**Complete Genome Sequence of the Ethanol-Producing *Zymomonas mobilis* subsp. *mobilis* Centrotpe ATCC 29191**

Andreas Desiniotis, Vassili K. Kouvelis, Karen Davenport, David Bruce, Chris Detter, Roxanne Tapia, Cliff Han, Lynne A. Goodwin, Tanja Woyke, Nikos C. Kyrpides, Milton A. Typas, and Katherine M. Pappas

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

*Zymomonas mobilis* is an ethanologenic bacterium that has been studied for use in biofuel production. Of the sequenced *Zymomonas* strains, ATCC 29191 has been described as the phenotypic centrotpe of *Zymomonas mobilis* subsp. *mobilis*. ATCC 29191 was isolated in Kinshasa, Congo, from palm wine fermentations. This strain is reported to be a robust levulin producer, while in recent years it has been employed in studies addressing *Z. mobilis* respiration. Here we announce the finishing and annotation of the ATCC 29191 genome, which comprises one chromosome and three plasmids.

*Z. mobilis* is an alphaproteobacterium that ferments sugars to ethanol and carbon dioxide to almost-perfect yields (14, 15). A comparative analysis of several strains of the species is under way at the U.S. DOE Joint Genome Institute, in collaboration with the University of Athens (http://www.jgi.gov). Strain ATCC 29191 (26; NCIMB 11199) was included in the analysis, as it is regarded the most representative strain of *Z. mobilis* subsp. *mobilis* (4). ATCC 29191 was isolated from Zairian *Elaeis* sp. sap fermentations (18), and it is superior to other *Z. mobilis* strains in levulin (polyfructan) production. It is also highly comparable to the biotechnologically important fast-growing *Zymomonas* sp. ATCC 31821 derivatives in overall growth and yields, particularly on sucrose substrates (16). Notably, energy metabolism in *Z. mobilis*, i.e., electron transport and oxidative phosphorylation, has been intensely studied with this strain (9, 10, 17).

Total and plasmid DNA from ATCC 29191 were prepared as described previously (12) and used for whole-genome shotgun sequencing at the DOE JGI, using a combination of Illumina (2) and 454 (11) technologies. For this, we constructed an Illumina GAII library (generating 6,353,828 sequence reads, totaling 228.7 Mb) and two 454 libraries, a Titanium standard and a paired-end library with a 22-kb average insert size (515,697 and 168,806 reads, respectively, totaling 121.6 Mb) (http://www.jgi.doe.gov). The 454 and Illumina data were assembled with Newbler version 2.3 and Velvet version 0.7.63, correspondingly (20). Final data integration made use of parallel phrap, version SPS 4.24 (High Performance Software, LLC). Illumina data were used to correct base errors and increase consensus quality using Polisher (A. Lapidus, unpublished data). Misassemblies were corrected by using gapResolution or Dupfinisher (8; C. Han, unpublished data), or sequencing bridging PCR fragments. Gaps between contigs were closed by editing in Consed (5, 6, 7), by PCR and by Bubble PCR primer walks (J.-F. Cheng, unpublished data). A total of 217 additional reactions and 2 Shatter libraries closed gaps and raised the quality to 0.00 errors per 10 kb. The final assembly was based on 38.3× and 115.8× genome coverage for the 454 and Illumina data, respectively. Coding gene prediction, functional gene assignment, and tRNA/rRNA gene identification were conducted as described before (13). Genome structure comparisons relied on ACT (3), BLASTN (1), and MegaBLAST (21).

ATCC 29191 contains a circular chromosome of 1,961,307 bp and three plasmids, p29191_1 to p29191_3, of 18,350 bp, 14,947 bp, and 13,742 bp, respectively (GC contents of 46.21% and of 41.02%, 42.19%, and 44.21%, correspondingly). The entire genome has 1,765 protein-coding genes, 51 tRNA genes, and 3 rRNA gene clusters.

The ATCC 29191 genome is 95,057 bp smaller than that of reference strain ATCC 31821 (ZM4) (19) and shares an average 97% identity with it. Synteny is retained for the largest part, with the exception of local shuffling, plausibly due to transposase presence (27 annotated transposases in total). Forty-one genes—a assigned, pseudogenes, or hypothetical—are unique to ATCC 29191 compared to ZM4, whereas for the latter, 115 genes are unique. Many of the ATCC 29191-unique genes are located on three stretches bearing the lowest resemblance to ZM4 (coordinates 110991 to 115129, 1298096 to 1304030, and 1861489 to 1864804), including helicase, transporter, and tellurium resistance genes. The plasmids harbor replicon maintenance, metabolism, regulation, transposition, and DNA restriction/modification genes.

**Nucleotide sequence accession numbers.** The ATCC 29191 genome was assigned GenBank accession numbers CP003704 for the chromosome and CP003705 to CP003707 for the plasmids.

**ACKNOWLEDGMENTS**

We sincerely acknowledge all DOE-JGI collaborators who contributed to the sequencing, assembly, and automated annotation.

Work at the DOE JGI is financed by the U.S. DOE Office of Science, contract number DE-AC02-05CH11231. K.M.P. thanks the NKUA Research Committee for award 70/4/7809. A.D. thanks the Vienna Medical Society/Sanofi for the Wilhelm-Auerswald-Preis 2012 award.

**REFERENCES**