



# First Insights into the Genome Sequence of the Alkaliphilic Thermotolerant Bacterium *Clostridium thermoalcaliphilum* JW/YL23-2<sup>T</sup>

 Anja Poehlein,<sup>a</sup> Alexandra Berg,<sup>b</sup> Gina Welsing,<sup>b</sup>  Rolf Daniel<sup>a</sup>

Genomic and Applied Microbiology & Göttingen Genomics Laboratory, Institute of Microbiology and Genetics, Georg-August University of Göttingen, Göttingen, Germany<sup>a</sup>; Members of the Applied Bioinformatics in Microbiology Course of the Microbiology and Biochemistry M.Sc./Ph.D. program, Georg-August University of Göttingen, Göttingen, Germany<sup>b</sup>

**ABSTRACT** *Clostridium thermoalcaliphilum* is an obligate anaerobic and rod-shaped bacterium isolated from sewage sludge. It is an alkaliphilic thermotolerant organism and utilizes sucrose, glucose, fructose, maltose, cellobiose, amino acids, and Casamino Acids as substrates. The draft genome comprises 2.031 Mbp and 2,027 predicted protein-coding genes.

*Clostridium thermoalcaliphilum* is rod-shaped and motile, with 2 to 12 peritrichous flagella. It was isolated from the Atlanta municipal sewage plant (Atlanta, GA, USA) together with its close relative, *Clostridium paradoxum*. A 16S rRNA-based analysis revealed a 2% evolutionary distance between the two species (1).

Genomic DNA of *C. thermoalcaliphilum* JW/YL23-2<sup>T</sup> was isolated using the MasterPure complete DNA purification kit, as recommended by the supplier (Epicentre, Madison, WI, USA). Illumina paired-end sequencing libraries were generated from the extracted DNA according to the protocol of the manufacturer (Illumina, San Diego, CA, USA). Sequencing was performed with a MiSeq instrument and MiSeq reagent kit version 3, as recommended by the manufacturer (Illumina). Sequencing resulted in 2,691,200 paired-end reads that were trimmed using Trimmomatic version 0.36 (2). Genome assembly with SPAdes version 3.10.0 (3) resulted in 51 contigs (>500 bp) and an average coverage of 272×. For validation of the assembly, Qualimap version 2.1 was used (4). The size of the draft genome and the G+C content were 2.031 Mb and 30.99%, respectively. Compared to other clostridial species, the genome size of *C. thermoalcaliphilum* is relatively small. The software tool Prokka (5) was used for automatic gene prediction and automatic annotation. The draft genome contains 12 rRNA genes, 63 tRNA genes, 1 transfer-messenger RNA (tmRNA) gene, 1,563 protein-coding genes with predicted function, and 464 genes encoding hypothetical proteins. The genome of *C. thermoalcaliphilum* harbors all necessary genes coding for proteins involved in the degradation of glycine and betaine. The corresponding gene cluster shows similarity to the corresponding ones identified in *Sporomusa ovata* (6), *Peptoclostridium acidaminophilum* (7), and *Clostridium tepidiprofundii* (8). We could also identify a proline reductase gene cluster exhibiting structural similarity to the cluster found in *P. litorale* (9). Moreover, we detected all genes (*selABCD*), including the tRNA, necessary for incorporation of selenocysteine into proteins (10). In addition, a putative Na<sup>+</sup>-translocating F-type ATPase was present in the *C. thermoalcaliphilum* genome. In *Clostridium paradoxum*, a homologue of this F-type ATPase was analyzed. It was shown that this enzyme functions strictly as a Na<sup>+</sup> exporter to establish an electrochemical gradient for driving cellular processes (11). Further, we detected that a ferredoxin-NAD<sup>+</sup> oxidoreductase (*Rnf*) is encoded by the genome of *C. thermoalcaliphilum*. This enzyme is putatively

Received 24 March 2017 Accepted 27 March 2017 Published 18 May 2017

**Citation** Poehlein A, Berg A, Welsing G, Daniel R. 2017. First insights into the genome sequence of the alkaliphilic thermotolerant bacterium *Clostridium thermoalcaliphilum* JW/YL23-2<sup>T</sup>. *Genome Announc* 5:e00368-17. <https://doi.org/10.1128/genomeA.00368-17>.

**Copyright** © 2017 Poehlein et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Rolf Daniel, [rdaniel@gwdg.de](mailto:rdaniel@gwdg.de).

involved in sodium efflux (11, 12). *C. thermoalcaliphilum* utilizes sucrose, glucose, fructose, maltose, cellobiose, and Casamino Acids as substrates. The fermentation products comprise the acids acetate and lactate (1). Correspondingly, putative genes coding for acetate kinase, phosphate acetyltransferase, and L-lactate dehydrogenase were present in the genome of *C. thermoalcaliphilum*.

**Accession number(s).** This whole-genome shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession no. [MZGW00000000](https://doi.org/10.1093/genomeA.00734-13). The version described here is version MZGW01000000.

## ACKNOWLEDGMENTS

We thank Melanie Heinemann for technical support. The work was partly supported by the Bundesministerium für Bildung und Forschung (BMBF). The funders (BMBF) had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

## REFERENCES

- Li Y, Engle M, Weiss N, Mandelco L, Wiegel J. 1994. *Clostridium thermoalcaliphilum* sp. nov., an anaerobic and thermotolerant facultative alkaliphile. *Int J Syst Bacteriol* 44:111–118. <https://doi.org/10.1099/00207713-44-1-111>.
- Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>.
- Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 19:455–477. <https://doi.org/10.1089/cmb.2012.0021>.
- García-Alcalde F, Okonechnikov K, Carbonell J, Cruz LM, Götz S, Tarazona S, Dopazo J, Meyer TF, Conesa A. 2012. Qualimap: evaluating next-generation sequencing alignment data. *Bioinformatics* 28:2678–2679. <https://doi.org/10.1093/bioinformatics/bts503>.
- Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30:2068–2069. <https://doi.org/10.1093/bioinformatics/btu153>.
- Poehlein A, Gottschalk G, Daniel R. 2013. First insights into the genome of the Gram-negative, endospore-forming organism *Sporomusa ovata* strain H1 DSM 2662. *Genome Announc* 1(5):e00734-13. <https://doi.org/10.1128/genomeA.00734-13>.
- Poehlein A, Andreesen JR, Daniel R. 2014. Complete genome sequence of the amino acid-utilizing *Eubacterium acidaminophilum* al-2 (DSM 3953). *Genome Announc* 2(3):e00573-14. <https://doi.org/10.1128/genomeA.00573-14>.
- Poehlein A, Friedrich I, Krüger L, Daniel R. 2016. First insights into the genome of the moderately thermophilic bacterium *Clostridium tepidiprofundii* SG 508<sup>T</sup>. *Genome Announc* 4(3):e00379-16. <https://doi.org/10.1128/genomeA.00379-16>.
- Poehlein A, Alghaithi HS, Chandran L, Chibani CM, Davydova E, Dharmotharan K, Ge W, Gutierrez-Gutierrez DA, Jagirdar A, Khonsari B, Nair KPPR, Daniel R. 2014. First insights into the genome of the amino acid-metabolizing bacterium *Clostridium litorale* DSM 5388. *Genome Announc* 2(4):e00754-14. <https://doi.org/10.1128/genomeA.00754-14>.
- Sawers G, Heider J, Zehelein E, Böck A. 1991. Expression and operon structure of the *sel* genes of *Escherichia coli* and identification of a third selenium-containing formate dehydrogenase isoenzyme. *J Bacteriol* 173:4983–4993. <https://doi.org/10.1128/jb.173.16.4983-4993.1991>.
- Ferguson SA, Keis S, Cook GM. 2006. Biochemical and molecular characterization of a Na<sup>+</sup>-translocating F1Fo-ATPase from the thermoalkaliphilic bacterium *Clostridium paradoxum*. *J Bacteriol* 188:5045–5054. <https://doi.org/10.1128/JB.00128-06>.
- Lancaster WA, Utturkar SM, Poole FL, Klingeman DM, Elias DA, Adams MWW, Brown SD. 2016. Near-complete genome sequence of *Clostridium paradoxum* strain JW-YL-7. *Genome Announc* 4(3):e00229-16. <https://doi.org/10.1128/genomeA.00229-16>.