Complete genome sequence of *Capnocytophaga ochracea* type strain (VPI 2845\(^T\))

Konstantinos Mavromatis\(^1\), Sabine Gronow\(^2\), Elizabeth Saunders,\(^{1,3}\) Miriam Land,\(^{1,4}\) Alla Lapidus\(^1\), Alex Copeland\(^1\), Tijana Glavina Del Rio\(^1\), Matt Nolan\(^1\), Susan Lucas\(^1\), Feng Chen\(^1\), Hope Tice\(^1\), Jan-Fang Cheng\(^1\), David Bruce,\(^{1,3}\) Lynne Goodwin,\(^{1,3}\) Sam Pitluck\(^1\), Amrita Pati\(^1\), Natalia Ivanova\(^1\), Amy Chen,\(^5\) Krishna Palaniappan,\(^5\) Patrick Chain\(^1,6\), Loren Hauser,\(^{1,4}\) Yun-Juan Chang,\(^1,4\), Cynthia C. Jeffries,\(^{1,4}\) Thomas Brettin,\(^{1,3}\) John C. Detter,\(^{1,3}\) Cliff Han,\(^{1,3}\) James Bristow\(^1\), Markus Göker\(^2\), Manfred Rohde\(^3\), Jonathan A. Eisen,\(^{1,8}\) Victor Markowitz,\(^5\) Nikos C. Kyrpides,\(^1\) Hans-Peter Klenk\(^2\), and Philip Hugenholtz*\(^1\)

1 DOE Joint Genome Institute, Walnut Creek, California, USA
2 DSMZ - German Collection of Microorganisms and Cell Cultures GmbH, Braunschweig, Germany
3 Los Alamos National Laboratory, Bioscience Division, Los Alamos, New Mexico USA
4 Oak Ridge National Laboratory, Oak Ridge, TN, USA
5 Biological Data Management and Technology Center, Lawrence Berkeley National Laboratory, Berkeley, CA, 94720, USA
6 Lawrence Livermore National Laboratory, Livermore, California 94550, USA
7 Helmholtz Centre for Infection Research, Braunschweig, Germany
8 University of California Davis Genome Center, Davis, California, USA

*Corresponding author: Philip Hugenholtz

### Keywords

gliding, capnophilic, periodontitis, gingivitis, *Flavobacteriaceae*

### Abstract

*Capnocytophaga ochracea* (Prévot *et al.* 1956) Leadbetter *et al.* 1982 is the type species of the genus *Capnocytophaga*. It is of interest because of its location in the *Flavobacteriaceae*, a genomically yet uncharted family within the order *Flavobacteriales*. The species grows as fusiform to rod shaped cells which tend to form clumps and are able to move by gliding. *C. ochracea* is known as a capnophilic organism with the ability to grow under anaerobic as well as under aerobic conditions (oxygen concentration larger than 15%), here only in the presence of 5% CO\(_2\). Strain VPI 2845\(^T\), the type strain of the species, is portrayed in this report as a gliding, Gram-negative bacterium, originally isolated from a human oral cavity. Here we describe the features of this organism, together with the complete genome sequence, and annotation. This is the first completed genome sequence from the flavobacterial genus *Capnocytophaga*, and the 2,612,925 bp long single replicon genome with its 2193 protein-coding and 59 RNA genes is a part of the *Genomic Encyclopedia of Bacteria and Archaea* project.

### Introduction

*Capnocytophaga ochracea* strain VPI 2845\(^T\) (DSM 7271 = ATCC 27872 = JCM 12966, and other strain collections) is the type strain of the species, which represents the type species of the genus *Capnocytophaga*. *C. ochracea* was first described by Prévot *et al.* [1] as ‘*Fusiformis nucleatus* var. *ochraceus*’ and later renamed by Leadbetter *et al.* [2]. Other synonyms for *C. ochracea* are ‘*Bacteroides oralis* var. *elongatus*’ [3], ‘*Bacteroides ochraceus*’ (basonym) [4] and "*Ristella ochraceus*" (sic) [5]. The organism is of significant interest for its...
position in the tree of life where the genus *Capnocytophaga* (8 species) is located within the large family of the *Flavobacteriaceae*. First, Leadbetter *et al.* placed the genus *Capnocytophaga* (Fig. 1) in the family of the *Cytophagaceae* within the order *Cytophagales* [6] which was emended in 2002 by the Subcommittee on the taxonomy of *Flavobacterium* and *Cytophaga*-like bacteria of the International Committee on Systematics of Prokaryotes [7]. *C. ochracea* is most often found in association with animal and human hosts. In general, it is a normal inhabitant of the human mouth and other nonoral sites. *C. ochracea* is associated with juvenile and adult periodontitis [8, 9] and may cause severe infections in immunocompromised as well as in immunocompetent patients [10-12]. Among these are endocarditis, endometritis, osteomyelitis, abscesses, peritonitis, and keratitis.

Here we present a summary classification and a set of features for *C. ochracea* VPI 2845\textsuperscript{T} (Table 1), together with the description of the complete genomic sequencing and annotation.

**Classification and features of organism**

Genbank lists 16S rRNA sequences for only a few cultivated strains belonging to *C. ochraceae*, all of them isolated from human oral sources (e.g. U41351, U41353, DQ012332). Phylotypes (sequences form uncultivated bacteria) closely linked to *C. ochracea* also originate in almost all cases from human oral samples collected from European, American, Asian and African samples (AF543292, AF543298, AY278613, AM420149, AY429469, FJ470418), except for two bacterial clones isolated from *Strongylocentrotus intermedicus* (sea urchin) in the Sea of Japan (EU432412, EU432438), and from *Oncorhynchus mykiss* (rainbow trout) caught in Scotland (AM179907). Screening of environmental genomic samples and surveys reported at the NCBI BLAST server indicated no closely related phylotypes (>91% sequence identity) that can be linked to the species or genus.

Figure 1 shows the phylogenetic neighborhood of *C. ochracea* VPI 2845\textsuperscript{T} in a 16S rRNA based tree. All four 16S rRNA gene copies in the genome of strain VPI 2845\textsuperscript{T} are identical, but differ by two nucleotides from the previously published 16S rRNA sequence (U41350) generated from ATCC 27872.

**Figure 1.** Phylogenetic tree highlighting the position of *C. ochracea* VP 2845\textsuperscript{T} relative to the other type strains within the genus *Capnocytophaga* and to selected type strains of other genera within the *Flavobacteriaceae*. The tree was inferred from 1405 aligned characters [13, 14] of the 16S rRNA gene sequence under the maximum likelihood criterion [15] and rooted with *Joostella* and *Galibacter*. The branches are scaled in terms of the expected number of substitutions per site. Numbers above branches are support values from 1000 bootstrap replicates if larger than 60%. Lineages with type strain genome sequencing projects registered in GOLD [16] are shown in blue, published genomes in bold.
C. ochracea is Gram-negative, has no flagellae and is motile by gliding (Fig. 2). Cells are pigmented and the name 'ochracea' is derived from the yellow colour shown by harvested cell mass [6]. It is a catalase- and oxidase-negative species. C. ochracea is usually susceptible to a number of antibiotic substances, however, resistance is increasing in this species [17, 18]. Furthermore, C. ochracea is known to possess an immunosuppressive factor [19]. All strains of C. ochracea are capable of fermenting glucose, sucrose, maltose and mannose, whereas most strains ferment amygdalin, fructose, galactose, lactose and raffinose [20]. The optimal growth temperature is 37°C. Nitrate is reduced to nitrite, and dextran, glycogen, starch and aesculin are hydrolysed by most strains. Indole is not produced. Acetic and succinic acid are the main metabolic end products of fermentation [6].

**Figure 2.** Scanning electron micrograph of C. ochracea VPI 2845T

Analysis of amino acids and amino sugars of the peptidoglycan revealed that glucosamine, muramic acid, D-glutamic acid, alanine, and diaminopimelic acid were the principal components and the peptidoglycan belongs to the Alγ-type. Serine and glycine were not found [21]. As in other Capnocytophaga strains, the fatty acid pattern of strain C. ochracea VPI 2845T is dominated by iso-branched chain saturated fatty acids i-C_{15:0} (63.5%), C_{18:2} (8.1%) and i-3OH C_{17:0} (13.8%) [17, 22, 23]. Phosphatidylethanolamine and an ornithine-amino lipid were identified as dominating polar lipids, as well as lesser amounts of lysophosphatidylethanolamine [24]. In addition, the unusual sulfonolipid capnine (2-amino-3-hydroxy-15-methylhexadecane-1-sulfonic acid) was identified as major cell wall component [25].

**Table 1.** Classification and general features of C. ochracea VPI 2845T in accordance to the MIGS recommendations [26]

| MIGS ID | Property | Term | Evidence code |
|---------|----------|------|---------------|
| Current classification | Domain | Bacteria | C[27] |
| | Phylum | 'Bacteroidetes' | C[7] |
| | Class | Flavobacteria | C[7] |
| | Order | Flavobacteriales | C[7] |
| | Suborder | Flavobacteriales | |
| | Family | Flavobacteriaceae | |
Genus *Capnocytophaga*  
Species *Capnocytophaga ochacea*  
Type strain VPI 2845

Gram stain negative  
Cell shape fusiform rods  
Motility gliding  
Sporulation non-sporulating  
Temperature range mesophile  
Optimum temperature 30-37°C  
Salinity nonhalophile

**MIGS-22** Oxygen requirement capnophilic; aerobic or anaerobic with at least 5% CO$_2$  
Carbon source glucose, maltose, lactose, sucrose  
Energy source chemoorganotroph, carbohydrates

**MIGS-6** Habitat human oral cavity  
**MIGS-15** Biotic relationship Free living  
**MIGS-14** Pathogenicity opportunistic pathogen  
**MIGS-11** Biosafety level 2  
**MIGS-13** Isolation human oral cavity

**MIGS-4** Geographic location Gerenzano, Italy  
**MIGS-5** Sample collection time about 1956  
**MIGS-4.1** Latitude – Longitude not reported  
**MIGS-4.2** Depth not reported  
**MIGS-4.4** Altitude not reported

**a)** Evidence code types – (R)eported for the purpose of this specific publication, directly observed by one of the authors or acknowledged person or institution for the living isolated sample, (C)ited: a direct report exists in the literature, or (I)nferred: not directly observed for the living, isolated sample, but based on a personally accepted property for this species, or anecdotal communication.

**b)** A general mapping of these evidence codes to those evidence codes ([http://www.geneontology.org/GO.evidence.shtml](http://www.geneontology.org/GO.evidence.shtml)) used by the Gene Ontology project [26] is: R= IDA; C=TAS; and I= NAS.

### Genome sequencing and annotation information

**Genome project history**
This organism was selected for sequencing on the basis of its phylogenetic position, and is part of the *Genomic Encyclopedia of Bacteria and Archaea* project. The genome project is deposited in the Genomes OnLine Database [10] and the complete genome sequence in GenBank (CP001632). Sequencing, finishing and annotation were performed by the DOE Joint Genome Institute (JGI). A summary of the project information is shown in Table 2.

| **Table 2. Genome sequencing project information** |
|-----------------------------------------------|

| **MIGS ID** | **Property**       | **Term** |
|------------|--------------------|---------|
| MIGS-31    | Finishing quality  | Finished|
| MIGS-28    | Genomic libraries  | used    |
|            | Two Sanger libraries: 6.5kb | |
Growth conditions and DNA isolation

*C. ochracea* VPI 2845, DSM 7271, was grown under anaerobic conditions in DSMZ medium 340 plus 0.1% NaHCO3 (*Capnocytophaga* Medium, available through www.dsmz.de) at 37°C. DNA was isolated from 1-1.5 g of cell paste using Qiagen Genomic 500 DNA Kit (Qiagen, Hilden, Germany) with a modified protocol for cell lysis using more lysozyme (1.6x) and a prolonged incubation time (60 minutes) at 37°C.

Genome sequencing and assembly

The genome was sequenced using a combination of Sanger and 454 sequencing platforms. All general aspects of library construction and sequencing performed at the JGI can be found at http://www.jgi.doe.gov/. 454 pyrosequencing reads were assembled using the Newbler assembler version 1.1.02.15 (Roche). Large Newbler contigs were broken into 2919 overlapping fragments of 1000bp and entered into assembly as pseudo-reads. The sequences were assigned quality scores based on Newbler consensus q-scores with modifications to account for overlap redundancy and to adjust inflated q-scores. A hybrid 454/Sanger assembly was made using the parallel phrap assembler (High Performance Software, LLC). Possible mis-assemblies were corrected with Dupfinisher or transposon bombing of bridging clones [29]. Gaps between contigs were closed by editing in Consed, custom primer walk or PCR amplification. 226 Sanger finishing reads were produced to close gaps, to resolve repetitive regions, and to raise the quality of the finished sequence. The error rate of the completed genome sequence is less than 1 in 100,000. Together all sequence types provided 35.1 x coverage of the genome.

Genome annotation

Genes were identified using Prodigal [30] as part of the Oak Ridge National Laboratory genome annotation pipeline, followed by a round of manual curation using JGI’s GenePRIMP pipeline (http://geneprimp.jgi-psf.org) [31]. The predicted CDSs were translated and used to search the National Center for Biotechnology Information (NCBI) nonredundant database, UniProt, TIGRFam, Pfam, PRIAM, KEGG, COG, and InterPro databases. Additional gene prediction analysis and functional annotation was performed within the Integrated Microbial Genomes (IMG-ER) platform (http://img.jgi.doe.gov/er) [32].

Genome properties
The genome is 2,612,925 bp long and comprises one circular chromosome with a 39.6% GC content (Tab. 3). Of the 2252 genes predicted, 2193 were protein coding genes, and 59 RNAs; 22 pseudogenes were also identified. 61.7% of the genes were assigned with a putative function while the remaining are annotated as hypothetical proteins. The distribution of genes into GOGs functional categories is presented in Table 4.

Table 3. Genome Statistics

| Attribute                        | Value   | % of Total |
|----------------------------------|---------|------------|
| Genome size (bp)                 | 2,612,925 |           |
| DNA Coding region (bp)           | 2,293,132 | 87.76%     |
| DNA G+C content (bp)             | 1,034,404 | 39.59%     |
| Number of replicons              | 1       |            |
| Extrachromosomal elements        | 0       |            |
| Total genes                      | 2252    | 100.00%    |
| RNA genes                        | 59      | 0.85%      |
| rRNA operons                     | 4       |            |
| Protein-coding genes             | 2193    | 97.38%     |
| Pseudo genes                     | 22      | 0.98%      |
| Genes with function prediction   | 1390    | 61.72%     |
| Genes in paralog clusters        | 207     | 9.19%      |
| Genes assigned to COGs           | 1330    | 59.06%     |
| Genes assigned Pfam domains      | 1379    | 61.23%     |
| Genes with signal peptides       | 862     | 38.28%     |
| Genes with transmembrane helices | 471     | 20.91%     |
| CRISPR repeats                   | 1       |            |
Figure 3. Graphical circular map of the genome. From outside to the center: Genes on forward strand (color by COG categories), Genes on reverse strand (color by COG categories), RNA genes (tRNAs green, sRNAs red, other RNAs black), GC content, GC skew.

Table 4. Number of genes associated with the 21 general COG functional categories

| Code | COG counts and percentage of protein-coding genes | Description                        |
|------|--------------------------------------------------|-------------------------------------|
|      | Genome value | % of total                          |
| J    | 134 | 6.1 | Translation |
| A    | 0  | 0.0 | RNA processing and modification |
| K    | 55 | 2.5 | Transcription |
| L    | 83 | 3.8 | Replication, recombination and repair |
| B    | 0  | 0.0 | Chromatin structure and dynamics |
| D    | 19 | 0.9 | Cell cycle control, mitosis and meiosis |
Y 0 0.0 Nuclear structure
V 34 1.6 Defense mechanisms
T 35 1.6 Signal transduction mechanisms
M 158 7.2 Cell wall/membrane biogenesis
N 7 0.3 Cell motility
Z 0 0.0 Cytoskeleton
W 0 0.0 Extracellular structures
U 35 1.6 Intracellular trafficking and secretion
O 61 2.8 Posttranslational modification, protein turnover, chaperones
C 69 3.1 Energy production and conversion
G 97 4.4 Carbohydrate transport and metabolism
E 90 4.1 Amino acid transport and metabolism
F 56 2.6 Nucleotide transport and metabolism
H 84 3.8 Coenzyme transport and metabolism
I 53 2.4 Lipid transport and metabolism
P 80 3.6 Inorganic ion transport and metabolism
Q 25 1.1 Secondary metabolites biosynthesis, transport and catabolism
R 145 6.6 General function prediction only
S 100 4.6 Function unknown
- 863 39.4 Not in COGs

Acknowledgements
We would like to gratefully acknowledge the help of Sabine Welnitz for growing C. ochracea cultures and Susanne Schneider for DNA extraction and quality analysis (both at DSMZ). This work was performed under the auspices of the US Department of Energy’s Office of Science, Biological and Environmental Research Program, and by the University of California, Lawrence Berkeley National Laboratory under contract No. DE-AC02-05CH11231, Lawrence Livermore National Laboratory under Contract No. DE-AC52-07NA27344, and Los Alamos National Laboratory under contract No. DE-AC02-06NA25396, well as German Research Foundation (DFG) INST 599/1-1.

References

1. Prévot AR, Tardieux P, Joubert L, de Cadore F. Recherches sur Fusiformis nucleatus (Knorr) et son pouvoir pathogène pour l'homme et les animaux. Ann Instiut Pasteur (Paris) 1956, 91:787-98.

2. Leadbetter ER et al. Validation List N° 8. Int J Syst Bacteriol 1982, 32:266-8.

3. Loesche WJ, Socransky SS, Gibbons RJ. Bacteroides oralis, proposed new species isolated from the oral cavity of man. J Bacteriol 1964, 88:1327-9.

4. Holdeman LV, Moore WEC. Bacteroides. In: Anaerobe Laboratory Manual, Virginia Polytechnic Institute Anaerobe Laboratory, Blacksburg, VA, USA, 1972.

5. Sebald M. Etudes sur les bactéries anaérobies gram-négatives asporulée. Imprimerie Barnéoud S. A., Lavel, France, 1962.
6. Leadbetter ER, Holt SC, Socransky SS. *Capnocytophaga*: new genus of Gram-negative gliding bacteria I. General characteristics, taxonomic considerations and significance. *Arch Microbiol* 1979, **122**:9-16.

7. Bernadet J-F, Nakagawa Y, Holmes B, for the Subcommittee on the taxonomy of *Flavobacterium* and *Cytophaga*-like bacteria of the International Committee on Systematics of Prokaryotes. Proposed minimal standards for describing new taxa of the family *Flavobacteriaceae* and emended description of the family. *Int J Sys Evol Microbiol* 2002, **52**:1049-70.

8. Holt SC, Simpson JL, Leadbetter ER. Some characteristics of "gliding" bacteria isolated from human dental plaque. *J Dent Res* 1975, **54**:ABS L208.

9. Newman MG, Weiner MS, Angel I, Grinenko V, Karge HJ. Predominant cultivable microbiota of the gingival crevice in "supernormal" patients. *J Dent Res* 1977, **56**, B121.

10. Desai SS, Harrison RA, Murphy MD. *Capnocytophaga ochracea* causing severe sepsis and purpura fulminans in an immunocompetent patient. *J Infect* 2007, **54**:107–9.

11. Bonatti H, Rossboth DW, Nachbaur D, Fille M, Aspöck C, Hend I, Hourmont K, White L, Malnick H, Allerberger FJ. A series of infections due to *Capnocytophaga* spp. in immunosuppressed and immunocompetent patients. *Clin Microbiol Infect* 2003, **9**:380–387.

12. Duong M, Besancenot JF, Neuwirth C, Buisson M, Chavanet P, Portier H. Vertebral osteomyelitis due to *Capnocytophaga species* in immunocompetent patients: report of two cases and review. *Clin Infect Dis* 1996, **22**:1099–101.

13. Lee C, Grasso C, Sharlow MF. Multiple sequence alignment using partial order graphs. *Bioinformatics* 2002, **18**:452-64.

14. Castresana J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol* 2000, **17**:540-52.

15. Stamatakis A, Hoover P, Rougemont J. A rapid bootstrap algorithm for the RAxML web-servers. *Syst Biol* 2008, **57**:758-71.

16. Liolios K, Mavromatis K, Tavernarakis N, Kyrpides NC. The Genomes OnLine Database (GOLD) in 2007: status of genomic and metagenomic projects and their associated metadata *Nucleic Acids Res* 2008, **36**:D475-9.

17. Wang HK, Chen YC, Teng LJ, Hung CC, Chen ML, Du SH, Pan HJ, Hsueh PR, Chang SC. Brain abscess associated with multidrug-resistant *Capnocytophaga ochracea* infection. *J Clin Microbiol* 2007, **45**:645–7.

18. Rosenau A, Cattier B, Gousset N, Harriau P, Philippon A, Quentin R. *Capnocytophaga ochracea*: characterization of a plasmid-encoded extended-spectrum TEM-17 beta-lactamase in the phylum Flavobacter-bacteroides. *Antimicrob Agents Chemother* 2000, **44**:760–2.
19. Ochiai K, Senpuku H, Kurita-Ochiai T. Purification of immunosuppressive factor from \textit{Capnocytophaga ochracea}. \textit{J Med Microbiol} 1998, 47:645–7.

20. Socransky SS, Holt SC, Leadbetter ER, Tanner ACR, Savitt E, Hammond BF. \textit{Capnocytophaga}: new genus of Gram-negative gliding bacteria. III. Physiological characterization. \textit{Arch Microbiol} 1979, 122:29-33.

21. Hanagata T. Chemical structure and immunological activities of peptidoglycan isolated from \textit{Capnocytophaga} species. Kanagawa Shigaku 1990, 25:316-25.

22. Dees SB, Karr DE, Hollis D, Moss CW. Cellular fatty acids of \textit{Capnocytophaga} species. \textit{J Clin Microbiol} 1982, 16:779-83.

23. Vandamme P, Vancanneyt M, Van Belkum A, Segers P, Quint WGV, Kersters K, Paster BJ, Dewhirst FE. Polyphasic analysis of strains of the genus \textit{Capnocytophaga} and Centers for Disease Control group DF-3. \textit{Int J Sys Bacteriol} 1996, 46:782-91.

24. Holt SC, Doundowlakis J, Takacs BJ. Phospholipid composition of \textit{Capnocytophaga} compared with \textit{Sporocytophaga}. \textit{Infect Immun} 1979, 26:305-10.

25. Godchaux W 3rd, Leadbetter ER. \textit{Capnocytophaga} spp. contain sulfonolipids that are novel in procaryotes. \textit{J Bacteriol} 1980, 144:592-602.

26. Field D, Garrity G, Gray T, Morrison N, Selengut J, \textit{et al.} Towards a richer description of our complete collection of genomes and metagenomes: the “Minimum Information about a Genome Sequence” (MIGS) specification. \textit{Nat Biotechnol} 2008, 26:541-7.

27. Cavalier-Smith T. The neomuran origin of archaeabacteria, the negibacterial root of the universal tree and bacterial megaclassification. \textit{Int J Syst Evol Microbiol} 2002, 52:7-76.

28. Biological Agents: Technical rules for biological agents www.baua.de TRBA 466.

29. Han CS, Chain P. Finishing repeat regions automatically with Dupfinisher. \textit{In}: Proceeding of the 2006 international conference on bioinformatics & computational biology. Hamid R. Arabnia & Homayoun Valafar (eds), CSREA Press. June 26-29, 2006:141-6.

30. \texttt{http://compbio.ornl.gov/prodigal/}

31. Pati et al. GenePRIMP: A Gene Prediction Improvement Pipeline for microbial genomes. \textit{in preparation} 2009.

32. Markowitz VM, Mavromatis K, Ivanova NN, Chen I-MA, Kyrpides NC. Expert Review of Functional Annotations for Microbial Genomes. \textit{Bioinformatics} 2009, in press.