Genome Sequence of the Fish Brain Bacterium *Clostridium tarantellae*

Luca Bano,a Matthias Kiel,b Gabriele Sales,c Andrew C. Doxey,d Michael J. Mansfield,d Haleluya T. Wami,b Marco Schiavone,e Ornella Rossetto,a Marco Pirazzini,e Ulrich Dobrindt,b Cesare Montecuccob,f

a Microbiology Laboratory, Istituto Zooprofilattico Sperimentale delle Venezie, Villorba, Italy
b Institute of Hygiene, University of Muenster, Muenster, Germany
c Department of Biology, University of Padua, Padua, Italy
d Department of Biology, University of Waterloo, Waterloo, Canada
e Department of Biomedical Sciences, University of Padua, Padua, Italy
f CNR Institute of Neuroscience, Padua, Italy

ABSTRACT *Eubacterium tarantellae* was originally cultivated from the brain of fish affected by twirling movements. Here, we present the draft genome sequence of *E. tarantellae* DSM 3997, which consists of 3,982,316 bp. Most protein-coding genes in this strain are similar to genes of *Clostridium* bacteria, supporting the renaming of *E. tarantellae* as *Clostridium tarantellae*.

The brain is generally considered a bacterium-free organ that can be infected only after alteration of the blood-brain barrier. However, Udey et al. (1) reported the unique isolation of a bacterium from the brain of apparently healthy fish displaying a rapid turning-around movement (twirling). This behavioral change is most likely to be the consequence of an alteration of the brain circuits that control body movements. Accordingly, this bacterium was named *Eubacterium tarantellae*, described as a non-spore-forming anaerobe and suggested to be the first of a novel class of anaerobic fish pathogenic bacteria (2). No further characterization was reported, though it was isolated from fish larval gut flora (3) and recently associated with a human case of joint septic arthritis (4).

*E. tarantellae* DSM 3997 was cultivated in synthetic thioglycolate broth under anaerobiosis. DNA was automatically extracted (Microlab STARlet; Hamilton, Switzerland) using a MagMAX total nucleic acid isolation kit (Ambion/Life Technologies, USA). Whole-genome sequencing was performed using an Illumina MiSeq sequencing platform. To prepare 500-bp paired-end libraries of *E. tarantellae*, we used the Nextera XT DNA library preparation kit (Illumina, USA). Libraries were sequenced using V2 sequencing chemistry. Quality trimming of 2,896,670 raw reads (1,448,335 pairs) was performed using Sickle v1.33 (https://github.com/najoshi/sickle) with the options –l 30 and –q 28 and resulted in 2,213,220 single reads corresponding to 1,106,610 paired-end reads. De novo assembly of the processed reads was performed with SPAdes v3.10.1 using default parameters (5). The assembly resulted in 713 contigs, validated using QualiMap v2.1, with an *N*₅₀ contig size of 11,440 bp. The draft genome consists of 3,979,886 bp with a 25.8% GC content and an average coverage of 149×. The genome assembly was estimated to be 98.5% complete using CheckM (6) with only 5% redundancy.

Automatic gene prediction was performed using PGAP v4.10 (www.ncbi.nlm.nih.gov/genome/annotation_prok/) (7). The draft genome encodes 3,069 predicted protein-coding genes, including 2,274 genes with a predicted function and 795 genes coding for hypothetical proteins. Additionally, 11 rRNA and 72 tRNA genes were predicted. Furthermore, one contig includes a clustered regularly interspaced short
FIG 1 Comparison of the genome content of clostridial species based on the presence/absence of genes. The pan-genome presence/absence analysis of 71 Clostridium strains was performed using Roary v3.12.0 with default parameters. The generated tree was visualized using FriPan (https://github.com/drpowell/FriPan).
palindromic repeat (CRISPR)-associated (Cas) operon including putative genes homologous to Cas3, Cas5, Cas7, and Cas8b/Csh1 from other clostridia.

Phylogenetic analysis of taxonomic markers (8) showed that the *E. tarantellae* genome is closely related to that of *Clostridium perfringens*, followed by those of other *Clostridium* species. Pan-genome analysis based on the presence/absence of genes in publicly available clostridial genome sequences and in the *C. tarantellae* DSM 3997 genome was performed with Roary v3.12.0 (8) with default parameters and corroborated that *C. perfringens* and *C. tarantellae* are indeed closely related (Fig. 1). Due to the peculiar association between *E. tarantellae* and fish brain/movement alteration, we searched the genome for potential neurotoxins possibly responsible for inducing twirling using the virulence factor database (VFDB) (9). No genes related to clostridial neurotoxins were found. However, a gene encoding α-perfringolysin-like pore-forming toxin from *C. perfringens* not known to induce alteration of movements is present. In addition, several putative toxins similar to the *Clostridium septicum* epsilon toxin ETX were identified. Furthermore, three leukocidins/hemolysins similar to the beta-channel-forming toxin NetE of *C. perfringens* and some clostridial-type phospholipases were predicted to be encoded in the *E. tarantellae* genome. Small cysteine-rich proteins that encode possible antimicrobial peptides or peptide toxins were also identified (e.g., CEFHDAMA_03104 and CEFHDAMA_00788) and found to be conserved in other clostridia. In addition, though *E. tarantellae* was not found to sporulate (1, 2), the genes encoding proteins involved in the execution of the sporogenes program are present. The high similarity of the present genome to those of clostridial species strongly supports the suggestion that *Eubacterium tarantellae* should be renamed *Clostridium tarantellae* (10).

**Data availability.** This whole-genome shotgun project has been deposited at DDBJ/ENA/GenBank under the accession number WHJC00000000. The version described in this paper is version WHJC01000000. The raw reads are available at the Sequence Read Archive under the accession number PRJNAS87720.

**ACKNOWLEDGMENTS**

We thank K. Tegelkamp (Münster) for excellent technical support. This work was supported by the University of Padua, by the IZTV, and by the German Research Foundation (grants SFB 1009 and TP 805). We thank the reviewers for their improvements to the paper.

**REFERENCES**