, Nierman WC, Lee J, Szczypinski B, Fraser CM: Complete genome sequencing of the broad host range vibriophage KVP40: comparative genomics of a T4-related bacteriophage. J Bacteriol 2003;185:5220–5233.
- Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ: Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 1997;25:3389–3402.
- Rice P, Longden I, Bleasby A: EMBOSS: the European Molecular Biology Open Software Suite. Trends Genet 2000;16:276–277.
- Durbin R, Dear S: Base qualities help sequencing software. Genome Res 1998;8:161–162.
- Altschul SF, Wootton JC, Gertz EM, Agarwala R, Morgulis A, Schäffer AA, Yu YK: Protein database searches using compositionally adjusted substitution matrices. FEBS J. 2005;272:5101–5109.
- Pearson WR, Lipman DJ: Improved tools for biological sequence comparison. Proc Natl Acad Sci USA 1988;85:2444–2448.
- Kwan T, Liu J, DuBow M, Gros P, Pelletier J: The complete genomes and proteomes of 27 *Staphylococcus aureus* bacteriophages. Proc Natl Acad Sci USA 2005;102:5174–5179.
- Pedulla ML, Ford ME, Houtz JM, Karthikeyan T, Wadsworth C, Lewis JA, Jacobs-Sera D, Falbo J, Gross J, Pannunzio NR, Brucker W, Kumar V, Kandasamy J, Keenan L, Bardarov S, Kriakov J, Lawrence JG, Jacobs WR Jr, Hendrix RW, Hatfull GF: Origins of highly mosaic mycobacteriophage genomes. Cell 2003;113:171–182.
- Desplats C, Dez C, Tétart F, Eleaume H, Krisch HM: Snapshot of the genome of the pseudo-T-even bacteriophage RB49. J Bacteriol 2002;184:2789–2804.
- Hardies SC, Comeau AM, Serwer P, Suttle CA: 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 2003;310:359–371.
- Brüssow H, Desiere F: Comparative phage genomics and the evolution of Siphoviridae: insights from dairy phages. Mol Microbiol 2001;39:213–222.
- Hendrix RW, Smith MC, Burns RN, Ford ME, Hatfull GF: Evolutionary relationships among diverse bacteriophages and prophages: all the world's a phage. Proc Natl Acad Sci USA 1999;96:2192–2197.
- Juhala RJ, Ford ME, Duda RL, Youton A, Hatfull GF, Hendrix RW: Genomic sequences of bacteriophages HK97 and HK022: pervasive genetic mosaicism in the lambdoid bacteriophages. J Mol Biol 2000;299:27–51.