JOURNAL OF BACTERIOLOGY, Nov. 2009, p. 7140–7141 Vol. 191, No. 22
0021-9193/09/$12.00 doi:10.1128/JB.01084-09

GENOME ANNOUNCEMENTS

Complete Genome Sequence of the Ethanol Producer *Zymomonas mobilis* NCIMB 11163▽

Vassili N. Kouvelis,[1] Elizabeth Saunders,[2] Thomas S. Brettin,[2] David Bruce,[3] Chris Detter,[2] Cliff Han,[2] 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 17 August 2009/Accepted 9 September 2009

***Zymomonas mobilis* is an ethanol-producing alphaproteobacterium currently considered a major candidate organism for bioethanol production. Here we report the finished and annotated genome sequence of *Z. mobilis* subsp. *mobilis* strain NCIMB 11163, a British ale-infecting isolate. This is the first *Z. mobilis* strain whose genome, chromosomal and plasmid, is presented in its entirety.**

Zymomonas mobilis is a bacterium vigorously studied as a platform organism for bioethanol production in North America and other parts of the world. *Z. mobilis* converts sugars such as glucose or sucrose into ethanol and carbon dioxide to almost theoretical yields and to rates higher than those of yeasts (17). Genetically engineered strains that ferment pentoses in addition to naturally utilized hexoses also hold great promise for use in lignocellulosic biomass degradations (5, 22). Besides ethanol, *Z. mobilis* can produce other high-value chemicals such as sorbitol, levan, or phenylacetylcarbinol and has attracted interest for its unusual membrane steroid content (11). Lastly, *Zymomonas* is regarded as a safe organism and is even used for medicinal purposes (12, 20), which further facilitates its employment in large-scale biotechnological endeavors.

The chromosomal sequence of the *Z. mobilis* subsp. *mobilis* industrial strain ATCC 31821 (ZM4) was recently published (19). Here we announce the first entire genome sequence of a *Z. mobilis* subsp. *mobilis* strain, that of the United Kingdom-originating strain NCIMB 11163 (B70) (20). Total DNA from NCIMB 11163 (16) was used for whole-genome shotgun sequencing at the U.S. DOE Joint Genome Institute. For this, an 8.7-kb DNA library and 454 and Solexa reads were used (http://www.jgi.doe.gov). Draft assemblies were based on 8,551 Sanger reads and 454 pyrosequencing to 20× coverage, whereas the Phred/Phrap/Consed software package was used for sequence assembly and quality assessment (6, 7, 9; http://www.phrap.com). After the shotgun stage, reads were assembled with parallel Phrap (High Performance Software, LLC), and misassemblies were corrected with Dupfinisher (10) or transposon bombing of bridging clones (Epicentre Biotechnologies, Madison, WI). A total of 144 primer walk reactions, five transposon bomb libraries, 53 PCR end reads, and two PCR shatter libraries were necessary to close gaps, resolve repetitive regions, and raise the quality of the finished sequence. The completed genome sequence of NCIMB 11163 was based on 11,048 reads, with an error rate of less than 6 bp out of 100,000 bp.

Open reading frame prediction and annotation were performed using Prodigal (http://compbio.ornl.gov/prodigal/) and BLAST (1); tRNAscan-SE and RNAmmer (14, 15) were used for tRNA and rRNA recognition, respectively. Functional assignment of genes was performed by searching translated open reading frames against sequences in the SPTR (TrEMBL) (2), Pfam (8), TIGRFAMs (18), COG (21), and KEGG (13) databases.

Z. mobilis NCIMB 11163 contains a single, circular chromosome of 2,124,771 bp and three plasmids, p11163_1, p11163_2, and p11163_3 of 53,380 bp, 40,818 bp, and 4,551 bp, respectively. The overall GC content of the chromosome is 46.83%, whereas those of the plasmids are 42.32%, 43.80%, and 36.37%, respectively. The entire genome of NCIMB 11163 contains 1,884 protein-encoding genes and 51 tRNA and nine rRNA genes, which are chromosomally located.

The chromosome of NCIMB 11163 is 68,355 bp larger than that of ZM4 (GenBank accession number NC_006526) (19) and colinear at its largest part with that of ZM4 (genome structure comparisons were performed using ACT) (3). It bears several unique regions, among which are two genomic islands of ca. 25 and 79 kb, with no detectable nucleotide homology to same-species sequences and high regional similarity to chromosomal stretches of *Paracoccus denitrificans* PD1222 (GenBank accession number CP000489.1), *Xanthobacter autotrophicus* Py2 (GenBank accession number CP000781.1), and *Gluconacetobacter diazotrophicus* PAl 5 (GenBank accession number CP001189.1). Genome plasticity in NCIMB 11163 is further indicated by the presence of a type IV secretion system on the 79-kb island, syntenous to the

* 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 18 September 2009.

Agrobacterium tumefaciens Ti (IncRh1) conjugal *trb* system (4), and also by multiple transposase and phage-related genes.

In plasmids, housekeeping genes implicated in replication, active partitioning, and plasmid addiction are recognized, as well as genes involved in metabolism, transport, regulation, transposition, and DNA modification. Most notably, p11163_1 bears an arsenical resistance operon inserted in a type II secretion locus, whereas p11163_2, otherwise homologous to the 41-kb ZM4 plasmid (GenBank accession number AY057845), harbors a unique ca. 12-kb CRISPR insertion that interrupts nucleotide colinearity with the aforementioned replicon.

Nucleotide sequence accession numbers. The genome sequence was deposited in GenBank, with accession number CP001722 for the chromosome and accession numbers CP001723, CP001724, and CP001725 for the plasmids.

We acknowledge the JGI personnel that participated in sequencing, assembly, and automated annotation. Special thanks to Lynne Goodwin and Nikos Kyrpides for help and support.

K.M.P. acknowledges the NKUA Research Committee for providing award 70/4/7809.

REFERENCES

1. **Altschul, S. F., T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman.** 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. **25:**3389–3402.
2. **Boeckmann, B., A. Bairoch, R. Apweiler, M. C. Blatter, A. Estreicher, E. Gasteiger, M. J. Martin, K. Michoud, C. O'Donovan, I. Phan, S. Pilbout, and M. Schneider.** 2003. The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. Nucleic Acids Res. **31:**365–370.
3. **Carver, T. J., K. M. Rutherford, M. Berriman, M. A. Rajandream, B. G. Barrell, and J. Parkhill.** 2005. ACT: the Artemis Comparison Tool. Bioinformatics **21:**3422–3423.
4. **Christie, P. J.** 2001. Type IV secretion: intercellular transfer of macromolecules by systems ancestrally related to conjugation machines. Mol. Microbiol. **40:**294–305.
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., J. Tate, J. Mistry, P. C. Coggill, S. J. Sammut, H.-R. Hotz, G. Ceric, K. Forslund, S. R. Eddy, E. L. L. Sonnhammer, and A. Bateman.** 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. **Hermans, M. A., B. Neuss, and H. Sahm.** 1991. Content and composition of hopanoids in *Zymomonas mobilis* under various growth conditions. J. Bacteriol. **173:**5592–5595.
12. **Jeffries, T. W.** 2005. Ethanol fermentation on the move. Nature **23:**40–41.
13. **Kanehisa, M., M. Araki, S. Goto, M. Hattori, M. Hirakawa, M. Itoh, T. Katayama, S. Kawashima, S. Okuda, T. Tokimatsu, and Y. Yamanishi.** 2008. KEGG for linking genomes to life and the environment. Nucleic Acids Res. **36:**D480–D484.
14. **Lagesen, K., P. F. Hallin, E. A. Rodland, H. H. Staerfeldt, T. Rognes, and D. W. Ussery.** 2007. RNAmmer: consistent annotation of rRNA genes in genomic sequences. Nucleic Acids Res. **35:**3100–3108.
15. **Lowe, T. M., and S. R. Eddy.** 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. **25:**955–964.
16. **Pappas, K. M., I. Galani, and M. A. Typas.** 1997. Transposon mutagenesis and strain construction in *Zymomonas mobilis*. J. Appl. Microbiol. **82:**379–388.
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., D. H. Haft, T. Davidsen, A. Ganapathy, M. Gwinn-Giglio, W. C. Nelson, A. R. Richter, and O. White.** 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., H. Chong, H. S. Park, K.-O. Yoon, C. Jung, J. J. Kim, J. H. Hong, H. Kim, J.-H. Kim, J.-I. Kil, C. J. Park, H.-M. Oh, J.-S. Lee, S.-J. Jin, H.-W. Um, H.-J. Lee, S.-J. Oh, J. Y. Kim, H. L. Kang, S. Y. Lee, K. J. Lee, and H. S. Kang.** 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. **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.