

Research Article

Exception to the Rule: Genomic Characterization of Naturally Occurring Unusual *Vibrio cholerae* Strains with a Single Chromosome

Gary Xie,¹ Shannon L. Johnson,¹ Karen W. Davenport,¹ Mathumathi Rajavel,² Torsten Waldminghaus,³ John C. Detter,¹ Patrick S. Chain,¹ and Shanmuga Sozhamannan^{4,5}

¹Los Alamos National Laboratory, Biosciences Division, Genome Science, Los Alamos, NM 87545, USA

²School of Computer, Mathematical and Natural Sciences, Morgan State University, Baltimore, MD 21251, USA

³LOEWE Centre for Synthetic Microbiology-SYNMIKRO, Philipps-Universität Marburg, Hans-Meerwein-Str. 6, 35032 Marburg, Germany

⁴Tauri Group, LLC, Alexandria, VA 22310, USA

⁵Defense Biological Product Assurance Office, 110 Thomas Johnson Drive, Frederick, MD 21702, USA

Correspondence should be addressed to Shanmuga Sozhamannan; shanmuga.sozhamannan.ctr@mail.mil

Received 21 February 2017; Revised 15 June 2017; Accepted 22 June 2017; Published 29 August 2017

Academic Editor: Graziano Pesole

Copyright © 2017 Gary Xie et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The genetic make-up of most bacteria is encoded in a single chromosome while about 10% have more than one chromosome. Among these, *Vibrio cholerae*, with two chromosomes, has served as a model system to study various aspects of chromosome maintenance, mainly replication, and faithful partitioning of multipartite genomes. Here, we describe the genomic characterization of strains that are an exception to the two chromosome rules: naturally occurring single-chromosome *V. cholerae*. Whole genome sequence analyses of NSCV1 and NSCV2 (natural single-chromosome vibrio) revealed that the Chr1 and Chr2 fusion junctions contain prophages, IS elements, and direct repeats, in addition to large-scale chromosomal rearrangements such as inversions, insertions, and long tandem repeats elsewhere in the chromosome compared to prototypical two chromosome *V. cholerae* genomes. Many of the known cholera virulence factors are absent. The two origins of replication and associated genes are generally intact with synonymous mutations in some genes, as are *recA* and mismatch repair (MMR) genes *dam*, *mutH*, and *mutL*; MutS function is probably impaired in NSCV2. These strains are ideal tools for studying mechanistic aspects of maintenance of chromosomes with multiple origins and other rearrangements and the biological, functional, and evolutionary significance of multipartite genome architecture in general.

1. Introduction

The genetic endowment of the majority of bacteria (~90%) is inherited in a single chromosome of their respective genomes [1]. An exception to this rule is the presence of two chromosomes in *V. cholerae*, the causative agent of cholera disease [2, 3]. In fact, all tested genera/species in the family *Vibrionaceae* possess two chromosomes [4]. The two *V. cholerae* chromosomes (Chr1 and Chr2) are of unequal sizes (~2.96 Mb and ~1.07 Mb, resp.). Most of our knowledge on the control of chromosome maintenance of multipartite genomes is derived from studies on

V. cholerae [5–7]. It was postulated that Chr2 is derived from plasmid based on its plasmid-like origin and evolved into a secondary chromosome by acquiring additional layers of regulation for its replication [3, 8]. Although Chr1 encodes the majority of the housekeeping genes and is considered as the main chromosome, Chr2 also harbors essential genes beside many genes with unknown functions [9–11]. The genes on Chr1 and Chr2 are differentially expressed in specific niches. For example, when the bacterium was grown midexponentially in rabbit ileal loops, it showed expression of many more genes of Chr2 than those expressed in aerobically grown cells in rich

medium and harvested at the midexponential phase [12]. Majority of these are probably important niche-specific genes and hence expressed preferentially in ileal loop. The results were similar when the bacteria were collected from stools of cholera patients [13].

Chr1 replication follows the traditional *E. coli* paradigm in that the replication origin, *ori1*, contains multiple DnaA boxes where DnaA binds to initiate DNA replication [14–16]. A *V. cholerae* Chr1 minireplicon can replicate in *E. coli* without the need for *V. cholerae*-transacting factors but merely using *E. coli* proteins such as DnaA. Similarly, *V. cholerae ori1* could functionally substitute *E. coli* replication origin *oriC* [11, 14, 16].

The *V. cholerae ori2* resembles those of low-copy-number plasmids such as P1 and F in that it contains an array of repeats (iterons) where Chr2-specific initiator protein, RctB, binds to unwind the DNA for *ori2* firing [14, 17]. In *V. cholerae*, Chr1 initiates at the onset of the replication period while initiation of Chr2 is delayed and occurs only when 2/3 of the Chr1 replication has already been completed. Because Chr2 is 1/3 the size of Chr1, both chromosomes terminate their replication roughly at the same time [18, 19]. It was found recently that a site, termed *crtS* (Chr2 replication triggering site), present on Chr1 triggers *ori2* initiation when it is replicated [20, 21].

The two chromosomes of *V. cholerae* are longitudinally arranged in the cell [22]. While Chr1 appears to be spread along the entire longitudinal axis of the cell, Chr2 is restricted to the younger half of the cell. In newborn cells, Chr1 extends from the old pole to the new pole and Chr2 extends from midcell to the new pole [22]. The differential positioning of the chromosomes within the cell is accompanied by a distinct segregation choreography owing to chromosome-specific ParAB/parS-based segregation systems. [22–24]. One of the last steps in chromosome segregation before cell division involves the resolution of dimeric chromosomes that are frequently produced by homologous recombination between sister chromatids following DNA damage [25]. In *V. cholerae*, dimers of Chr1 and Chr2, are resolved by the action of the same machinery, XerC and XerD site-specific recombinases at the *dif* sites (*dif1* and *dif2*), located in the *ter* regions of Chr1 and Chr2, respectively [26]. While the *ter* regions of both chromosomes replicate at the same time point within the cell cycle, Chr1 sister termini are held together at midcell much longer than Chr2 sister termini [22]. The MatP/matS system was found to impede separation of Chr1 sister termini and restrict movement of Chr2 sister termini to allow processing by FtsK right before cell division [27]. In addition, a nucleoid-associated, FtsZ-binding protein termed SlmA has been shown to be required for blocking septal ring assembly. SlmA is a DNA-associated division inhibitor that is directly involved in preventing Z ring assembly on portions of the membrane surrounding the nucleoid [28].

Chr2 is indispensable for viability of the cell since elimination of Chr2 is lethal due to multiple “suicidal” toxin-antitoxin systems encoded by Chr2 [29–31]. *V. cholerae* with single chromosomes has been created by genetic engineering to fuse the two natural chromosomes [32]. Chromosomal fusions in *V. cholerae* were also isolated as suppressor

mutations for a deletion of the *dam* (DNA adenine methyltransferase) gene [33]. Dam is essential for replication of Chr2 because the initiator protein RctB binds to the target sites in *ori2* only when they are methylated [14, 16, 32]. The only way to replicate Chr2 without a functional *ori2*, that is, unmethylated *ori2*, was a Chr1 and Chr2 chromosomal fusion that permits replication of the entire chromosome from *ori1* [33]. In *dam*-suppressor mutants, Chr1 and Chr2, fusions had occurred either by homologous recombination between IS elements or site-specific recombination between *dif* sites. Interestingly, fusion occurred preferentially in the terminus regions of the chromosomes [33]. Recently, it was found that chromosomal fusion in *V. cholerae* can also occur as a suppressor of Δ *crtS* mutations [20].

Until recently, evidence for a naturally occurring *Vibrio* with fused chromosomes was missing. In an attempt to assess the genomic diversity of non-O1/non-O139 *V. cholerae*, a whole genome mapping strategy was applied on a well-defined, geographically, and temporally diverse strain collection, the Sakazaki serogroup type strains [34]. In that study, the whole genome map data on 91 of the 206 serogroup type strains supported the hypothesis that *V. cholerae* has an unprecedented genetic and genomic structural diversity with very few clonal complexes. Fortuitously, chromosomal fusions in two unusual strains that possess a single chromosome instead of the two chromosomes usually found in *V. cholerae* were discovered which was further confirmed by pulse field gel electrophoresis (PFGE) [35].

Earlier, we reported the whole genome sequencing and generation of a gapless single-contig sequence of the two unusual single-chromosome strains, 1154-74 (serogroup O49) and 10432-62 (serogroup O27), hereafter referred to as NSCV1 and NSCV2, respectively [36]. In the current paper, we report further genome sequence analyses of NSCV1 and NSCV2. We delineate the Chr1 and Chr2 fusion junctions, other structural anomalies such as indels, inversions and duplications, salient features of their gene content and the origins of replication, and their potential activity. Furthermore, we analyze the genes involved in replication of the multiple origins in the same chromosome. Thus, the primary focus of this paper is to lay the foundation for future studies on functional and mechanistic aspects of chromosome and structural maintenance in these unusual *V. cholerae* strains.

2. Materials and Methods

2.1. Bacterial Strains and Growth Conditions. *V. cholerae* 1154-74 is a strain isolated in India in 1974 from a diarrheal sample, and it belongs to the O49 serogroup. *V. cholerae* 10432-62 is a strain isolated in the Philippines in 1962 from a diarrheal sample, and it belongs to the O27 serogroup. These strains are referred to as NSCV1 and NSCV2, respectively, in this manuscript. The corresponding Los Alamos National Labs (LANL) Sequencing Center designations are VAAO49 and VABO27, and the old locus tags in the Gen Bank entries carry these designations. The bacterial strains were grown in LB medium at 37°C using normal standard laboratory protocols. Genomic DNAs for sequencing

were extracted using Qiagen genomic DNA extraction kits (Qiagen Inc.).

2.2. Whole Genome Sequencing Method. The draft genomes of *Vibrio cholerae* strains NSCV1 (1154-74) and NSCV2 (10432-62) were generated at the LANL Genome Science Group using a combination of Illumina [37] and 454 technologies [38]. For each genome, we constructed and sequenced an Illumina GAiiX shotgun library, a 454 Titanium standard library and a paired end 454 library, and a Pacific Biosciences RS long read library (P4-C2 chemistry). The 454 Titanium standard data and the 454 paired end data were assembled together with Newbler, version 2.3-PreRelease-6/30/2009. The Newbler consensus sequences were computationally shredded into 2 kb overlapping fake reads (shreds). Illumina sequencing data was assembled with VELVET, version 1.0.13 [39], and the consensus sequence was computationally shredded into 1.5 kb overlapping fake reads (shreds). We integrated the 454 Newbler consensus shreds, the Illumina VELVET consensus shreds, the PacBio subreads, and the read pairs in the 454 paired end library using parallel phrap, version SPS - 4.24 (High Performance Software, LLC). Illumina data were used to correct potential base errors and increase consensus quality using the software Polisher developed at Joint Genome Institute (JGI) [40]. Possible misassemblies were corrected using gapResolution [41], or Dupfinisher [42]. Completed genome assemblies were compared to optical maps to ensure consensus. Each final assembly consisted of a single ca 4.1 Mb chromosome.

2.3. Genome Annotation. Assembled genomes were annotated using a modified version of the IGS Annotation Engine (released by the University of Maryland, Institute for Genome Sciences at the School of Medicine) on an Ergatis workflow manager. Annotated genomes are available in GenBank under accession numbers NSCV1-1154-74: NZ_CP010811.1 and NSCV2-10432-62: NZ_CP010812.1. Post assembly genome sequence analysis was done using the CLC Genomic Workbench version 6.5.

Annotation of the assembled genome sequence was also carried out with genome annotation systems at LANL EDGE server <https://bioedge.lanl.gov/> [43] and RAST server [44]. A combined gene prediction strategy was applied by means of the GLIMMER 2.0 system and the CRITICA program suite [45] along with post processing by the RBSfinder tool [46]. tRNA genes were identified with tRNAscan-SE [47]. The deduced proteins were functionally characterized by automated searches in public databases, including SWISS-PROT and TrEMBL [48], Pfam [49], TIGRFAM [50], InterPro [51], and KEGG [52]. Each gene was functionally classified by assigning clusters of orthologous group (COG) number and corresponding COG category [53].

2.4. Genomic Comparison. Comparative genome analyses of NSCV1 and NSCV2 to other *V. cholerae* strains MS6, O1 biovar El Tor str. N16961, and O139 MO10 were performed by using a set of tools available at LANL and RAST servers. Homology searches were conducted at the nucleotide and

amino acid sequence level using BLAST [54]. To obtain a list of orthologs from bacteroidetes genomes, a perl script that determines bidirectional best hits was written. For example, genes g and h are considered orthologs if h is the best BLASTP hit for g and vice versa. *E* values of 10^{-15} were acceptable. A gene is considered strain specific if it has no hits with an *E* value of 10^{-5} or less. The genome comparisons at the nucleotide level were carried out with genome alignment tools, such as MUMmer2 [55], NUCmer [56], the Artemis Comparison Tool (ACT) [57], and WebAct (<http://www.webact.org/WebACT/home>) at Imperial College, London.

The MUMmer package [58] programs nucmer, repeat_match, and exact_tandems were used for analysis of repeat regions. To identify long inexact genomic repeats, the nucmer program was used with the options `-maxmatch` and `-nosimplify`. The resulting dotplot was obtained using `mummerplot`. The `repeat_match` and `exact_tandems` programs were run with default arguments. Tandem repeats finder [59] and inverted repeats finder [60] were run with recommended default arguments to identify tandem and inverted repeats, respectively. Phage islands and putative genomic islands were identified using PHAST (PHAGE Search Tool) [61] and Islandviewer [62], respectively. The circular genome diagram (Figure S3 available online at <https://doi.org/10.1155/2017/8724304>) was drawn using the DNAPlotter [63]. Blast Ring Image Generator (BRIG) software was used to generate genome comparison views presented in Figure S2 [64]. Multiple genome alignment shown in Figure S1 was created using progressive MAUVE [65].

3. Results and Discussion

3.1. Whole Genome Sequencing of *V. cholerae* Strains NSCV1-(1154-74/VAA_O49 and NSCV2-(10432-62/VAB_O27). Whole genome mapping (WGM) also known as optical mapping and PFGE data provided evidence of Chr1 and Chr2 fusions in *V. cholerae* strains NSCV1 (1154-74) and NSCV2 (10432-62) [35]. We performed whole genome sequencing of the two *V. cholerae* strains in order to understand the *molecular mechanisms* that might have led to Chr1 and Chr2 fusions in these strains [36]. We generated Roche 454 (21× and 29×), Illumina (2178× and 507×), and Pac Bio RS (350× and 190×) sequence data at the indicated coverages for NSCV1 and NSCV2, respectively, and used a hybrid assembly approach that resulted in a single-contig, gapless genome (Table S1). Further resolution of the ambiguous regions was carried out using Sanger-based primer walking according to Las Alamos National Labs (LANL) genome finishing pipeline yielding a final, finished, and closed single-contig sequence [66]. Consistent with the WGM results and in contrast to the previous genomic reports of *V. cholerae* and its close neighbors [2–4], these two sequences assembled into genomes consisting of a single chromosome of approximately the same size as the two expected chromosomes added together. All other genomic statistics for NSCV1 and NSCV2 appear to be highly similar to other *V. cholerae* genomes currently available in GenBank (Table S1). The genomes of NSCV1 and NSCV2 are comprised of a single circular chromosome of 3,928,357

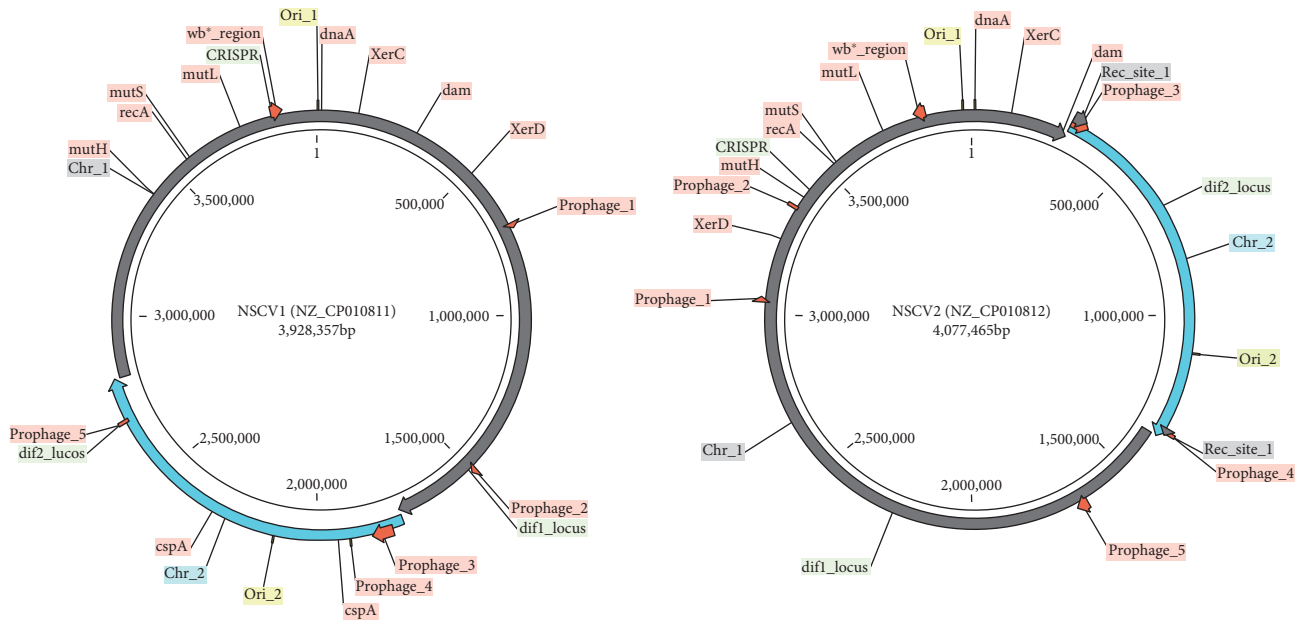


FIGURE 1: Circular genome maps of NSCV1 (1154-74_VAAO49) and NSCV2 (10432-62_VABO27). Fusion of Chr1 (dark grey) to Chr2 (blue) is shown in the circle at the respective locations. Various unique features such as prophages and the origins of replication- and replication-associated genes are indicated around the circles.

and 4,074,462 bp in length with an overall G + C content of 47.73% and 47.66%, respectively (Figure 1). Following the sequencing convention, the nucleotide position 1 was placed upstream of the *dnaA* gene (VAA_049_1 and VAB_027_1) encoding the chromosomal replication initiator protein. Chr2 portion of NSCV1 and NSCV2 spans from genome positions 1,720,246–2,772,395 (1,052,150 bps) and 297,221–1,375,947 (1,078,727 bps), respectively. All other NSCV1 and NSCV2 genome statistics are presented in Table S1.

3.2. Whole Genome Sequence Comparison of NSCV1 and NSCV2 to Other *V. cholerae*. DNA sequence of *V. cholerae* strain MS6 is the closest match to NSCV1 and NSCV2 chromosomes in public genome databases based on whole genome sequence homology [67]. MS6 is a *V. cholerae* O1 strain with a novel genetic background designated in Thailand–Myanmar and isolated from a stool sample of a diarrheal patient [68]. NSCV1 and NSCV2 genome sequences exhibit a high degree of synteny with MS6, N16961, MO10, and TSY216 as visualized in the Mauve alignment despite genetic rearrangements such as insertions and inversions across the chromosome (Figure S1) [69]. A pairwise Megablast comparison of NSCV1 and NSCV2 genomes to four pathogenic *V. cholerae* (N16961, MO10, MS6, and TSY 216) was performed. At 85%, identity > 88–94% is shared between NSCV1 and NSCV2 compared to N16961, MS6, or MO10 genomes and 74% and 80% is shared between NSCV1 and NSCV2 compared to TSY216, respectively. The unique regions varied from 5 to 10% (Table S2).

To further delineate the differences between NSCV1 and NSCV2 against all other genomes, the locations and gene content of the unique regions were determined (Table S3) and displayed as BRIG view (Figure S2). In the whole genome

comparison of NSCV1 and NSCV2 to *V. cholerae* strains MS6, N16961, and TSY216, large regions (~10 kb or more) missing in comparator strains are highlighted in the outer circle of the BRIG view (Figure S2). As expected, many features (e.g., prophages) that are unique to NSCV1 and NSCV2 are absent in the pathogenic strains and similarly, many of the known virulence regions (e.g., CTX and VPI) that are present in pathogenic *V. cholerae* such as N16961 are absent in NSCV1 and NSCV2. In Chr2 segment, the super integron region is present in NSCV1 and NSCV2 with many subtle indels/variations.

3.3. Comparison of Genomic Content of NSCV1 and NSCV2 to Pathogenic *V. cholerae*. Further analysis of the genomic content of NSCV1 and NSCV2 was performed by comparing the genome annotations. We were interested to see if there are any differences in the genetic content that might explain the mechanism of genomic fusion. Overall, the number of CDS encoded by NSCV1 and NSCV2 is very similar to other *V. cholerae* (Table S1). NSCV1 and NSCV2 have 2759 CDS in common with all other genomes compared here. Comparison of the genomes of NSCV1 and NSCV2 with MS6, N16961, and MO10 (an epidemic *V. cholerae* strain belonging to the O139 serogroup) revealed that the four organisms have approximately 3188/3259 genes in common, depending on whether NSCV1 or NSCV2 is used as the query, respectively (Figure 2). There are 173/304 CDSs unique to NSCV1 and NSCV2, respectively.

3.4. Genomic Islands, Prophages, and Other Mobile Genetic Elements. In the NSCV1 and NSCV2 chromosomes, 49 and 20 genomic islands amounting to 677, 625 bps and 205, 247 bps or 16.32% and 4.94% of the entire genome, respectively, were detected. Clustered regularly, interspaced short

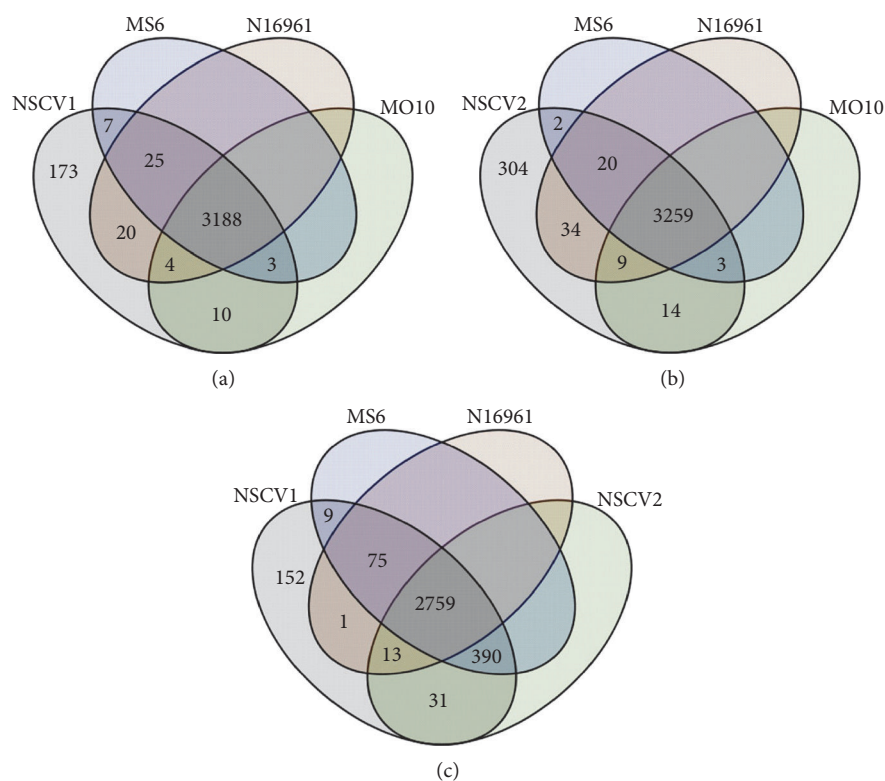


FIGURE 2: Comparison of the genomic content of NSCV1 (a), NSCV2 (b), and NSCV1 and NSCV2 (c) to various other genomes. Venn diagram showing the number of NSCV1 or NSCV2 predicted CDS with significant homology ($1e^{-5}$) with the predicted products of the near neighbors: *V. cholerae* MS6, N16961 (serogroup O1 biovar El Tor), and O139 MO10 (serogroup O139). The number outside the circles (173 or 304) represents the number of NSCV1 or NSCV2 CDS that does not have significant homologs in the three strains compared. Conserved genes among them were defined by whole-genome pairwise sequence comparisons using the sequence-based comparison tool in RAST.

palindromic repeat (CRISPR) element is a widely found defense mechanism of prokaryotes against entry of foreign DNA including plasmids and phages [70]. In the NSCV1 and NSCV2 chromosomes, one CRISPR element was located at 3,606,088–3,606,203 bps and 3,572,464–3,572,579 bps in NSCV1 and NSCV2 genomes, respectively. The NSCV1 and NSCV2 genomes contained 58 and 55 mobile genetic elements, respectively, that encode phage integrases, transposases, and site-specific recombinase (Figure S3), compared to 46 mobile element genes in *V. cholerae* MS6. NSCV1 has 5 prophages located at (1) 685,729–702,622; (2) 1,458,739–1,476,484; (3) 1,756,187–1,818,403; (4) 1,879,405–1,883,642; and (5) 2,643,793–2,650,802. Prophages 3 and 5 are at the immediate boundary region and inside of the Chr2 insertion locus. Only prophage 3 at the 5' end of Chr2 insertion appears to be intact and encodes 56 CDS. NSCV2 has 5 prophages located at (1) 3,113,341–3,130,219; (2) 3,416,918–3,428,588; (3) 299,327–349,041; (4) 1,359,430–1,369,981; and (5) 1,676,903 to 1,714,350. Only prophage 5 is intact whereas the other four are defective or incomplete. Prophages 3 and 4 are at the immediate flanking region of the Chr2 insertion junction (Figure 1).

3.5. Virulence and Antimicrobial Resistance Genes. *V. cholerae* strains NSCV1 and NSCV2 were isolated from patients exhibiting atypical cholera symptoms [34]. It is of interest to

see if these strains carry any of the usual *V. cholerae* virulence factors and if not, what other virulence factors or toxins might be present that would explain the diarrheal symptoms. The major virulence factor of *V. cholerae*, the cholera toxin encoded by *ctxAB* genes, is not found in NSCV1; however, other toxin genes *ace* and *zot* and RTX toxin cluster can be found, in addition to *toxRS* genes. The toxin genes are absent in NSCV2. NSCV1 and NSCV2 each has 57 genes encoding putative resistance to antibiotics and toxic compounds such as colicin V, bacteriocin, cobalt-zinc-cadmium, copper homeostasis, fluoroquinolones, and multidrug resistance efflux pumps.

3.6. O-Antigen Regions. NSCV1 and NSCV2 belong to serogroups O49 and O27, respectively, and as seen in other serogroups of *V. cholerae*, the O-antigen biosynthesis genes are encoded in a cluster (*wb**) on Chr1 part of the respective genomes from 3,777,887–3,815,081 (CDS VAAO49_3428-VAAO49_3463) and 3,888,708–3,920,841 (CDS VABO27_3582-VABO27_3611) of NSCV1 and NSCV2, respectively. Also, as seen in other *wb** regions, the boundary regions of this cluster are highly conserved whereas the region between the conserved genes is highly divergent. More specifically, in the NSCV1 *wb** region, 2 segments of 2974 bps and 6212 bps at the left end and another segment of 5733 bps at the right end, respectively, are 96% and 94% identical to those in the

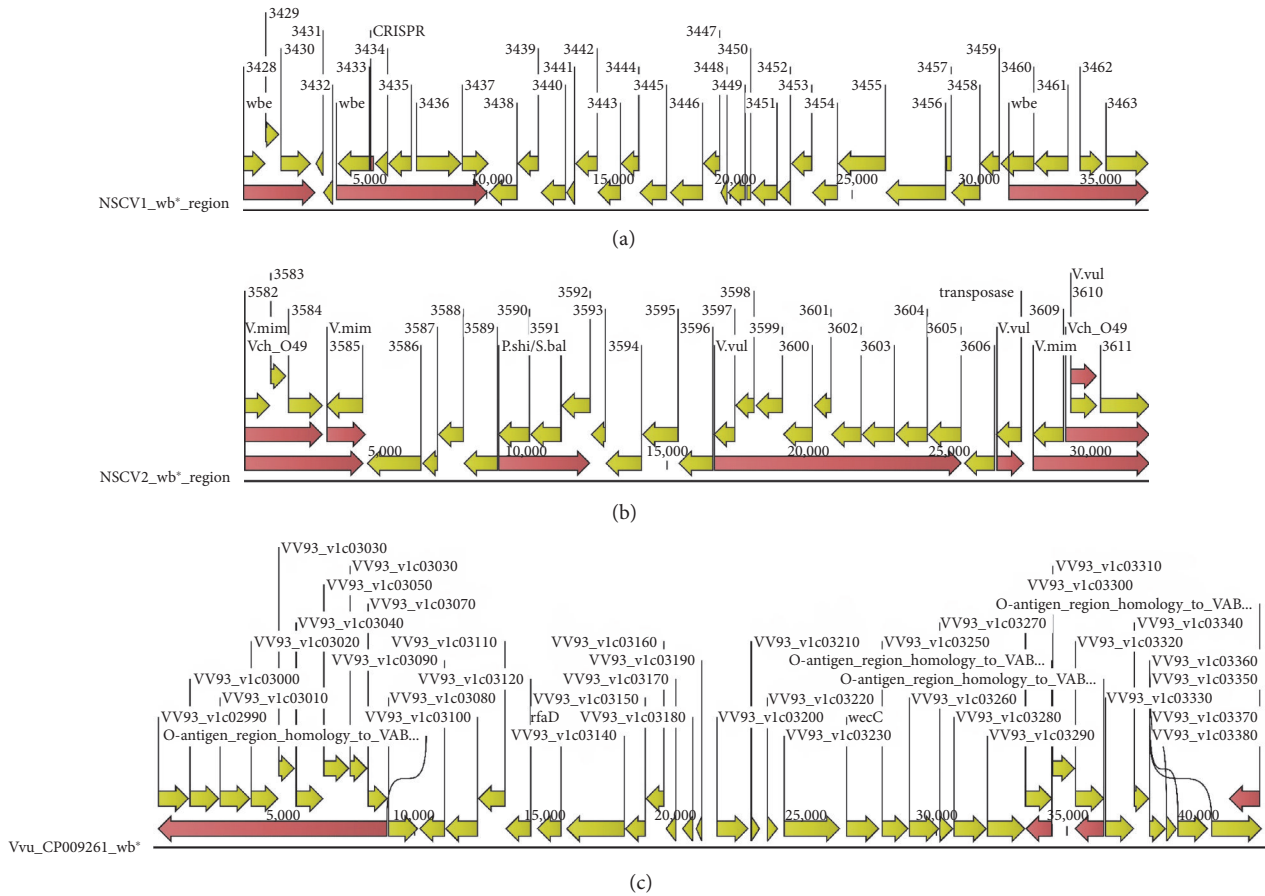


FIGURE 3: Genetic maps of O-antigen regions (wb^*) of NSCV1 and NSCV2. The various genes and their orientations are indicated by the arrows. Homologous regions to other serogroups are indicated by the red arrows. The *V. vulnificus* wb^* cluster that has extensive homology to NSCV2 wb^* cluster is shown in (c).

wbe region (O1 serogroup). A blast analysis of the wb^* region of NSCV2 revealed segments that are similar to *V. vulnificus* (GenBank Accession # CP009261) gene cluster (around 42 kb) with 4 segments that are >65–85% identical, ranging in length from ~1.0 kb to 8.831 kb to the NSCV2 wb^* cluster. In addition, identities to *V. mimicus* at 83–92% in segments of 4176 bps, 2789 bps, and 1410 bps and to *Plesiomonas shigelloides* and *Shewanella baltica* at ~80% (3281 bps), and to NSCV1 at 94–96% (4325, and 3007 bps) identities were also found as depicted in Figure 3.

3.7. Identification of Large Tandem Repeats and Inversion. A closer analysis of the WGMs (whole genome maps) of NSCV1 and NSCV2 strains revealed a large duplication in the general region of 1240 and 1506 kb on a reference genome (M66–2) coordinates based on optical restriction maps [35]. Experimental WGM data is further supported by in silico WGM generated from whole genome sequence data of NSCV1 and NSCV2 (Figure S4). Sequence read mapping results showed that there is a large duplication of 200 kbs and 70 kbs regions in NSCV1 and NSCV2, respectively, as evidenced by >1 \times , but <2 \times , depth of average coverage of the genome at 1,503,484 to 1,718,861 bps in NSCV1 and 2,221,105 to 2,300,924 bps in NSCV2 compared to that in the rest of the genome (Figure S5). Manual

validation indicated that the 200 kbs and 70 kbs regions indeed are fragmental duplications, existing among subpopulation of NSCV1 and NSCV2 cells, respectively. Both 1 copy and 2 copy versions of genome assembly have enough evidence supporting variations among population. These duplicated sequences spanned from positions 1,503,484 to 1,718,861 (unit 1) and from 1,718,862 to 1,934,269 bp (unit 2) in NSCV1 and from positions 2,221,105 to 2,300,924 (unit 1) and from 2,300,925 to 2,380,745 (unit 2) in NSCV2 in genome sequences with repeats included (Figure S6). The duplicated sequences span a total length of 430,785 bp and 159,640 bp on Chr1 region and encode 177 and 59 duplicated genes in NSCV1 and NSCV2, respectively. It should be noted that the two copies of repeat units are not identical. There is a 31 bp indel between the two copies in NSCV1, which resides at the locus of VAA_049_1509, encoding exonuclease family protein. There are 5 single bp indels between the two copies in NSCV2. Although tandem repeats are present at different locations on NSCV1 and NSCV2, there is a 40 kb DNA segment that overlapped at the 5' end of NSCV1 and NSCV2 repeats with 98% identities (Figure S7). However, the tandem duplications could not be verified by PCR of the junctions since the orientation of the two copies cannot be ascertained in the assembly. Hence, we have decided to keep the genome sequence with a single

copy as the ref seq (GenBank submission) for all the analyses. This long duplication anomaly could not be unequivocally resolved with the existing sequence data. In addition, there is a large inversion from 1,477,189 to 3,484,983 bps in NSCV2 compared to MS6 chromosome 1 region from 612,067 to 2,541,958 bps (Figure S1b).

3.8. Mapping the Fusion Junctions and Possible Mechanism of Genome Fusion. Chromosomal fusions were observed previously in suppressor mutants resulting from a genetic screen [33]. These fusions were analyzed in detail and found to occur via two different mechanisms, either by homologous recombination at IS elements or by site-specific recombination at *dif* sites [32] or as transient chromosomal fusions in $\Delta crtS$ suppressor mutants [20]. In this study, we analyzed the fusion sites of NSCV1 and NSCV2 in detail to see if any one of these mechanisms also led to chromosomal fusion in the natural single-chromosome isolates. The circular single chromosomes of NSCV1 and NSCV2 are shown in Figure 1, and along with N16961, Chr1- and Chr2-concatenated circular maps are shown in Figure S2, and the genome sequence statistics are provided in Table S1. The insertion sites on Chr1 backbone (as an acceptor) and Chr2 backbone (as a donor) are not the same for NSCV1 and NSCV2 (see below). Artemis Comparison Tool views (zoomed in view) of the fusion junctions are presented in Figures S8 and S9. Further description of the hypothetical events leading to the fusions is provided below.

3.9. Chr1 and Chr2 Fusion Junctions in *V. cholerae* Strain NSCV1. A closer examination of the Chr1-Chr2 insertion junction in NSCV1 to Chr1 and Chr2 sequences of MS6 indicates that multiple events probably occurred before the final NSCV1 was derived. In an NSCV1-like intermediate, on Chr1 between VAA049_1545 and VAA049_2500, there probably was an insertion of MS6_A0562- and MS6_A0272-like CDS (Chr2 CDS) (Figure 4 (top panel)). On Chr2, there was an insertion of a prophage to the right of MS6_A0272 (Figure 4 (top panel)). In the next step, Chr2 with the new prophage was inserted into Chr1 via the homologies between the two CDS (VAA049_1594 versus MS6_A0272 and VAA049_2432 versus MS6_A0562), and upon the resolution of the cointegrate, one copy each of the two homologous CDS along with intervening sequences was deleted (Figure 4 (bottom panel)). Thus, Chr2 in NSCV1 spans from 1,714,319 to 2,772,395 bps (including its flanking unique region, Figure 4 (bottom panel)) or from 1,757,711 to 2,722,145 bps (including boundary of Chr2 homologous region) (Figure 4 (bottom panel)) spanning NSCV1 CDS VAA049_1594–2432. This recombination event likely occurred between MS6_1029 and MS6_1028 or between 1,168,144 and 1,168,021 bp in MS6-like Chr1 background and inserted from the breakpoint between MS6_A0272 and MS6_A0562 or between 303,999 and 446,488 bp on MS6-like Chr2 background (Figure 4 (bottom panel)). There are 2 prophages (3 and 5) closer to the boundaries of Chr2 insertion locus. Flanking the Chr2 insertion sites, there are unique regions of 37,425 bps and 50,250 bps in length at 5' end and 3' end, respectively. The role of any of these unique regions

in the recombination event or the precise mechanism of recombination cannot be ascertained with available whole genome sequence data in public databases.

3.10. Chr1 and Chr2 Fusion Junctions in *V. cholerae* Strain NSCV2. A closer examination of the insertion junctions indicates that an NSCV2-like intermediate probably possessed MS6_A0925- and MS6_A0924-like Chr2 CDS on Chr1 flanked by 2 prophages (Figure 5 (top panel)). In that intermediate strain, Chr2 was inserted via recombination of the homologous segments (VAB027_307 versus MS6_A0924 and VAB027_1228 versus MS6_A0925). As a result, Chr2 spans from 337,822 bps to 1,351,337 bps (without its flanking prophage regions between CDS VAB027_307-VAB027_1228) or from 297,221 bps to 1,375,947 bps (including prophage regions between CDS VAB027_275-VAB027_1225). This recombination event likely occurred between 2,643,219 and 2,643,208 bp or between MS6_2336 and MS6_2335 on MS6 Chr1 backbone and inserted between MS6_A0924 and MS6_A0925 break points (852,904–852,973) on MS6 Chr2 backbone (Figure 5 (bottom panel)). The insertion boundary is flanked by 245 bp repeats located at 337,552 to 337,796 bp and 1,351,093 to 1,351,337 bps with 96% identities, then flanked by 2 prophages. In the flanking regions of 2 prophages, there are 12 bp identical repeats (caccgcagggtg) (297,210–297,221 bps and 1,375,947–1,375,958 bps). The 245 bp repeat also overlaps 155 bps with VAB027_1228 that encodes L-threonine 3-dehydrogenase.

As mentioned above for NSCV1, the role of any of these unique regions in the recombination event or the mechanism of recombination that resulted in chromosomal fusion cannot be ascertained with available *V. cholerae* whole genome sequence data. Since the immediate predecessor or recombination intermediate strains are not known, it is difficult to decipher the exact events that led to the NSCV1 and NSCV2. For example, the exact mechanism of recombination (site specific versus generalized recombination) or the precise recombination cross-over points cannot be predicted. It also appears that there is more than one event before the final NSCV1 and NSCV2 arose. The presence of prophages at the insertion junctions leads us to speculate that the Chr2 insertion events in NSCV1 and NSCV2 appear to be mediated by generalized homologous recombination via homologies provided by prophages or parts/genes homologous to prophage genes and mobile gene elements present in both Chr1 and Chr2 in the precursor strain.

3.11. Sequence Analysis of the Origins of Replication and Associated Genes. The *ori1* and *ori2* origins of replications in NSCV1 and NSCV2 chromosomes were identified by a homology to the Chr1 and Chr2 origins of MS6. The *ori1* is colocalized with the genes (*rpmH*, *dnaA*, *dnaN*, *recF*, and *gyrA*) often found near the *oriC* in prokaryotic genomes, and origin of the location corresponded with GC nucleotide skew (G – C/G + C) analysis as illustrated in Figure S2 and Figure S3. Based on these data, we assigned base-pair 1 in an intergenic region located in the putative *ori1*. Similarly, *ori2* was located based on sequence homology to the *ori2* in other prototypical *V. cholerae* genomes. The genome

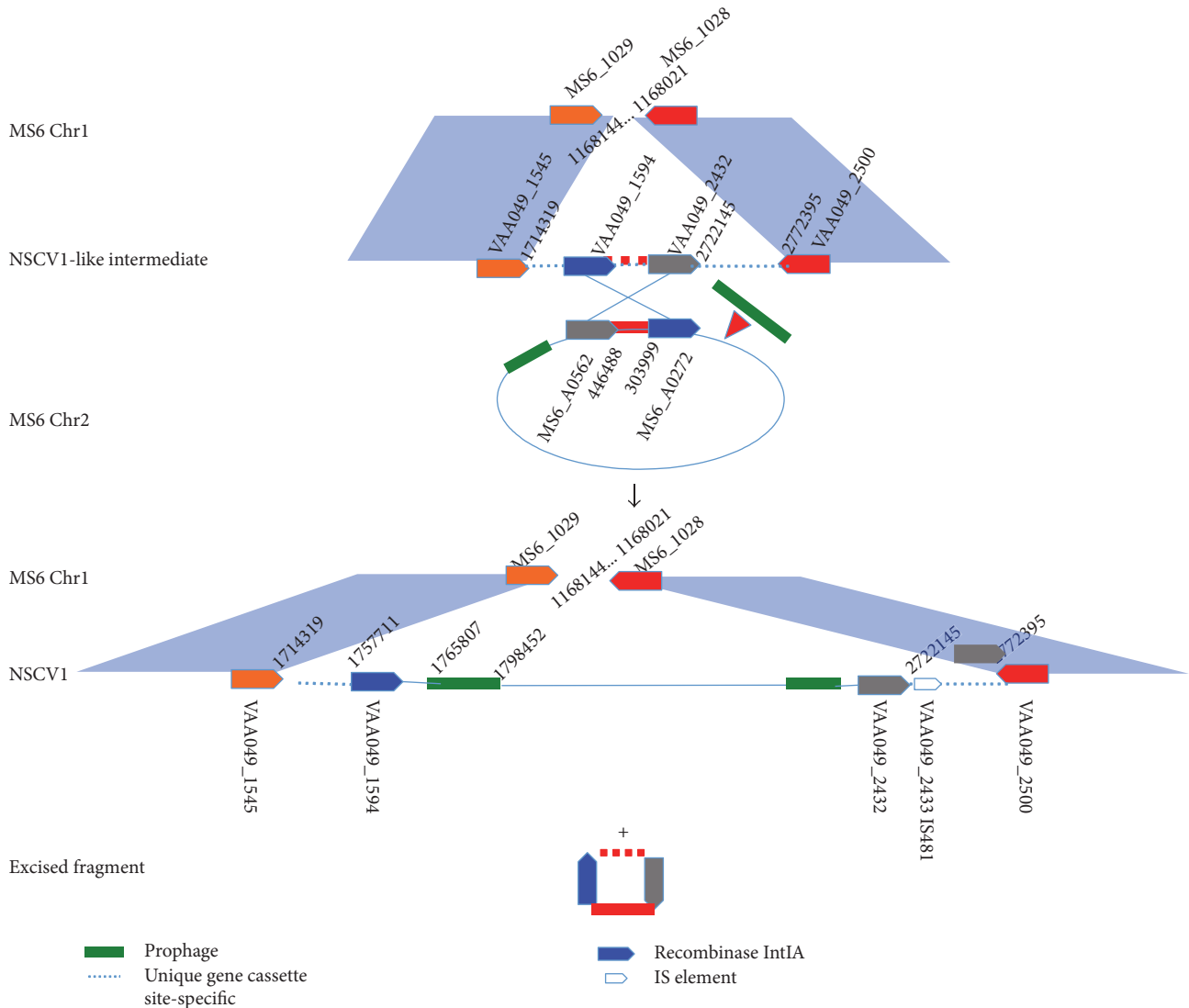


FIGURE 4: Putative recombination event that resulted in Chr1 and Chr2 fusions in NSCV1. Top panel: MS6 Chr1 CDS at the fusion junction and the same location in a putative NSCV1-like intermediate and the MS6 Chr2 circle and the codons at the fusion junction are depicted. The cross-over region between Chr1 and Chr2 is indicated by the long X. A prophage insertion event to the right fusion junction prior or post to Chr 2 fusion is depicted by a green line. Bottom panel: The recombination products after fusion event are shown with NSCV 1 in the middle and an excision product that contains one copy of the cross-over genes with the intervening sequences at the bottom.

positions of the *ori*, their orientation, and genes within the *ori* are indicated in Table 1. A list of putative *dif* sites (chromosome dimer resolution sites) and their locations are provided in Table S4. A previous study had found chromosome fusions in *V. cholerae* as suppressors of impaired Chr2 replication [33]. As mentioned above, some of these fusions had occurred by site-specific recombination of the two *dif* sites of Chr1 and Chr2, respectively. The positions of the intact *dif* sites on the fused chromosomes in NSCV1 and NSCV2 support the notion that this was not the mechanism how fusion occurred in these strains (Table S4). It remains to be seen which of the two *dif* sites on NSCV1 and NSCV2 chromosomes is active and used for chromosome dimer resolution. This might depend on the molecular mechanism involved in positioning the respective XerCD recombinase at the *dif* site.

There is extensive genetic conservation in the *ori1* and *ori2* of NSCV1 (Figure 6) and NSCV2 (Figure 7) compared to a prototypical reference genome such as N16961. A closer look at the sequences of the origin regions including the genes within the origin of replication indicated that there is no significant indels between the respective origins compared to a reference genome such as N16961. However, a number of SNPs were found in the origins of replication. The nucleotide and amino acid changes at the origins of NSCV1 and NSCV2 are presented in Table S5. It remains to be analyzed experimentally if both origins on the fused chromosomes are active in replication.

3.12. Sequence Analyses of Replication Associated and Mismatch Repair Genes That May Be Potentially Involved in Single-Chromosome Maintenance. Earlier studies have

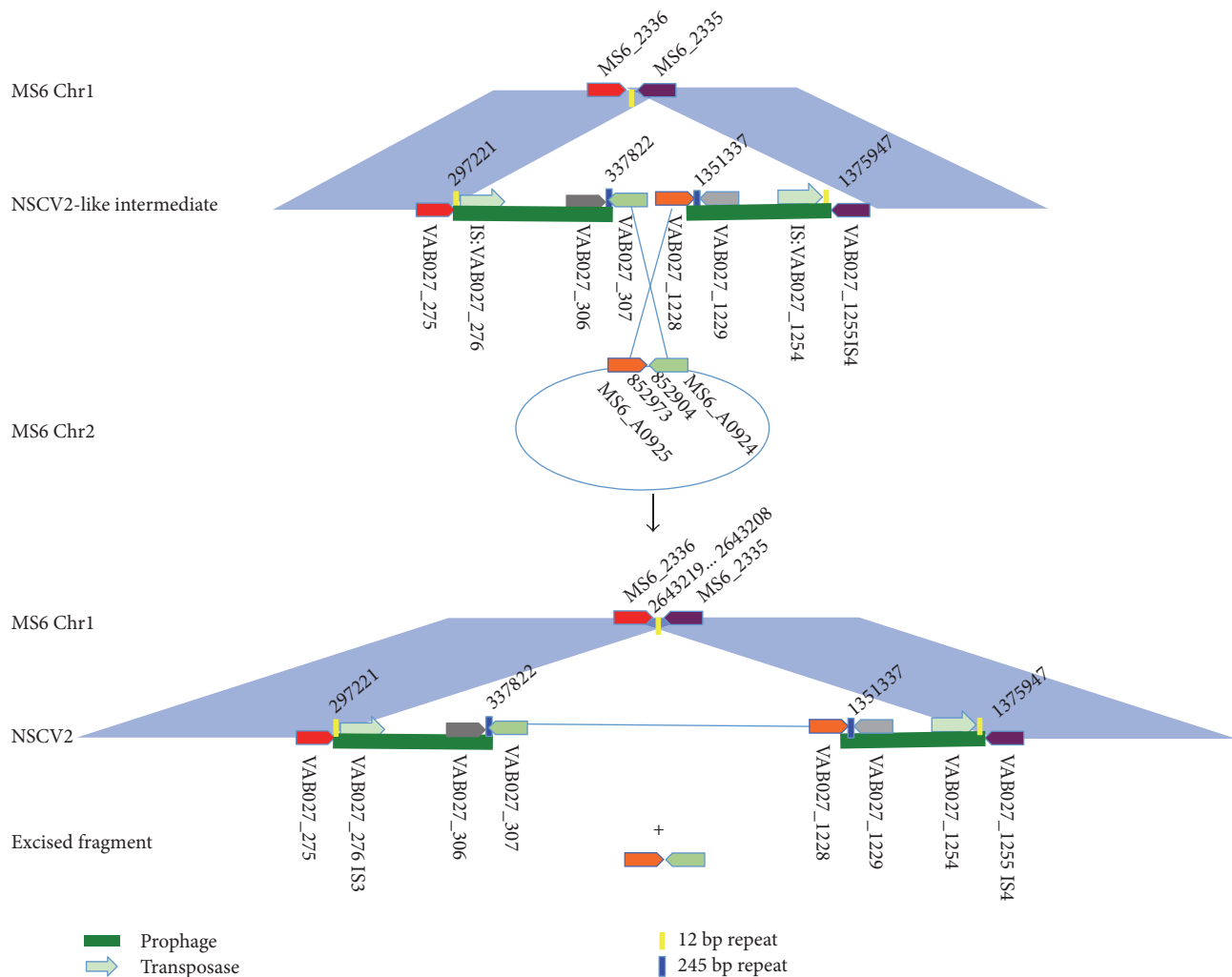


FIGURE 5: Putative recombination event that resulted in Chr1 and Chr2 fusions in NSCV2. Top panel: MS6 Chr1 CDS at the fusion junction and the same location in a putative NSCV2-like intermediate and the MS6 Chr2 circle and the codons at the fusion junction are depicted. The cross-over region between Chr1 and Chr2 is indicated by the long X. Bottom panel: The recombination products after fusion event are shown with NSCV2 in the middle and an excision product that contains one copy of the cross-over genes with the intervening sequences at the bottom.

indicated that the *dam* gene is essential for the viability of *V. cholerae*, and depletion of DNA adenine methyl transferase (Dam protein) leads to successful spontaneous fusion of the two chromosomes [33]. Hence, we examined the genetic status of DNA adenine methylase gene (*dam*) in natural single-chromosome *V. cholerae*, NSCV1, and NSCV2 and found the *dam* gene to be intact. We also inspected the status of RecA and the MMR genes since they have been implicated in maintenance of chromosomal rearrangements such as large tandem repeats, inversion, and fusion. The nucleotide and amino acid changes in various replication and MMR genes in NSCV1 and NSCV2 are presented in Table S6. RecA-mediated homologous recombination was probably involved in the fusion event, and the fact that the resolution of the single chromosome into 2 chromosomes has not been observed during normal growth conditions indicates that the RecA in NSCV1 and NSCV2 is nonfunctional or altered due to the presence of multiple SNPs, of which one leads to single-amino acid

change in RecA protein (Y305C). Preliminary data indicate that NSCV1 and NSCV2 are probably recombination deficient since construction of recombinants via natural transformation has been unsuccessful. The MMR genes are generally intact except for *mutS*. The *mutS* gene is impaired, that is, deletion of amino acid residues 132–150, in NSCV2. Among the other mismatch repair system genes, *mutH* and *mutL* have single-nucleotide polymorphisms that lead to amino acid changes in the respective proteins. Other replication-associated proteins such as DnaA, ParAB, and XerC have amino acid changes whereas XerD is intact (Table S6). Preliminary data indicate that the *ori1* is active in both strains suggesting a functional DnaA protein. However, the effect of these protein alterations in ParAB, XerC, RecA, and MMR proteins on recombination, replication, and maintenance of chromosomal fusions and other large-scale genome rearrangements described above awaits more stringent functional studies including cloning and intra-/interspecific complementation and these studies are underway.

TABLE 1: Genome locations of origins of replication and replication-associated genes.

Ori1 and Ori2 and the genes within Ori												
1154-74_O49						10432-62_O27						
Locus ID	Gene/ Locus	Start	End	Size	NI6961_O1 Strand	Orientation	Locus ID	Start	End	Size	10432-62_O27 Strand	Orientation
Rep_ori_ Chr_1		2956820	2961149/1	806	5136	Clockwise	Rep_ori_ Chr_1	4039130	4044265	5136		Clockwise
VC2772	<i>parB</i>	2956823	2957704	882	Complement	Clockwise	VAB027_RS18795	4039133	4040014	882	Complement	Clockwise
VC2773	<i>parA</i>	2957731	2958504	774	Complement	Clockwise	VAB027_RS18800	4040041	4040814	774	Complement	Clockwise
VC2774	<i>gidB</i>	2958519	2959151	633	Complement	Clockwise	VAB027_RS18805	4040829	4041461	633	Complement	Clockwise
VC2775	<i>gidA</i>	2959151	2961046	1896	Complement	Clockwise	VAB027_RS18810	4041461	4043356	1896	Complement	Clockwise
VC0001	<i>Ori_1</i>	2961047	2961149	371	474		<i>Ori_1</i>	4043357	4043830	474		
VC0002	<i>hypo</i>	235	402	168	Complement	Clockwise	VAB027_RS18815	4043694	4043861	168	Complement	Clockwise
VC0002	<i>mioC</i>	372	806	435	Complement	Clockwise	VAB027_RS18800	2269384	2269818	435	Complement	Clockwise
Rep_ori_ Chr_2		1069696	1072315/1	3191	5811	Clockwise	Rep_ori_ Chr_2	1115023	1120832	5810		Counterclockwise
VCA1113		1068927	1069958	1032	Complement	Clockwise	VAB027_RS05240	1120570	1121601	1032	Complement	Clockwise
VCA1114	<i>parB</i>	1070018	1070989	972	Complement	Clockwise	VAB027_RS05235	1119539	1120510	972	Complement	Clockwise
VCA1115	<i>parA</i>	1070997	1072220	1224	Complement	Clockwise	VAB027_RS05230	1118314	1119531	1218	Complement	Clockwise
VCA0001	<i>rcfA</i>	112	246	135	Complement	Clockwise	<i>rcfA</i>	1117968	1118102	135	Complement	Clockwise
VCA0002	<i>Ori_2</i>	247	1133	887			<i>Ori_II</i>	1117081	1117967	887		
VCA0002	<i>rcfB</i>	1134	3110	1977			VAB027_RS05225	1115104	1117080	1977		Complement
Replication-associated genes												
1154-74_O49						10432-62_O27						
Locus ID	Gene	Start	End	Size	NI6961_O1 Strand	Orientation	Locus ID	Start	End	Size	10432-62_O27 Strand	Orientation
VC2626	<i>dam</i>	2796912	2797745	834	Complement	Clockwise	VAB027_RS01350	282403	283236	834	Complement	Clockwise
VC0543	<i>recA</i>	574696	575760	1065	Complement	Clockwise	VAB027_RS16670	3602557	3603621	1065	Complement	Clockwise
VC0668	<i>mutH</i>	715847	716512	666	Complement	Clockwise	VAB027_RS16030	3461932	3462597	666	Complement	Clockwise
VC0345	<i>mutL</i>	367072	369033	1962	Complement	Clockwise	VAB027_RS17585	3784248	3786209	1962	Complement	Clockwise
VC0535	<i>mutS</i>	565832	568420	2589	Complement	Clockwise	VAB027_RS016705	3609897	3612475	2580	Complement	Clockwise
VC0128	<i>xerC</i>	122005	122940	936	Complement	Clockwise	VAB027_RS00575	114596	115531	936	Complement	Clockwise
VC2419	<i>xerD</i>	2593375	2594283	909	Complement	Clockwise	VAB027_RS15355	3316085	3316993	909	Complement	Clockwise
VC0012	<i>dnaA</i>	7397	8815	1419			VAB027_RS00005	38	1441	1404		
							IGR of					
	<i>dif-1</i>	1564104	1564131	28	Complement	Clockwise	VAB027_RS10790-	2301152	2301179	28	Complement	Clockwise
							VAB027_RS10795					
							IGR of					
	<i>dif-2</i>	507983	508010	28	Complement	Clockwise	VAB027_RS03005-	664645	664672	28	Complement	Clockwise
							VAB027_RS03010					

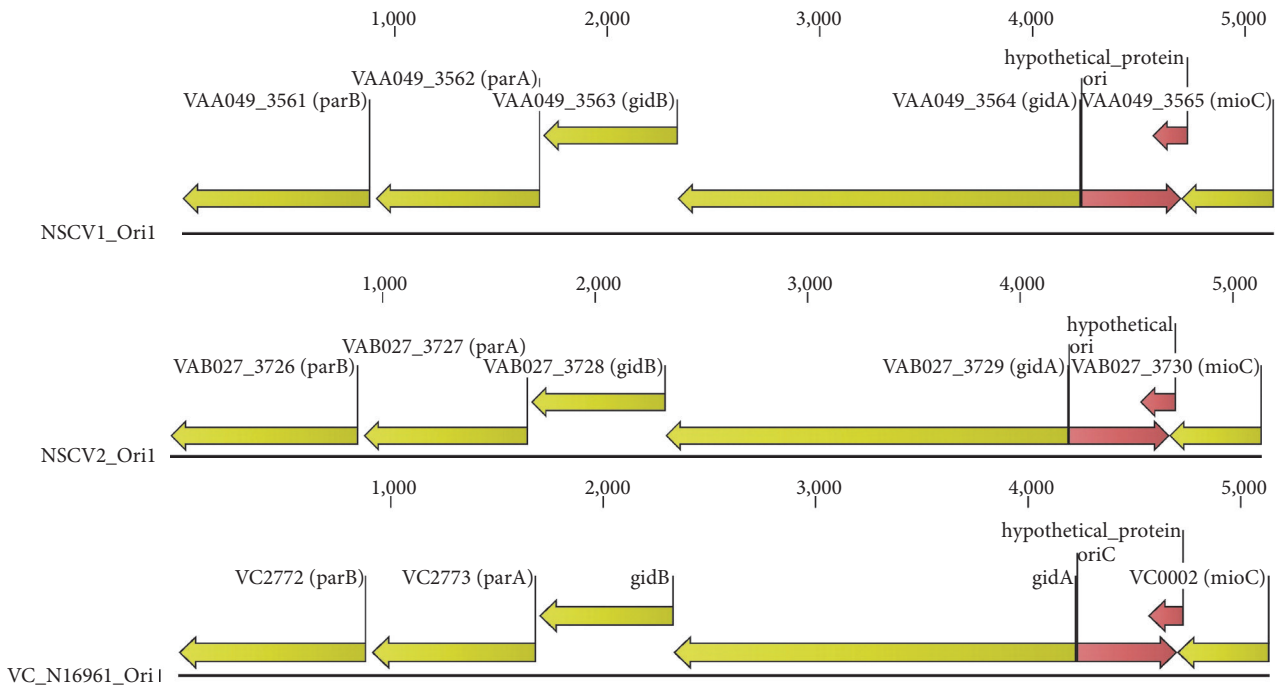


FIGURE 6: Genetic organization of *ori1* of NSCV1 and NSCV2 in comparison to the respective *ori* in N16961. The old locus tags with known gene designations have been used to indicate the ORFs. The physical *ori* and unannotated ORFs are indicated by red arrows.

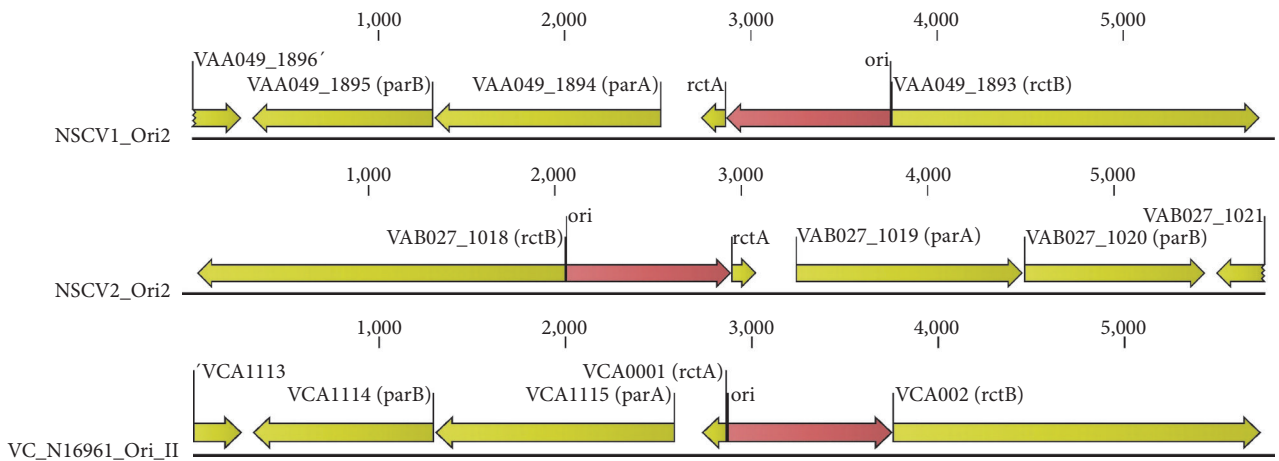


FIGURE 7: Genetic organization of *ori2* of NSCV1 and NSCV2 in comparison to the respective *ori* in N16961. The old locus tags with known gene designations have been used to indicate the ORFs. The physical *ori* is indicated by a red arrow.

4. Conclusions

All species of the genus *Vibrio* known to date harbor two chromosomes. Here, we present an exception to this rule by describing the genomic architecture of two natural *V. cholerae* isolates with one fused chromosome. For many years, the quest to understand why and how *Vibrios* evolved their bipartite genomes remains enigmatic. The strains described here appeared to have taken an evolutionary path backwards and might be instrumental in future unraveling of long-standing questions on chromosome biology in *Vibrios*. One fascinating question to address is whether the replication of the fused chromosome is dominated by one of the

subchromosomes or if they share the two replicons. If the latter is true then this would be the first example of a bacterial chromosome with two active replication origins. A tantalizing idea that has been proposed is that Chr2 is actually a “parasitic replicon.” The concept of selfish replication origins has been established recently [71]. It could be that the *V. cholerae* strains described here are the result of a selfish replicon taking over the “host.” Currently, studies are underway to begin to address these questions to unravel the mystery of a single chromosome with multiple replication origins. A second question that is worth exploring pertains to how the chromosomal fusions are maintained and if and how or under what conditions they

ever revert into two separate chromosomes. Many of the genes involved in these functions have suffered changes. It remains to be seen if NSCV1 and NSCV2 strains are defective or functionally altered in any of these genes in order to maintain the chromosomal fusions and other structural alterations.

Abbreviations

NSCV: Natural single-chromosome vibrio
 WGM: Whole genome map (optical map)
 WGS: Whole genome sequence
 PFGE: Pulse field gel electrophoresis
 LANL: Los Alamos National Labs
 JGI: Joint Genome Institute
 BRIG: Blast Ring Image Generator.

Additional Points

Availability of Supporting Data. Whole genome sequences are available in GenBank under accession numbers NSCV1-1154-74: NZ_CP010811.1 and NSCV2-10432-62: NZ_CP010812.1.

Conflicts of Interest

The authors declare no conflict of interests.

Acknowledgments

This work was supported by a grant of the Deutsche Forschungsgemeinschaft (Grant no. WA 2713/4-1) to Torsten Waldminghaus. This article is approved by LANL for unlimited release (LA-UR-17-21600).

References

- [1] A. A. Prozorov, "Additional chromosomes in bacteria: properties and origin," *Mikrobiologiya*, vol. 77, no. 4, pp. 437–447, 2008.
- [2] M. Trucksis, J. Michalski, Y. K. Deng, and J. B. Kaper, "The *Vibrio cholerae* genome contains two unique circular chromosomes," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 95, no. 24, pp. 14464–14469, 1998.
- [3] J. F. Heidelberg, J. A. Eisen, W. C. Nelson et al., "DNA sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*," *Nature*, vol. 406, no. 6795, pp. 477–483, 2000.
- [4] K. Okada, T. Iida, K. Kita-Tsukamoto, and T. Honda, "Vibrios commonly possess two chromosomes," *Journal of Bacteriology*, vol. 187, no. 2, pp. 752–757, 2005.
- [5] E. S. Egan, M. A. Fogel, and M. K. Waldor, "Divided genomes: negotiating the cell cycle in prokaryotes with multiple chromosomes," *Molecular Microbiology*, vol. 56, no. 5, pp. 1129–1138, 2005.
- [6] J. K. Jha, J. H. Baek, T. Venkova-Canova, and D. K. Chattoraj, "Chromosome dynamics in multichromosome bacteria," *Biochimica et Biophysica Acta*, vol. 1819, no. 7, pp. 826–829, 2012.
- [7] M. E. Val, A. Soler-Bistué, M. J. Bland, and D. Mazel, "Management of multipartite genomes: the *Vibrio cholerae* model," *Current Opinion in Microbiology*, vol. 22, pp. 120–126, 2014.
- [8] T. Venkova-Canova and D. K. Chattoraj, "Transition from a plasmid to a chromosomal mode of replication entails additional regulators," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 108, no. 15, pp. 6199–6204, 2011.
- [9] D. E. Cameron, J. M. Urbach, and J. J. Mekalanos, "A defined transposon mutant library and its use in identifying motility genes in *Vibrio cholerae*," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 105, no. 25, pp. 8736–8741, 2008.
- [10] M. C. Chao, J. R. Pritchard, Y. J. Zhang et al., "High-resolution definition of the *Vibrio cholerae* essential gene set with hidden Markov model-based analyses of transposon-insertion sequencing data," *Nucleic Acids Research*, vol. 41, no. 19, pp. 9033–9048, 2013.
- [11] H. D. Kamp, B. Patimalla-Dipali, D. W. Lazinski, F. Wallace-Gadsden, and A. Camilli, "Gene fitness landscapes of *Vibrio cholerae* at important stages of its life cycle," *PLoS Pathogens*, vol. 9, no. 12, article e1003800, 2013.
- [12] Q. Xu, M. Dziejman, and J. J. Mekalanos, "Determination of the transcriptome of *Vibrio cholerae* during intrainestinal growth and midexponential phase in vitro," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 100, no. 3, pp. 1286–1291, 2003.
- [13] D. S. Merrell, S. M. Butler, F. Qadri et al., "Host-induced epidemic spread of the cholera bacterium," *Nature*, vol. 417, no. 6889, pp. 642–645, 2002.
- [14] E. S. Egan and M. K. Waldor, "Distinct replication requirements for the two *Vibrio cholerae* chromosomes," *Cell*, vol. 114, no. 4, pp. 521–530, 2003.
- [15] S. Duigou, K. G. Knudsen, O. Skovgaard, E. S. Egan, A. Løbner-Olesen, and M. K. Waldor, "Independent control of replication initiation of the two *Vibrio cholerae* chromosomes by DnaA and RctB," *Journal of Bacteriology*, vol. 188, no. 17, pp. 6419–6424, 2006.
- [16] G. Demarre and D. K. Chattoraj, "DNA adenine methylation is required to replicate both *Vibrio cholerae* chromosomes once per cell cycle," *PLoS Genetics*, vol. 6, no. 5, article e1000939, 2010.
- [17] S. Duigou, Y. Yamaichi, and M. K. Waldor, "ATP negatively regulates the initiator protein of *Vibrio cholerae* chromosome II replication," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 105, no. 30, pp. 10577–10582, 2008.
- [18] T. Rasmussen, R. B. Jensen, and O. Skovgaard, "The two chromosomes of *Vibrio cholerae* are initiated at different time points in the cell cycle," *The EMBO Journal*, vol. 26, no. 13, pp. 3124–3131, 2007.
- [19] C. Stokke, T. Waldminghaus, and K. Skarstad, "Replication patterns and organization of replication forks in *Vibrio cholerae*," *Microbiology*, vol. 157, Part 3, pp. 695–708, 2011.
- [20] M. E. Val, M. Marbouty, F. de Lemos Martins et al., "A checkpoint control orchestrates the replication of the two chromosomes of *Vibrio cholerae*," *Science Advances*, vol. 2, no. 4, article e1501914, 2016.
- [21] J. H. Baek and D. K. Chattoraj, "Chromosome I controls chromosome II replication in *Vibrio cholerae*," *PLoS Genetics*, vol. 10, no. 2, article e1004184, 2014.
- [22] A. David, G. Demarre, L. Muresan, E. Paly, F. X. Barre, and C. Possoz, "The two cis-acting sites, parS1 and oriC1,

- contribute to the longitudinal organisation of *Vibrio cholerae* chromosome I,” *PLoS Genetics*, vol. 10, no. 7, article e1004448, 2014.
- [23] J. Livny, Y. Yamaichi, and M. K. Waldor, “Distribution of centromere-like parS sites in bacteria: insights from comparative genomics,” *Journal of Bacteriology*, vol. 189, no. 23, pp. 8693–8703, 2007.
- [24] Y. Yamaichi, M. A. Fogel, S. M. McLeod, M. P. Hui, and M. K. Waldor, “Distinct centromere-like parS sites on the two chromosomes of *Vibrio* spp,” *Journal of Bacteriology*, vol. 189, no. 14, pp. 5314–5324, 2007.
- [25] C. Lesterlin, F. X. Barre, and F. Cornet, “Genetic recombination and the cell cycle: what we have learned from chromosome dimers,” *Molecular Microbiology*, vol. 54, no. 5, pp. 1151–1160, 2004.
- [26] M. E. Val, S. P. Kennedy, M. El Karoui, L. Bonné, F. Chevalier, and F. X. Barre, “FtsK-dependent dimer resolution on multiple chromosomes in the pathogen *Vibrio cholerae*,” *PLoS Genetics*, vol. 4, no. 9, article e1000201, 2008.
- [27] G. Demarre, E. Galli, L. Muresan et al., “Differential management of the replication terminus regions of the two *Vibrio cholerae* chromosomes during cell division,” *PLoS Genetics*, vol. 10, no. 9, article e1004557, 2014.
- [28] T. G. Bernhardt and P. A. de Boer, “SlmA, a nucleoid-associated, FtsZ binding protein required for blocking septal ring assembly over chromosomes in *E. coli*,” *Molecular Cell*, vol. 18, no. 5, pp. 555–564, 2005.
- [29] J. Yuan, Y. Yamaichi, and M. K. Waldor, “The three *vibrio cholerae* chromosome II-encoded ParE toxins degrade chromosome I following loss of chromosome II,” *Journal of Bacteriology*, vol. 193, no. 3, pp. 611–619, 2011.
- [30] Y. Yamaichi, M. A. Fogel, and M. K. Waldor, “Par genes and the pathology of chromosome loss in *Vibrio cholerae*,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 104, no. 2, pp. 630–635, 2007.
- [31] N. Iqbal, A. M. Guérout, E. Krin, F. Le Roux, and D. Mazel, “Comprehensive functional analysis of the 18 *Vibrio cholerae* N16961 toxin-antitoxin systems substantiates their role in stabilizing the superintegron,” *Journal of Bacteriology*, vol. 197, no. 13, pp. 2150–2159, 2015.
- [32] M. E. Val, O. Skovgaard, M. Ducos-Galand, M. J. Bland, and D. Mazel, “Genome engineering in *Vibrio cholerae*: a feasible approach to address biological issues,” *PLoS Genetics*, vol. 8, no. 1, article e1002472, 2012.
- [33] M. E. Val, S. P. Kennedy, A. J. Soler-Bistué et al., “Fuse or die: how to survive the loss of dam in *Vibrio cholerae*,” *Molecular Microbiology*, vol. 91, no. 4, pp. 665–678, 2014.
- [34] T. Shimada, E. Arakawa, K. Itoh et al., “Extended serotyping scheme for *Vibrio cholerae*,” *Current Microbiology*, vol. 28, no. 3, pp. 175–178, 1994.
- [35] C. Chapman, M. Henry, K. A. Bishop-Lilly et al., “Scanning the landscape of genome architecture of non-O1 and non-O139 *Vibrio cholerae* by whole genome mapping reveals extensive population genetic diversity,” *PLoS One*, vol. 10, no. 3, article e0120311, 2015.
- [36] S. L. Johnson, A. Khiani, K. A. Bishop-Lilly et al., “Complete genome assemblies for two single-chromosome *Vibrio cholerae* isolates, strains 1154-74 (serogroup O49) and 10432-62 (serogroup O27),” *Genome Announcements*, vol. 3, no. 3, 2015.
- [37] S. Bennett, “Solexa Ltd,” *Pharmacogenomics*, vol. 5, no. 4, pp. 433–438, 2004.
- [38] M. Margulies, M. Egholm, W. E. Altman et al., “Genome sequencing in microfabricated high-density picolitre reactors,” *Nature*, vol. 437, no. 7057, pp. 376–380, 2005.
- [39] D. R. Zerbino and E. Birney, “Velvet: algorithms for de novo short read assembly using de Bruijn graphs,” *Genome Research*, vol. 18, no. 5, pp. 821–829, 2008.
- [40] B. Foster, K. LaButti, S. Trong, C. Han, T. Brettin, and A. Lapidus, “POLISHER: A tool for using ultra short reads in genome sequence improvement,” 2009, - Report Number: LBNL-2792E Poster <https://pubarchive.lbl.gov/islandora/object/ir%3A153560>.
- [41] S. Trong, K. LaButti, B. Foster, C. Han, T. Brettin, and A. Lapidus, “Gap resolution: a software package for improving Newbler genome assemblies,” in *Sequencing, Finishing, Analysis in the Future Meeting*, Santa Fe, NM, 2009.
- [42] C. Han and P. Chain, “Finishing repeat regions automatically with Dupfinisher,” in *Proceedings of the 2006 International Conference on Bioinformatics & Computational Biology*, H. R. Arabnia and H. Valafar, Eds., pp. 141–146, CSREA Press, Las Vegas, NV, USA, 2006.
- [43] P. E. Li, C. C. Lo, J. J. Anderson et al., “Enabling the democratization of the genomics revolution with a fully integrated web-based bioinformatics platform,” *Nucleic Acids Research*, vol. 45, no. 1, pp. 67–80, 2017.
- [44] R. K. Aziz, D. Bartels, A. A. Best et al., “The RAST server: rapid annotations using subsystems technology,” *BMC Genomics*, vol. 9, p. 75, 2008.
- [45] A. C. McHardy, A. Goesmann, A. Pühler, and F. Meyer, “Development of joint application strategies for two microbial gene finders,” *Bioinformatics*, vol. 20, no. 10, pp. 1622–1631, 2004.
- [46] B. E. Suzek, M. D. Ermolaeva, M. Schreiber, and S. L. Salzberg, “A probabilistic method for identifying start codons in bacterial genomes,” *Bioinformatics*, vol. 17, no. 12, pp. 1123–1130, 2001.
- [47] T. M. Lowe and S. R. Eddy, “tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence,” *Nucleic Acids Research*, vol. 25, no. 5, pp. 955–964, 1997.
- [48] B. Boeckmann, A. Bairoch, R. Apweiler et al., “The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003,” *Nucleic Acids Research*, vol. 31, no. 1, pp. 365–370, 2003.
- [49] A. Bateman, E. Birney, L. Cerruti et al., “The Pfam protein families database,” *Nucleic Acids Research*, vol. 30, no. 1, pp. 276–280, 2002.
- [50] D. H. Haft, J. D. Selengut, and O. White, “The TIGRFAMs database of protein families,” *Nucleic Acids Research*, vol. 31, no. 1, pp. 371–373, 2003.
- [51] N. J. Mulder, R. Apweiler, T. K. Attwood et al., “The InterPro database, 2003 brings increased coverage and new features,” *Nucleic Acids Research*, vol. 31, no. 1, pp. 315–318, 2003.
- [52] M. Kanehisa and S. Goto, “KEGG: kyoto encyclopedia of genes and genomes,” *Nucleic Acids Research*, vol. 28, no. 1, pp. 27–30, 2000.
- [53] R. L. Tatusov, N. D. Fedorova, J. D. Jackson et al., “The COG database: an updated version includes eukaryotes,” *BMC Bioinformatics*, vol. 4, p. 41, 2003.
- [54] S. Altschul, T. L. Madden, A. A. Schäffer et al., “Gapped BLAST and PSI-BLAST: a new generation of protein

- database search programs," *Nucleic Acids Research*, vol. 25, pp. 3389–3402, 1997.
- [55] A. L. Delcher, A. Phillippy, J. Carlton, and S. L. Salzberg, "Fast algorithms for large-scale genome alignment and comparison," *Nucleic Acids Research*, vol. 30, no. 11, pp. 2478–2483, 2002.
- [56] S. Kurtz, A. Phillippy, A. L. Delcher et al., "Versatile and open software for comparing large genomes," *Genome Biology*, vol. 5, no. 2, p. R12, 2004.
- [57] T. J. Carver, K. M. Rutherford, M. Berriman, M. A. Rajandream, B. G. Barrell, and J. Parkhill, "ACT: the Artemis Comparison Tool," *Bioinformatics*, vol. 21, no. 16, pp. 3422–3423, 2005.
- [58] S. Kurtz, A. Phillippy, A. L. Delcher et al., "Versatile and open software for comparing large genomes," *Genome Biology*, vol. 5, article R12, 2004.
- [59] G. Benson, "Tandem repeats finder: a program to analyze DNA sequences," *Nucleic Acids Research*, vol. 27, pp. 573–580, 1999.
- [60] P. E. Warburton, J. Giordano, F. Cheung, Y. Gelfand, and G. Benson, "Inverted repeat structure of the human genome: the X-chromosome contains a preponderance of large, highly homologous inverted repeats that contain testes genes," *Genome Research*, vol. 14, pp. 1861–1869, 2004.
- [61] Y. Zhou, Y. Liang, K. H. Lynch, J. J. Dennis, and D. S. Wishart, "PHAST: a fast phage search tool," *Nucleic Acids Research*, vol. 39, pp. W347–W352, 2011.
- [62] B. K. Dhillon, M. R. Laird, J. A. Shay et al., "IslandViewer 3: more flexible, interactive genomic island discovery, visualization and analysis," *Nucleic Acids Research*, vol. 43, no. W1, pp. W104–W108, 2015.
- [63] T. Carver, N. Thomson, A. Bleasby, M. Berriman, and J. Parkhill, "DNAPlotter: circular and linear interactive genome visualization," *Bioinformatics*, vol. 25, no. 1, pp. 119–120, 2009.
- [64] N. F. Alikhan, N. K. Petty, N. L. Ben Zakour, and S. A. Beatson, "BLAST ring image generator (BRIG): simple prokaryote genome comparisons," *BMC Genomics*, vol. 12, p. 402, 2011.
- [65] A. E. Darling, B. Mau, and N. T. Perna, "ProgressiveMauve: multiple genome alignment with gene gain, loss and rearrangement," *PLoS One*, vol. 5, no. 6, article e11147, 2010.
- [66] P. S. Chain, D. V. Grafham, R. S. Fulton et al., "Genomics. Genome project standards in a new era of sequencing," *Science*, vol. 326, no. 5950, pp. 236–237, 2009.
- [67] K. Okada, M. Na-Ubol, W. Natakathung et al., "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*, vol. 9, no. 6, article e98120, 2014.
- [68] K. Okada, A. Roobthaisong, W. Swaddiwudhipong, S. Hamada, and S. Chantaroj, "Vibrio cholerae O1 isolate with novel genetic background, Thailand-Myanmar," *Emerging Infectious Diseases*, vol. 19, no. 6, pp. 1015–1017, 2013.
- [69] A. C. Darling, B. Mau, F. R. Blattner, and N. T. Perna, "Mauve: multiple alignment of conserved genomic sequence with rearrangements," *Genome Research*, vol. 14, no. 7, pp. 1394–1403, 2004.
- [70] R. Barrangou, C. Fremaux, H. Deveau et al., "CRISPR provides acquired resistance against viruses in prokaryotes," *Science*, vol. 315, no. 5819, pp. 1709–1712, 2007.
- [71] M. Hawkins, S. Malla, M. J. Blythe, C. A. Nieduszynski, and T. Allers, "Accelerated growth in the absence of DNA replication origins," *Nature*, vol. 503, no. 7477, pp. 544–547, 2013.