

***Vibrio cholerae* embraces two major evolutionary traits as revealed by targeted gene sequencing**

Kazuhisa Okada^{1,2*}, Warawan Wongboot¹, Siriporn Chantaroj³, Wirongrong Natakuathung¹, Amonrattana Roobthaisong¹, Watcharaporn Kamjumphol¹, Fumito Maruyama⁴, Taichiro Takemura⁵, Ichiro Nakagawa⁴, Makoto Ohnishi⁶, Shigeyuki Hamada^{1,2}

¹ Thailand-Japan Research Collaboration Center on Emerging and Re-emerging Infections (RCC-ERI), Nonthaburi, Thailand, ²Research Institute for Microbial Diseases, Osaka University, Suita, Osaka, Japan, ³National Institute of Health, Department of Medical Sciences, Ministry of Public Health, Nonthaburi, Thailand, ⁴ Department of Microbiology, Kyoto University, Sakyo-ku, Kyoto, Japan, ⁵Vietnam Research Station, Institute of Tropical Medicine, Nagasaki University, Hanoi, Vietnam, ⁶ Department of Bacteriology, National Institute of Infectious Diseases , Shinjuku-ku, Tokyo, Japan.

***Corresponding author:** Kazuhisa Okada

E-mail Address: kazuhisa@biken.osaka-u.ac.jp

Supplemental material

Table S1

Table S2

Figure S1

Figure S2

Figure S3

Figure S1. Dendrogram based on the *M* gene sequences of vibrios shows higher similarity within species than between species. Scale bars indicate nucleotide substitutions per site. *Vibrio cholerae* strains are highlighted in the light blue box.

Figure S2. Schematic representation of three types of gene arrangement found in the locus MS6_A0927 and hypothesized gene replacement of *M* and *LH* genes among *Vibrio cholerae* and *V. mimicus* strains. Gene arrangements among the eight strains grouped by the *M/LH* analysis are shown. Two *V. cholerae* strains N2_17 and 87395 showed a similar gene arrangement that contained both *M* (common M5) and *LH* (LH4 or LH19). Strain 87395 was phylogenetically most closely related with strain HE-09, which exhibited M5. The strains 87395, N56, N79, N80, and N83 exhibited identical LH4 and clustered in a deep branch of the dendrogram shown in Figure 2.

Figure S3. Multiplex PCR assay for rapid screening of six targeted genes including three virulence-related genes (*toxR*, *tcpA*, *ctxAB*), a seventh pandemic group-specific marker (VC2346), and *M/LH*. Lane M, 100-bp DNA ladder; lane 1, *V. cholerae* O1 El Tor strain N16961 (seventh cholera pandemic clone); lane 2, *V. cholerae* O1 classical strain 569B; lane 3, *V. cholerae* O1 El Tor strain MS6 (non-epidemic strain).

Table S1. Profiling of *Vibrio cholerae* and closely related species deposited in GenBank

Strain	Species*	Category	Year	Location	Serogroup	ctxA/B	tcpA	toxR	VC2346	LH	M	Subclade	Accession no.
O395	VC	C	1965	India	O1	+	+	+	-	+	-	LH1	NC_012582.1/3.1
RC27	VC	U	1991	Indonesia	O1	+	+	+	-	+	-	LH1	ADAI010000000
V52	VC	C	1968	Sudan	O37	+	+	+	-	+	-	LH1	AAKJ000000000
95412	VC	C	1997	Mexco	O1	+	+	+	-	+	-	LH1	APFM000000000
NCTC 8457	VC	C	1910	Saudi Arabia	O1	-	+	+	-	+	-	LH2	AAWD000000000
M66-2	VC	C	1937	Indonesia	O1	-	+	+	-	+	-	LH2	CP001233/4
MAK757	VC	C	1937	Celebes Island	O1	+	+	+	-	+	-	LH2	AAUS_020000000
BX330286	VC	E	1986	Australia	O1	+	+	+	-	+	-	LH2	ACIA000000000
B33	VC	C	2004	Beira, Mozambique	O1	+	+	+	+	-	-	LH2	ACHZ010000000
MJ-1236	VC	C	1994	Matlab, Bangladesh	O1	+	+	+	+	-	-	LH2	NC_012667/8
RC9	VC	C	1985	Kenya	O1	+	+	+	+	-	-	LH2	ACHX000000000
N16961	VC	C	1975	Bangladesh	O1	+	+	+	+	-	-	LH2	NC_002505/6
2011EL-301	VC	E	2011	Taganrog, Russia	O1	+	+	+	+	-	-	LH2	AJFN000000000
CP1041(14)	VC	C	2004	Zambia	O1	+	+	+	+	-	-	LH2	ALDE000000000
CP1030(3)	VC	C	2008	Mexco	O1	+	+	+	+	-	-	LH2	ALCZ000000000
CP1032(5)	VC	C	1991	Mexico	O1	+	+	+	+	-	-	LH2	ALDA010000000
CP1033(6)	VC	C	2000	Mexico	O1	+	+	+	+	-	-	LH2	AJRL000000000
CP1038(11)	VC	C	2003	Zimbabwe	O1	+	+	+	+	-	-	LH2	ALDC000000000
CP1040(13)	VC	C	2004	Zambia	O1	+	+	+	+	-	-	LH2	ALDD000000000
CP1042(15)	VC	C	2010	Thailand	O1	+	+	+	+	-	-	LH2	ALDF000000000
CP1044(17)	VC	C	1991	Peru	O1	+	+	+	+	-	-	LH2	ALDG000000000
CP1047(20)	VC	C	1995	Peru	O1	+	+	+	+	-	-	LH2	ALDI000000000
CP1048(21)	VC	C	2010	Bangladesh	O1	+	+	+	+	-	-	LH2	ALDJ000000000
CP1050(23)	VC	C	2010	Bangladesh	O1	+	+	+	+	-	-	LH2	ALDK000000000
H1	VC	C	2010	Haiti	O1	+	+	+	+	-	-	LH2	AKGH000000000
HC-06A1	VC	C	2010	Haiti	O1	+	+	+	+	-	-	LH2	AGUI000000000
HC-17A1	VC	C	2010	Haiti	O1	+	+	+	+	-	-	LH2	AJRN000000000
HC-17A2	VC	C	2010	Haiti	O1	+	+	+	+	-	-	LH2	AJSI000000000
HC-19A1	VC	C	2010	Haiti	O1	+	+	+	+	-	-	LH2	AGUJ000000000
HC-20A2	VC	C	2010	Haiti	O1	+	+	+	+	-	-	LH2	ALDL000000000
HC-21A1	VC	C	2010	Haiti	O1	+	+	+	+	-	-	LH2	AGUK000000000
HC-23A1	VC	C	2010	Haiti	O1	+	+	+	+	-	-	LH2	AGUM000000000
HC-28A1	VC	C	2010	Haiti	O1	+	+	+	+	-	-	LH2	AGUN000000000
HC-32A1	VC	C	2010	Haiti	O1	+	+	+	+	-	-	LH2	AGUO000000000
HC-33A2	VC	C	2010	Haiti	O1	+	+	+	+	-	-	LH2	AGUP000000000
HC-37A1	VC	C	2010	Haiti	O1	+	+	+	+	-	-	LH2	AJSJ000000000
HC-38A1	VC	C	2010	Haiti	O1	+	+	+	+	-	-	LH2	AHGF000000000
HC_39A1	VC	C	2010	Haiti	O1	+	+	+	+	-	-	LH2	AFOV000000000
HC_40A1	VC	C	2010	Haiti	O1	+	+	+	+	-	-	LH2	ALDM000000000
HC_41A1	VC	C	2010	Haiti	O1	+	+	+	+	-	-	LH2	AFOK000000000
HC-42A1	VC	C	2010	Haiti	O1	+	+	+	+	-	-	LH2	ALDN000000000
HC-43A1	VC	C	2010	Haiti	O1	+	+	+	+	-	-	LH2	ALDO000000000
HC_46A1	VC	C	2010	Haiti	O1	+	+	+	+	-	-	LH2	AGUQ000000000
HC_47A1	VC	C	2010	Haiti	O1	+	+	+	+	-	-	LH2	ALDO000000000
HC-48A1	VC	C	2010	Haiti	O1	+	+	+	+	-	-	LH2	ALDR000000000
HC-48B2	VC	C	2010	Haiti	O1	+	+	+	+	-	-	LH2	AFOL000000000
HC-49A2	VC	C	2010	Haiti	O1	+	+	+	+	-	-	LH2	AGUR000000000
HC-56A2	VC	C	2010	Haiti	O1	+	+	+	+	-	-	LH2	AFOM000000000
HC-57A2	VC	C	2010	Haiti	O1	+	+	+	+	-	-	LH2	ALDX000000000
HC-61A1	VC	C	2010	Haiti	O1	+	+	+	+	-	-	LH2	ALDZ000000000
HC-62A1	VC	C	2010	Haiti	O1	+	+	+	+	-	-	LH2	AGUS000000000
HC-62B1	VC	C	2010	Haiti	O1	+	+	+	+	-	-	LH2	AJRV000000000
HC-64A1	VC	C	2010	Haiti	O1	+	+	+	+	-	-	LH2	AJSO000000000
HC-65A1	VC	C	2010	Haiti	O1	+	+	+	+	-	-	LH2	AGUT000000000
HC-67A1	VC	C	2010	Haiti	O1	+	+	+	+	-	-	LH2	AGUU000000000
HC-68A1	VC	C	2010	Haiti	O1	+	+	+	+	-	-	LH2	AGUV000000000
HC-69A1	VC	C	2010	Haiti	O1	+	+	+	+	-	-	LH2	AGUW000000000
HC-71A1	VC	C	2010	Haiti	O1	+	+	+	+	-	-	LH2	AJSP000000000
HC-72A2	VC	C	2010	Haiti	O1	+	+	+	+	-	-	LH2	AGUX000000000
HC-77A1	VC	C	2010	Haiti	O1	+	+	+	+	-	-	LH2	AGUY000000000
HC-7A1	VC	C	2010	Haiti	O1	+	+	+	+	-	-	LH2	AJRW000000000
HC-80A1	VC	C	2010	Haiti	O1	+	+	+	+	-	-	LH2	AGVA000000000
HC-81A1	VC	C	2010	Haiti	O1	+	+	+	+	-	-	LH2	AGVB000000000
HC-81A2	VC	C	2010	Haiti	O1	+	+	+	+	-	-	LH2	AGVC000000000
HC-22A1	VC	C	2010	Haiti	O1	+	+	+	+	-	-	LH2	ALDA000000000
HC-70A1	VC	C	2010	Haiti	O1	+	+	+	+	-	-	LH2	AGUL000000000
HCUF01	VC	C	2010	Haiti	O1	+	+	+	+	-	-	LH2	AFON000000000
HFU_02	VC	C	2010	Haiti	O1	+	+	+	+	-	-	LH2	AFOO000000000
INDRE_91_1	VC	C	1991	Mexico	O1	+	+	+	+	-	-	LH2	AFOO000000000
116059	VC	C	1992	Brazil	O1	+	+	+	+	-	-	LH2	ADAK000000000
2009V-1046	VC	C	2009	USA	O1	+	+	+	+	-	-	LH2	APFJ000000000
2009V-1085	VC	C	2009	USA	O1	+	+	+	+	-	-	LH2	AHFY000000000
2009V-1096	VC	C	2009	USA	O1	+	+	+	+	-	-	LH2	AHFY000000000
2009V-1116	VC	C	2009	USA	O1	+	+	+	+	-	-	LH2	AHFZ000000000
2009V-1131	VC	C	2009	USA	O1	+	+	+	+	-	-	LH2	AHGA000000000
2010EL-1749	VC	C	2010	Cameroon	O1	+	+	+	+	-	-	LH2	AHGB000000000
2010EL-1786	VC	C	2010	Haiti	O1	+	+	+	+	-	-	LH2	AHGC000000000
2010EL-1792	VC	C	2010	Haiti	O1	+	+	+	+	-	-	LH2	NC_012667/8
2010EL-1798	VC	C	2010	Haiti	O1	+	+	+	+	-	-	LH2	AELJ000000000
2010EL-1961	VC	C	2010	Haiti	O1	+	+	+	+	-	-	LH2	AELI000000000
2010EL-2010H	VC	C	2010	Haiti	O1	+	+	+	+	-	-	LH2	AHGD000000000
2010EL-2010H	VC	C	2010	Haiti	O1	+	+	+	+	-	-	LH2	AHGE000000000
2010V-1014	VC	C	2010	USA	O1	+	+	+	+	-	-	LH2	AHGG000000000
2011EL-1089	VC	C	2010	Haiti	O1	+	+	+	+	-	-	LH2	AHGH000000000
2011EL-1133	VC	C	2011	Haiti	O1	+	+	+	+	-	-	LH2	AHGI000000000
2011EL-1137	VC	C	2009	South Africa	O1	+	+	+	+	-	-	LH2	AHGO000000000
2011V-1021	VC	C	2011	Dominican Republic	O1	+	+	+	+	-	-	LH2	AHKK000000000
3500-05	VC	C	2005	USA	O1	+	+	+	+	-	-	LH2	AHGL000000000
3546-06	VC	C	2006	USA	O1	+	+	+	+	-	-	LH2	AHGM000000000
3554-08	VC	C	2008	USA	O1	+	+	+	+	-	-	LH2	AHGN000000000
3582-05	VC	C	2005	USA	O1	+	-	+	+	-	-	LH2	AHGP000000000
AG-7404	VC	C	1991	Bangladesh	O1	+	+	+	+	-	-	LH2	APFN000000000
AG-8040	VC	C	1991	Bangladesh	O1	+	+	+	+	-	-	LH2	APFO000000000
C6706	VC	C	1991	Peru	O1	+	+	+	+	-	-	LH2	AHGQ000000000
CIRS101	VC	C	2002	Dhaka, Bangladesh	O1	+	+	+	+	-	-	LH2	ACWV000000000
EC-0009	VC	E	2010	Bangladesh	O1	+	+	+	+	-	-	LH2	APFP010000000
EC-0012	VC	E	2010	Bangladesh	O1	+	+	+	+	-	-	LH2	APFG000000000
EC-0051	VC	E	2011	Bangladesh	O1	+	+	+	+	-	-	LH2	APFR000000000
EDC-020	VC	E	2010	Bangladesh	O1	+	+	+	+	-	-	LH2	APFS000000000
EDC-022	VC	E	2010	Bangladesh	O1	+	+	+	+	-	-	LH2	APFT000000000
EM-1536	VC	E	2010	Bangladesh	O1	+	+	+	+	-	-	LH2	APFU000000000
EM-1626	VC	E	2010	Bangladesh	O1	+	+	+	+	-	-	LH2	APFV000000000
EM-1626	VC	E	2010	Bangladesh	O1	+	+	+	+	-	-	LH2	APFW000000000
EM-1727	VC	E	2011	Bangladesh	O1	+	+	+	+	-	-	LH2	APFX000000000
G4222	VC	C	2001	South Africa	O1	+	+	+	+	-	-	LH2	APFZ010000000
Nep-21106	VC	C	2003	Nepal	O1	+	+	+	+	-	-	LH2	APGB000000000
NHCC_09A	VC	C	2010	Bangladesh	O1	+	+	+	+	-	-	LH2	APGA010000000
NHCC-006C	VC	C	2010	Bangladesh	O1	+	+	+	+	-	-	LH2	APGB000000000
NHCC-010F	VC	C	2010	Bangladesh	O1	+	+	+	+	-	-	LH2	APGD000000000
Nep-21113	VC	C	2003	Nepal	O1	+	+	+	+	-	-	LH2	APGF000000000
VC1761	VC	C	2009	Malaysia	O1	+	+	+	+	-	-	LH2	AMBS000000000
MO10	VC	C	1992	Madras, India	O139	+	+	+	+	-	-	LH2	AAKF030000000
4260B	VC	C	1993	Bangladesh	O139	+	+	+	+	-	-	LH2	AMVL000000000
VC4370	VC	E	2008	Selangor, Malaysia	O139	+	+	+	+	-	-	LH2	AMBT000000000
MS6	VC	C	2008	Thailand	O1	+	+	+	-	-	+	M1	AP014524/5
P-18785	VC	C	2005	Russia	O1	-	+	+	-	-	+	M1	ANHS000000000
2740-80	VC	E	1980	US Gulf coast, USA	O1	-	+	+	-	-	+	M2	AAUT010000000
3569-08	VC	E	2008	US Gulf coast, USA	O1	+	+	+	-				

Table S2. Primers for multiplex PCR assay and targeted gene sequencing of the MS6_A0927

Target gene	Primer name	Nucleotide sequence (5'→3')	Amplicon size (bp)	Reference
Multiplex PCR				
<i>M</i>	metY-F	GCGTGAAACCGGAGATGATCC	353	(1)
	lysR-R	AGCGCAGAAGGTGTTACGCCA		(1)
<i>LH</i>	luxR-F	TAGCTCACCGCGAGCTCGTTG	521	(1)
	lysR-R	AGCGCAGAAGGTGTTACGCCA		(1)
<i>VC2346</i>	VC2346-F	CAACTGTGCTAGCAGTTGCC	243	This study
	VC2346-R	TCAGAGTCATAGATTGAAGC		This study
<i>ctxA-B</i>	ctxA-B-F	GTGTGTTGTGGTATTCTGCAC	702	This study
	ctxA-B-R	TCATGCAAGAGGAACTCAGAC		This study
<i>tcpA</i>	tcpA-F	ACCGGTCAAGAGGGTATGACA	110	This study
	tcpA-R	CTGCGAATCAATCGCACGCTG		This study
<i>toxR</i>	toxR-F	TGCGTAAGGTTATGTTTTCCC	844	This study
	toxR-R	GTTCGGATTAGGACACAACCTC		This study
Targeted gene sequencing				
MS6_A0927	MS6_A0926F	CGGATTTGCTGATAAAAAGCAGAAGA	c.a. 2,130 or	This study
	MS6_A0928R	GCGGTAAACTGCCTTCATTRACCAC	1,960 or over	This study
Sequence primer				
<i>M</i>	m_uni1-R	TCAATCCCTTGGCTTGGC		This study
	m_uni2-R	AGGATCATGAAGGCGTTCA		This study
	m_uni3-F	TCAACTACGCGATTCTGAC		This study
	m_uni4-F	GTACACCGAAGCCTTTGG		This study
<i>LH</i>	lh_uni1-R	ACCTCTTGGCTGTCAGGC		This study
	lh_uni2-F	ATCTTCTTTCAGGCAGCC		This study
	lh_uni3-F	GATAGTCCGCTTGCAGGC		This study
	lh_uni4-F	ATATTGCCCTATAACTCG		This study

Supplemental reference

- (1) Okada, K., Na-Ubol, M., Natakuathung, W., Roobthaisong, A., Maruyama, F., Nakagawa, I., Chantaroj, S., and Hamada, S. (2014). Comparative genomic characterization of a Thailand-Myanmar isolate, MS6, of *Vibrio cholerae* O1 El Tor, which is phylogenetically related to a "US Gulf Coast" clone. *PLoS One* 9, e98120.

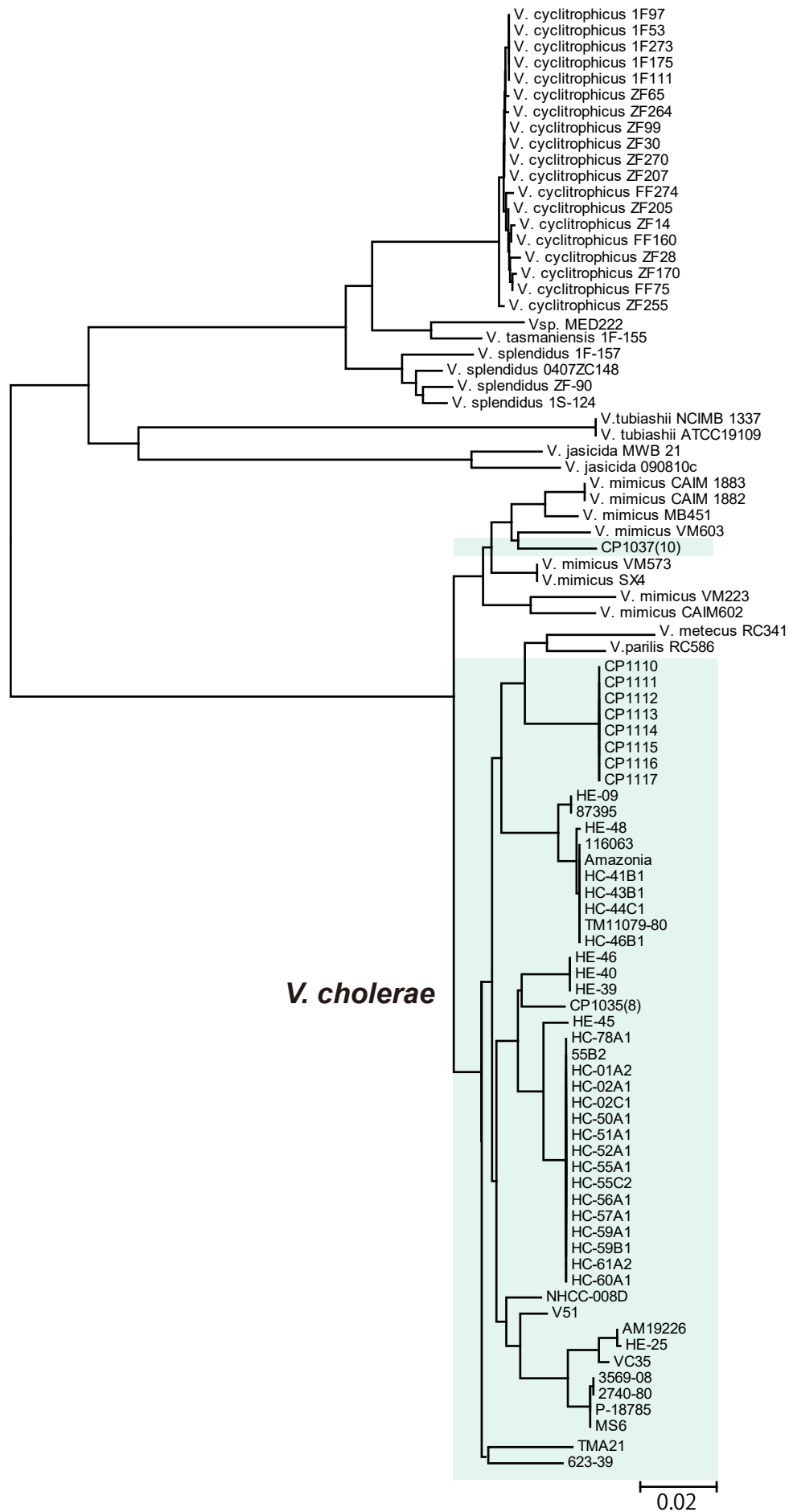


Figure S1. Dendrogram based on the *M* gene sequences of vibrios shows higher similarity within species than between species. Scale bars indicate nucleotide substitutions per site. *Vibrio cholerae* strains are highlighted in the light blue box.

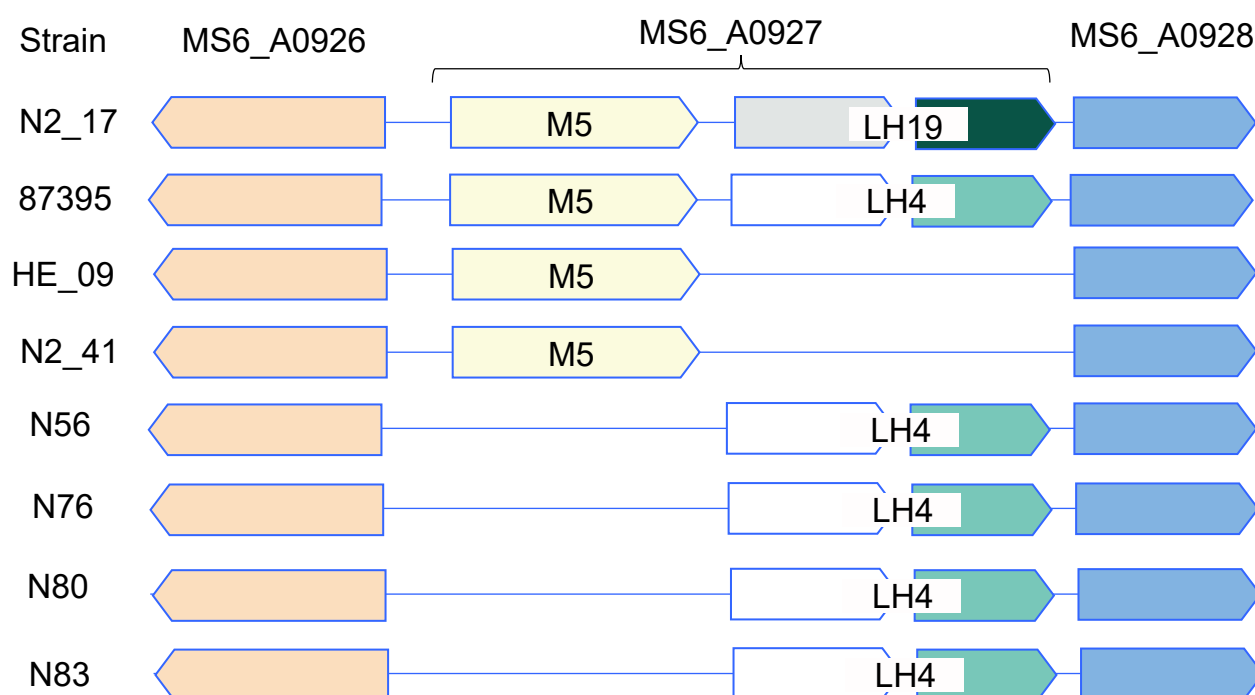


Figure S2. Schematic representation of three types of gene arrangement found in the locus MS6_A0927 and hypothesized gene replacement of *M* and *LH* genes among *Vibrio cholerae* and *V. mimicus* strains. Gene arrangements among the eight strains grouped by the *M/LH* analysis are shown. Two *V. cholerae* strains N2_17 and 87395 showed a similar gene arrangement that contained both *M* (common M5) and *LH* (LH4 or LH19). Strain 87395 was phylogenetically most closely related with strain HE-09, which exhibited M5. The strains 87395, N56, N79, N80, and N83 exhibited identical LH4 and clustered in a deep branch of the dendrogram shown in Figure 2.

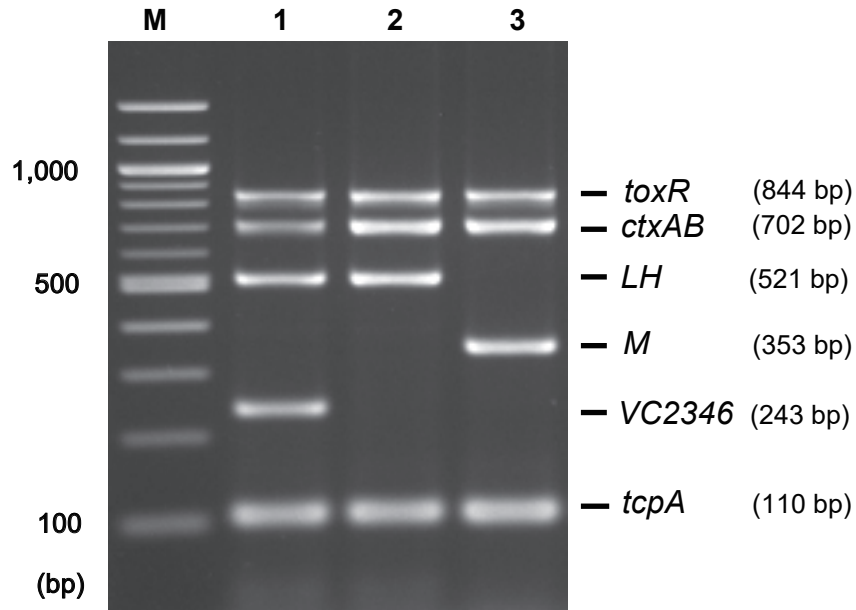


Figure S3. Multiplex PCR assay for rapid screening of six targeted genes including three virulence-related genes (*toxR*, *tcpA*, *ctxAB*), a seventh pandemic group-specific marker (*VC2346*), and *M/LH*. Lane M, 100-bp DNA ladder; lane 1, *V. cholerae* O1 El Tor strain N16961 (seventh cholera pandemic clone); lane 2, *V. cholerae* O1 classical strain 569B; lane 3, *V. cholerae* O1 El Tor strain MS6 (non-epidemic strain).