JOURNAL OF BACTERIOLOGY, Dec. 2010, p. 6492–6493 Vol. 192, No. 24
0021-9193/10/$12.00 doi:10.1128/JB.01023-10

Genome Sequence of the Polysaccharide-Degrading, Thermophilic Anaerobe *Spirochaeta thermophila* DSM 6192▽

Angel Angelov,[1]† Susanne Liebl,[1,2] Meike Ballschmiter,[1]‡ Mechthild Bömeke,[1,2]
Rüdiger Lehmann,[2] Heiko Liesegang,[2] Rolf Daniel,[1,2] and Wolfgang Liebl[1,3]*

Department of Genomic and Applied Microbiology[1] and Göttingen Genomics Laboratory,[2] Institute of Microbiology and Genetics, Georg-August-Universität, Grisebachstr. 8, D-37077 Goettingen, Germany, and Department of Microbiology, Technische Universität München, Emil-Ramann-Str. 4, D-85354 Freising, Germany[3]

Received 27 August 2010/Accepted 28 September 2010

***Spirochaeta thermophila* is a thermophilic, free-living anaerobe that is able to degrade various α- and β-linked sugar polymers, including cellulose. We report here the complete genome sequence of *S. thermophila* DSM 6192, which is the first genome sequence of a thermophilic, free-living member of the *Spirochaetes* phylum. The genome data reveal a high density of genes encoding enzymes from more than 30 glycoside hydrolase families, a noncellulosomal enzyme system for (hemi)cellulose degradation, and indicate the presence of a novel carbohydrate-binding module.**

The currently known members of the *Spirochaeta* genus are anaerobes or facultative anaerobes and have been isolated from a variety of aquatic habitats such as the sediments and the water columns of ponds, lakes, rivers, and oceans. Among the isolated free-living members of the genus, only *Spirochaeta thermophila* has been reported to be able to grow on cellulose and hemicellulose, the two main components of plant biomass (1). The type and components of the enzyme machinery that perform cellulose and xylan breakdown in *S. thermophila* are unknown at this point in time.

Genomic DNA of *S. thermophila* DSM 6192 was used to create a shotgun library (2 to 3 kbp) and a fosmid library (30 to 40 kbp). The Sanger sequencing data from plasmid and fosmid end sequencing was combined with that of a sequencing-by-synthesis reaction (454 Sequencing and Genome Sequencer FLX454; both from 454 Life Sciences, Roche Applied Science). The Sanger sequencing- and the 454 Sequencing-generated traces were assembled into contigs with the Newbler Assembler software (454 Life Sciences, Roche) and the PHRAP assembly tool (3, 4), respectively, and were edited with the Staden software package (10). Open reading frames (ORFs) likely to encode proteins were predicted by the YACOB software package (11). Automatic and manual annotation was accomplished with the ERGO annotation tool (Integrated Genomics, Arlington Heights, IL), which was refined by searches against Pfam, PROSITE, ProDom, and COGs databases. Searches for glycoside hydrolases (GH), glycosyl transferases (GT) and carbohydrate-binding modules (CBM) were performed using HMMER3 hmmscan (2) of the *S. thermophila* ORFs against a custom subset of the Pfam 24.0 database (5), containing 145 Pfam-A families.

The genome of *S. thermophila* DSM 6192 consists of a 2,472,645-bp large single circular chromosome and has an average G+C content of 61.9%. A total of 2,207 protein-coding genes were predicted, covering 93.5% of the chromosome. A tentative function could be assigned to 1,265 ORFs, and 309 ORFs had no detectable similarity to other proteins in the public databases. The genome contains 46 tRNA genes and 2 rRNA gene operons.

Existing biochemical data and our genome data for *S. thermophila* suggest the presence of an electron bifurcation pathway, which may be regarded as an additional means of energy conservation (6, 7, 9). In this pathway, both reduced electron carriers NADH and ferredoxin are coupled to the generation of H_2 by a heterotrimeric [FeFe] hydrogenase, as has been shown for the homologous enzyme from *Thermotoga maritima* (8).

S. thermophila is able to grow on various carbohydrate polymers, including microcrystalline cellulose, as the single carbon and energy source. The genome of *S. thermophila* was found to encode a large number of glycoside hydrolases (72 ORFs), the coding density for these enzymes being among the highest known (29.1 ORFs/Mbp). Metabolic pathways for the utilization of the main and accessory components of various α- and β-linked glycan polymers were reconstructed. The genome data indicate that cellulose and hemicellulose degradation in *S. thermophila* is accomplished by a noncellulosomal enzyme system.

From the genome and biochemical data, we have indications for a novel carbohydrate-binding module (module X) which was found in seven *S. thermophila* GH ORFs. The nearly identical module X-like sequences were found to be linked with different GH family domains. The lack of sequence homologs of module X in the public databases may be due to the small number of sequences currently known from free-living spirochetes.

---

* Corresponding author. Mailing address: Lehrstuhl für Mikrobiologie, Technische Universität München, Emil-Ramann-Straße 4, D-85354 Freising-Weihenstephan, Germany. Phone: 49-8161-715450. Fax: 49-8161-715475. E-mail: wliebl@wzw.tum.de.

† Present address: Department of Microbiology, Technische Universität München, Emil-Ramann-Straße 4, D-85354 Freising-Weihenstephan, Germany.

‡ Present address: NWG Weisse Biotechnologie, Universität Leipzig, Deutscher Platz 5b, D-04103 Leipzig, Germany.

▽ Published ahead of print on 8 October 2010.

Nucleotide sequence accession number. Genome information for the chromosome of *S. thermophila* DSM6192 has been deposited in GenBank under accession number CP001698.

We thank Beate Schumacher for providing excellent technical assistance and Markus Mientus, Christoph Loderer, and Susanne Pompei for their support.

REFERENCES

1. **Aksenova, H., F. Rainey, P. Janssen, G. Zavarzin, and H. Morgan.** 1992. *Spirochaeta thermophila* sp. nov., an obligately anaerobic, polysaccharolytic, extremely thermophilic bacterium. Int. J. Syst. Bacteriol. **42:**175–177.
2. **Eddy, S. R.** 2008. A probabilistic model of local sequence alignment that simplifies statistical significance estimation. PLoS Comput. Biol. **4:**e1000069.
3. **Ewing, B., and P. Green.** 1998. Base-calling of automated sequencer traces using phred. II. Error probabilities. Genome Res. **8:**186–194.
4. **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.
5. **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. Sonnhammer, and A. Bateman.** 2008. The Pfam protein families database. Nucleic Acids Res. **36:**D281–D288.
6. **Herrmann, G., E. Jayamani, G. Mai, and W. Buckel.** 2008. Energy conservation via electron-transferring flavoprotein in anaerobic bacteria. J. Bacteriol. **190:**784–791.
7. **Janssen, P. H., and H. W. Morgan.** 1992. Glucose catabolism by *Spirochaeta thermophila* RI 19.B1. J. Bacteriol. **174:**2449–2453.
8. **Schut, G. J., and M. W. W. Adams.** 2009. The iron-hydrogenase of *Thermotoga maritima* utilizes ferredoxin and NADH synergistically: a new perspective on anaerobic hydrogen production. J. Bacteriol. **191:**4451–4457.
9. **Seedorf, H., W. F. Fricke, B. Veith, H. Brüggemann, H. Liesegang, A. Strittmatter, M. Miethke, W. Buckel, J. Hinderberger, F. Li, C. Hagemeier, R. K. Thauer, and G. Gottschalk.** 2008. The genome of *Clostridium kluyveri*, a strict anaerobe with unique metabolic features. Proc. Natl. Acad. Sci. U. S. A. **105:**2128–2133.
10. **Staden, R., K. F. Beal, and J. K. Bonfield.** 2000. The Staden package. Methods Mol. Biol. **132:**115–130.
11. **Tech, M., and R. Merkl.** 2003. YACOP: enhanced gene prediction obtained by a combination of existing methods. In Silico Biol. **3:**441–451.