

Genome Sequence of the Clinical O4:K12 Serotype *Vibrio parahaemolyticus* Strain 10329[∇]

N. Gonzalez-Escalona,^{1*} E. A. Strain,² A. J. De Jesús,¹ J. L. Jones,³ and A. DePaola³

Division of Microbiology¹ and Division of Mathematics,² Center for Food Safety and Applied Nutrition, Food and Drug Administration, 5100 Paint Branch Parkway, College Park, Maryland, and Gulf Coast Seafood Laboratory, Division of Seafood Science and Technology, Food and Drug Administration, Dauphin Island, Alabama³

Received 8 April 2011/Accepted 25 April 2011

***Vibrio parahaemolyticus* is the leading cause of food-borne illnesses worldwide. Here, we report a draft genome of *V. parahaemolyticus* strain 10329 of the O4:K12 serotype. It belongs to the main U.S. West Coast clonal complex of *V. parahaemolyticus* (sequence type 36 [ST36]) causing oyster-associated human illness. It contains the virulence determinants *tdh* and *trh* but appears to infect at much lower doses than *V. parahaemolyticus* strains with these same determinants from other areas, such as the U.S. Gulf and Atlantic coasts.**

Vibrio parahaemolyticus is a natural inhabitant of coastal waters and is the leading cause of seafood-borne gastroenteritis (12). Since 1996, *V. parahaemolyticus* infections and outbreaks have increased in numbers throughout the world, caused mostly by strains belonging to the *V. parahaemolyticus* pandemic clonal complex (VpPCC) (4, 9–11). The emergence of this VpPCC has elevated public health concerns regarding worldwide spread of *V. parahaemolyticus*, previously uncharacteristic for this pathogen. Serology, the historic hallmark for *V. parahaemolyticus* epidemiologic surveillance, has been unreliable in tracking the spread of this VpPCC. In addition to the original O3:K6 serotype, at least 11 other sero-variants associated with the pandemic clonal strains have been identified (2). The existence of other clonal complexes (CC) of *V. parahaemolyticus* (CC36 and CC34) other than the VpPCC were recently observed among coastal U.S. strains (5). CC36 has been linked almost exclusively to outbreaks associated with the consumption of raw oysters harvested from the U.S. Pacific Coast since the 1970s (1). Whereas CC36 has been geographically restricted, it is similar to the VpPCC in that it displays multiple serotypes (O4:K12 and O12:K12) within the ancestor ST36 and consists of at least six different sequence types (STs).

So far, only one complete genome sequence of a O3:K6 *V. parahaemolyticus* strain, RIMD2210633 (GenBank accession numbers BA000031 and BA000032 for chromosomes I and II, respectively), which belongs to the VpPCC, has been reported and deposited in GenBank (8). This is a clinical isolate carrying *tdh* and lacking both the *trh* and the *ure* cluster genes, all of which are virulence markers. Recently, a few *V. parahaemolyticus* draft genomes have become available, those for prepandemic O3:K6 strain AQ3810 (carrying *tdh* and lacking *trh* and *ure*), believed to be an ancestral strain of the VpPCC (3), and prepandemic O3:K6 strain AQ4037, VpPCC O4:K68 strain

AN-5034, strain 16, strain K5030, and strain Peru-466 (<http://www.ncbi.nlm.nih.gov/genomes/lproks.cgi>). In the present publication, we announce a draft genome sequence for *V. parahaemolyticus* serotype O4:K12 isolated from a clinical case in Washington State (5). This strain belongs to CC36, is ST36, and carries the *tdh*, *trh*, and *ure* cluster genes (5).

Genomic DNA was isolated from overnight cultures using a Wizard genomic DNA purification kit (Promega, WI). The genomes were sequenced by BGI (BGI, China) using high-throughput pair-end Illumina sequencing technology (Illumina), according to the manufacturer's instructions at a 100× coverage. Sequencing libraries were constructed according to the manufacturer's instructions (Illumina). Short reads for 10329 were assembled into genomic sequences using SOAPdenovo, an in-house assembler based on the de Bruijn graph theory, and gaps were filled using mapping information (7). A total of 9 scaffolds were obtained, with a total length of 5,113,034 bp. A total of 4742 genes were discovered, with an average GC content of 46.32% per gene. The complete *V. parahaemolyticus* strain RIMD2210633 genome sequence was employed as a reference. Sequences were annotated using the NCBI Prokaryotic Genomes Automatic Annotation Pipeline (<http://www.ncbi.nlm.nih.gov/genomes/static/Pipeline.html>) (6).

A detailed report of a full comparative analysis between 10329 and the genomes of other available *V. parahaemolyticus* isolates will be included in a future publication.

Nucleotide sequence accession number. The draft genome sequence for *V. parahaemolyticus* strain 10329 is available in GenBank under accession number AFBW00000000.

This work was supported by FDA Foods Program Intramural Funds. We thank Eric W. Brown and Steven M. Musser for their support on our genomic efforts.

REFERENCES

1. Abbott, S. L., et al. 1989. Emergence of a restricted bioserovar of *Vibrio parahaemolyticus* as the predominant cause of *Vibrio*-associated gastroenteritis on the West Coast of the United States and Mexico. *J. Clin. Microbiol.* 27:2891–2893.
2. Ansaruzzaman, M., et al. 2005. Pandemic serovars (O3:K6 and O4:K68) of

* Corresponding author. Mailing address: FDA, CFSAN, 5100 Paint Branch Parkway HFS-712, College Park, MD 20740. Phone: (301) 436-1937. Fax: (301) 436-2644. E-mail: narjol.gonzalez-escalona@fda.hhs.gov.

[∇] Published ahead of print on 6 May 2011.

- Vibrio parahaemolyticus* associated with diarrhea in Mozambique: spread of the pandemic into the African continent. *J. Clin. Microbiol.* **43**:2559–2562.
3. **Boyd, E. F., et al.** 2008. Molecular analysis of the emergence of pandemic *Vibrio parahaemolyticus*. *BMC Microbiol.* **8**:110.
 4. **Gonzalez-Escalona, N., et al.** 2005. *Vibrio parahaemolyticus* diarrhea, Chile, 1998 and 2004. *Emerg. Infect. Dis.* **11**:129–131.
 5. **Gonzalez-Escalona, N., et al.** 2008. Determination of molecular phylogenetics of *Vibrio parahaemolyticus* strains by multilocus sequence typing. *J. Bacteriol.* **190**:2831–2840.
 6. **Klimke, W., et al.** 2009. The National Center for Biotechnology Information's Protein Clusters Database. *Nucleic Acids Res.* **37**:D216–D223.
 7. **Li, R., Y. Li, K. Kristiansen, and J. Wang.** 2008. SOAP: short oligonucleotide alignment program. *Bioinformatics* **24**:713–714.
 8. **Makino, K., et al.** 2003. Genome sequence of *Vibrio parahaemolyticus*: a pathogenic mechanism distinct from that of *V. cholerae*. *Lancet* **361**:743–749.
 9. **Martinez-Urtaza, J., et al.** 2005. Pandemic *Vibrio parahaemolyticus* O3:K6, Europe. *Emerg. Infect. Dis.* **11**:1319–1320.
 10. **Matsumoto, C., et al.** 2000. Pandemic spread of an O3:K6 clone of *Vibrio parahaemolyticus* and emergence of related strains evidenced by arbitrarily primed PCR and *toxRS* sequence analyses. *J. Clin. Microbiol.* **38**:578–585.
 11. **Nair, G. B., et al.** 2007. Global dissemination of *Vibrio parahaemolyticus* serotype O3:K6 and its serovariants. *Clin. Microbiol. Rev.* **20**:39–48.
 12. **Scallan, E., et al.** 2011. Foodborne illness acquired in the United States—major pathogens. *Emerg. Infect. Dis.* **17**:7–15.