Mutation Landscape of Base Substitutions, Duplications, and Deletions in the Representative Current Cholera Pandemic Strain

Wen Wei[1,2,†], Lifeng Xiong[3,†], Yuan-Nong Ye[1,4,†], Meng-Ze Du[1], Yi-Zhou Gao[1], Kai-Yue Zhang[1], Yan-Ting Jin[1], Zujun Yang[1], Po-Chun Wong[3], Susanna K.P. Lau[3,5], Biao Kan[6], Jun Zhu[7], Patrick C.Y. Woo[3,5,*], and Feng-Biao Guo[1,8,9,*]

[1]School of Life Science and Technology, University of Electronic Science and Technology of China, Chengdu, China

[2]School of Life Sciences, Chongqing University, China

[3]Department of Microbiology, Research Centre of Infection and Immunology, State Key Laboratory of Emerging Infectious Diseases, and Carol Yu Centre for Infection, The University of Hong Kong, China

[4]Bioinformatics and Biomedical Bigdata Mining Laboratory, Key Laboratory of Biology and Medical Engineering, Guizhou Medical University, Guiyang, China

[5]Collaborative Innovation Center for Diagnosis and Treatment of Infectious Diseases, The University of Hong Kong, China

[6]National Institute for Communicable Disease Control and Prevention, State Key Laboratory of Infectious Disease Prevention and Control, Beijing, China

[7]Department of Microbiology, Perelman School of Medicine, University of Pennsylvania

[8]Center for Informational Biology, University of Electronic Science and Technology of China, Chengdu, China

[9]Key Laboratory for NeuroInformation of the Ministry of Education, University of Electronic Science and Technology of China, Chengdu, China

[†]These authors contributed equally to this work.

***Corresponding authors:** E-mails: fbguo@uestc.edu.cn; pcywoo@hku.hk.

Accepted: July 27, 2018

Data deposition: This project has been deposited at the National Center for Biotechnology Information under the accession PRJNA383052.

Abstract

Pandemic cholera is a major concern for public health because of its high mortality and morbidity. Mutation accumulation (MA) experiments were performed on a representative strain of the current cholera pandemic. Although the base-pair substitution mutation rates in *Vibrio cholerae* (1.24×10^{-10} per site per generation for wild-type lines and 3.29×10^{-8} for mismatch repair deficient lines) are lower than that previously reported in other bacteria using MA analysis, we discovered specific high rates (8.31×10^{-8} site/generation for wild-type lines and 1.82×10^{-6} for mismatch repair deficient lines) of base duplication or deletion driven by large-scale copy number variations (CNVs). These duplication–deletions are located in two pathogenic islands, IMEX and the large integron island. Each element of these islands has discrepant rate in rapid integration and excision, which provides clues to the pandemicity evolution of *V. cholerae*. These results also suggest that large-scale structural variants such as CNVs can accumulate rapidly during short-term evolution. Mismatch repair deficient lines exhibit a significantly increased mutation rate in the larger chromosome (Chr1) at specific regions, and this pattern is not observed in wild-type lines. We propose that the high frequency of GATC sites in Chr1 improves the efficiency of MMR, resulting in similar rates of mutation in the wild-type condition. In addition, different mutation rates and spectra were observed in the MA lines under distinct growth conditions, including minimal media, rich media and antibiotic treatments.

Key words: *Vibrio cholerae*, mutation rate, pathogenic island, replication timing.

Introduction

Vibrio cholerae is the causative agent of the acute life-threatening diarrheal disease cholera. Since the early 19th century, seven cholera pandemics have originated from serogroup O1 clones (Kaper et al. 1995). Although both the most recent pandemics were caused by the El Tor biotype, which is a derivative of the classical O1 serotype, the seventh pandemic strain did not directly originate from the sixth pandemic strain but from a nonpathogenic strain (Hu et al. 2016). The acquisition of the *V. cholerae* toxin (CTX) prophage was a major event in the evolution of the current pandemic strains (Mutreja et al. 2011; Shah et al. 2014). A modest population change in the seventh pandemic strain occurred from wave 1 to wave 2 and wave 3, which led to CTXϕ phage recombination (Mutreja et al. 2011; Banerjee et al. 2014).

The genome of *V. cholerae* has two circular chromosomes, a primary chromosome (Chr1) of 3 Mbp and a secondary chromosome (Chr2) of 1 Mbp. Despite the distinct lengths, the replication of these chromosomes is synchronized during cell division, with Chr1 replicating before Chr2 (Egan et al. 2004). Chr2 is replicated at a later stage, and more importantly, its genes are expressed less than those of Chr1 during the exponential phase (Couturier and Rocha 2006). Therefore, Chr2 is expected to be under relaxed purifying selection, and to evolve more rapidly (Cooper et al. 2010; Morrow and Cooper 2012). However, the incorporation of mutations and recombination on Chr1 played an important role in the transition of the nonpathogenic strain into the seventh pandemic strain (Hu et al. 2016). Moreover, several subtle changes in CTX are considered to contribute the global dissemination of the disease (Safa et al. 2010; Kim et al. 2015). Most of the seventh pandemic strains harbor various arrays of CTX on Chr1, whereas wave 2 strains harbor atypical CTX on Chr2 (Lee et al. 2009; Chin et al. 2011; Kim et al. 2014). Accurately characterizing different types of mutations, especially for these highly dynamic elements, is critical to understand the mechanisms underlying the current changes in *V. cholerae* populations.

A comparison of the whole-genome sequences of historical *V. cholerae* strains can provide important information for understanding the evolution of the current cholera pandemic strains; however, the disappearance of many previous strains results in incomplete information when comparing historical strains. Here, we apply the unbiased mutation accumulation (MA) strategy to characterize the rate and molecular spectrum of genomic mutations accumulated by 84 lines of *V. cholerae* N16961, a representative wave 1 strain that was the first whole-genome-sequenced *V. cholerae* isolate. These MA experiments can provide clues about the direction of population changes and the evolution of new variants in strains from the current pandemic.

Materials and Methods

Bacterial Strains and Growth Conditions

The bacterial strains, plasmids and primers used in this study are listed in supplementary table S1, Supplementary Material online. All *V. cholerae* strains were derived from N16961. The *V. cholerae* and *E. coli* strains were grown on LB/LB agar (LBA) plates (Difco, BD) or defined M9 minimal medium (6.8 g/l Na_2HPO_4, 3 g/l KH_2PO_4, 0.5 g/l NaCl, 1 g/l NH_4Cl) and supplemented with 0.1 mM $CaCl_2$, 2 mM $MgSO_4$ and 0.5% glucose as the carbon source (minimal glucose medium). All bacterial cultures were grown at 37°C. The nonpolar deletion of the *mutL* gene on parental strain N16961 was constructed by homologous recombination using the suicide plasmid pCVD442 as described previously (Xiong et al. 2014, 2015). Briefly, the in-frame deletion pattern of *mutL* containing its 5′- and 3′-flanking regions was generated with the overlap PCR method and subcloned into the suicide vector pCVD442. The resulting plasmid was transferred into N16961 by bacterial conjugation with the *E. coli* SM10 λpir strain. Selection of the allelic replacement was performed as described previously (Xiong et al. 2017), and the mutant strain was confirmed by PCR using the primers LPW28811/28814 and inner primers LPW28815/28816, which are specific for the deleted sequence. All Δ*mutL* mutant strains were confirmed by PCR products using Sanger sequencing (ABI 3700).

Mutation Accumulation Protocol

A single colony of the wild-type (*WT*) and the Δ*mutL* (*L*) mutant strain were picked and grown overnight in LB medium, respectively. The culture was serially diluted and spread on an LB plate and incubated at 37 °C for 24 h, at which point 40 *WT* and 20 *L* MA lines a were initiated, each from a separate well-isolated colony. MA lines were passaged daily from a single randomly chosen colony. Each of the *WT* lines had been passaged for 210 days, and the *L* mutant line had been passaged for 30 days (supplementary fig. S1, Supplementary Material online).

To investigate the accumulated mutations in the pandemic strains in response to poor growth pressure, we transferred 30-day evolved *L* lines into minimal glucose media. Among the 20 *L* lines, 14 were randomly chosen. We picked a representative colony for each 14 *L* line and had performed the daily MA process on minimal glucose media for 30 days (supplementary fig. S1, Supplementary Material online). Ten of the 20 *L* lines were also randomly chosen for further rifampicin treatment. An overnight culture of a single colony was serially diluted and plated onto LBA containing 100 μg/ml of rifampicin. The rifampicin-resistant colonies were picked up and further streaked out on LBA plates with increasing concentrations of rifampicin (100, 200, and 400 μg/ml) for 10 days' passaging (supplementary fig. S1, Supplementary Material online).

Estimation of Daily Generations and Total Number of Generations

The generation time of each line was estimated by calculating the number of viable bacterial cells after 24 h incubation on LBA plates or minimal glucose media. From each line, four randomly selected colonies were inoculated in PBS, serially diluted, and spread plated. The plates were incubated for 24 h and the number of generations was calculated from the average number of viable cells of the four colonies. To estimate total number of generations over the duration of the MA experiment, we measured daily generations bimonthly for *WT* lines, and biweekly for *L*, *Lm* and *Lrif* lines. The total number of generations (*G*) was calculated from the equation $G = \sum d*g$, where g is the average of daily generations of MA lines in each measurement and d is the elapsed days before the next measurement. We used the sum of generations elapsed between each measurement over the course of MA study to calculate total number of generations. The generation numbers were 5, 985, 855, 726, and 283 for *WT*, *L*, *Lm*, and *Lrif* lines, respectively.

Genome-Wide Sequencing and Mutation Calling

Genomic DNA was extracted from 5 ml of overnight cultures using the Genomic DNA Buffer Set (QIAGEN). DNA concentration and purity were evaluated with a NanoDrop ND-1000 Spectrophotometer (Thermo Fisher Scientific). Next-generation sequencing for 85 samples (84 MA lines and a wild-type ancestor) was performed at the Beijing Genome Institute (BGI) using the Illumina 2500 platform (2 × 125 bp).

Adapters and low-quality sequences were cleaned by Trimmomatics (Bolger et al. 2014). Cleaned reads from each sample were mapped to the reference genome (GenBank accession: NC_002505 and NC_002506) with BWA (v0.5.9) (Li and Durbin 2010). On average, 99.92% of the chromosome positions in the reference genome were covered by the MA lines. We then removed duplicated reads and sorted the remaining reads with SAMtools (v0.1.19) (Li 2011).

The VarScan tool (Koboldt et al. 2012) was used to call candidate base-pair substitutions (BPSs) and small indels. The candidate variants with read depths <10 and phred scores <15 were removed. The putative mutations in the *WT/L* lines were determined using three additional steps: 1) variants at a given nucleotide position had to be supported by 80% of the reads covering the position, with two or more reads on each strand; 2) ancestral variants were identified from the differences between the reference genome and the wild-type ancestor. The putative mutations were compared against the ancestral variants; and 3) consensus variant calls present in more than one *WT/L* line were removed. The novel mutations in each *Lm* and *Lrif* line distinguished these lines from the original *L* line.

Overall BPS and indel mutation rates were calculated for each MA line using the equation: $\mu = m/(T*n)$. Here, μ denotes the mutation rate, m represents the number of BPSs or indels observed, T is the total number of generations elapsed per lineage, and n is the number of ancestral sites covered by sequencing reads with an average phred scores >15.

Large Duplication and Deletion Calling

A sliding window approach was adopted by sliding 500 bp windows in 250 bp increments. The number (depth) of reads in each window was normalized by the mean read depth of windows for a given MA line. We checked these normalized depths of windows for the MA ancestor and did not find a successive window of >2 kb that was significantly deviated from 1, suggesting that duplication and deletion did not occur in our MA seed. We used the normalized depth of each window in each MA line divided by this depth of the corresponding window in the ancestor to estimate the pairwise coverage ratios between all pairs of MA lines and the ancestor. Windows without variation supporting a duplication or deletion were filtered out. The cutoff of the deletion depth of the coverage ratio was 0.5, whereas the duplication depth of the coverage ratio was 1.5. To detect large-scale variation events, we required the duplication or deletion of a region of at least 5 kb in the MA experiments. These predictions were also validated by the CNV-seq tool (Xie and Tammi 2009) with a window size of 500 bp and a step size of 250 bp, indicating that high-confidence regions supporting duplication and deletion were obtained. Intra and interchromosomal exchanges were identified using breakdancer (Chen et al. 2009) with the requirement of a minimum depth of 10× and variant frequency of >10%.

For each MA line, overall duplication and deletion mutation rates were estimated using the equation: $\mu = m/(T*n)$. Here, μ represents the mutation rate, m represents the number of nucleotides that have been duplicated or deleted, T is the total number of generations propagated, and n is the number of ancestral sites covered by sequencing reads with an average phred scores >15.

Availability of Sequenced Data

All sequencing data that support this study have been deposited in the National Center for Biotechnology Information (Bioproject: PRJNA383052). These *V. cholerae* raw data are available under the following biosamples: SAMN06768116 (wild-type ancestor), SAMN06768351 (*WT* lines), SAMN06818706 (*L* lines), SAMN06806367 (*Lm* lines), and SAMN06820431 (*Lrif* lines).

Results

Spectra of Mutations

A clone of the MA ancestral seed was deeply sequenced to 291× to confirm that it belonged to the seventh pandemic

strain N16961. The small number of differences in the MA ancestral seed do not significantly change the characters associated with the seventh pandemic (supplementary notes, Supplementary Material online). We first performed two separate MA experiments in which wild-type lines (40 *WT* lines) and mismatch repair (MMR)-deficient lines (20 *L* lines) were propagated on LB agar for 7 months and 1 month, respectively (table 1). All MA lines were deep sequenced (supplementary table S2, Supplementary Material online). We identified 120 base-pair substitutions (BPSs) in the *WT* lines (supplementary table S3, Supplementary Material online), yielding a genome-wide BPS mutation rate of 1.24×10^{-10} per nucleotide per generation (fig. 1*A*). Moreover, 11 small indels were identified (supplementary table S4, Supplementary Material online), giving a small indel mutation rate of 1.14×10^{-11} per nucleotide per generation (fig. 1*A*). The mutation rate of pathogenic *V. cholerae* is nearly identical to that previously reported in nonpathogenic strains of *V. cholerae* (BPS rate: 1.07×10^{-10} and indel rate: 1.71×10^{-11}) (Dillon et al. 2017). This result suggests that pathogenic and nonpathogenic strains, at least in *V. cholerae*, may have similar population dynamics according to the drift-barrier hypothesis (Sung et al. 2012). The mutation rate of *V. cholerae* is slightly lower than that of other bacteria in previously reported MA analysis (Wei et al. 2014; Dillon et al. 2015; Long et al. 2015; Sung et al. 2015; Dettman et al. 2016).

The deletion of a core gene in the MMR system can reveal mutations generated by replicative polymerase not found in the *WT* lines (Foster et al. 2015). In total, 2,272 BPSs (fig. 1*A* and supplementary table S5, Supplementary Material online) and 192 small indels (supplementary table S6, Supplementary Material online) accumulated in the *L* lines. The observed BPS rate (per nucleotide per generation) of the *L* lines was $\sim 265 \times$ greater than the rate of the *WT* lines (Wilcoxon rank sum test, $P = 3.62 \times 10^{-10}$), and the indel mutation rate was $\sim 245 \times$ greater ($P = 1.04 \times 10^{-11}$). The BPS spectrum of the *L* lines was dominated by transitions (fig. 1*B*), significantly more than in the *WT* lines ($\chi^2 = 510.49$, $P < 2.2 \times 10^{-16}$).

The core *V. cholerae* genome accumulates approximately 3.3 BPSs per year (Mutreja et al. 2011), similar to the average BPS rate of the *WT* lines over the course of this experiment. An average of 113.6 BPSs accumulated in the *L* lines over the course of 1 month is equivalent to approximately three decades of mutation in a natural genome. A multi-locus sequence typing (MLST) analysis (Jolley and Maiden 2010) of *V. cholerae* indicated that the BPSs accumulated in the MA lines did not lead to any MLST switching (fig. 1*C*). Previous studies have shown that population changes of *V. cholerae* have been highly dynamic and unusual among bacterial pathogens (Banerjee et al. 2014); however, the MLST analysis results indicate that phenotypic evolution is an inherently slow process, at least in the absence of strong selection.

Different Replication Fidelities and Errors between Chromosomes

Based on the number of ancestral sites observed in MA lines, the expected BPS percentage for Chr1 is 73.4% (supplementary table S2, Supplementary Material online). The BPS between Chr1 (1,722) and Chr2 (550) in the MMR-defective lines was significantly greater than expected percentage ($\chi^2 = 6.58$, $P = 0.01$). However, this percentage in the *WT* lines (Chr1: 88 vs. Chr2: 32) was not significantly different from expected ($\chi^2 = 0.0004$, $P = 0.984$), suggesting that Chr1 has a higher innate rate of mutational lesions than Chr2, because the major mechanism underlying replication error repair has been deleted in the *L* lines.

The MMR machinery searches for an available GATC motif, which is the nearest hemimethylated GATC site located on the replication origin-distal site of the mismatch, to initiate removal of the nascent strand containing the mismatch (Wang and Hays 2004, 2006). Chr1 has more GATC sites than Chr2 (Chr1: 14, 205 vs. Chr2: 4, 763, $\chi^2 = 20.96$, $P = 4.67 \times 10^{-6}$); thus, MMR is expected to have higher repair efficiency on Chr1 in the *WT* lines. Mutations on Chr1 in the *L* lines were closer to the nearest origin-distal GATC motif than expected at random (fig. 2*A*, Monte Carlo simulation, $P = 0.004$), whereas the Chr2 mutations were not (fig. 2*A*, $P = 0.832$). These results are consistent with previous studies reporting that the efficacy of MMR is reduced with increasing distance to the nearest available GATC site (Hasan and Leach 2015). The majority of mutational lesions in the vicinity of GATC sites may be repaired by the MMR system; mutations are randomly distributed around GATC sites in the *WT* lines (fig. 2*B*).

Chr1 is replicated earlier and expressed more highly than Chr2; thus, it has experienced greater purifying selection and tends to be more highly conserved (Cooper et al. 2010; Morrow and Cooper 2012). However, more mutations accumulated on Chr1 in the MMR-defective lines. We speculate that the greater density of GATC sites on Chr1 enhances the efficiency of MMR repair, thereby reducing the Chr1 mutation rate.

Context-Dependent Mutations Contribute the Difference of Mutation Rates between Two Chromosomes

We investigated 64 possible genomic triplets with respect to the 5′ and 3′ sites of the permutations between Chr1 and Chr2. The triplets of two chromosomes displayed a similar mutation pattern in the *L* lines (supplementary table S7, Supplementary Material online, Pearson's correlation, $r = 0.820$, $P < 2.20 \times 10^{-16}$). The observed BPSs of three triplets on Chr1, including 5′G[C→X]C3′ ($\chi^2 = 5.855$, $P = 0.016$; X is any other nucleotide), 5′T[G→X]G3′ ($\chi^2 = 3.880$, $P = 0.049$) and 5′G[A→X]G3′ ($\chi^2 = 3.865$, $P = 0.049$), were significantly greater than those on Chr2, relative to the expected difference of the triplet counts between two

Table 1
Summary of MA Lines

Lines	No. of Lines	Generations Per Line	Average Coverage (%)	Average Depth	No. of BPSs	BPS Rate	No. of Indels	Indel Rate
WT	40	5,985	99.97	176×	120	1.24×10^{-10}	11	1.14×10^{-11}
L	20	855	99.86	196×	2,272	3.29×10^{-8}	192	2.79×10^{-9}
Lm	14	726	99.91	296×	193	4.70×10^{-9}	31	7.57×10^{-10}
Lrif	10	279	99.87	307×	558	4.90×10^{-8}	75	6.58×10^{-9}

chromosomes. Furthermore, three of four R[R/Y→X] R/Y triplets on Chr1 (supplementary table S8, Supplementary Material online, R is A or G and Y is C or T), including 5′R[R→X] R3′ ($\chi^2 = 5.377$, $P = 0.020$), 5′R[Y→X] Y3′ ($\chi^2 = 5.005$, $P = 0.025$) and 5′R[Y→X]R3′ ($\chi^2 = 4.124$, $P = 0.042$), differed significantly from those triplets on Chr2. There was no significant difference for the four Y[R/Y→X] R/Y triplets. The 5′ neighboring nucleotide A or G may drive higher mutation rates on Chr1. The frequency of 5′S[S→X] S3′ on Chr1 was also significantly higher than that on Chr2 (supplementary table S9, Supplementary Material online, $\chi^2 = 7.495$, $P = 0.006$; S is C or G). However, in no case was the context-specific mutation rate of Chr1 higher than Chr2 in the *WT* lines (supplementary tables S10–S12, Supplementary Material online). This could be due lack of statistical power resulting from the small number of mutations in the *WT* lines. Taken together, our observations suggest that asymmetric contextual mutation patterns affect variation in mutation rate between two chromosomes, but only in the absence of MMR.

Mutation Rate and Replication Timing

Both Chr1 replichores were divided into 30 windows of 50 kb and the two Chr2 replichores were divided into ten windows of 50 kb. We combined windows that were equidistant from the *oriC* on the two opposing replichores. Thus, 30 combined windows of 100 kb for Chr1 and ten combined windows of 100 kb for Chr2 were used in the following analysis. Because almost all sites were covered by sequencing reads that passed filters (>99.9%), we used the number of sites as the denominator to estimate the mutation rate for each combined window. The mutation rates of the *L* lines were not uniformly distributed along either chromosome ($\chi^2 = 110.96$, $P = 7.95 \times 10^{-9}$, df = 39). As shown in figure 3*A*, significant mutation peaks were observed near the replication termini of Chr2 ($P < 0.05$) and around the replication midpoint for Chr1 ($P < 0.05$). Chromosome replication arrests near the replication termini (Mirkin and Mirkin 2007; Stamatoyannopoulos et al. 2009), which could result in an increased risk of exposure as single-stranded DNA and lead to elevated endogenous DNA damage (Lindahl 1993; Yang et al. 2008). Mutation rates near the Chr1 replication termini also increased, but not significantly ($P < 0.1$).

The replication midpoint on Chr1 includes regions that regulate replication initiation, such as the *crtS* locus (Val et al. 2016). Initiation of Chr2 replication is triggered when the replication fork approaches the *crtS* locus (Val et al. 2016). Moreover, as a result of a requirement for rapid growth, another round of Chr1 replication starts when the replication fork approaches the midpoint (Srivastava and Chattoraj 2007; Stokke et al. 2011). The number of replication forks increases from 4 to 16 on LB medium at this end of Chr2 replication initiation. According to the model of replication timing of *V. cholerae* growing on LB medium (Stokke et al. 2011), the 12 novel replication forks may take 2–3 min to be initiated, during which time the movement of the replication fork(s) is arrested. These complex initiation events potentially elevate mutation rates around the replication midpoint of Chr1, because they slow down the local replication efficiency and cause intrinsic damage from replication fork arrest.

Several known intrinsic factors, including differential expression between chromosomes, affect mutation rates. Our analyses show that none of these factors can explain the different mutation rates between the two *V. cholerae* chromosomes (supplementary notes, Supplementary Material online). However, the difference in mutation rates between chromosomes disappears when the Chr1-related peak is excluded ($\chi^2 = 0.350$, $P = 0.554$). Thus, increased mutation rates in Chr1 of the *L* lines are associated with the region around the replication midpoint. Several GATC hotspot regions are located near the replication midpoint of Chr1 (supplementary fig. S2*A*, Supplementary Material online). These GATC hotspots may improve the capacity of MMR to repair mutations in the region around the Chr1 replication midpoint. We did not observe a prominent peak of mutation rate around this region in the *WT* lines (fig. 3*B*). Mutation rates of the *WT* lines are uniformly distributed along both chromosomes ($\chi^2 = 38$, $P = 0.515$, $\nu = 39$). Because power is lower due to the smaller number of mutations in the *WT* lines, we used a greater BPS number to reestimate variation in mutation rates within chromosomes in the *WT* lines. Using the observed BPS number of *L* lines (2,272 BPSs), the expected BPS numbers of the two chromosomes were calculated from the ratio observed in the *WT* lines (Chr1: 1,666 and Chr2: 606). The prominent peak of mutation rate around the replication midpoint of Chr1 was not observed (supplementary fig. S2*B*, Supplementary Material online), suggesting that power is

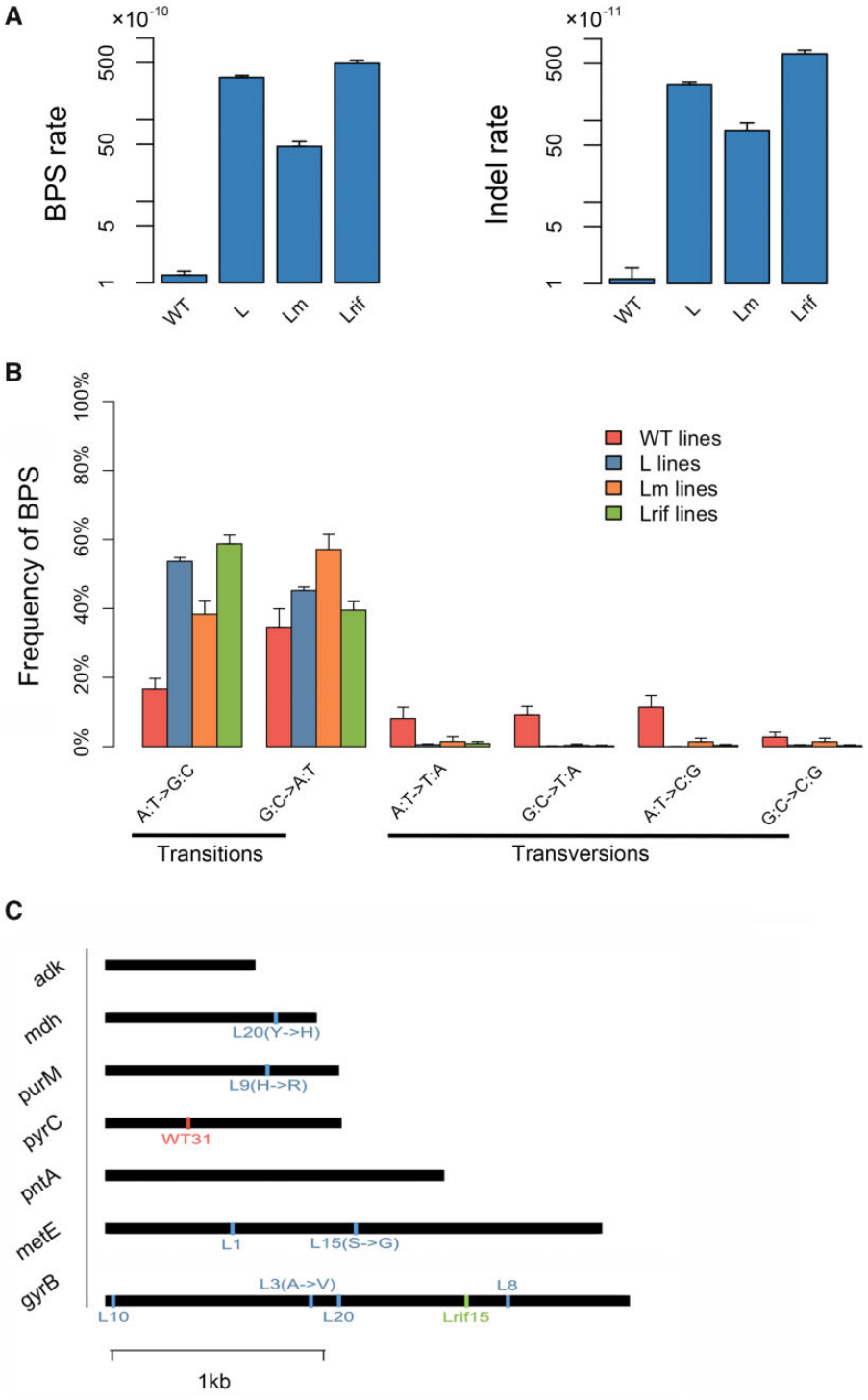

Fig. 1.—Base-pair substitution (BPS) and small insertion–deletion (indel) mutations for *V. cholerae* lines. Four separate mutation accumulation (MA) experiments were performed. Three experiments were propagated independently on LB medium and consisted of wild-type lines (40 *WT* lines) and Δ*mutL* lines treated without (20 *L* lines) or with rifampicin (10 *Lrif* lines). The fourth experiment generated a set of MA lines cultured on minimal glucose medium (14 *Lm* lines). (*A*) Overall BPS and indel mutation rates per base-pair per generation. The bars denote the mean mutation rates for four MA experiments and the error bars show the standard error of the mean (SEM) for each experiment. (*B*) Frequency of BPS mutations of each transition and transversion. The bars denote the mean frequency of mutations for six BPS groups, and the error bars show the SEM for each group. (*C*) Mutation analysis on the seven housekeeping genes used for multilocus sequence typing (MLST). A total of ten sites were involved in nucleotide mutations in the seven housekeeping genes, four of which changed amino acids.

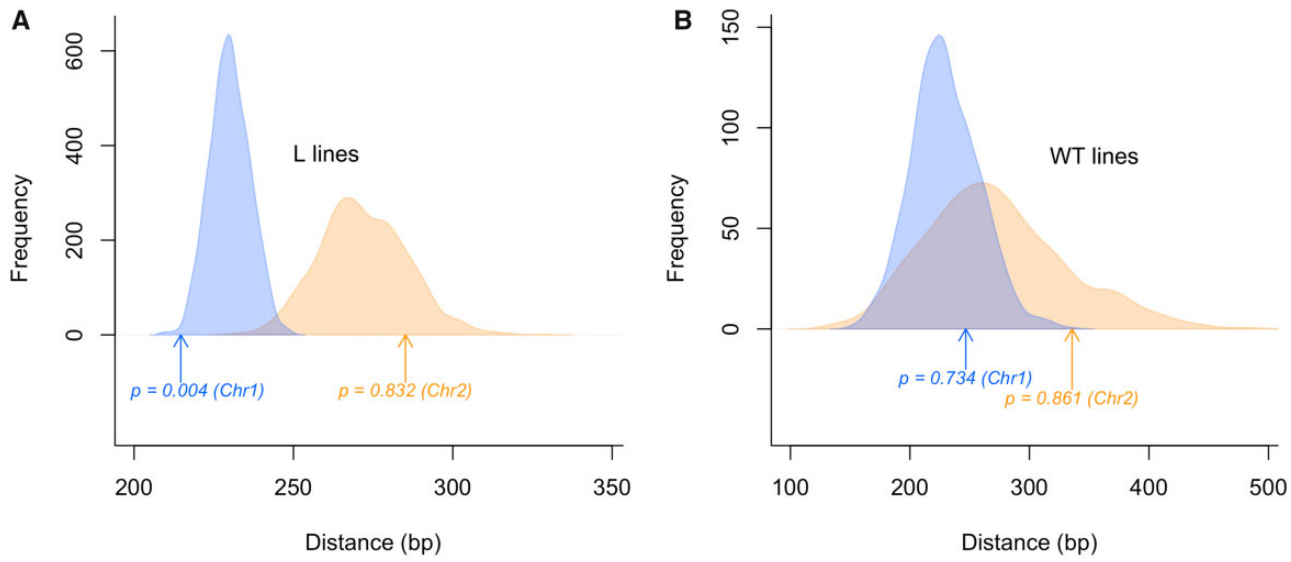

Fig. 2.—Distance between mutation and the first available GATC site for (*A*) *L* and (*B*) *WT* lines. Distribution of average random distance was estimated by Monte Carlo simulations with 10,000 replications. For each simulation, random mutation sites were picked according to a number of real mutation sites. Arrows denote the mean distance between real mutation sites and the origin-distal GATC site. *P* values denote a significance level at which the real distance is closer than random expectations with 10,000 simulations.

unlikely to explain the lack of higher mutation rate around the replication midpoint in the wild-type study. However, the later-replicating regions do have a remarkable increase of mutation rates (supplementary fig. S2*B*, Supplementary Material online). Mismatch repair system may improve its efficiency in early-replicating regions because the number of GATC site increases in the first half of the replichore, especially in early-replicating replication midpoint (supplementary fig. S2*A*, Supplementary Material online).

Rapid Duplication–Deletion in Genomic Islands

Few large duplication–deletion events have been observed in previous bacterial MA studies in unstressed populations because these events are rare in stable environmental conditions. Interestingly, a total of six and four large (>5 kb) genomic duplications were found in the *WT* and *L* lines, and three and four large deletions were accumulated in the *WT* and *L* lines (fig. 4*A* and supplementary table S13, Supplementary Material online). The duplication rate of the *L* lines (4.89×10^{-7} per site per generation) was $\sim 10\times$ greater than that of the *WT* lines (5.02×10^{-8}), and the deletion rate of the *L* lines (1.33×10^{-6}) was $\sim 40\times$ greater than that of the *WT* lines (3.29×10^{-8}). However, these differences between the two lines was smaller than the difference in the BPS and small indel events. All 17 of these duplication–deletions were located in two genomic islands (GIs): VCGI5 (=IMEX) and VCGI8.

VCGI5 is an array of prophages and includes protoxin-linked cryptic (TLC), CTX and RS1, which was integrated into Chr1 by the *Xer* site (Hassan et al. 2010; Banerjee et al. 2014). In this study, eight of the 40 (20.0%) *WT* lines and five of the 20 (25.0%) *L* lines accumulated duplications or deletions in VCGI5 (fig. 4*B* and supplementary table S13, Supplementary Material online). Prophage copies of TLC, CTX-RS1, and CTX increased in six, one and two MA lines, respectively, indicating that prophage recombination occurs frequently. New CTXϕ phages can be produced from the early integrated CTX prophage segments on the core chromosomes of the current pandemic strains (Moyer et al. 2001; Huber and Waldor 2002; Martinez et al. 2015), and they may also be integrated onto the chromosome using the original *Xer* recombination site (Huber and Waldor 2002). Our results suggest that CTX prophages as well as TLC and RS1 prophages are able to produce new viral particles and subsequent irreversible integration. Breakdancer (Chen et al. 2009) detected no intra or interchromosomal phage recombinations in the MA lines. Thus, these new phages appear to have used the *Xer* site to integrate into the genome. The average duplication ratio of CTX was 85.5% and the ratio of TLC was 173.5% in those duplication-gained lines. TLCϕ is almost always the first integrated element of an IMEX array, and its integration is a prerequisite for CTXϕ integration (Hassan et al. 2010). This dependent relationship may explain why the TLCϕ integration occurred more frequently than the CTXϕ integration in the MA lines.

Prophage deletion was less frequent than prophage integration. The prophage was deleted in only three MA lines, all of them involved in CTX loss. A joint loss of TLC-CTX was also observed in one of the three MA lines. The excision of CTX could be promoted by the *Xer* recombination reaction, which results in a joint excision of the TLC-CTX event

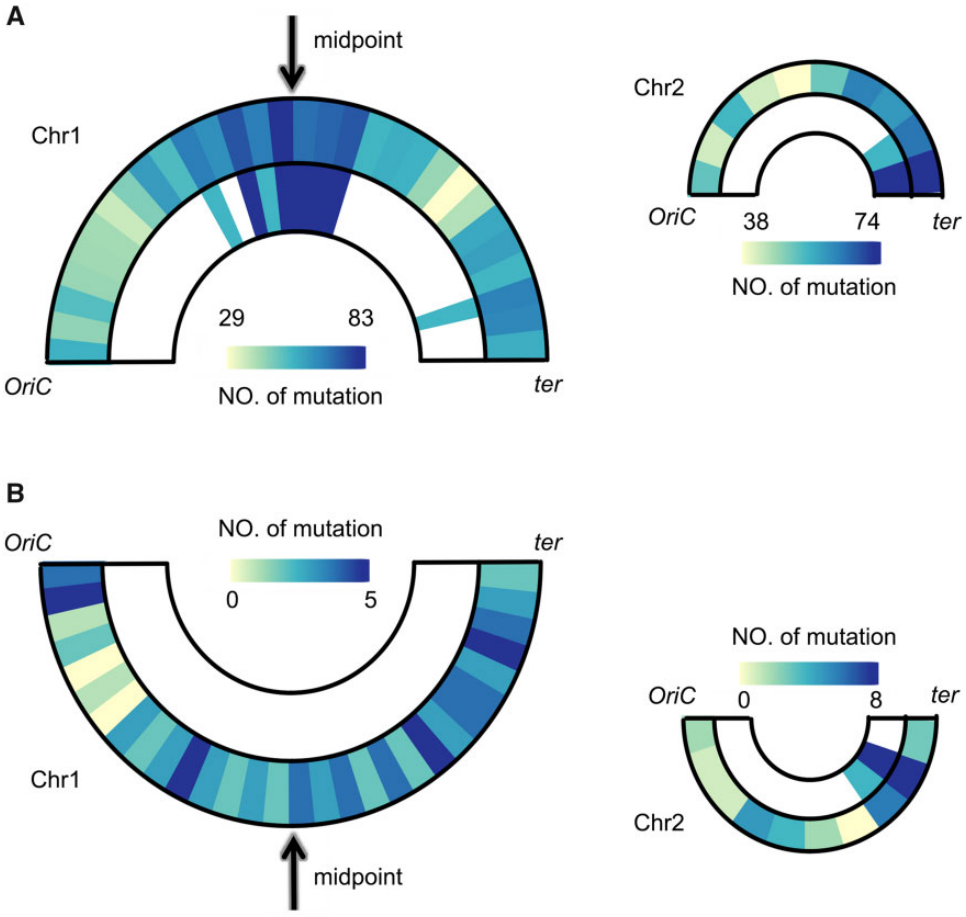

Fig. 3.—Relationship between mutation rates and replication timing in the (*A*) *L* and (*B*) *WT* lines. The replication *oriC* and *ter* loci were obtained from DoriC (Gao et al. 2013). Each replichore of each chromosome was divided into 50 kb windows. We combined windows that were equidistant from the *oriC* on the two opposing replichores. Finally, 30 and ten 100 kb windows were obtained for Chr1 and Chr2, respectively. The number of genomic sites was used as the denominator to estimate the BPS rate of each 100 kb window. The outer semicircle shows the BPS rate of each sliding window, and the colored frames in the inner semicircle denote a significance level at which the mutation rate is greater than random expectations with 10,000 Monte Carlo simulations (turquoise: $P < 0.1$, dark blue: $P < 0.05$).

(Midonet et al. 2014). Although we found a joint loss of TLC-CTX, the complete elimination of CTX (average loss ratio = 98.9%) and partial excision of TLC elements (49.9%) implies that this joint event may not be simply mediated by the single-step excision of TLC.

In addition to VCGI5, we also observed one large-scale duplication and three large-scale deletions in the large integron island VCGI8 (fig. 4*B* and supplementary table S13, Supplementary Material online) on Chr2. This island harbors 98 pathogen-associated genes, although there is no direct evidence that VCGI8 is responsible for the worldwide spread of the disease. In the *L* lines, two deletions of 60 kb and 21 kb and a 10 kb duplication were observed, which resulted in different large-scale mutational event in the pathogenicity island (PAI). In contrast, only one deletion of 8 kb was observed in the *WT* lines. VCGI8 homologs were present only in the five pandemic strains, suggesting that the PAI was recently acquired by pandemic-related *V. cholerae* (fig. 5*A*). Obvious differences were observed in the sequences of the VCGI8 homologs in these pandemic strains, indicating that complex recombination and deletion events had occurred (fig. 5*B*). Moreover, VCGI8 showed greater BPS rates in the *L* lines (fig. 5*C*). Therefore, the duplication–deletions in VCGI8 that were observed in the MA lines may result from the intrinsic instability of redundant integrons in this PAI.

The GIs of pathogenic bacteria originate from horizontal gene transfer and these integrated GIs could evolve rapidly for surviving and functioning in host genome (Dobrindt et al. 2004). Most of the GIs show similar BPS rates compared with random expectations (fig. 5*C*). The BPS mutation rate was much lower than the duplication–deletion mutation rate in the GIs. Phage and PAI transfers are important drivers of genomic diversification between closely related bacteria and allow for rapid adaptation to new hosts, which primarily occurs by reasserting the pathogenic-association factor repertoire. We suggest that rapid assemblage and elimination in

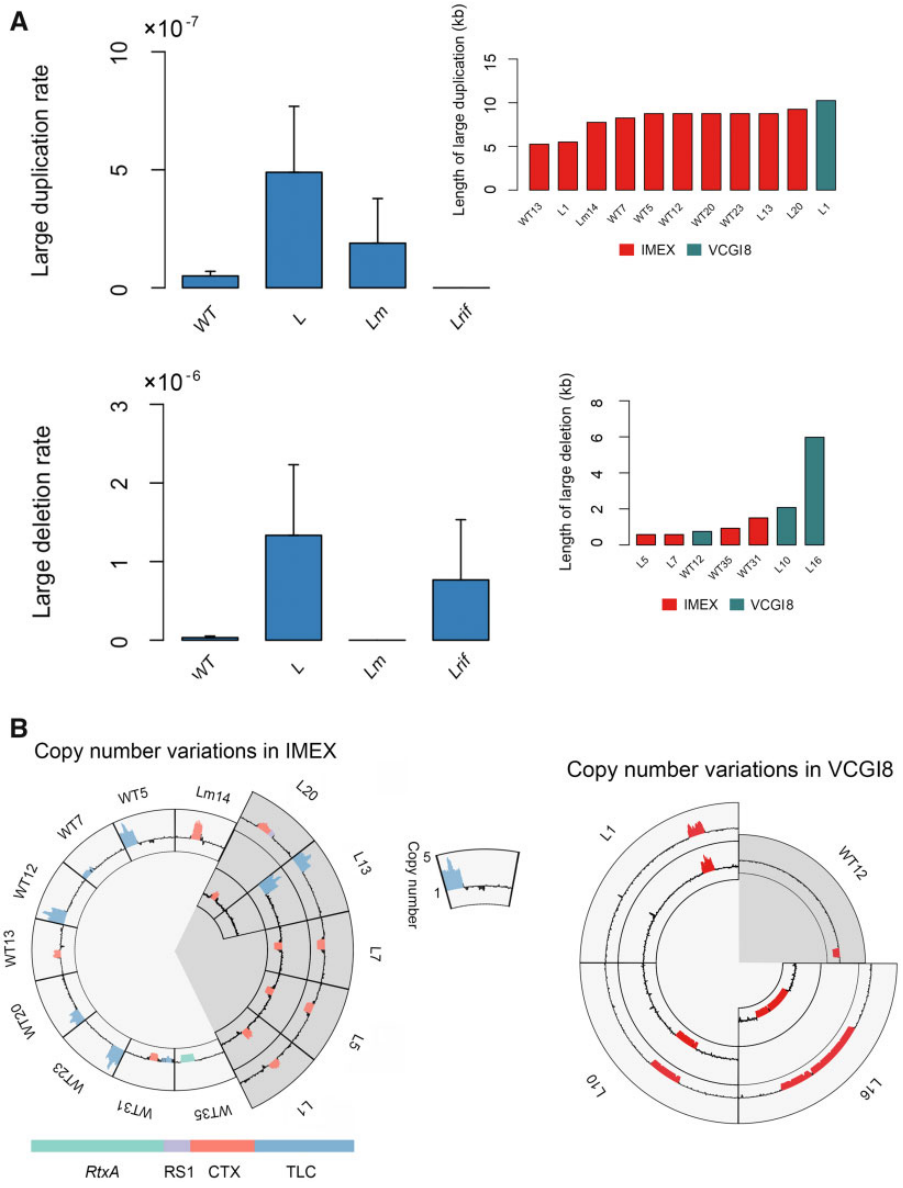

Fig. 4.—Large-scale duplication and deletion mutations in *V. cholerae* lines. Nine genomic islands (GIs) in the genome of *V. cholerae* N16961 were identified by integrated methods (Dhillon et al. 2015; Wei et al. 2017) (supplementary notes and supplementary table S14, Supplementary Material online). All large-scale mutations are significantly enriched on two GI regions. (*A*) Overall large-scale duplication and deletion mutation rates per site per generation (Left), and length of large-scale mutations in IMEX (red) and VCGI8 (dark green) (Right). The bars denote the mean mutation rates for four MA experiments, and the error bars show the SEM for each experiment. For the *Lm* and *Lrif* lines, the extra copies removed from the *L* line are not shown in the right chart. (*B*) Duplication–deletion pattern in IMEX (Left) and VCGI8 (Right). A sliding window approach comparing all pairwise groups of the MA lines and the ancestor was adopted using 500 bp windows and sliding with 250 bp increments. All windows containing a mutation are colored. Outer rings represent de novo variations in the MA line, whereas middle and inner rings represent copy number patterns corresponding to *Lm* and *Lrif* lines, respectively.

PAIs, at least in VCGI5 and VCGI8, rather than intrinsically low rates of BPS mutations allows for the constant acquisition of different PAIs and can even cause new genotype variants.

Lower Mutation Rates in Nutrient-Poor Media

We investigated the growth rate on LB medium both for the wild-type ancestor and 30th day colony of each *L* line, and observed that the previously accumulated mutations in the *L* lines did not significantly influence the growth rates (supplementary fig. S3*A*, Supplementary Material online, Wilcoxon rank sum test, all $P > 0.4$). To investigate the effects of these accumulated mutations, especially the large duplication–deletions, in the pandemic strains in response to novel growth conditions, we transferred *L* lines into minimal glucose medium. These *L* lines displayed very similar growth rates in the

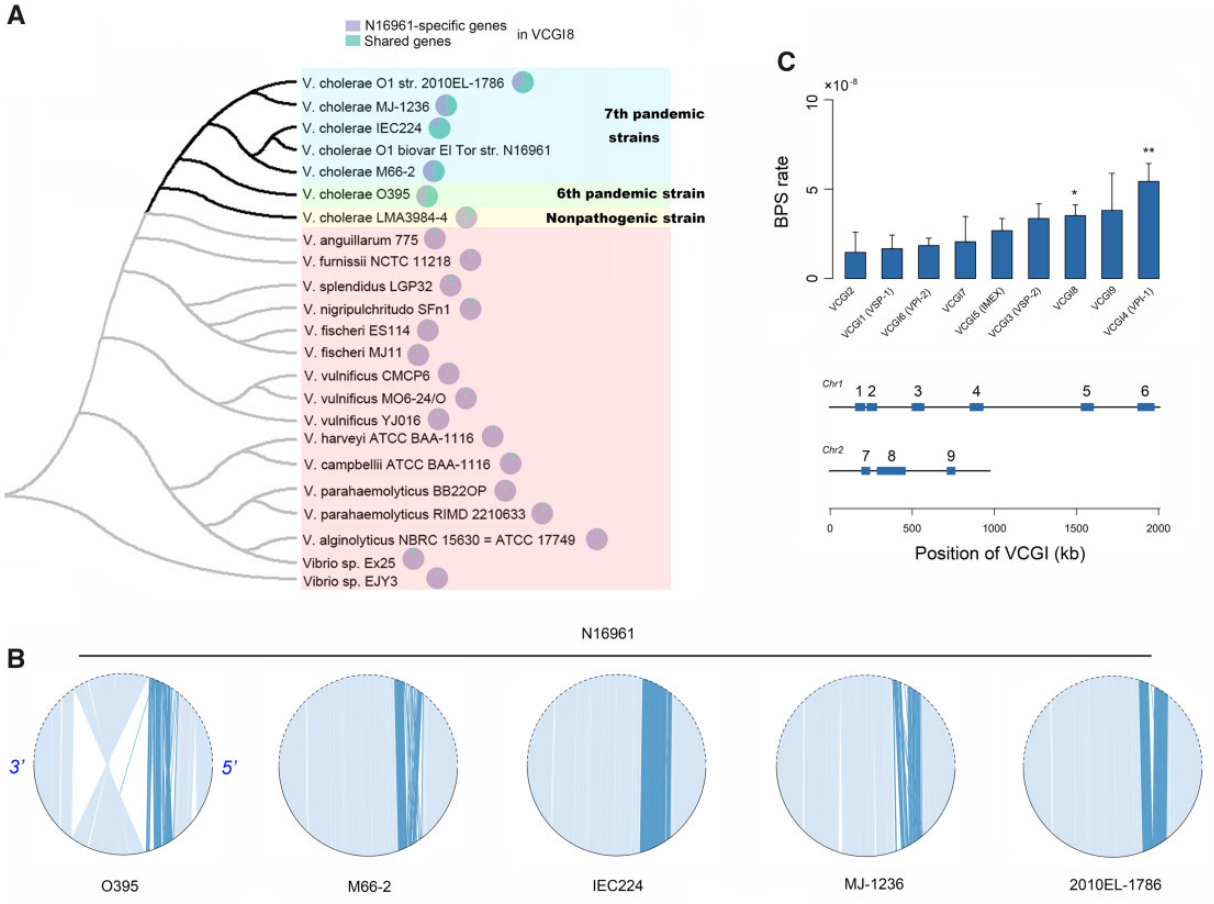

Fig. 5.—Alterations in genomic islands. (*A*) Evolutionary tree based on the whole-genome analysis by CVtree (http://tlife.fudan.edu.cn/cvtree/), which includes one sixth pandemic strain (O395), one seventh prepandemic-related strain (M66-2), three seventh pandemic strains of wave 2 (IEC224 and MJ-1236) and wave 3 (2010EL-1786), one nonpathogenic strain (LMA3984-4), and one strain from each of the other 16 *Vibrio* species. The pie chart represents the gene percentage between the N16961-specific genes (lavender) and the genes shared with the leaf species (light green). The PAI homologs, which share at least 40% of the VCGI8 genes, were only observed in the five pandemic-related strains. Shared genes were identified using the best reciprocal hit (RBH) from a genome-against-genome BLASTP search with a cut-off of *E*-value $< 10^{-6}$, identity $> 30\%$ and coverage $> 80\%$. (*B*) Gene alignments of N16961 and five pandemic strains. The upper semicircle is the small chromosome of N16961, whereas the lower semicircles are for Chr2 of the five other pandemic strains. The alignments of VCGI8 are marked by dark blue. (*C*) BPS rates of nine VCGIs per site per generation in the *L* lines (top), and genomic positions (bottom) of these VCGIs. The bars denote the mean mutation rates for the GIs and the error bars show the SEM for each GI. Asterisks denotes significance levels at which the mutation rate is greater than random expectations, which was estimated by randomly selecting mutation sites in accordance with the number of BPS mutations in these lines with 10,000 Monte Carlo simulations ($*P < 0.05$, $**P < 0.01$).

novel condition (24–25 generations per day). A total of 14 randomly selected *L* lines were passaged on the minimal glucose medium for an additional 30 days of MA (*Lm* lines, supplementary table S2, Supplementary Material online). The generation time did not differ between the 1st and 30th day of *Lm* lines (supplementary fig. S3*B*, Supplementary Material online, Wilcoxon rank sum test, all $P > 0.4$).

Of the 1,478 BPSs previously accumulated on the 14 selected *L* lines, ten BPSs were not detected in the *Lm* lines. Although the ancestors of the *Lm* lines and the corresponding sequenced *L* lines both originated from a single clone of each 29th day *L* line, they are expected to accumulate several new BPSs from the divergence of 30th day (3.79 BPSs per day per line). We observed one, three and six missing BPSs in *Lm*1, *Lm*9, and *Lm*13, respectively (supplementary table S15, Supplementary Material online). These missing BPSs may result from the divergence of 30th day rather than strong selection or reverse mutation under new environment because daily generations of the three lines as compared with other *L* lines did not decline significantly when transferred into the poor medium (*L*1: 24.3, *L*9: 24.1, *L*13: 24.7, and others: average of 24.5 generations per day). A total of 193 novel BPSs and 31 small indels were accumulated (supplementary tables S16 and S17, Supplementary Material online). None of the

duplication–deletions originating in the *L* lines were absent in the *Lm* lines (fig. 4*B*). Furthermore, *Lm*14 gained a CTX prophage duplication (fig. 4).

Compared with the number of mutations accumulated in the *L* lines per day, the novel mutations observed per day in the *Lm* lines decreased markedly (BPS: $\chi^2 = 78.21$, $P < 2.2 \times 10^{-16}$; indel: $\chi^2 = 4.62$, $P = 0.03$). *L* lines (average generation of 28.5 divisions per day) have shorter generation times than *Lm* lines (average generation of 24.2 divisions per day). If most mutations are associated with DNA replication, the lower rate in the *Lm* lines could simply be a function of their having undergone fewer rounds of replication (Weller and Wu 2015). When scaled per-cell division rather than per-day, BPS (Wilcoxon rank sum test, $P = 1.44 \times 10^{-9}$) and indel rates ($P = 1.01 \times 10^{-6}$) are still lower in the *Lm* lines relative to the *L* lines (fig. 1*A*), which indicates that the lower mutation rate in the *Lm* lines is due to less error-prone rather than simply due to fewer rounds of DNA replication.

The ratio of BPSs between chromosomes does not differ significantly from the random expectation (Chr1: 147 vs. Chr2: 46, $\chi^2 = 0.614$, $P = 0.433$). Moreover, the effect of increasing mutation rates around replication midpoint was reduced. (supplementary fig. S4*A*, Supplementary Material online). The number and complexity of replication forks decrease during the cell cycle when the bacteria are growing on poor medium, thus reducing the competition of dNTPs among the different replication forks and decreasing the opportunity of replication forks to congest in the midpoint (Srivastava and Chattoraj 2007; Stokke et al. 2011). Although the *Lm* lines exhibit relatively low-efficiency growth, fewer replication errors were observed in the *Lm* lines.

Rifampicin-Induced Mutation Behaviors

Antibiotic treatment may induce an SOS response (Fijalkowska et al. 1997; Napolitano et al. 2000) and enhance mutations in bacterial pathogens (Long et al. 2016; Levin-Reisman et al. 2017). To estimate the mutation events of *V. cholerae* in response to antibiotic stress, we randomly treated ten *L* lines with a broad-spectrum antibiotic (rifampicin). The ten *L* lines were treated with sublethal concentrations of rifampicin for 10 days (*Lrif* lines, supplementary table S2, Supplementary Material online), after which they had accumulated 558 new BPSs and 75 new indels (supplementary tables S18 and S19, Supplementary Material online). Extra copies of TLC found in *L*13 were absent in the *Lrif*13 line (fig. 4*B*). The BPS rate of the *Lrif* lines was ~1.5 × greater than that of the *L* lines (fig. 1*A*, Wilcoxon rank sum test, $P = 0.002$). The BPSs of resistance protein (6) and nonresistance proteins (552) in the *Lrif* lines was similar to those in the *L* lines (13 vs. 2,259) ($\chi^2 = 1.030$, $P = 0.310$), indicating that antibiotic treatment enhances mutagenesis without resistance pressure. The BPSs of the PAI and non-PAI was similar between the two kinds of MA lines (*L* lines: 175 vs. 2,097, *Lrif* lines: 37 vs. 521, $\chi^2 = 0.596$, $P = 0.440$). Therefore, the antibiotic treatment does not particularly accelerate the mutagenesis of PAIs. Although the BPS mutations between the chromosomes in the *Lrif* lines resembled the rates in the *L* lines ($\chi^2 = 0.094$, $P = 0.759$), antibiotics decrease the relative high mutation rate around the Chr1 replication midpoint in the *Lrif* lines but increase it around the Chr2 replication midpoint (supplementary fig. S4*B*, Supplementary Material online).

Discussion

Spontaneous Variation and Genetic Diversity in Strains of the Current Cholera Pandemic

We systematically investigated the genome-wide mutational process in a strain from the seventh cholera pandemic. *Vibrio* populations have high diversity, which could be driven primarily by BPS mutation (Sawabe et al. 2009); however, we observed that the BPS mutation rate was much lower than the copy number variation (CNV) rate in *V. cholerae* MA lines. Moreover, the BPS mutation rates we observed were lower than those in previously reported bacteria (Wei et al. 2014; Dillon et al. 2015; Long et al. 2015; Sung et al. 2015; Dettman et al. 2016). High nucleotide diversity at silent sites in sequenced *V. cholerae* genomes suggests a large effective population size for this species (Sung et al. 2016), which may have led to increased power of natural selection to reduce the BPS mutation rate (Lynch et al. 2016).

Duplication/deletion mutations (or CNVs) are not randomly distributed in the *V. cholerae* genome. High rates of duplication/deletion mutations were observed on the two genomic islands (the IMEX array and the VCGI8). The IMEX array is highly variable in pathogenic *V. cholerae*, which is associated with global dissemination of the current cholera (Kim et al. 2015). These results suggest that large-scale structural variation such as CNVs may accumulate rapidly during short-term bacterial evolution. Large-scale integrations and deletions may be highly deleterious and subject to stronger purifying selection than BPS mutations, because large changes in copy number of genomic elements can affect transcriptional regulation and gene dosage (Itsara et al. 2010). However, phage and PAI transfer are important drivers of genomic diversification between closely related bacteria and allow rapid adaptation to new hosts, notably by reassorting the pathogenic-association factor repertoire. Gain of IMEX elements has been important in the origin, evolution, and transition to pandemicity of the seventh pandemic strains. We observed rapid accumulation of structural variations in prophages, including TLC, CTX, and RS1, which may provide genetic materials to improve evolution of IMEX elements. These mutational processes may explain that CTX prophage was eliminated in the wave 2 strains of the current cholera pandemic and IMEX elements was rearranged in the wave 3 strains of the current

cholera pandemic (Safa et al. 2010; Banerjee et al. 2014; Kim et al. 2015). CTX prophage harbors the principal cholera toxin genes, and its types are categorized primarily by *rstR* and *ctxB* genotype and further subdivided by BPSs in *rstA* and *rstB* genes (Kim et al. 2015). However, we did not observe any mutation in these four genes in all MA lines. The high rate of CNVs of IMEX elements, rather than the intrinsically low rate of BPS mutation, better explains the steady acquisition of variant IMEX elements.

Complex Genome Replication and Potential Effects on *V. cholerae* Evolution

The pattern of genome replication differs among *V. cholerae* cells with different growth rates, as well as between the two chromosomes (Stokke et al. 2011). These differences affect the mutational process both globally and locally. First, the global mutation rate increases when cells are grown on rich nutrient media (*L* lines), perhaps as a consequence of the greater complexity of replication forks. We hypothesize that competition for dNTPs among the different replication forks may increase the likelihood of endogenous mismatch damage, leading to a higher genome-wide mutation rate.

Second, the increased mutation rate on Chr1 is associated with replication of the region around the *crtS* locus, which functions in triggering the replication of Chr2. At high growth rates, at the same time the replication of Chr2 is trigged, a new round of replication of Chr1 begins (Stokke et al. 2011). The initiation of multiple replication forks is potentially time-consuming and could lead to elevated local mutations rates on Chr1. In addition, VPI-1, which is located in this mutation hotspot, experiences an increase in mutation rate (fig. 5C, Random test, $P = 0.0049$). The main gene clusters responsible for the pathogenicity of cholera are associated with the production of VPI-1 and the IMEX element CTX prophage. The VPI-1 carries genes encoding the synthesis of a type IV pilus, the bacterial structure involved in the reception of CTXϕ phage, and the colonization of the intestine of human host (Banerjee et al. 2014). The spontaneous high rate of BPS for VPI-1 and of CNV for CTX could provide a fascinating model of coevolution because of their associated functions in pandemic cholera.

Replication errors will increase in the replication midpoint of Chr1. Chr1 plays a more important role in maintaining the essential functions in *V. cholerae* as well as its pathogenicity (Cooper et al. 2010; Kim et al. 2015; Hu et al. 2016). Accordingly, Chr1 is under stronger purifying selection and evolves more slowly than Chr2. A high density of GATC sites at mutational hotspots represents a potentially effective method of increasing the efficiency of the MMR system. GATC sites are overrepresented on Chr1 and particularly near the Chr1 replication midpoint. The rates of mutation are similar between two chromosomes of the *WT* lines, consistent with a significant role of GATC sites in the modulation of mutation rate by MMR.

Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

Acknowledgments

This work was supported by the Program for New Century Excellent Talents in University (NCET-11-0059), Science Strength Promotion Programme of UESTC the National Natural Science Foundation of China (31470068, 31501063, and 31660320), the Fundamental Research Funds for the Central Universities of China (ZYGX2016J117 and ZYGX2015Z006), the Tianjin Natural Science Foundation of China (16JCQNJC09600), and the Postdoctoral Science Foundation of China (2015M580211).

Author Contributions

F.B.G. conceived the study and devised the experiments. W.W. performed the bioinformatics analyses and Y.N.Y. double-checked them. L.X. performed the wet experiments and P.C.Y.W. supervised them. W.W. and F.B.G. analyzed and discussed the results. M.Z.D., Y.Z.G., K.Y.Z., and Y.T.J. took part in downloading data and double-checking some parts of the results. P.C.W. assisted to perform the MA experiments. B.K. and J.Z. contributed bacterial strain and reagents. W.W. and F.B.G. wrote, and F.B.G, P.C.Y.W., L.X, S.K.P.L., and Z.Y. revised the paper. All authors approved the paper.

Literature Cited

Banerjee R, Das B, Balakrish Nair G, Basak S. 2014. Dynamics in genome evolution of *Vibrio cholerae*. Infect Genet Evol. 23:32–41.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30(15):2114–2120.

Chen K, et al. 2009. BreakDancer: an algorithm for high-resolution mapping of genomic structural variation. Nat Methods 6(9):677–681.

Chin CS, et al. 2011. The origin of the Haitian cholera outbreak strain. N Engl J Med. 364(1):33–42.

Cooper VS, Vohr SH, Wrocklage SC, Hatcher PJ. 2010. Why genes evolve faster on secondary chromosomes in bacteria. PLoS Comput Biol. 6(4):e1000732.

Couturier E, Rocha EP. 2006. Replication-associated gene dosage effects shape the genomes of fast-growing bacteria but only for transcription and translation genes. Mol Microbiol. 59(5):1506–1518.

Dettman JR, Sztepanacz JL, Kassen R. 2016. The properties of spontaneous mutations in the opportunistic pathogen *Pseudomonas aeruginosa*. BMC Genomics 17:27.

Dhillon BK, et al. 2015. IslandViewer 3: more flexible, interactive genomic island discovery, visualization and analysis. Nucleic Acids Res. 43(W1):W104–W108.

Dillon MM, Sung W, Lynch M, Cooper VS. 2015. The rate and molecular spectrum of spontaneous mutations in the GC-rich multichromosome genome of *Burkholderia cenocepacia*. Genetics 200(3):935–946.

Dillon MM, Sung W, Lynch M, Sebra R, Cooper VS. 2017. Genome-wide biases in the rate and molecular spectrum of spontaneous mutations in *Vibrio cholera* and *Vibrio fischeri*. Mol Biol Evol. 34(1):93–109.
Dobrindt U, Hochhut B, Hentschel U, Hacker J. 2004. Genomic islands in pathogenic and environmental microorganisms. Nat Rev Microbiol. 2(5):414–424.
Egan ES, Lobner-Olesen A, Waldor MK. 2004. Synchronous replication initiation of the two *Vibrio cholerae* chromosomes. Curr Biol. 14(13):R501–R502.
Fijalkowska IJ, Dunn RL, Schaaper RM. 1997. Genetic requirements and mutational specificity of the *Escherichia coli* SOS mutator activity. J Bacteriol. 179(23):7435–7445.
Foster PL, Lee H, Popodi E, Townes JP, Tang H. 2015. Determinants of spontaneous mutation in the bacterium *Escherichia coli* as revealed by whole-genome sequencing. Proc Natl Acad Sci USA. 112(44):E5990–E5999.
Gao F, Luo H, Zhang CT. 2013. DoriC 5.0: an updated database of oriC regions in both bacterial and archaeal genomes. Nucleic Acids Res. 41(Database issue):D90–D93.
Hasan AM, Leach DR. 2015. Chromosomal directionality of DNA mismatch repair in *Escherichia coli*. Proc Natl Acad Sci USA. 112(30):9388–9393.
Hassan F, Kamruzzaman M, Mekalanos JJ, Faruque SM. 2010. Satellite phage TLCphi enables toxigenic conversion by CTX phage through dif site alteration. Nature 467(7318):982–985.
Hu D, et al. 2016. Origins of the current seventh cholera pandemic. Proc Natl Acad Sci USA. 113(48):E7730–E7739.
Huber KE, Waldor MK. 2002. Filamentous phage integration requires the host recombinases XerC and XerD. Nature 417(6889):656–659.
Itsara A, et al. 2010. *De novo* rates and selection of large copy number variation. Genome Res. 20(11):1469–1481.
Jolley KA, Maiden MC. 2010. BIGSdb: scalable analysis of bacterial genome variation at the population level. BMC Bioinformatics 11:595.
Kaper JB, Morris JG, Jr, Levine MM. 1995. Cholera. Clin Microbiol Rev. 8(1):48–86.
Kim EJ, Lee CH, Nair GB, Kim DW. 2015. Whole-genome sequence comparisons reveal the evolution of *Vibrio cholerae* O1. Trends Microbiol. 23(8):479–489.
Kim EJ, Lee D, Moon SH, Lee CH, Kim DW. 2014. CTX prophages in *Vibrio cholerae* O1 strains. J Microbiol Biotechnol. 24(6):725–731.
Koboldt DC, et al. 2012. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome Res. 22(3):568–576.
Lee JH, et al. 2009. Classification of hybrid and altered *Vibrio cholerae* strains by CTX prophage and RS1 element structure. J Microbiol. 47(6):783–788.
Levin-Reisman I, et al. 2017. Antibiotic tolerance facilitates the evolution of resistance. Science 355(6327):826–830.
Li H. 2011. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics 27(21):2987–2993.
Li H, Durbin R. 2010. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics 26(5):589–595.
Lindahl T. 1993. Instability and decay of the primary structure of DNA. Nature 362(6422):709–715.
Long H, et al. 2015. Background mutational features of the radiation-resistant bacterium *Deinococcus radiodurans*. Mol Biol Evol. 32(9):2383–2392.
Long H, et al. 2016. Antibiotic treatment enhances the genome-wide mutation rate of target cells. Proc Natl Acad Sci USA. 113(18):E2498–E2505.
Lynch M, et al. 2016. Genetic drift, selection and the evolution of the mutation rate. Nat Rev Genet. 17(11):704–714.
Martinez E, Paly E, Barre FX. 2015. CTXphi replication depends on the histone-like HU protein and the UvrD helicase. PLoS Genet. 11(5):e1005256.
Midonet C, Das B, Paly E, Barre FX. 2014. XerD-mediated FtsK-independent integration of TLCphi into the *Vibrio cholerae* genome. Proc Natl Acad Sci USA. 111(47):16848–16853.
Mirkin EV, Mirkin SM. 2007. Replication fork stalling at natural impediments. Microbiol Mol Biol Rev. 71(1):13–35.
Morrow JD, Cooper VS. 2012. Evolutionary effects of translocations in bacterial genomes. Genome Biol Evol. 4(12):1256–1262.
Moyer KE, Kimsey HH, Waldor MK. 2001. Evidence for a rolling-circle mechanism of phage DNA synthesis from both replicative and integrated forms of CTXphi. Mol Microbiol. 41(2):311–323.
Mutreja A, et al. 2011. Evidence for several waves of global transmission in the seventh cholera pandemic. Nature 477(7365):462–465.
Napolitano R, Janel-Bintz R, Wagner J, Fuchs RP. 2000. All three SOS-inducible DNA polymerases (Pol II, Pol IV and Pol V) are involved in induced mutagenesis. EMBO J. 19(22):6259–6265.
Safa A, Nair GB, Kong RY. 2010. Evolution of new variants of *Vibrio cholerae* O1. Trends Microbiol. 18(1):46–54.
Sawabe T, et al. 2009. Mutation is the main driving force in the diversification of the *Vibrio splendidus* clade. Microbes Environ. 24(4):281–285.
Shah MA, et al. 2014. Genomic epidemiology of *Vibrio cholerae* O1 associated with floods, Pakistan, 2010. Emerg Infect Dis. 20(1):13–20.
Srivastava P, Chattoraj DK. 2007. Selective chromosome amplification in *Vibrio cholerae*. Mol Microbiol. 66(4):1016–1028.
Stamatoyannopoulos JA, et al. 2009. Human mutation rate associated with DNA replication timing. Nat Genet. 41(4):393–395.
Stokke C, Waldminghaus T, Skarstad K. 2011. Replication patterns and organization of replication forks in *Vibrio cholerae*. Microbiology 157(Pt 3):695–708.
Sung W, et al. 2016. Evolution of the insertion-deletion mutation rate across the tree of life. G3 (Bethesda) 6(8):2583–2591.
Sung W, et al. 2015. Asymmetric context-dependent mutation patterns revealed through mutation-accumulation experiments. Mol Biol Evol. 32(7):1672–1683.
Sung W, Ackerman MS, Miller SF, Doak TG, Lynch M. 2012. Drift-barrier hypothesis and mutation-rate evolution. Proc Natl Acad Sci USA. 109(45):18488–18492.
Val ME, et al. 2016. A checkpoint control orchestrates the replication of the two chromosomes of *Vibrio cholerae*. Sci Adv. 2(4):e1501914.
Wang H, Hays JB. 2006. Analysis of mismatch repair in human nuclear extracts. Curr Protoc Toxicol. 277: 26136–26142.
Wang H, Hays JB. 2004. Signaling from DNA mispairs to mismatch-repair excision sites despite intervening blockades. EMBO J. 23(10):2126–2133.
Wei W, et al. 2017. Zisland Explorer: detect genomic islands by combining homogeneity and heterogeneity properties. Brief Bioinform. 18(3):357–366.
Wei W, et al. 2014. SMAL: a resource of spontaneous mutation accumulation lines. Mol Biol Evol. 31(5):1302–1308.
Weller C, Wu M. 2015. A generation-time effect on the rate of molecular evolution in bacteria. Evolution 69(3):643–652.
Xie C, Tammi MT. 2009. CNV-seq, a new method to detect copy number variation using high-throughput sequencing. BMC Bioinformatics 10:80.
Xiong L, et al. 2014. Arginine deiminase pathway is far more important than urease for acid resistance and intracellular survival in *Laribacter hongkongensis*: a possible result of *arc* gene cassette duplication. BMC Microbiol. 14:42.

Xiong L, et al. 2015. Molecular characterization of arginine deiminase pathway in *Laribacter hongkongensis* and unique regulation of arginine catabolism and anabolism by multiple environmental stresses. Environ Microbiol. 17(11):4469–4483.

Xiong L, et al. 2017. Laribacter hongkongensis anaerobic adaptation mediated by arginine metabolism is controlled by the cooperation of FNR and ArgR. Environ Microbiol. 19(3):1266–1280.

Yang Y, Sterling J, Storici F, Resnick MA, Gordenin DA. 2008. Hypermutability of damaged single-strand DNA formed at double-strand breaks and uncapped telomeres in yeast *Saccharomyces cerevisiae*. PLoS Genet. 4(11):e1000264.

Associate editor: Charles Baer