



# Draft Genome Sequence of *Pediococcus pentosaceus* Strain PP16CC, Isolated from Oyster *Crassostrea corteziensis*

 Julio A. Hernandez-Gonzalez,<sup>a</sup>  Ricardo Vazquez-Juarez,<sup>b</sup>  Jose Manuel Vazquez-Guillen,<sup>c</sup>  Carlos Rangel-Davalos,<sup>a</sup>  
 Cristina Rodriguez-Padilla,<sup>c</sup>  Maurilia Rojas<sup>a</sup>

<sup>a</sup>Laboratorio de Ciencia y Tecnología de los Alimentos, Universidad Autónoma de Baja California Sur, La Paz, B.C.S., Mexico

<sup>b</sup>Laboratorio de Genómica y Bioinformática, Centro de Investigaciones Biológicas del Noroeste, La Paz, B.C.S., Mexico

<sup>c</sup>Laboratorio de Inmunología y Virología, Facultad de Ciencias Biológicas, Universidad Autónoma de Nuevo León, Monterrey, Nuevo León, Mexico

**ABSTRACT** *Pediococcus pentosaceus* strain PP16CC comes from the intestine of *Crassostrea corteziensis*. A 1.82-Mbp draft genome of this strain was assembled using A5-miseq from illumina reads, resulting in 4 contigs and 1,856 predicted protein coding genes. Additionally, 23 proteins belonging to various glycosyl hydrolase families and 6 prophage regions were identified.

*Crassostrea corteziensis* is a potential alternative in oyster farming of Mexican Pacific (1). However, a main problem has been diseases (2); therefore, probiotic *Pediococcus pentosaceus* has been studied to improve oyster survival during seed production (3).

*P. pentosaceus* belongs to the *Lactobacillaceae* family, and in a core genome phylogenetic analysis of this family, it forms a monophyletic clade with the heterofermentative *Lactobacillus plantarum* group and the pediococci (4). The pangenome and core genome of *P. pentosaceus* have 7,938 and 1,240 genes, respectively, and the main differences between strains were found in carbohydrate metabolism and horizontally transferred DNA (5).

*P. pentosaceus* strain PP16CC was isolated on May 2011 using MRS agar at room temperature in anaerobiosis from *C. corteziensis* harvested from the Pacific Ocean and preserved with glycerol at  $-85^{\circ}\text{C}$ . DNA of the strain was extracted using the Wizard genomic DNA purification kit (Promega, USA). The quality and quantity of DNA were determined via the Quant-iT PicoGreen double-stranded DNA (dsDNA) assay kit on the Qubit 2.0 fluorometer (both from Thermo Fisher Scientific, USA). The DNA library was prepared using the Nextera DNA Flex library prep kit (Illumina) and sequenced with the MiSeq reagent kit v2 (300 cycles), yielding 627,419 paired-end reads with an average length of 151 bp. Reads were quality filtered and assembled with the A5-miseq pipeline v. 20160825 (6) and SSPACE v. 3.0 (7), resulting in a draft genome with a G+C content of 37.0235%, a total length of 1,820,443 bp in 4 contigs ( $N_{50}$  527,937 bp), and an average coverage of 84-fold.

Taxonomy was established by the Microbial Genome Atlas (MiGA) (8), determined that the PP16CC strain belongs to the genus *Pediococcus* ( $P = 0.00153$ ) and to the species *P. pentosaceus* ( $P = 0.0198$ ). And its closest relatives are *P. pentosaceus* GCA\_004354495.1 (98.86% average nucleotide identity [ANI]) and *P. pentosaceus* GCA\_001437285.1 (98.83% ANI).

The contig order was obtained via mauve contig mover (9) using the genome of *P. pentosaceus* ATCC 25745 as a reference (NC\_008525.1). The genome annotation was performed using PGAP v. 6.0 (10) and predicted 1,794 coding sequences, 7 rRNA genes, 52 tRNA genes, and 3 noncoding RNA genes.

Glycosyl hydrolase (GH) enzymes were annotated with the dbCAN metaserver (11, 12) using the carbohydrate-active-enzyme database (13), identifying 23 proteins belonging to GH families, of which some were repeated. Seven of eight families that Jiang et al. (5) found in most *P. pentosaceus* genomes were GH1 (3.2.1.86), GH25 (3.2.1.17), GH73 (Unspecified NA), GH65 (2.4.1.8), GH2 (3.2.1.23), GH126 (NA), and GH13\_29 (3.2.1.93); in addition, 5 families

**Editor** Irene L. G. Newton, Indiana University, Bloomington

**Copyright** © 2022 Hernandez-Gonzalez et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](#).

Address correspondence to Maurilia Rojas, mrojas@uabcs.mx.

The authors declare no conflict of interest.

**Received** 28 April 2022

**Accepted** 19 August 2022

**Published** 13 September 2022

**TABLE 1** General characteristics of the prophage region

Region	Region length (Kbp)	Completeness	Score	Total no. of proteins	Region position	Accession no. for the most common phage
1	14.8	Incomplete	20	20	Scaffold2 102192–117013	<a href="#">NC_031036</a> (3)
2	25.9	Intact	130	31	Scaffold2 122233–148190	<a href="#">NC_024387</a> (11)
3	29	Incomplete	40	12	Scaffold2 284517–313522	<a href="#">NC_023719</a> (2)
4	26.5	Intact	150	35	Scaffold2 425474–452037	<a href="#">NC_004746</a> (6)
5	24.1	Incomplete	30	21	Scaffold2 452054–476230	<a href="#">NC_019489</a> (2)
6	51.8	Intact	150	75	Scaffold3 236382–288269	<a href="#">NC_004305</a> (14)

were identified, namely, GH31 (3.2.1.177), GH78 (3.2.1.40), GH43\_26 (3.2.1.-), GH23 (NA), and GH170 (NA).

Integrated prophages were predicted using PHASTER (14), six prophage regions, three intact and three incomplete were identified (Table 1). The phages with the highest number of proteins (values in parenthesis, Table 1) similar to prophage regions 2, 4, and 6 were *Listeria* virus LP101, *Lactococcus* phage P335 *sensu lato*, and *Lactobacillus* phage phig1e, respectively. All three belong to the *Siphoviridae* family. Software packages were used with default parameters.

**Data availability.** The whole-genome shotgun project for *P. pentosaceus* PP16CC was deposited at DDBJ/ENA/GenBank ([JALCZR000000000](#)) and under BioProject number [PRJNA814659](#), BioSample number [SAMN26563464](#), and SRA [SRR18292892](#).

#### ACKNOWLEDGMENT

We thank CONACYT for fellowship 480473.

#### REFERENCES

- Castillo-Durán A, Chávez-Villalba J, Arreola-Lizárraga A, Barraza-Guardado R. 2010. Comparative growth, condition, and survival of juvenile *Crassostrea gigas* and *C. corteziensis* oysters cultivated in summer and winter. *Cienc Mar* 36:29–39. <https://doi.org/10.7773/cm.v36i1.1623>.
- Guo X, Ford SE. 2016. Infectious diseases of marine mollusks and host responses as revealed by genomic tools. *Philos Trans R Soc B* 371:20150206. <https://doi.org/10.1098/rstb.2015.0206>.
- Savin-Amador M, Rojas-Contreras M, Arce-Amézquita PM, Rangel-Dávalos C, Vázquez-Juárez R. 2021. *Lactobacillus* strains isolated from oysters improve the production of *Crassostrea gigas* larvae. *Lat Am J Aquat Res* 49:551–564. <https://doi.org/10.3856/vol49-issue4-fulltext-2672>.
- Zheng J, Wittouck S, Salvetti E, Franz CM, Harris H, Mattarelli P, O'Toole PW, Pot B, Vandamme P, Walter J, Watanabe K, Wuys S, Felis GE, Gänzle MG, Lebeer S. 2020. A taxonomic note on the genus *Lactobacillus*: description of 23 novel genera, emended description of the genus *Lactobacillus* Beijerinck 1901, and union of *Lactobacillaceae* and *Leuconostocaceae*. *Int J Syst Evol Microbiol* 70:2782–2858. <https://doi.org/10.1099/ijsem.0.004107>.
- Jiang J, Yang B, Ross RP, Stanton C, Zhao J, Zhang H, Chen W. 2020. Comparative genomics of *Pediococcus pentosaceus* isolated from different niches reveals genetic diversity in carbohydrate metabolism and immune system. *Front Microbiol* 11:253. <https://doi.org/10.3389/fmicb.2020.00253>.
- Coil D, Jospin G, Darling AE. 2015. A5-miseq: an updated pipeline to assemble microbial genomes from Illumina MiSeq data. *Bioinformatics* 31:587–589. <https://doi.org/10.1093/bioinformatics/btu661>.
- Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W. 2011. Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics* 27:578–579. <https://doi.org/10.1093/bioinformatics/btq683>.
- Rodríguez-R LM, Gunturu S, Harvey WT, Rosselló-Mora R, Tiedje JM, Cole JR, Konstantinidis KT. 2018. The Microbial Genomes Atlas (MiGA) webserver: taxonomic and gene diversity analysis of Archaea and Bacteria at the whole genome level. *Nucleic Acids Res* 46:W282–W288. <https://doi.org/10.1093/nar/gky467>.
- Rissman AI, Mau B, Biehl BS, Darling AE, Glasner JD, Perna NT. 2009. Reordering contigs of draft genomes using the mauve aligner. *Bioinformatics* 25:2071–2073. <https://doi.org/10.1093/bioinformatics/btp356>.
- Tatusova T, DiCuccio M, Badretdin A, Chetverin V, Nawrocki EP, Zaslavsky L, Lomsadze A, Pruitt KD, Borodovsky M, Ostell J. 2016. NCBI Prokaryotic Genome Annotation Pipeline. *Nucleic Acids Res* 44:6614–6624. <https://doi.org/10.1093/nar/gkw569>.
- Yin Y, Mao X, Yang J, Chen X, Mao F, Xu Y. 2012. dbCAN: a Web resource for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res* 40:W445–W451. <https://doi.org/10.1093/nar/gks479>.
- Zhang H, Yohe T, Huang L, Entwistle S, Wu P, Yang Z, Busk PK, Xu Y, Yin Y. 2018. dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res* 46:W95–W101. <https://doi.org/10.1093/nar/gky418>.
- Drula E, Garron ML, Dogan S, Lombard V, Henrissat B, Terrapon N. 2022. The carbohydrate-active enzyme database: functions and literature. *Nucleic Acids Res* 50:D571–D577. <https://doi.org/10.1093/nar/gkab1045>.
- Arndt D, Grant JR, Marcu A, Sajed T, Pon A, Liang Y, Wishart DS. 2016. PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Res* 44:W16–W21. <https://doi.org/10.1093/nar/gkw387>.