

Genome Sequence of the Ethanol-Producing *Zymomonas mobilis* subsp. *mobilis* Lectotype Strain ATCC 10988[∇]

Katherine M. Pappas,^{1*} Vassili N. Kouvelis,¹ Elizabeth Saunders,² Thomas S. Brettin,² David Bruce,³ Chris Detter,² Mariya Balakireva,² Cliff S. Han,² Giannis Savvakis,¹ Nikos C. Kyrpides,³ and Milton A. Typas¹

Department of Genetics & Biotechnology, Faculty of Biology, University of Athens, Panepistimiopolis, Athens 15701, Greece¹; DOE Joint Genome Institute, Bioscience Division, Los Alamos National Laboratory, Los Alamos, New Mexico 87545²; and DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, California 94598³

Received 26 May 2011/Accepted 22 June 2011

***Zymomonas mobilis* ATCC 10988 is the type strain of the *Z. mobilis* subsp. *mobilis* taxon, members of which are some of the most rigorous ethanol-producing bacteria. Isolated from *Agave cactus* fermentations in Mexico, ATCC 10988 is one of the first *Z. mobilis* strains to be described and studied. Its robustness in sucrose-substrate fermentations, physiological characteristics, large number of plasmids, and overall genomic plasticity render this strain important to the study of the species. Here we report the finishing and annotation of the ATCC 10988 chromosomal and plasmid genome.**

Zymomonas mobilis is a natural fermentative bacterium that has been studied for decades due to its remarkable ethanologenic properties. Among isolates deriving from various parts of the world, ATCC 10988, identified in the 1920s in Mexican pulque fermentations (initially called “*Thermobacterium mobile*”) (17), is one of the longest-studied *Zymomonas* strains and has served for most biochemical and physiological works on the organism. ATCC 10988 exhibits fermenting performances that equal that of industrial strain CP4 in sucrose-rich media (15) and has also attracted attention due to its split genome: it bears multiple different plasmids, six of which are reported in this work, while the smallest have been cloned and sequenced in earlier works (2, 10, 12). Most species-specific vectors constructed and used for *Z. mobilis*, indeed, derive from the small multicopy plasmids of ATCC 10988 (16, 19, 21).

Total and plasmid DNA from ATCC 10988 was prepared (11) and used for whole-genome shotgun sequencing at the US DOE Joint Genome Institute after a 8.9-kb DNA library was constructed. Draft assemblies were based on 8,950 reads, providing 5.4× coverage. In addition to Sanger sequencing, 454 pyrosequencing was done to 20× coverage. The Phred/Phrap/Consed software package was used for sequence assembly and quality assessment, while after the shotgun stage, misassemblies and gaps were corrected as previously reported (7). The completed genome of ATCC 10988 contains 10,920 reads, achieving an average of 6-fold sequence coverage per base and an error rate of less than 1 in 100,000.

Coding gene prediction and annotation were performed using Prodigal (<http://compbio.ornl.gov/prodigal/>) and BLAST (1); tRNA and rRNA genes were identified with tRNAscan-SE and RNAmmer (8, 9). Functional gene assignment was based on the SPTR (TrEMBL) (3), Pfam (5), TIGRFAMs (13),

COG (18), and KEGG (6) databases. Genome structure comparisons relied on ACT (4), BLASTN (1), and MegaBLAST (22).

ATCC 10988 contains a single, circular chromosome of 2,021,773 bp and six plasmids reported in this work, p10988_1 to p10988_6, of 32,479 bp, 32,283 bp, 31,692 bp, 18,461 bp, 4,023 bp, and 2,750 bp, respectively. The chromosome and plasmids have GC contents of 46.22% and 43.46%, 45.41%, 43.23%, 41.79%, 37.63% and 41.31%, correspondingly. The entire genome contains 1,875 protein-encoding genes and 48 tRNA and 6 rRNA genes (two rRNA clusters) in the chromosome.

The chromosome is 34,590 bp smaller than that of ATCC 31821 (ZM4) (20), with an average of 98% sequence identity. Four large syntenic units, shuffled in order and direction compared to ZM4, are discerned (ranging from 732 kb to 179 kb). In their junctions identical transposase genes are met. In total, 36 transposase genes are found in this organism, a number far exceeding those for the *Z. mobilis* strains sequenced so far. Additionally, 54 indels are found scattered, the largest of which is 25.3 kb in size and corroborates previous microarray findings comparing ZM4 to ATCC 10988 (14). The plasmids harbor genes implicated in replicon sustenance and in transport, regulation, transposition, DNA restriction/modification, and metabolism. They share variable regions of homology, albeit no full-length identity, with plasmids from previously sequenced *Z. mobilis* strains.

Nucleotide sequence accession numbers. The ATCC 10988 genome received GenBank accession numbers CP002850 for the chromosome and CP002851 to CP002856 for the six plasmids.

We acknowledge collaborators at JGI that participated in the sequencing, assembly, and automated annotation processes. Special thanks to Tanja Woyke and Lynne Goodwin for help with the project and support.

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.

* 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 1 July 2011.

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. Arvanitis, N., et al. 2000. Characterization and replication properties of the *Zymomonas mobilis* ATCC 10988 plasmids pZMO1 and pZMO2. *Plasmid* **44**:127–137.
3. Boeckmann, B., et al. 2003. The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic Acids Res.* **31**:365–370.
4. Carver, T. J., et al. 2005. ACT: the Artemis Comparison Tool. *Bioinformatics* **21**:3422–3423.
5. Finn, R. D., et al. 2008. The Pfam protein families database. *Nucleic Acids Res.* **36**(Suppl. 1):D281–D288.
6. Kanehisa, M., et al. 2008. KEGG for linking genomes to life and the environment. *Nucleic Acids Res.* **36**:D480–D484.
7. Kouvelis, V. N., et al. 2011. Genome sequence of the ethanol-producing *Zymomonas mobilis* subsp. *pomaceae* lectotype strain ATCC 29192. *J. Bacteriol.* **193**:5050–5051.
8. Lagesen, K., et al. 2007. RNAMmer: consistent annotation of rRNA genes in genomic sequences. *Nucleic Acids Res.* **35**:3100–3108.
9. 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.
10. Misawa, N., K. Nakamura, and K. Kitamura. 1985. Three 1.7 kilobase pair plasmids in *Zymomonas mobilis* NRRL B-806. *Agric. Biol. Chem.* **49**:2769–2771.
11. Pappas, K. M., I. Galani, and M. A. Typas. 1997. Transposon mutagenesis and strain construction in *Zymomonas mobilis*. *J. Appl. Microbiol.* **82**:379–388.
12. Scordaki, A., and C. Drainas. 1987. Analysis of natural plasmids of *Zymomonas mobilis* ATCC 10988. *J. Gen. Microbiol.* **133**:2547–2556.
13. 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.
14. Seo, J.-S., et al. 2005. The genome sequence of the ethanologenic bacterium *Zymomonas mobilis* ZM4. *Nat. Biotechnol.* **23**:63–68.
15. Skotnicki, M. L., K. J. Lee, D. E. Tribe, and P. L. Rogers. 1981. Comparison of ethanol production by different *Zymomonas* strains. *Appl. Environ. Microbiol.* **41**:889–893.
16. Sprenger, G. A., M. A. Typas, and C. Drainas. 1993. Genetics and genetic engineering of *Zymomonas mobilis*. *World J. Microbiol. Biotechnol.* **9**:17–24.
17. Swings, J., and J. De Ley. 1977. The biology of *Zymomonas*. *Bacteriol. Rev.* **41**:1–46.
18. Tatusov, R. L., E. V. Koonin, and D. J. Lipman. 1997. A genomic perspective on protein families. *Science* **278**:631–637.
19. Tonomura, K., T. Okamoto, M. Yasui, and H. Yanase. 1986. Shuttle vectors for *Zymomonas mobilis*. *Agric. Biol. Chem.* **50**:805–808.
20. Yang, S., et al. 2009. Improved genome annotation for *Zymomonas mobilis*. *Nat. Biotechnol.* **27**:893–894.
21. Zhang, M., C. Eddy, K. Deanda, M. Finkelstein, and S. Picataggio. 1995. Metabolic engineering of a pentose metabolism pathway in ethanologenic *Zymomonas mobilis*. *Science* **267**:240–243.
22. Zhang, Z., S. Schwartz, L. Wagner, and W. Miller. 2000. A greedy algorithm for aligning DNA sequences. *J. Comput. Biol.* **7**:203–214.