JOURNAL OF BACTERIOLOGY, Sept. 2011, p. 5049–5050 Vol. 193, No. 18
0021-9193/11/$12.00 doi:10.1128/JB.05273-11

Genome Sequence of the Ethanol-Producing *Zymomonas mobilis* subsp. *pomaceae* Lectotype Strain ATCC 29192[▽]

Vassili N. Kouvelis,[1] Karen W. Davenport,[2] Thomas S. Brettin,[2] David Bruce,[3] Chris Detter,[2] Cliff S. Han,[2] Matt Nolan,[3] Roxanne Tapia,[2] Agni Damoulaki,[1] Nikos C. Kyrpides,[3] Milton A. Typas,[1] and Katherine M. Pappas[1]*

Department of Genetics & Biotechnology, Faculty of Biology, University of Athens, Panepistimiopolis, Athens 15701, Greece[1]; DOE Joint Genome Institute, Bioscience Division, Los Alamos National Laboratory, Los Alamos, New Mexico 87545[2]; and DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, California 94598[3]

Received 12 May 2011/Accepted 29 June 2011

***Zymomonas mobilis* is an alphaproteobacterium studied for bioethanol production. Different strains of this organism have been hitherto sequenced; they all belong to the *Z. mobilis* subsp. *mobilis* taxon. Here we report the finished and annotated genome sequence of strain ATCC 29192, a cider-spoiling agent isolated in the United Kingdom. ATCC 29192 is the lectotype of the second-best-characterized subspecies of *Z. mobilis*, *Z. mobilis* subsp. *pomaceae*. The nucleotide sequence of ATCC 29192 deviates from that of *Z. mobilis* subsp. *mobilis* representatives, which justifies its distinct taxonomic positioning and proves particularly useful for comparative and functional genomic analyses.**

Zymomonas mobilis is a bacterium highly promising for bioethanol production. The most robust *Z. mobilis* strains are ATCC 31821 variants—Brazilian strains ZM4 and CP4—that convert glucose to ethanol and carbon dioxide to almost theoretical yields (5, 17) and, wild-type or engineered, are involved in applications. These strains belong to the *Z. mobilis* subsp. *mobilis* taxon, and, along with other members of the subspecies—i.e., ATCC 10988 and NCIMB 11163—they have been sequenced to completion (12, 16, 19, 22). ATCC 29192 is the type strain of the *Z. mobilis* subsp. *pomaceae* taxon—formerly called *Z. anaerobia* (4)—which is the second-best known for the organism. ATCC 29192 was isolated in Bristol, United Kingdom, as a cider sickness-causing organism and, together with other members of its group, is known to exhibit distinct traits compared to its *Z. mobilis* subsp. *mobilis* counterparts—low oxygen tolerance, increased nutritional requirements, inability to utilize sucrose, low DNA hybridization relatedness, and deviant proteome electropherograms (20).

Total DNA and plasmid DNA from ATCC 29192 were prepared separately (15) and used for whole-genome shotgun sequencing at the US DOE Joint Genome Institute, via a 9.7-kb library construction (2 kb for the plasmid sample). Draft assemblies were based on 17,196 reads (9.4-fold coverage) and made use of the Phred/Phrap/Consed package (6, 7, 9). Misassemblies were corrected with Dupfinisher (10) or transposon bombing of bridging clones. Gaps were closed by editing in Consed, primer walks, and PCR amplifications. The completed sequence contains 19,257 reads, achieving a 9.5-fold and 4.9-fold coverage per base for the chromosome and plasmids, respectively (less than 1 in 100,000 errors).

Open reading frame (ORF) prediction made use of Prodigal (http://compbio.ornl.gov/prodigal/) and BLAST (1); tRNA and rRNA recognition made use of tRNAscan-SE and RNAmmer (13, 14). For functional gene assignment, translated ORFs were compared to those of the SPTR (TrEMBL) (2), Pfam (8), TIGRFAMs (18), COG (21), and KEGG (11) databases.

ATCC 29192 contains a 1,989,865-bp circular chromosome and two plasmids, p29192_1 and p29192_2, of 37,387 and 34,161 bp, respectively, with GC contents of 44.09%, 40.96%, and 44.00%, correspondingly. It has 1,777 protein-coding genes and 51 tRNA and 9 rRNA genes (3 rRNA clusters), chromosomally located. The chromosome of ATCC 29192 is 66,498 bp smaller than that of ATCC 31821 (ZM4) (22) and shares an average of 73% nucleotide identity.

Chromosomal structure comparisons between ATCC 29192 and ZM4 were performed using ACT (3), BLASTN (1), and MegaBLAST (23). Synteny is retained for the largest part, except for three regions bearing least resemblance to ZM4 (coordinates 1122645 to 1131366, 1241177 to 1312137, and 1508142 to 1523712). These include a Fe-only nitrogenase operon, reductase genes, electron transport complex genes, and a molybdenum ABC transporter operon, providing evidence for functional differences between the two strains. Based on ACT, 138 genes were unique for ATCC 29192, whereas for ZM4, 292 genes were unique. In plasmids, replication and stabilization genes were recognized, as well as genes involved in transport, regulation, transposition, and DNA modification. p29192_1 carries a CRISPR repeat region, while p29192_2 notably harbors a gene coding for a zinc-binding alcohol dehydrogenase, which hints to the importance of extrachromosomal DNA in this organism.

Nucleotide sequence accession numbers. The ATCC 29192 genome received GenBank accession numbers CP002865, CP002866, and CP002867 for the chromosome and two plasmids, in order of size.

* Corresponding author. Mailing address: Department of Genetics & Biotechnology, Faculty of Biology, University of Athens, Panepistimiopolis, Ilissia, Athens 15701, Greece. Phone: 30-210-7274-340. Fax: 30-210-7274-318. E-mail: kmpappas@biol.uoa.gr.

▽ Published ahead of print on 8 July 2011.

We are grateful to the JGI personnel who participated in the sequencing, assembly, and automated annotation processes. Special

thanks to Tanja Woyke and Lynne Goodwin for program and project managing.

Work at JGI is financed by the U.S. DOE Office of Science, contract no. DE-AC02-05CH11231. K.M.P. acknowledges the NKUA Research Committee for providing award 70/4/7809.

REFERENCES

1. **Altschul, S. F., et al.** 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. **25:**3389–3402.
2. **Boeckmann, B., et al.** 2003. The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. Nucleic Acids Res. **31:**365–370.
3. **Carver, T. J., et al.** 2005. ACT: the Artemis Comparison Tool. Bioinformatics **21:**3422–3423.
4. **De Ley, J., and J. Swings.** 1976. Phenotypic description, numerical analysis and a proposal for improved taxonomy and nomenclature of the genus *Zymomonas* Kluyver and van Niel 1936. Int. J. Syst. Bacterol. **26:**146–147.
5. **Dien, B. S., M. A. Cotta, and T. W. Jeffries.** 2003. Bacteria engineered for fuel ethanol production: current status. Appl. Microbiol. Biotechnol. **63:**258–266.
6. **Ewing, B., and P. Green.** 1998. Base-calling of automated sequencer traces using phred. II. Error probabilities. Genome Res. **8:**186–194.
7. **Ewing, B., L. Hillier, M. C. Wendl, and P. Green.** 1998. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. Genome Res. **8:**175–185.
8. **Finn, R. D., et al.** 2008. The Pfam protein families database. Nucleic Acids Res. **36**(Suppl. 1)**:**D281–D288.
9. **Gordon, D., C. Abajian, and P. Green.** 1998. Consed: a graphical tool for sequence finishing. Genome Res. **8:**195–202.
10. **Han, C. S., and P. Chain.** 2006. Finishing repeat regions automatically with Dupfinisher, p. 141–146. *In* H. R. Arabnia and H. Valafar (ed.), Proceedings of the 2006 international conference on bioinformatics and computational biology. CSREA Press, Las Vegas, NV.
11. **Kanehisa, M., et al.** 2008. KEGG for linking genomes to life and the environment. Nucleic Acids Res. **36:**D480–D484.
12. **Kouvelis, V. N., et al.** 2009. Complete genome sequence of the ethanol producer *Zymomonas mobilis* NCIMB 11163. J. Bacteriol. **191:**7140–7141.
13. **Lagesen, K., et al.** 2007. RNammer: consistent annotation of rRNA genes in genomic sequences. Nucleic Acids Res. **35:**3100–3108.
14. **Lowe, T. M., and S. R. Eddy.** 1997. tRNAscan-SE: a program for improved detection of tRNA genes in genomic sequence. Nucleic Acids Res. **25:**955–964.
15. **Pappas, K. M., I. Galani, and M. A. Typas.** 1997. Transposon mutagenesis and strain construction in *Zymomonas mobilis*. J. Appl. Microbiol. **82:**379–388.
16. **Pappas, K. M., et al.** 2011. Genome sequence of the ethanol-producing *Zymomonas mobilis* subsp. *mobilis* lectotype strain ATCC 10988. J. Bacteriol. **193:**5052–5053.
17. **Rogers, P. L., Y. J. Jeon, K. J. Lee, and H. G. Lawford.** 2007. *Zymomonas mobilis* for fuel ethanol and higher value products. Adv. Biochem. Eng. Biotechnol. **108:**263–288.
18. **Selengut, J. D., et al.** 2007. TIGRFAMs and Genome Properties: tools for the assignment of molecular function and biological process in prokaryotic genomes. Nucleic Acids Res. **35:**D260–D264.
19. **Seo, J.-S., et al.** 2005. The genome sequence of the ethanologenic bacterium *Zymomonas mobilis* ZM4. Nat. Biotechnol. **23:**63–68.
20. **Swings, J., and J. De Ley.** 1977. The biology of *Zymomonas*. Bacteriol. Rev. **41:**1–46.
21. **Tatusov, R. L., E. V. Koonin, and D. J. Lipman.** 1997. A genomic perspective on protein families. Science **278:**631–637.
22. **Yang, S., et al.** 2009. Improved genome annotation for *Zymomonas mobilis*. Nat. Biotechnol. **27:**893–894.
23. **Zhang, Z., S. Schwartz, L. Wagner, and W. Miller.** 2000. A greedy algorithm for aligning DNA sequences. J. Comput. Biol. **7:**203–214.