Complete Genome Sequence of the Dog Commensal and Human Pathogen Capnocytophaga canimorsus Strain 5

Pablo Manfredi,1 Marco Pagni,2 and Guy R. Cornelis1,*

Biozentrum, University of Basel, Basel, Switzerland,1 and Vital-IT Group, Swiss Institute of Bioinformatics, Lausanne, Switzerland2

Received 21 July 2011/Accepted 25 July 2011

Capnocytophaga canimorsus is a commensal Gram-negative bacterium, originally isolated from a dog’s mouth, that causes septicemia in humans. C. canimorsus has the unusual ability to feed on host cells, including phagocytes. This capacity depends on surface-exposed glycan-foraging systems. Here we present the first complete genome sequence of a C. canimorsus strain (Cc5).

Capnocytophaga canimorsus are commensal Gram-negative bacteria that are present in the mouths of dogs and that cause septicemia in humans who have become infected as a result of contact with a dog (24). Capnocytophaga is exclusively and commonly found in the oral cavities of mammals. This genus belongs to the family Flavobacteriaceae in the order Bacteroidales (3). These organisms escape the innate immune defenses of the host (19, 22, 23) and have the unusual ability to feed on the glycan moieties of glycoproteins present on the surfaces of mammalian cells, including phagocytes (14). This property is based on a sialidase and specialized complexes of the Sus family, mainly composed of surface-exposed lipoproteins (15, 16, 21). Here, we present the genome sequence of C. canimorsus Cc5, a strain that was isolated from a human septicemia patient (22). It is the first report of a C. canimorsus genome sequence.

Genome sequencing included pair-end reads from 4-kb- and 40-kb-insert libraries, a run of 454 pyrosequencing and a set of Illumina microreads (3.5, 0.5, 34, and 49× coverage, respectively). 454 reads were condensed into pseudoreads using Newbler and assembled with Sanger data with Phred/Phrnap/Consed (6–8). After gap closure, microreads were aligned onto the chromosome to increase coverage using MAQ (12). CDSs were predicted with Glimmer 3.02 (5), and start codons were reevaluated considering bacterial frequencies of different start codons, potential signal peptides computed by LipoP (10), and CDS alignments with their best Blast hit against the nonredundant database (1) (http://www.ncbi.nlm.nih.gov/). CDSs were screened with InterProScan (26) and PRIAM (4) for functional annotation. For poorly characterized CDSs, PSI-BLAST (1) was used with uniref90 for matrix computation and Swiss-Prot, TREMBL (2), or STRING orthologous groups (9) for data retrieval. Noncoding RNAs were inferred from Rfam using INFERNAL (20), RNAMMER (11), and tRNAscan-SE (13). The origin of replication was determined on the basis of the cumulative GC skew and on the occurrence of degenerated DnaA box clusters.

The genome of Cc5 consists of a single circular replicon of 2,571,406 bp with a G+C content of 36.11%, and it contains 2,405 CDSs. This genome size is similar to that of the human-hosted species Capnocytophaga ochracea (NC_013162, 2.6 Mb) (17) but noticeably smaller than that of other members of the Bacteroidetes, such as the free-living species Flavobacterium johnsoniae (6.1 Mb) (18) or the commensal species Bacteroides thetaiotaomicro (6.25 Mb) (25). The Cc5 genome contains 46 tRNAs, three sets of rRNA, an RNase P, two tmRNAs, a TFP riboswitch, and an SRP, and it contains one CRISPR region. It does not encode any type III, IV, or VI secretion system, which are commonly linked to pathogenesis. Consistently with the presence of several Sus-like systems, the Cc5 genome contains 206 lipoprotein genes (8.5% of total CDSs), some potentially acquired from eukaryotes or Gram-positive organisms. This high proportion of lipoprotein genes is unusual among the Eubacteria but typical of the Bacteroidetes. The genome encodes the LolACDE lipoprotein export system but, as for all members of the Bacteroidetes studied to date, lacks a LolB homolog.

Nucleotide sequence accession number. The annotated genome sequence of Cc5 was deposited in GenBank under accession number CP002115.

We thank H. Shin and M. Mally for their early contribution to this project. This project was supported by the Swiss National Science Foundation (grant 3100A0-128659) and the Novartis Stiftung.

REFERENCES


* Corresponding author. Mailing address: Biozentrum, Klingelbergstrasse 50, CH-4056, Basel. Phone: (41) 61 267 21 10. Fax: (41) 61 267 21 18. E-mail: guy.cornelis@unibas.ch.