Genome Sequence of the Ruminal Bacterium *Megasphaera elsdenii*

Hans Marx,[1] Alexandra Bettina Graf,[2,4] Nadine Elpida Tatto,[4] Gerhard Günther Thallinger,[3,4]
Diethard Mattanovich,[1,4]* and Michael Sauer[1,4]

*Department of Biotechnology, BOKU University of Natural Resources and Life Sciences, Muthgasse 18, 1190 Vienna, Austria[1];
School of Bioengineering, University of Applied Sciences, FH-Campus Wien, Muthgasse 56, 1190 Vienna, Austria[2];
Institute for Genomics and Bioinformatics, Graz University of Technology, Petersgasse 14, A-8010 Graz,
Austria[3]; and Austrian Centre of Industrial Biotechnology, Muthgasse 11, 1190 Vienna,
and Petersgasse 14, A-8010 Graz, Austria[4]*

Megasphaera elsdenii **is a Gram-negative ruminal bacterium. It is being investigated as a probiotic supplement for ruminants as it may provide benefits for energy balance and animal productivity. Furthermore, it is of biotechnological interest due to its capability of producing various volatile fatty acids. Here we report the complete genome sequence of *M. elsdenii* DSM 20460, the type strain for the species.**

The anaerobic Gram-negative coccus *Megasphaera elsdenii* is found in cattle, sheep, and other ruminants. Elsden et al. (2, 3) were the first to isolate this strain, and they have already described its capability to produce a variety of volatile fatty acids. This organism is interesting for the chemical industry as a possible biocatalyst, but characterization of its metabolism is also important for understanding the function of the rumen. The spectrum of short-chain carboxylic acids that is produced depends largely on the carbon source used by the microorganism. However, the metabolic pathways which underlie these observations have been unclear until now. Furthermore, carboxylic acids can also be used as carbon sources. For example, lactic acid is among the most preferred substrates of *M. elsdenii*. This renders this bacterium a very beneficial member of the rumen community. With the uptake of lactic acid, *M. elsdenii* can relieve acidosis, a dreaded condition of livestock (1).

The genome of *Megasphaera elsdenii* was sequenced with a combination of next-generation sequencing methods. A first-draft assembly (Roche 454 GS, FLX Titanium; 773,553 reads with a total of 164.6 Mb; 68-fold coverage) generated with Newbler 2.5.3 consisted of 56 contigs, which could be joined into 1 scaffold. To improve the quality of the sequence by eliminating the 454 sequencing errors in homopolymer stretches, the genome was subsequently sequenced using the Illumina paired-end method (HiSeq 2000; 13,481,796 reads with a total of 1.35 Gb; 558-fold coverage). The Illumina reads were aligned to the already-assembled scaffold with the Genomics Workbench 4.7.1 program (CLC, Aarhus, Denmark). The final consensus sequence was derived by counting instances of each nucleotide at a position and then letting the majority decide the nucleotide in the consensus sequence.

The annotation was performed using Prodigal gene finder (5), tRNAscan-SE 1.21 (7), and RNAmmer 1.2 (6). Additionally, the origin of replication was predicted with OriginX (11), and the genome was scanned against Rfam to find other small RNA species. Functional annotation of the predicted genes was performed using a reciprocal best-hit strategy (8) against a group of phylogenetically related organisms. In addition, the putative proteins were searched against Uniprot, Clusters of Orthologous Groups (COG) (9), Pfam (4), and Superfam (10) databases.

The draft genome includes 2,474,718 bases, with a GC content of 53%. The number of putative genes totals 2,220, with an average GC content in the coding regions of 54%. There are seven instances of the ribosomal 5S-23S-16S cluster, and 64 predicted tRNAs of which one is a pseudogene. In the functional annotation of the predicted genes, 72% had an ortholog (determined by reciprocal blast), and 94% had a homolog in the Uniprot database. Furthermore, for 88% a significant protein family was found when searching against Pfam, and for 75% a significant protein superfamily was found when searched against the Superfamily database. The sequences annotated by COG fell into 18 of the 25 functional COG classes (C to M, O to T, and V).

Nucleotide sequence accession number. The genome sequence for *M. elsdenii* DSM 20460 has been deposited at EMBL under the accession number HE576794.

REFERENCES

1. **Aikman, P. C., P. H. Henning, D. J. Humphries, and C. H. Horn.** 2011. Rumen pH and fermentation characteristics in dairy cows supplemented with Megasphaera elsdenii NCIMB 41125 in early lactation. J. Dairy Sci. **94:**2840–2849.
2. **Elsden, S. R., F. M. Gilchrist, D. Lewis, and B. E. Volcani.** 1956. Properties of a fatty acid forming organism isolated from the rumen of sheep. J. Bacteriol. **72:**681–689.
3. **Elsden, S. R., F. M. Gilchrist, D. Lewis, and B. E. Volcani.** 1951. The formation of fatty acids by a Gram-negative coccus. Biochem. J. **49:**lxix–lxx.
4. **Finn, R. D., et al.** 2010. The Pfam protein families database. Nucleic Acids Res. **38:**D211–D22.
5. **Hyatt, D., et al.** 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics **11:**119.
6. **Lagesen, K., et al.** 2007. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res. **35:**3100–3108.
7. **Lowe, T. M., and S. R. Eddy.** 1997. tRNAscan-SE: a program for improved

* Corresponding author. Mailing address: Department of Biotechnology, BOKU University of Natural Resources and Life Sciences, Muthgasse 18, 1190 Vienna, Austria. Phone: 43 1 47654-6200. Fax: 43 1 3697-615. E-mail: diethard.mattanovich@boku.ac.at.

detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. **25:**955–964.

8. **Moreno-Hagelsieb, G., and K. Latimer.** 2008. Choosing BLAST options for better detection of orthologs as reciprocal best hits. Bioinformatics **24:**319–324.

9. **Tatusov, R. L., et al.** 2003. The COG database: an updated version includes eukaryotes. BMC Bioinformatics **4:**41.

10. **Wilson, D., M. Madera, C. Vogel, C. Chothia, and J. Gough.** 2007. The SUPERFAMILY database in 2007: families and functions. Nucleic Acids Res. **35:**D308–D313.

11. **Worning, P., L. J. Jensen, P. F. Hallin, H. H. Staerfeldt, and D. W. Ussery.** 2006. Origin of replication in circular prokaryotic chromosomes. Environ. Microbiol. **8:**353–361.