# Finished Genome of *Zymomonas mobilis* subsp. *mobilis* Strain CP4, an Applied Ethanol Producer

Vassili N. Kouvelis,[a] Hazuki Teshima,[b] David Bruce,[b] Chris Detter,[b] Roxanne Tapia,[b] Cliff Han,[b] Vassileia-Olga Tampakopoulou,[a] Lynne Goodwin,[b] Tanja Woyke,[b] Nikos C. Kyrpides,[c] Milton A. Typas,[a] Katherine M. Pappas[a]

Department of Genetics & Biotechnology, Faculty of Biology, National and Kapodistrian University of Athens, Panepistimiopolis, Athens, Greece[a]; DOE Joint Genome Institute, Bioscience Division, Los Alamos National Laboratory, Los Alamos, New Mexico, USA[b]; DOE Joint Genome Institute, Walnut Creek, California, USA[c]

*Zymomonas mobilis* subsp. *mobilis* is one of the most rigorous ethanol-producing organisms known to date, considered by many to be the prokaryotic alternative to yeast. The two most applied *Z. mobilis* subsp. *mobilis* strains, ZM4 and CP4, derive from Recife, Brazil, and have been isolated from sugarcane fermentations. Of these, ZM4 was the first *Z. mobilis* representative strain to be sequenced and analyzed. Here, we report the finishing of the genome sequence of strain CP4, which is highly similar but not identical to that of ZM4.

Address correspondence to Katherine M. Pappas, kmpappas@biol.uoa.gr.

**Z**ymomonas mobilis subsp. *mobilis* strain CP4, formerly known as *Z. mobilis* var. *recifensis* (1), is a most aerotolerant, quickly growing, and ethanol-yielding *Z. mobilis* strain (2, 3). CP4 and its kin strain *Z. mobilis* subsp. *mobilis* ZM4 originate from the same source at Recife, Brazil (4, 5), and are known to reach theoretical maxima of ethanol production when grown on glucose substrates (6). Both strains have undergone numerous independent genetic manipulations in order to be optimized for ligninocellulosic biomass fermentations (3, 6–8). Despite their wide applications, controversy exists as to whether these strains are different or identical; although they have distinct plasmid profiles (5), they are currently considered the same strain by bacterial repositories, including the American Type Culture Collection (ATCC 31821 [http://www.lgcstandards-atcc.org/products/all/31821]). The finishing of the genome sequence of CP4 unequivocally proves that CP4 is not identical to ZM4.

Total DNA from CP4 was prepared as described previously (9). The genome sequence was generated at the Department of Energy (DOE) Joint Genome Institute (JGI) (http://www.jgi.doe.gov/) using a combination of Sanger, Illumina (10), and 454 technologies (11). To this end, a Sanger library (average insert size of 6.6 ± 1.6 kb), a 454 Titanium standard library (395,449 reads), two paired-end 454 libraries (average insert sizes of 4.6 ± 1.1 bp and 25.1 ± 6.3 kb), and an Illumina GAII shotgun library were constructed, generating sequence reads totaling 2.1 Mb, 202.0 Mb, and 222 Mb, respectively. The 454 and Illumina data were assembled with Newbler version 2.3 and Velvet version 0.7.63, respectively (12). The Sanger reads, the 454 Newbler consensus shreds, the read pairs in the 454 paired-end library, and the Illumina Velvet consensus shreds were integrated using parallel Phrap version SPS-4.24 (High Performance Software, LLC). The software Consed (13–15) was used for finishing. The Illumina data were used to increase consensus quality using the software Polisher developed at JGI (A. Lapidus, unpublished data). Possible misas-semblies were corrected using Gap Resolution (C. Han, unpublished data), dupFinisher (16), or sequencing bridging PCR fragments after subcloning. The gaps between contigs were closed by editing in Consed, by PCR, and by Bubble PCR primer walks (J.-F. Cheng, unpublished data). The final assembly is based on data providing coverage of the genome up to 106-fold. Coding gene prediction, functional gene assignment, and tRNA/rRNA identification were determined as described before (17). Genome structure comparisons relied on ACT (18), BLASTn (19), and Mega-BLAST (20).

The genome of CP4 comprises a circular chromosome of 1,998,637 bp and five plasmids, pCP4_1 to pCP4_5, of 36,892 bp, 33,915 bp, 32,400 bp, 30,952 bp, and 30,440 bp, respectively (G+C contents of 46.24% for the chromosome and of 42.39%, 42.28%, 43.69%, 43.70%, and 42.68% for the five plasmids, respectively). It has 1,860 protein-coding genes, 48 tRNA genes, and 2 rRNA gene clusters.

The CP4 genome is 57,727 bp smaller than that of ZM4 (21) and shares syntenic units that locally reach 99% identity. However, four stretches within the CP4 chromosome totaling 20,452 bp (coordinates 110280 to 121208, 1243209 to 1246778, 1259989 to 1262157, and 1552432 to 1554025) and 18 genes are unique to the strain compared to ZM4; conversely, 15 regions totaling 74,674 bp and 58 genes are unique to ZM4. The CP4 plasmids harbor housekeeping and accessory genes (host-beneficial or other), as well as genes of phage origin, located on pCP4_1.

**Nucleotide sequence accession numbers.** The CP4 genome was assigned GenBank accession no. CP006818 for the chromosome and CP006891 to CP006895 for the plasmids.

## ACKNOWLEDGMENTS

## REFERENCES

1. **Swings J, De Ley J.** 1977. The biology of *Zymomonas*. Bacteriol. Rev. **41**:1–46.
2. **Skotnicki ML, Lee KJ, Tribe DE, Rogers PL.** 1981. Comparison of ethanol production by different *Zymomonas* strains. Appl. Environ. Microbiol. **41**:889–893.
3. **Dien BS, Cotta MA, Jeffries TW.** 2003. Bacteria engineered for fuel ethanol production: current status. Appl. Microbiol. Biotechnol. **63**:258–266. http://dx.doi.org/10.1007/s00253-003-1444-y.
4. **Gonçalves de Lima O, De Araújo JM, Schumacher IE, Cavalcanti Da Silva E.** 1970. Estudos de microrganismos antagonistas presentes nas bebidas fermentadas usadas pelo povo do Recife. I. Sôbre uma variedade de *Zymomonas mobilis* (Lindner) (1928). Rev. Inst. Antibiot, Univ. Recife **10**:3–15.
5. **Yablonsky MD, Goodman AE, Stevnsborg N, Gonçalves de Lima O, Falcão de Morais JO, Lawford HG, Rogers PL, Eveleigh DE.** 1988. *Zymomonas mobilis* CP4: a clarification of strains via plasmid profiles. J. Biotechnol. **9**:71–80. http://dx.doi.org/10.1016/0168-1656(88)90016-8.
6. **Rogers PL, Jeon YJ, Lee KJ, Lawford HG.** 2007. *Zymomonas mobilis* for fuel ethanol and higher value products. Adv. Biochem. Engin. Biotechnol. **108**:263–288. http://dx.doi.org/10.1007/10_2007_060.
7. **Zhang M, Eddy C, Deanda K, Finkelstein M, Picataggio S.** 1995. Metabolic engineering of a pentose metabolism pathway in ethanologenic *Zymomonas mobilis*. Science **267**:240–243. http://dx.doi.org/10.1126/science.267.5195.240.
8. **Joachimsthal EL, Rogers PL.** 2000. Characterization of a high-productivity recombinant strain of *Zymomonas mobilis* for ethanol production from glucose/xylose mixtures. Appl. Biochem. Biotechnol. **84–86**:343–356.
9. **Pappas KM, Galani I, Typas MA.** 1997. Transposon mutagenesis and strain construction in *Zymomonas mobilis*. J. Appl. Microbiol. **82**:379–388. http://dx.doi.org/10.1046/j.1365-2672.1997.00376.x.
10. **Bennett S.** 2004. Solexa Ltd. Pharmacogenomics **5**:433–438. http://dx.doi.org/10.1517/14622416.5.4.433.
11. **Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen YJ, Chen Z, Dewell SB, Du L, Fierro JM, Gomes XV, Godwin BC, He W, Helgesen S, Ho CH, Irzyk GP, Jando SC, Alenquer ML, Jarvie TP, Jirage KB, Kim JB, Knight JR, Lanza JR, Leamon JH, Lefkowitz SM, Lei M, Li J, Lohman KL, Lu H, Makhijani VB, McDade KE, McKenna MP, Myers EW, Nickerson E, Nobile JR, Plant R, Puc BP, Ronan MT, Roth GT, Sarkis GJ, Simons JF, Simpson JW, Srinivasan M, Tartaro KR, Tomasz A, Vogt KA, Volkmer GA, Wang SH, Wang Y, Weiner MP, Yu P, Begley RF, Rothberg JM.** 2005. Genome sequencing in microfabricated high-density picolitre reactors. Nature **437**:376–380. http://dx.doi.org/10.1038/nature03959.
12. **Zerbino DR, Birney E.** 2008. Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs. Genome Res. **18**:821–829. http://dx.doi.org/10.1101/gr.074492.107.
13. **Ewing B, Green P.** 1998. Base-calling of automated sequencer traces using Phred. II. Error Probabilities. Genome Res. **8**:186–194.
14. **Ewing B, Hillier L, Wendl MC, Green P.** 1998. Base-calling of automated sequencer traces using Phred. I. Accuracy Assessment. Genome Res. **8**:175–185. http://dx.doi.org/10.1101/gr.8.3.175.
15. **Gordon D, Abajian C, Green P.** 1998. Consed: a graphical tool for sequence finishing. Genome Res. **8**:195–202. http://dx.doi.org/10.1101/gr.8.3.195.
16. **Han CS, Chain P.** 2006. Finishing repeat regions automatically with Dupfinisher, p 141–146. *In* Arabnia HR, Valafar H (ed), Proceedings of the 2006 International Conference on Bioinformatics and Computational Biology. CSREA Press, Las Vegas, NV.
17. **Pappas KM, Kouvelis VN, Saunders E, Brettin TS, Bruce D, Detter C, Balakireva M, Han CS, Savvakis G, Kyrpides NC, Typas MA.** 2011. Genome sequence of the ethanol-producing *Zymomonas mobilis* subsp. *mobilis* lectotype strain ATCC 10988. J. Bacteriol. **193**:5051–5052. http://dx.doi.org/10.1128/JB.05395-11.
18. **Carver TJ, Rutherford KM, Berriman M, Rajandream MA, Barrell BG, Parkhill J.** 2005. ACT: the Artemis Comparison Tool. Bioinformatics **21**:3422–3423. http://dx.doi.org/10.1093/bioinformatics/bti553.
19. **Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ.** 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. **25**:3389–3402. http://dx.doi.org/10.1093/nar/25.17.3389.
20. **Zhang Z, Schwartz S, Wagner L, Miller W.** 2000. A greedy algorithm for aligning DNA sequences. J. Comput. Biol. **7**:203–214. http://dx.doi.org/10.1089/10665270050081478.
21. **Yang S, Pappas KM, Hauser LJ, Land ML, Chen GL, Hurst GB, Pan C, Kouvelis VN, Typas MA, Pelletier DA, Klingeman DM, Chang YJ, Samatova NF, Brown SD.** 2009. Improved genome annotation for *Zymomonas mobilis*. Nat. Biotechnol. **27**:893–894. http://dx.doi.org/10.1038/nbt1009-893.