ISSN 1022-7954, Russian Journal of Genetics, 2016, Vol. 52, No. 9, pp. 914–925. © Pleiades Publishing, Inc., 2016.
Original Russian Text © N.I. Smirnova, T.A. Kul'shan', E.Yu. Baranikhina, Ya.M. Krasnov, D.A. Agafonov, V.V. Kutyrev, 2016, published in Genetika, 2016, Vol. 52, No. 9, pp. 1029–1041.

GENETICS
OF MICROORGANISMS

Genome Structure and Origin of Nontoxigenic Strains of *Vibrio cholerae* of El Tor Biovar with Different Epidemiological Significance

N. I. Smirnova, T. A. Kul'shan', E. Yu. Baranikhina, Ya. M. Krasnov, D. A. Agafonov, and V. V. Kutyrev

Russian Research Anti-Plague Institute Microbe, Saratov, 410005 Russia
e-mail: rusrapi@microbe.ru

Received November 12, 2015; in final form, November 27, 2015

Abstract—Intraspecies genetic differentiation of nontoxigenic strains of *Vibrio cholerae* of El Tor biovar containing one of the key pathogenicity genes, *tcpA*, is studied along with the phylogenetic relationships between these strains and toxigenic isolates. Comparative analysis of the whole genome nucleotide sequences demonstrates for the first time that *ctxA*$^-$*tcpA*$^+$ strains vary considerably and can be clustered into two separate groups, the CTXφ$^-$RS1φ$^+$VPI$^+$VSP$^+$/CTXφ$^-$RS1φ$^-$VPI$^+$VSP$^+$ isolates and the CTXφ$^-$RS1φ$^-$VPI$^+$VSP$^-$ isolates, differing in their epidemiological significance. In the course of model experiments, it is established that nontoxigenic potentially epidemic CTXφ$^-$RS1φ$^+$VPI$^+$VSP$^+$/CTXφ$^-$RS1φ$^-$VPI$^+$VSP$^+$ isolates are derivatives of toxigenic strains. The results of whole genome SNP analysis of 35 *Vibrio cholerae* strains confirm these data and indicate genetic remoteness of nontoxigenic CTXφ$^-$RS1φ$^-$VPI$^+$VSP$^-$ strains both from the potentially epidemic strains and from the toxigenic isolates. It is found that the genomes of the CTXφ$^-$RS1φ$^-$VPI$^+$VSP$^-$ strains contain unique SNPs which are characteristic of them alone. The new data on the structure of the genome of nontoxigenic strains with different epidemiological significance may be further used for their genetic differentiation.

Keywords: Vibrio cholerae, PCR, whole genome sequencing, SNP analysis
DOI: 10.1134/S1022795416060120

INTRODUCTION

The seventh cholera pandemic which has already continued for more than 50 years has spread to many regions of the Southeast Asia, Africa, and Latin America [1]. The territory of Russian Federation is not endemic for cholera and the emerging epidemic complications of this highly infectious disease are associated with its importation from the countries of the former Soviet Union and from the far abroad helped by increased international ties. For instance, during the period from the 2001 to the 2014, nine cases of cholera importation were registered in Russia with two of them leading to cholera outbreaks (Kazan in 2001 and Rostov oblast in 2005) [2, 3]. A person with cholera or a vibriocarrier is a common source of infection, as well as water and food contaminated by them. The causative agent of the current cholera pandemic is *Vibrio cholera* O serogroup 1 of El Tor biovar, whose genome is made up of two circular chromosomes [4]. The genome, in addition to the core genes, contains various mobile genetic elements, encoding the traits associated with virulence, adaptation, and persistence in different ecological niches, as well as with antibiotic resistance. Among them, there are virulence prophages (CTXφ, abbreviated from cholera toxin, and RS1φ, abbreviated from repeat sequence), pathogenicity islands (VPI-1 and VPI-2, from *Vibrio* pathogenicity island), pandemicity islands (VSP-I and VSP-II, from *Vibrio* seventh pandemic island), and the environmental persistence island (EPI) [5–8], which determine basic pathogenic and adaptive properties of the strains. The cholera causative agent is able to survive outside the organism of its biological host (human) and naturally occurs in aquatic ecosystems [9–11]. Once in the digestive tract, *Vibrio cholerae* bacteria colonize the mucous membrane of the small intestine using the toxin-coregulated pili (TCP) which are present on the bacterial cell surface. The second key step in the course of the cholera infectious process is the production of cholera toxin (CT) by *Vibrio cholerae*, a complex compound consisting of two subunits, A and B, which causes profuse diarrhea, the typical cholerogenic reaction [9]. Production of TCP and CT is controlled by the *tcpA-F* gene cluster located on VPI-1 (41.3 kb) and the *ctxAB* operon contained in the CTXφ prophage (6.9 kb), respectively. By their functional significance, the *tcpA-F* and *ctxAB* genes are assigned to the key pathogenicity genes, which are

absolutely indispensable for development of the cholera infectious process [5, 11]. This gives grounds to assume that toxigenic *V. cholerae* strains belonging to the El Tor biovar, which contain VPI-1 and the CTXφ prophage in their genome, are epidemic strains [12]. Such strains are usually isolated from the patients with cholera. Moreover, a relatively short time ago, the genomes of the indicated strains revealed the presence of the VSP-I and VSP-II pandemicity islands (16.0 kb and 29.6 kb, respectively), which are considered as accessory genetic markers for toxigenic epidemic strains [6].

Currently, toxigenic strains of *V. cholerae* of El Tor biovar encompass not only the typical strains which have caused the current cholera pandemic but also the genetic variants which emerged about 25 years ago. The principal difference between these two groups of strains consists in their CTXφ prophages containing different alleles for the *ctxB* gene encoding the B subunit of cholera toxin. The typical strains bear the *ctxB3* allele, while the genetic variants bear the *ctxB1* allele. The occurrence of the *ctxB1* gene allele characteristic of the cholera causative agent belonging to the other biovar, the classical one, in the genome of *V. cholerae* of El Tor biovar is the result of horizontal gene transfer in the natural populations of *V. cholerae*.

At the same time, nontoxigenic strains lacking both the CTXφ prophage and pathogenicity and pandemicity islands are quite often isolated from the environment. Since such strains are not able to cause epidemics, they are considered to be nonepidemic strains [12, 17]. At the present time, to differentiate between the strains with different epidemiological significance, the *ctxA* and *tcpA* genes which encode the A subunit of CT and the main subunit of TCP (TcpA) and are a part of the CTXφ prophage and VPI-1, respectively, are commonly used. Toxigenic strains possess the *ctxA* and *tcpA* genes, while nontoxigenic strains lack them [18, 19].

Apart from the two groups of *V. cholerae* discussed above, the strains of *V. cholerae* of El Tor biovar whose genomes contain an incomplete set of mobile elements bearing the key pathogenicity genes have been detected in waters of surface reservoirs [19, 20]. Among them, especially noteworthy are the nontoxigenic strains bearing the genes for toxin-coregulated pili ($ctxA^-tcpA^+$). It became known that $ctxA^-tcpA^+$ strains exist in nature in 1998 after they were discovered by Faruque et al. [21], and until recently they were considered to be potentially epidemic. The reason for accepting this level of epidemiological significance for the $ctxA^-tcpA^+$ strains was the fact that TCPs not only are necessary for colonization of the small intestine but also serve as a receptor for the CTXφ prophage [22]. This means that nontoxigenic $ctxA^-tcpA^+$ strains may acquire the genes for cholera toxin encoded by the prophage through the process of phage conversion, the evident consequence of this event being the emergence of toxigenic epidemic strains.

Isolation of $ctxA^-tcpA^+$ strains from water is not uncommon not only on the territories endemic for cholera but also in the Russian Federation [3, 23]. However, the structure of the genome of these strains is poorly studied and the origin of these strains still remains an enigma. Another issue also remains unsolved whether all $ctxA^-tcpA^+$ strains have retained in their genomes the remaining key pathogenicity and epidemicity (pandemicity) genes characteristic of toxigenic isolates. At the same time, the development of modern techniques of molecular biology makes it possible to acquire new knowledge on the genome structure of different $ctxA^-tcpA^+$ strains, which may be further used to reevaluate the epidemiological significance of these strains and will shed light on their origin. It thus appears to be very important to study in detail the structure of the genome of nontoxigenic strains bearing the *tcpA* gene and to understand the genetic relationships between these strains and the toxigenic ones.

The aim of the current work is to perform a molecular genetic analysis of a number of nontoxigenic strains of *V. cholerae* O subgroup 1 of El Tor biovar bearing the *tcpA* gene and to study their origin on the basis of the comparison of full genome nucleotide sequences.

MATERIALS AND METHODS

Bacterial strains, culturing conditions. Seventeen *V. cholerae* El Tor biovar strains isolated from people and water were used in the work (Table 1). The strains were obtained from the State Collection of Pathogenic Bacteria of the Russian Research Anti-Plague Institute Microbe, where they were stored lyophilized. For culturing bacteria we used Luria-Bertani (LB) broth and agar. All procedures were carried out in accordance with the currently accepted health and hygiene standards [24].

Generation of nontoxigenic strains. Cells of the strains grown for 18 h at 37°C on LB agar were used to prepare a suspension with a cell concentration of 10^9 cfu/mL; 0.5 ml of the suspension was introduced into flasks containing 50 mL of sterilized river water, which were incubated at 8°C. Plating of 0.1 mL of contaminated river water on LB agar and incubation until isolated colonies appeared was repeated at intervals of 4–6 days during the whole time of the study (45 days). Then ten randomly selected isolated colonies for each strain were plated onto LB agar to accumulate an amount of bacterial cells sufficient for further isolation of DNA and PCR-assisted analysis.

Polymerase chain reaction was performed as previously described [13]. Briefly, the reaction mixture contained 1× PCR buffer (2.5 μL of 10× PCR buffer), 2.0 μL of $MgCl_2$, 2.5 μL of a mixture of dNTPs,

Table 1. Strains of *Vibrio cholerae* O serogroup 1 of El Tor biovar used in the study

Strain	Place and year of isolation	Source	Presence of *ctxA* and *tcpA* genes according to the PCR analysis data	
			ctxA	*tcpA*
Natural strains				
M299	Turkmenistan, 1965	Human	–	+
M139	Turkmenistan, 1965	Water	–	+
M888	Astrakhan, 1970	Human	+	+
M887	″	″	+	+
M893	″	″	+	+
M943	″	″	+	+
M890	″	″	+	+
M1056	″	″	+	+
M905	″	″	+	+
M996	″	″	+	+
M641	Astrakhan, 1972	″	+	+
M1395	Astrakhan, 1981	Water	–	+
M1399	Astrakhan, 1982	″	–	+
P18899	Murmansk, 2006	Human	+	+
M1501	Kalmyk Republic, 2011	″	–	+
M1518	Kalmyk Republic, 2012	Water	–	+
M1524	Kalmyk Republic, 2013	″	–	+
Experimentally obtained strains				
M888D, derivative of M888			–	+
P18899D, derivative of P18899			–	+

0.05 μL of each of the oligonucleotide primers, and 0.1 U of Taq DNA polymerase. DNA amplification was carried out in a BIS-N thermal cycler with heated lid (Russia). Amplification products were analyzed by electophoretic separation in 1.5–2% agarose gel made using 1×TBE buffer with the addition of 0.5 μg/mL ethidium bromide. Commercial DNA size marker GenRuler TM 100 bp DNA Ladder (MBI Fermentas, Lithuania) was used as a molecular weight control. The results of electrophoretic separation were documented with the aid of the VersaDoc gel documentation system (Bio-Rad, United States) and Quantity One v 4.6.9 software (Bio-Rad, United States). The oligonucleotide primers [25–29] used in the current work are listed in Table 2.

Southern blot hybridization with digoxigenine-labeled probe was performed as described in the manual for molecular cloning [30]. The *ctxA* gene amplicon with the size of 564 bp obtained by PCR was used as a probe designated CT. Chromosome DNA of the strains under study was fragmented with the aid of *Pst*I restriction endonuclease.

Genomic DNA isolation and purification from the bacterial suspension was performed using the commercially available Axy Prep Bacterial Genomic DNA Miniprep Kit in accordance with the manufacturer's instructions. Cells were preliminarily treated with sodium merthiolate to the final dilution of 1 : 10000 (0.01%) and preheated at 56°C for 30 min.

Whole genome sequencing and SNP typing. DNA sequencing libraries were prepared from 0.5–1 μg of genomic DNA in accordance with the manufacturer's recommendations and submitted to next generation sequencing on an Ion PGM (Ion Torrent) genetic analyzer. The obtained single reads were corrected with the aid of Ion Torrent Suite Software v 4.4 and de novo assembled using Newbler GS Assembler v. 2.6 software. With the aid of the DNAStar Lasergene v. 11.2 software package (DNASTAR Inc., United States) and Mega 6 software (http://www.megasoftware.net/), the whole genome nucleotide sequences for 17 *V. cholerae* strains de novo assembled in contigs with the lengths of up to 510 kb were mapped successively onto the reference genome nucleotide sequences of the N16961 strain, against which the automated

Table 2. Nucleotide sequences for the oligonucleotide primers used in the study

Mobile elements tested	Primer name	Nucleotide sequences (5'–3') for forward (F) and reverse (R) primers	Amplified fragment length, bp	Reference
CTXφ	*ctxA*-1 *ctxA*-2	F:CGGGCAGATTCTAGACCTCCTG R:CGATGATCTTGGAGCATTCCCAC	564	[25]
	ctxB-1 *ctxB*-2	F:ATGATTAAATTAAAATTTGG R:TTAATTTGCCATACTAATTG	354	[25]
	zot-1 *zot*-2	F:TCGCTTAACGATGGCGCGTTTT R:AACCCCGTTTCACTTCTACCCA	947	[25]
	ace-1 *ace*-2	F:TAAGGATGTGCTTATGATGGACACC R:CCGTGATGAATAAAGATACTCATAGG	289	[25]
	orfU-1 *orfU*-2	F:CAAAATGAGCATGGCGGC R:CCCATTGTGCAATCGGTGT	721	Calculated by the authors
	cep-1 *cep*-2	F:CAGAACAATTGCCCCCACCAC R:AAGCACGCTTTCACTCGGGG	162	The same
	rstA-2 *rstA*-1	F:ATCGTCGTGAATTTCCTTAG R:GCGATTAGTCTCTGAGCC	400	»
	rstB-1 *rstB*-2	F:CGTGATGGGTCTTCTGGGTC R:TGGTGCCTCTCATTCTGAAG	240	»
	rstR-1 *rstR*-2	F:AAGACTCAGGCACAAATG R:CTCTTCTTCATCTAACTGTTG	216	»
RS1φ	*rstC-1* *rstC-2*	F:AACAGCTACGGGCTTATTC R:TGAGTTGCGGATTTAGGC	238	[26]
VPI-1	*tcpAEl-1* *tcpAEl-2*	F:GAAGAAGTTTGTAAAAGAAGAACAC R:GAAAGGACCTTCTTTCACGTTG	471	[27]
	toxT-1 *toxT-2*	F:ATAACTTTACGTGGATGGC R:AAAATCAGTGATACAATCG	933	[25]
	aldA-1 *aldA-2*	F:TTTTCTTGATTGTTAGGATGC R:ATTCTTCTGAGGATTGCTGAT	885	[28]
	mop-1 *mop-2*	F:GAAACATCCCATGAGTTGCG R:CGGACATATAGGGCCAGTCA	778	[29]
VPI-2	*nanH-1* *nanH-2*	F:GACAGTCCAGCCAAACAG R:CGTTAGCGTTGTTAGCCTC	1900	[29]
VSP-I	*VC0175-1* *VC0175-2*	F:GGAACGTTCAGCAATGCGT R:CTGGCAGATGCTCGAATGAT	586	[29]
	VC0185-1 *VC0185-2*	F:TGCCCCATCCATCCCATA R:CAAGAAGCGGAGAGATCGC	781	[29]
VSP-II	*VC0497-F* *VC0497-R*	F:TGGAAAGAAGAGCGTTACTGC R:CCCTGTTGATGATGTGATTTG	320	Calculated by the authors
	VC0502-F *VC0502-R*	F:CTGTGATTCGGGCTTTATCGG R:GCGTAAACTGAGCCAATAAGC	761	The same
	VC0514-F *VC0514-R*	F:CTTGATGGAGCGGAGAAAAC R:CGATGAATAGCCTGTTGAAC	604	»

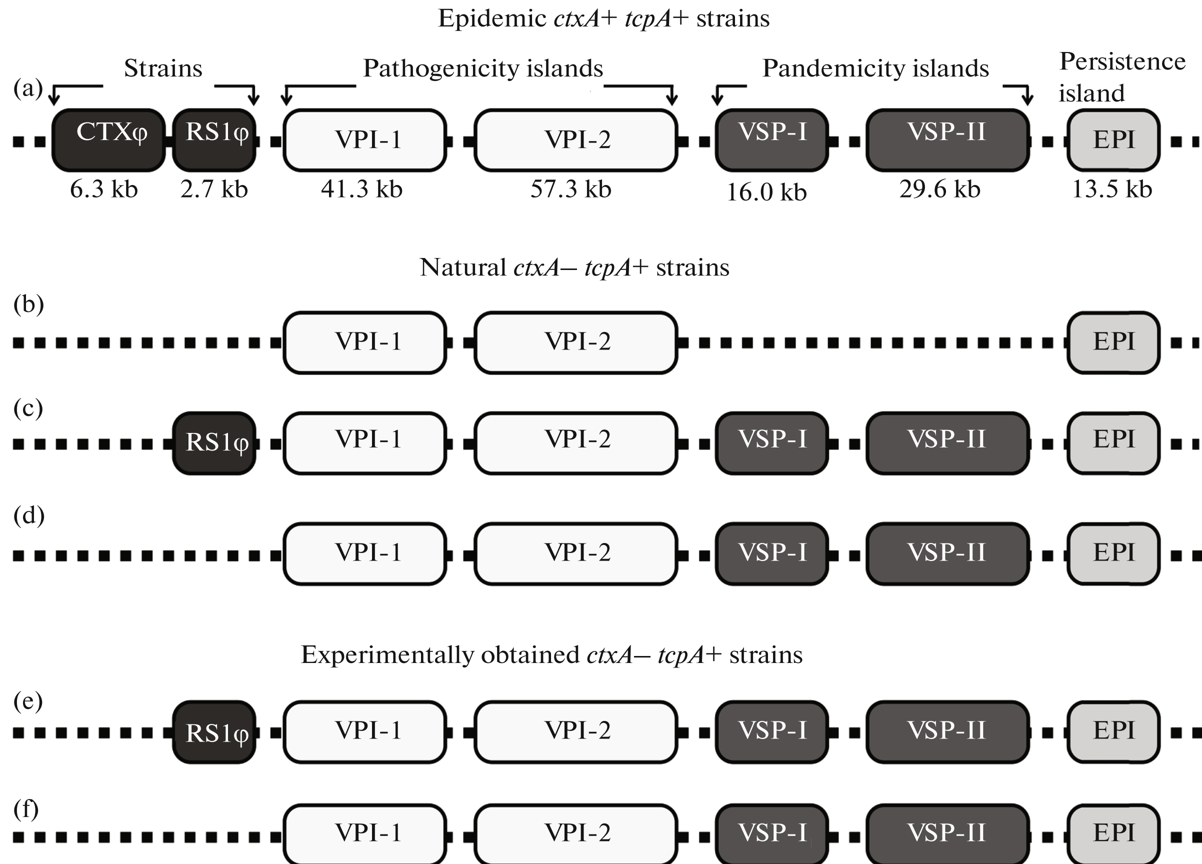

Fig. 1. Schematic representation of the structure of the genome of toxigenic ($ctxA^+tcpA^+$) and nontoxigenic ($ctxA^-tcpA^+$) strains of *V. cholerae* El Tor biovar with different epidemiological significance (a–f, see text).

search for and identification of SNPs was performed. Core SNPs in the genomes of the group of strains under study were identified using Wombac 2.0 software (http://www.bioinformatics.net.au/software.wombac.shtml). The results were deposited in the BioNumerics 7.1 database (Applied Maths, Belguim). For the phylogenetic tree construction, we used “Maximum parsimony tree” method with categorical coefficient.

RESULTS

Comparative Analysis of the Genomes of V. cholerae $ctxA^-tcpA^+$ Strains

As the first step of our work, we determined the nucleotide sequences of whole genomes for nine *V. cholerae* El Tor biovar strains, two of which (M888 and M18899) according to the results of PCR analysis appeared to be toxigenic epidemic strains, while seven other strains (M299, M139, M1395, M1399, M1501, M1518, and M1524) were nontoxigenic potentially epidemic (Table 1). The comparison of the nucleotide sequences revealed that, unlike the toxigenic strains, all the examined nontoxigenic strains indeed lacked the CTXφ prophage containing the genes for cholera toxin in their genomes, but retained VPI-1 containing the genes encoding TCP along with a number of other mobile elements associated with pathogenicity and persistence (VPI-2 and EPI). At the same time, it emerged that the nontoxigenic strains under study also lacked the VSP-I and VSP-II pandemicity islands and the RS1φ prophage containing genes involved in the control of CT biosynthesis and located in the close proximity to CTXφ (Table 3, Figs. 1a and 1b). It is worth noting that all these strains were isolated at different times on the territories of Russia and Turkmenistan not endemic for cholera (Table 1).

To summarize, the results of sequencing showed that the $ctxA^-tcpA^+$ strains under study possess the same set of mobile elements, $CTXφ^-RS1φ^-VPI\text{-}1^+VPI\text{-}2^+VSP\text{-}I^-VSP\text{-}II^-$. For the sake of convenience, the genotype of these strains will be further referred to as $CTXφ^-RS1φ^-VPI^+VSP^-$. It is clear that genomic differences between theses strains and the toxigenic ones consist not only in the absence in the former of the CTXφ prophage. The differences proved to be more pro-

Table 3. Molecular genetic characteristics of toxigenic and nontoxigenic *V. cholerae* El Tor biovar strains

Strain	Place and year of isolation	Mobile elements tested					
		CTXφ	RS1φ	VPI-1	VPI-2	VSP-I	VSP-II
Toxigenic *ctxA*$^+$*tcpA*$^+$ strains							
M888	Astrakhan, 1970	+	+	+	+	+	+
M818	Saratov, 1970	+	+	+	+	+	+
N16961*	Bangladesh, 1971	+	+	+	+	+	+
M1030	Turkmenistan, 1972	+	+	+	+	+	+
M1275	Dagestan, 1993	+	+	+	+	+	+
M1293	Dagestan, 1994	+	+	+	+	+	+
P17644	Achinsk, 1997	+	+	+	+	+	+
M1344	Kazan, 2001	+	+	+	+	+	+
RND18826*	Tver, 2005	+	+	+	+	+	+
P18899	Murmansk, 2006	+	+	+	+	+	+
2010EL*	Haiti, 2010	+	+	+	+	+	+
RND6878	Moscow, 2012	+	+	+	+	+	+
L3265	Moscow, 2010	+	+	+	+	+	+
RND19187*	Moscow, 2010	+	+	+	+	+	+
RND81*	Rostov-on-Don, 2014	+	+	+	+	+	+
Nontoxigenic *ctxA*$^-$*tcpA*$^+$ strains							
GP60*	India, 1973	–	–	+	+	+	+
GP106*	Germany, 1975	–	–	+	+	+	+
GP143*	Bahrain, 1978	–	+	+	+	+	+
V5*	India, 1989	–	–	+	+	+	+
NHCM017*	Bangladesh, 2001	–	+	+	+	+	+
4661*	Bangladesh, 2001	–	+	+	+	+	+
MBRN14*	India, 2004	–	–	+	+	+	+
4122*	Vietnam, 2007	–	–	+	+	+	+
M299	Turkmenistan, 1965	–	–	+	+	–	–
M139	Turkmenistan, 1965	–	–	+	+	–	–
M1395	Astrakhan, 1981	–	–	+	+	–	–
M1399	Astrakhan, 1982	–	–	+	+	–	–
A213*	Georgia, 1984	–	–	+	+	–	–
P18785*	Rostov, 2005	–	–	+	+	–	–
M1501	Kalmyk Republic, 2011	–	–	+	+	–	–
I-1471*	Altai, 2011	–	–	+	+	–	–
M1518	Kalmyk Republic, 2012	–	–	+	+	–	–
M1524	Kalmyk Republic, 2013	–	–	+	+	–	–
M888D**	derivative of M888	–	+	+	+	+	+
P18899D**	derivative of P18899	–	–	+	+	+	+

* Indicated are natural strains whose whole genome nucleotide sequences were obtained from GenBank.
** Experimentally obtained nontoxigenic *ctxA*$^-$*tcpA*$^+$ strains.

found, as the strains under study lacked at least three other DNA regions containing the RS1φ prophage and VSP-I and VSP-II (Figs. 1a and 1b).

In order to study the prevalence of nontoxigenic CTXφ$^-$ RS1φ$^-$ VPI$^+$ VSP$^-$ isolates in other regions of Russia and throughout the world, we analyzed the generally available full genome sequences for 11 *ctxA*$^-$ *tcpA*$^+$ strains of *V. cholerae* of El Tor biovar from the NCBI GenBank database. It appeared that among these strains nine nontoxigenic isolates obtained in the regions endemic for cholera contained not only VPI-1 and VPI-2 pathogenicity islands but also VSP-I and VSP-II, three of them bearing the RS1φ prophage as well (Table 3, Figs. 1c and 1d). And only two strains isolated in Georgia and Russia (Altai) lacked the extended DNA regions encompassing CTXφ and RS1φ prophages and VSP-I and VSP-II pandemicity islands in their genomes.

In summary, the results of the analysis of the nucleotide sequences of the full genomes for 18 natural nontoxigenic *ctxA*$^-$*tcpA*$^+$ strains previously considered potentially epidemic made it possible to observe for the first time their genetic heterogeneity and to cluster them into two groups. The isolates included with the first group are distinguished from the toxigenic strains by the lack of just one or two DNA regions with the size of 6.9 bp (CTXφ prophage) or 9.6 bp (CTXφ and RS1φ prophages) (Figs. 1a, 1c, and 1d). At the same time for the isolates in the second group, this difference is much more significant, these strains lacking at least four DNA regions (the two prophages, CTXφ and RS1φ, as well as both VSP-I and VSP-II pandemicity islands) with the total size of 54.2 kb (Fig. 1d). These findings lead to an important conclusion that the epidemiological significance may not be identical for the strains in the two discussed groups, as was thought previously. Apparently, only those nontoxigenic *ctxA*$^-$*tcpA*$^+$ strains which lack only one or two prophages associated with CT production in the genome may be potentially epidemic. These strains are likely to serve as a natural reservoir out of which a toxigenic population may emerge as a result of CTXφ prophage acquisition by phage conversion in the regions endemic for cholera. On the other hand, those nontoxigenic *ctxA*$^-$*tcpA*$^+$ strains whose genome lacks both prophages and pandemicity islands are of epidemiological significance, but in another sense. They are nonepidemic since the acquisition of extended DNA regions to establish a toxigenic population is a highly improbable event under natural conditions.

It is important to note that almost all CTXφ$^-$VPI$^+$VSP$^+$ strains which have lost or retained the second virulence prophage RS1φ, namely, seven out of the eight studied, were detected only in the regions endemic for cholera, where there are stable infection reservoirs in surface water basins. In Russia, such strains have not been detected. The presented data give grounds to speculate that nontoxigenic CTXφ$^-$VPI$^+$VSP$^+$ strains may possibly derive from toxigenic ones through adaptive changes in the genomes of the latter under aquatic environmental conditions. These changes may consist in the loss of a number mobile elements dispensable for *V. cholera* survival under new environmental conditions, one of them possibly being the CTXφ prophage with CT genes [16].

Assessment of the Possibility for Nontoxigenic CTXφ$^-$ RS1φ$^+$VPI$^+$VSP$^+$ and CTXφ$^-$RS1φ$^-$VPI$^+$VSP$^+$ Strains to Originate from Toxigenic Strains

To confirm our assumption that nontoxigenic CTXφ$^-$RS1φ$^-$VPI$^+$VSP$^+$ and CTXφ$^-$RS1φ$^-$VPI$^+$VSP$^+$ isolates originated as a result of genome reorganization in toxigenic strains in an aqueous environment, we performed model experiments using nine clinical *V. cholerae ctxA*$^+$*tcpA*$^+$ isolates (M888, M887, M943, M996, M893, M890, M1056, M905, M641, and P18899). Using Southern blot hybridization, we preliminarily established the presence of a single copy of CTXφ prophage in the genome of each isolate. In addition, it was known from the data of PCR analysis that the chromosomes of these strains contain the RS1φ prophage, VPI-1 and VPI-2, VSP-I and VSP-II pandemicity islands, and EPI environmental persistence island. We used river water as the surrounding medium since rivers are one of the locations where *V. cholerae* can be found in nature. We assumed that one of the stressing factors inducing genome reorganization could be deficiency of nutrition components in water. Indeed, PCR analysis of 150–170 isolated colonies of each strain grown on nutrient agar revealed that, at certain time intervals in the populations of six strains (M888, M887, M996, M890, M641, and P18899), there appeared clones lacking the *ctxA*$^-$ gene. The time periods over which *ctxA*$^-$ clones appeared in the populations under study varied for different strains and ranged from 4 to 45 days. The proportion of *ctxA*$^-$ clones varied from 4.3% to 85.0% of all the clones examined depending on the strain. Subsequent verification of the identified *ctxA*$^-$ clones by PCR showed that all of them had lost the CTXφ prophage. Their genomes lacked not only the *ctxA* gene but also other genes encoded by the prophage, namely, *ctxB, zot, ace, orfU, cep, rstA, rstB*, and *rstR*, which made it possible to come to such a conclusion. At the same time, we did not detect the loss by these clones of the genes encoded by the VPI-1 (*tcpA, toxT, aldA*, and *mop*) and VPI-2 (*nanH*) pathogenicity islands, RS1φ prophage (*rstC*), and VSP-I (vc0175 and vc0185) and VSP-II (vc0497, vc0502, and vc0514) pandemicity islands. The only exception was the P18899 strain, which lost apart from the CTXφ prophage also the second prophage, RS1φ.

To confirm the data obtained, we determined the full genome nucleotide sequence for two nontoxigenic

strains designated as M888D and P18899D (GenBank accession nos. LPXQ00000000 and LAKN00000000, respectively) derived from the toxigenic M888 and P18899 strains, respectively (GenBank accession nos. LKBH00000000 and LAKM00000000, respectively). The comparison of the genomes of the nontoxigenic strains with the genomes of the original ones made it possible to conclude that the differences between the isogenic strains are caused directly by the loss by the toxigenic strains of either the CTXφ prophage (in the case of M888D) or both the CTXφ and RS1φ prophages (in the case of P18899D) (Figs. 1e and 1f).

In summary, we have for the first time experimentally proved the origination of nontoxigenic $CTX\varphi^-RS1\varphi^+VPI^+VSP^+$ and $CTX\varphi^-RS1\varphi^-VPI^+VSP^+$ strains from toxigenic ones when exposed to an aqueous environment. It was demonstrated that emergence of nontoxigenic isolates is a result of the loss of the CTXφ prophage encompassing CT genes by toxigenic strains. In some cases, such reorganization of the genome is accompanied by the loss of the RS1φ prophage as well.

Phylogenetic Relationships between the $ctxA^-tcpA^+$ Strains with Different Epidemiological Significance

The phylogeny of the genetically different strains was studied on the basis of SNP analysis. The analysis involved 20 nontoxigenic *V. cholerae* El Tor biovar strains retaining pandemicity islands ($CTX\varphi^-RS1\varphi^+VPI^+VSP^+$ and $CTX\varphi^-RS1\varphi^-VPI^+VSP^+$) and lacking them ($CTX\varphi^-RS1\varphi^-VPI^+VSP^-$) along with 15 toxigenic isolates (Table 3). The comparison of the full genome nucleotide sequences for the 35 strains indicated above with the reference nucleotide sequence for the *V. cholerae* El Tor biovar strain N16961 revealed 12 901 single nucleotide substitutions (SNPs) in 1575 core genes and 772 intergenic regions on both chromosomes constituting the genome. On the basis of the analysis of single polymorphic nucleotides in the genomes of these strains, we reconstructed a dendrogram representing the phylogenetic relationships between them (Fig. 2). On the phylogenetic tree, one can clearly see two distinct phylogenetic groups, which differ from each other in 4367 SNPs on average. The first group includes 15 toxigenic strains and all nontoxigenic strains (10 isolates in total), the only difference of which from the former consists in their lacking one (CTXφ) or two (CTXφ and RS1φ) prophages. Of particular importance is the presence in this group of the nontoxigenic M888D and P18899D strains whose origination from toxigenic strains was experimentally proved. This indicates the close phylogenetic relationship between the nontoxigenic $CTX\varphi^-RS1\varphi^+VPI^+VSP^+$ and $CTX\varphi^-RS1\varphi^-VPI^+VSP^+$ strains and the toxigenic strains and gives support to the data on the origin of the former obtained by us. The toxigenic and nontoxigenic strains within the first group share on the whole a considerable number of single nucleotide substitutions and differ from each other by just 1 to 137 SNPs.

The second group consists of ten nontoxigenic strains characterized by the absence of not only both prophages, CTXφ and RS1φ, but also the two pandemicity islands, VSP-I and VSP-II. Nearly 3000 unique SNPs have been detected in the genomes of these strains, which have not been found in any isolate from the first cluster and are, therefore, their characteristic genetic markers. For example, the *rtxA* gene encoding multifunctional autoprocessed RTX cytotoxin (from Repeats in toxin) [31] contains two nonsynonymous single nucleotide substitutions (G/A in position 6901 and A/G in position 8731), which are absent in all the examined strains in the first group. Phylogenetic remoteness of the strains of this group from the toxigenic isolates may be indirect evidence of their independent origin. Moreover, the level of genetic heterogeneity of the nontoxigenic strains in the second group is significantly higher than the level of heterogeneity of those in the first group. This is indicated by the fact that the strains within the second group differ from each other by 2935–4360 SNPs, while the difference between the strains in the first group ranges from 1 to 137 SNPs. These data make it possible to subdivide the second group into at least five subgroups (Fig. 2, group *2*, subgroups a, b, c, d, e).

To summarize, the results of the whole genome SNP analysis of 35 strains of *V. cholerae* El Tor biovar suggest, first, that there are two distinct phylogenetic groups of nontoxigenic $ctxA^-tcpA^+$ strains, each possessing a characteristic set of single nucleotide substitutions. Second, the detected close genetic relationship between the strains within the first group may be indicative of toxigenic isolates being the source of nontoxigenic $CTX\varphi^-RS1\varphi^+VPI^+VSP^+$ strains in nature. This observation may be of great value for epidemiological studies. The origin of the nontoxigenic $CTX\varphi^-RS1\varphi^-VPI^+VSP^-$ isolates constituting the second phylogenetic group, which are likely to be nonepidemic, remains unknown. The unique SNPs identified in these strains might serve as markers making it possible to differentiate between nontoxigenic $ctxA^-tcpA^+$ strains with different epidemiological significance.

DISCUSSION

The analysis of the full genome nucleotide sequences for 18 natural nontoxigenic *V. cholerae* El Tor biovar strains and two experimentally obtained isolates containing the key pathogenicity gene $tcpA^+$ (Table 3) made it possible to show for the first time their genetic heterogeneity. It became clear that the $ctxA^-tcpA^+$ strains under study, which were earlier considered as potentially epidemic, fall into two separate groups, differing in the structure of the genomes of the constituent strains (Fig. 2, *1* and *2*). The first

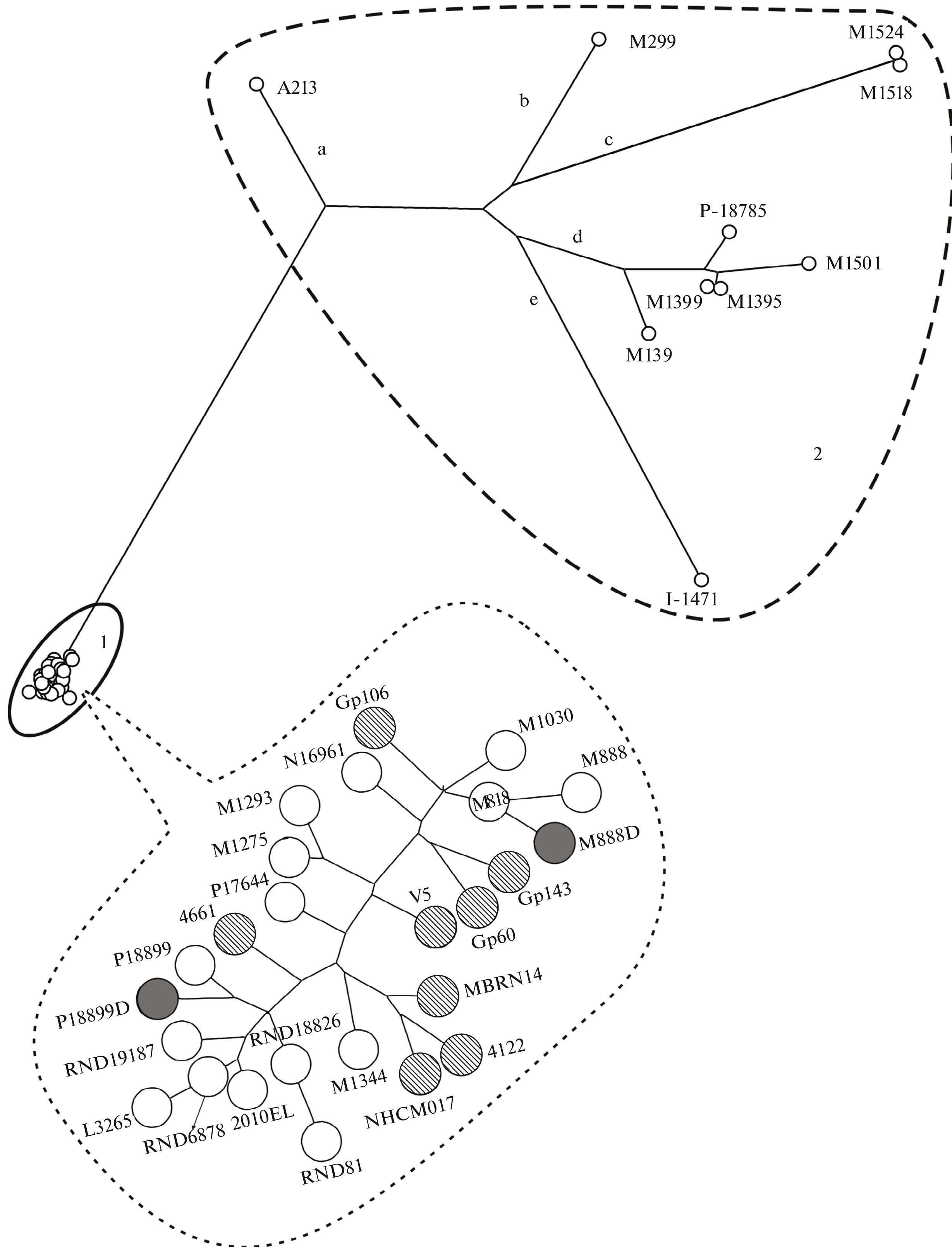

Fig. 2. Phylogenetic tree for toxigenic (*ctxA*$^+$*tcpA*$^+$) and nontoxigenic (*ctxA*$^-$*tcpA*$^+$) strains of *V. cholerae* El Tor biovar constructed on the basis of the results of the whole genome SNP analysis. BioNumerics v 7.1 software, maximum parsimony tree method. (*1*) Phylogenetic group including natural toxigenic strains (while ovals), natural nontoxigenic potentially epidemic strains (shaded ovals), and experimentally obtained nontoxigenic potentially epidemic strains (gray ovals); (*2*) phylogenetic group including nontoxigenic nonepidemic strains. Letters at the base of branches indicate subgroups of strains.

group is formed by the strains which, in contrast to toxigenic strains, lack relatively small DNA regions in the genome which contain structural genes for cholera toxin (CTXφ prophage) and, in some cases, genes involved in the control of production of this key pathogenicity factor (RS1φ prophage). The second group includes the strains whose genomes lack not only the CTXφ and RS1φ prophages but also the VSP-I and VSP-II pandemicity islands.

The observed differences in the genome structure are of the utmost value for the assessment of epidemiological significance of nontoxigenic strains isolated from the water of surface reservoirs in the course of environmental monitoring. $CTX\varphi^-RS1\varphi^+VPI^+VSP^+$ and $CTX\varphi^-RS1\varphi^-VPI^+VSP^+$ strains may actually be potentially epidemic owing to the presence in their genome of both the genes encoded by pathogenicity islands and a full set of genes encoded by pandemicity islands contributing to their adaptation to different environmental conditions [5, 6]. The potentially epidemic status for these strains, which were isolated mostly on the territories endemic for cholera, is determined by the actual probability of acquisition by them of the cholera toxin genes as a part of the CTXφ prophage through phage conversion and, as a consequence, the development of toxigenic strains [32]. This course of events appears quite possible if such strains occur together with toxigenic ones in biofilms on biotic or abiotic surfaces [33]. Thus, nontoxigenic strains with the described genotype may serve as a supply of material for the development of a toxigenic epidemic population when toxigenic strains circulate in the environment. The possibility of such an event was earlier experimentally proved [21–23].

On the other hand, the second group of nontoxigenic $CTX\varphi^-RS1\varphi^-VPI^+VSP^-$ strains identified by us is not epidemic, since to develop a toxigenic population these strains have to undergo a considerable reorganization of their genome, namely, to acquire from a donor four different DNA regions with the total size of 54.2 kb containing prophages and pandemicity islands. The possibility of such reorganization occurring in nature has not yet been experimentally demonstrated and is in fact extremely unlikely.

It is important to note that differentiation of nontoxigenic *ctxA*$^-$*tcpA*$^+$ strains isolated in the course of environmental monitoring by their epidemiological significance on the basis of the differences in their genomes revealed here is of great importance to epidemiologists, since it will make it possible to differentiate the amount of epidemic countermeasures.

The results of the model experiments which demonstrated that, when a genetically homogeneous population of a toxigenic strain [17] is placed under external environmental conditions, some strains after a certain period of time can form a heterogeneous population which includes nontoxigenic clones formed as a result of the loss of one of the virulence prophages, CTXφ, or both of them, CTXφ and RS1φ, are of interest by themselves. These results firmly support the previously obtained data indicating that stress increases genetic diversity in prokaryotes [34]. Moreover, the data indicate that the genome of a typical strain can undergo reorganization in response to stress factors which leads to the loss of the virulence prophage (or prophages), whose genes are not vital for the bacterial cell survival under certain environmental conditions. Secondly, we have for the first time reported the experimental verification for the origination of potentially epidemic strains from toxigenic isolates. The results are important for understanding the mechanisms of formation of such strains and the reasons underlying it.

The analysis of phylogenetic relationships based on studying single nucleotide polymorphisms in the genomes of 35 strains made it possible to show genetic similarities and differences between toxigenic and nontoxigenic strains. It was demonstrated for the first time here that nontoxigenic *ctxA*$^-$*tcpA*$^+$ strains cluster on the phylogenetic tree into two separate groups differing in their epidemiological significance according to the rate of the intraspecies variability of their full genome nucleotide sequences. Nontoxigenic $CTX\varphi^-RS1\varphi^+VPI^+VSP^+$ and $CTX\varphi^-RS1\varphi^-VPI^+VSP^+$ strains fall in the same group as toxigenic strains, thus indicating a close genetic relationship between these potentially epidemic strains and the toxigenic ones and providing support to our data on their origin.

As for the $CTX\varphi^-RS1\varphi^-VPI^+VSP^-$ strains, it was demonstrated that they form a separate (second) group on the phylogenetic tree and clearly diverge from both the nontoxigenic potentially epidemic strains and the toxigenic isolates. This observation most likely implies that they originated independently from toxigenic strains. We identified for the first time single nucleotide substitutions unique to this second phylogenetic group, which may be used as genetic markers to differentiate nontoxigenic isolates with different epidemiological significance. Also of interest is the observation that the genomes of nontoxigenic strains from different phylogenetic groups are characterized by a different level of heterogeneity.

In summary, the data presented and the results of their analysis clearly indicate that the development of modern molecular biology techniques has provided fundamentally new facilities making it possible to obtain novel data on intraspecies reorganizations of the *V. cholera* genome in the course of evolution. The new data obtained on the genome structure of nontoxigenic strains with different genotypes may be of significant value for further studies of the effects which genomic reorganizations may exert on the phenotypic properties of the pathogen, including those underlying its virulence and persistence in various ecological niches. In addition, identification of marker SNPs in the full nucleotide sequences of the genomes of non-

toxigenic strains from the second phylogenetic group will allow proceeding to development of new diagnostic means of differentiation of the natural *V. cholera* strains with different epidemiological significance on the genetic level.

ACKNOWLEDGMENTS

We are grateful to Yu.V. Lozovskii, a member of the Department of Microbiology, for his assistance in the preparation of this manuscript.

REFERENCES

1. Rahaman, M.H., Islam, T., Colwell, R.R., and Alam, M., Molecular tools in understanding the evolution of *Vibrio cholerae*, *Front. Microbiol.*, 2015, vol. 6, p. 1040. doi 10.3389/fmicb.2015.01040
2. Onishchenko, G.G., Moskvitina, E.A., Kologorov, A.I., et al., Cholera outbreak in Kazan in 2001, *Probl. Osobo Opasnykh Infekts.*, 2001, vol. 2, no. 82, pp. 15–26.
3. Onishchenko, G.G., Lomov, Yu.M., Moskvitina, E.A., et al., Cholera caused by *Vibrio cholerae* O1 *ctxAB*$^-$ *tcpA*$^+$, *Zh. Mikrobiol. Epidemiol. Immunobiol.*, 2007, no. 1, pp. 23–29.
4. Heidelberg, J.F., Eisen, J.A., Nelson, W.C., et al., DNA sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*, *Nature*, 2000, vol. 406, no. 6795, pp. 477–483. doi 10.1038/35020000
5. Smirnova, N.I. and Kutyrev, V.V., Evolution of the cholera causative agent, *Mol. Mikrobiol. Virusol.*, 2004, no. 4, pp. 3–13.
6. Dziejman, M., Balon, E., Boydet, D., et al., Comparative genomic analysis of *Vibrio cholerae*: genes that correlate with cholera endemic and pandemic disease, *Proc. Natl. Acad. Sci. U.S.A.*, 2002, vol. 99, no. 3, pp. 1556–1561. doi 10.1073/pnas.042667999
7. Karaolis, D.K.R., Johnson, J.A., Bailey, C.C., et al., A *Vibrio cholerae* pathogenicity island associated with epidemic and pandemic strains, *Proc. Natl. Acad. Sci. U.S.A.*, 1998, vol. 95, no. 6, pp. 3134–3139.
8. Marsh, J.W. and Taylor, R.K., Genetic and transcriptional analyses of the *Vibrio cholerae* mannose-sensitive hemagglutinin type 4 pilus gene locus, *J. Bacteriol.*, 1999, vol. 181, no. 4, pp. 1110–1117.
9. Kaper, J.B., Morris, J.G., and Levine, M.M., Cholera, *Clin. Microbiol. Rev.*, 1995, vol. 8, no. 1, pp. 48–89.
10. Farhadi, T., Nezafat, N., and Ghasemi, Y., *In silico* phylogenetic analysis of *Vibrio cholerae* isolates based on thee housekeeping genes, *Int. J. Comput. Biol. Drug Des.*, 2015, vol. 8, no. 1, pp. 62–74. doi 10.1504/IJCBDD.2015.068789
11. Kahler, A.M., Haley, B.J., Chen, A., et al., Environmental surveillance for toxigenic *Vibrio cholerae* in surface waters of Haiti, *Am. J. Trop. Med. Hyg.*, 2015, vol. 92, no. 125, pp. 118–125. doi 10.4269/ajtmh.13-0601
12. Staley, J.T., Boome, D.R., Bremmer, P.J., et al., *Berges Manual Systematic Bacteriology: the Proteobacteria,* part B: *The Gamma Proteobacteria*, Springer-Verlag, 2005, vol. 2.
13. Smirnova, N.I., Zadnova, S.P., Agafonov, D.A., et al., Comparative molecular-genetic analysis of mobile elements in natural strains of cholera agent, *Russ. J. Genet.*, 2013, vol. 49, no. 9, pp. 898–908.
14. Nair, G.B., Faruque, S.M., Bhuiyan, N.A., et al., New variants of *Vibrio cholerae* O1 biovar El Tor with attributes of the classical biovar from hospitalized patients with acute diarrhea in Bangladesh, *J. Clin. Microbiol.*, 2002, vol. 40, no. 9, pp. 3296–3299. doi 10.1128/JCM.40.9.3296-3299.2002
15. Safa, A., Nair, G.B., and Kong, R.Y.C., Evolution of new variants of *Vibrio cholerae* O1, *Trends Microbiol.*, 2010, vol. 18, no. 1, pp. 46–54. doi 10.1016/j.tim.2009.10.003
16. Bhattacharia, D., Dey, S., and Pazhani, G.P., *Vibrio cholerae* O1 El Tor variant and emergence of Haitian *ctxB* variant in the strains isolated from South India, *Med. Microbiol. Immunol.*, 2015, vol. 4. doi 10.1007/s00430-015-0433-y
17. Gintsburg, A.L., Bryukhanov, A.F., Yanishevskii, N.V., et al., Application of gene probe method for the detection of epidemically dangerous strains of cholera, *Mol. Genet., Mikrobiol., Virusol.*, 1987, no. 11, pp. 8–13.
18. Chakraborty, S., Mukhopadhyay, A.K., Bhadra, R.K., et al., Virulence genes in environmental strains of *Vibrio cholerae*, *Appl. Environ. Microbiol.*, 2000, vol. 66, no. 9, pp. 4022–4028.
19. Smirnova, N.I., Kostromitina, E.A., Osin, A.V., and Kutyrev, V.V., Genomic variability of *Vibrio cholerae* El Tor biovariant strains, *Vestn. Ross. Akad. Med. Nauk*, 2005, no. 7, pp. 19–26.
20. Mohaptra, S.S., Mantri, C.K., Bhotra, T., and Singh, D.V., Characteristics of *Vibrio cholerae* O1 isolated from water of the River Ganga, Varanasi, India, *Ind. J. Med. Microbiol.*, 2015, vol. 33, no. 4, pp. 507–515. doi 10.4103/0255-0857.167327
21. Faruque, S.M., Asadulghani Saha, M.N., et al., Analysis of clinical and environmental strains of nontoxigenic *Vibrio cholerae* for susceptibility to CTXφ: molecular basis for origination of new strains with epidemic potential, *Infect. Immun.*, 1998, vol. 66, no. 12, pp. 5819–5825.
22. Waldor, M.K. and Mekalanos, J.J., Lysogenic conversion by a filamentous bacteriophage encoding cholera toxin, *Science*, 1996, vol. 272, no. 5270, pp. 1910–1914. doi 10.1126/science272.5270.1910
23. Gridneva, L.G., Musatov, Yu.S., Gromova, T.V., et al., Results of monitoring over and biological properties of *Vibrio cholerae* isolated from ambient environment objects in the Khabarovsk territory, Probl. Osobo Opasnykh Infekts., 2014, no. 1, pp. 121–125.
24. *Bezopasnost' raboty s mikroorganizmami I-II grupp patogennosti (opasnosti)* (The Safe Handling of Microorganisms of I-II Pathogenicity Groups (Hazard)), *Sanitary and Epidemiological Rules 1.3.3118-132013*, 2013.
25. Shchelkanova, E.Yu., Goryaev, A.A., and Smirnova, N.I., Genomic variations of CTXφ prophage of *Vibrio cholerae* biovar El tor induced by Tn5-Mob transposone, *Mol. Genet., Microbiol., Virusol.*, 2008, no. 3, article 111.
26. Waldor, M.K., Rubin, E.J., Pearson, G.D., et al., Regulation, replication, and integration functions of the

Vibrio cholerae CTXφ are encoded by region RS2, *Mol. Microbiol.*, 1997, vol. 24, no. 5, pp. 917–926.

27. Keasler, S.P. and Hall, R.H., Detecting and biotyping *Vibrio cholerae* O1 with multiplex polymerase chain reaction, *Lancet*, 1993, vol. 341, no. 8861, p. 1661.
28. Parsot, C. and Mekalanos, J.J., Expression of the *Vibrio cholerae* gene encoding aldehyde dehydrogenase is under control of ToxR, the cholera toxin transcriptional activator, *J. Bacteriol.*, 1991, vol. 173, no. 9, pp. 2842–2851.
29. Osin, A.V., Nefedov, K.S., Yeroshenko, G.A., and Smirnova, N.I., Comparative genomic analysis of *Vibrio cholera* El Tor preseventh and seventh pandemic strains isolated in various periods, *Russ. J. Genet.*, 2005, vol. 41, no. 1, pp. 1–10.
30. Maniatis, T., Fritsch, E.F., and Sambrook, J., *Molecular Cloning: a Laboratory Manual*, Cold Spring Harbor: Cold Spring Harbor Lab., 1982.
31. Dolores, J. and Satchell, K.J., Analysis of *Vibrio cholerae* genome sequences reveals unique *rtxA* variants in environmental strains and an *rtxA*-null mutation in recent altered El Tor isolates, *Mbio.*, 2013, vol. 4, no. 2, pp. 1–9. doi 10.1128/mBio.00624-12
32. Fan, F. and Kan, B., Survival and proliferation of the lysogenic bacteriophage CTXφ in *Vibrio cholerae*, *Virol. Sin.*, 2015, vol. 30, no. 1, pp. 19–25. doi 10.1007/s12250-014-3550-7
33. Meibom, K.L., Blokesch, M., Dolganov, N.A., et al., Chitin induces natural competence in *Vibrio cholerae*, *Science*, 2005, vol. 310, no. 5755, pp. 1824–1827. doi 10.1126/science.1120096
34. Velkov, V.V., New insights into the molecular mechanisms of evolution: stress increases genetic diversity, *Mol. Biol.* (Moscow), 2002, vol. 36, no. 2, pp. 209–215.

Translated by E. Martynova