Complete genome sequence of the gliding freshwater bacterium *Fluviicola taffensis* type strain (RW262[T])

Tanja Woyke[1], Olga Chertkov[1], Alla Lapidus[1], Matt Nolan[1], Susan Lucas[1], Tijana Glavina Del Rio[1], Hope Tice[1], Jan-Fang Cheng[1], Roxanne Tapia[1,2], Cliff Han[1,2], Lynne Goodwin[1,2], Sam Pitluck[1], Konstantinos Liolios[1], Ioanna Pagani[1], Natalia Ivanova[1], Marcel Huntemann[1], Konstantinos Mavromatis[1], Natalia Mikhailova[1], Amrita Pati[1], Amy Chen[3], Krishna Palaniappan[3], Miriam Land[1,4], Loren Hauser[1,4], Evelyne-Marie Brambilla[5], Manfred Rohde[6], Romano Mwirichia[7], Johannes Sikorski[5], Brian J. Tindall[5], Markus Göker[5], James Bristow[1], Jonathan A. Eisen[1,7], Victor Markowitz[4], Philip Hugenholtz[1,9], Hans-Peter Klenk[5], and Nikos C. Kyrpides[1*]

[1] DOE Joint Genome Institute, Walnut Creek, California, USA
[2] Los Alamos National Laboratory, Bioscience Division, Los Alamos, New Mexico, USA
[3] Biological Data Management and Technology Center, Lawrence Berkeley National Laboratory, Berkeley, California, USA
[4] Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA
[5] DSMZ - German Collection of Microorganisms and Cell Cultures GmbH, Braunschweig, Germany
[6] HZI – Helmholtz Centre for Infection Research, Braunschweig, Germany
[7] Jomo Kenyatta University of Agriculture and Technology, Kenya
[8] University of California Davis Genome Center, Davis, California, USA
[9] Australian Centre for Ecogenomics, School of Chemistry and Molecular Biosciences, The University of Queensland, Brisbane, Australia

*Corresponding author: Nikos C. Kyrpides

Fluviicola taffensis O'Sullivan *et al*. 2005 belongs to the monotypic genus *Fluviicola* within the family *Cryomorphaceae*. The species is of interest because of its isolated phylogenetic location in the genome-sequenced fraction of the tree of life. Strain RW262[T] forms a monophyletic lineage with uncultivated bacteria represented in freshwater 16S rRNA gene libraries. A similar phylogenetic differentiation occurs between freshwater and marine bacteria in the family *Flavobacteriaceae*, a sister family to *Cryomorphaceae*. Most remarkable is the inability of this freshwater bacterium to grow in the presence of Na$^+$ ions. All other genera in the family *Cryomorphaceae* are from marine habitats and have an absolute requirement for Na$^+$ ions or natural sea water. *F. taffensis* is the first member of the family *Cryomorphaceae* with a completely sequenced and publicly available genome. The 4,633,577 bp long genome with its 4,082 protein-coding and 49 RNA genes is a part of the **G**enomic **E**ncyclopedia of **B**acteria and **A**rchaea project.

Introduction

Strain RW262[T] (= DSM 16823 = NCIMB 13979) is the type strain of the species *Fluviicola taffensis*, which is the type species of the monotypic genus *Fluviicola* [1], affiliated with the family *Cryomorphaceae* [2]. The genus name is derived from the Latin words *fluvius*, meaning 'river' and *-cola* meaning 'inhabitant, dweller', yielding the Neo-Latin word *Fluviicola*, the *river dweller* [1,3]. The species epithet is derived from the Neo-Latin

word *taffensis*, referring to the place where the type strain has been isolated, the river Taff (Wales, UK) [1,3]. The family *Cryomorphaceae* belongs to the class *Flavobacteria* which contains many species that probably play an integral role for the flow of carbon and energy in the marine environment [4].

Flavobacteria are the major decomposers of high-molecular-mass organic matter in sea water [5].

　　　　　　　　　　　　　　　　　　　　　　The Genomic Standards Consortium

Phylogenetically the family *Cryomorphaceae* is located between the families *Flavobacteriaceae* and *Bacteroidaceae* [2] and currently comprises the genera *Brumimicrobium*, *Cryomorpha* and *Crocinitomix* [2], *Owenweeksia* [6], *Wandonia* [7], *Fluviicola* [1] and *Lishizhenia* [8]. The family *Cryomorphaceae* exhibits the greatest degree of phenotypic similarity to the family *Flavobacteriaceae* [9] and includes species with a mostly rod-like to filamentous morphology; cells are usually non-motile or move by gliding and often contain carotenoid pigments [1,2]. All members of the *Cryomorphaceae* are strictly aerobic or facultatively anaerobic (fermentative) with a chemoheterotrophic metabolism [1,2] and often have complex growth requirements for sea water salts, organic compounds as sole nitrogen sources, yeast extract and vitamins for growth [2]. To date no further isolates of *F. taffensis* have been reported. Here we present a summary classification and a set of features for *F. taffensis* RW262^T, together with the description of the complete genomic sequencing and annotation.

Classification and features

A representative genomic 16S rRNA sequence of *F. taffensis* RW262^T was compared using NCBI BLAST [10] under default settings (e.g., considering only the high-scoring segment pairs (HSPs) from the best 250 hits) with the most recent release of the Greengenes database [11] and the relative frequencies of taxa and keywords (reduced to their stem [12]) were determined, weighted by BLAST scores. The most frequently occurring genera were *Brumimicrobium* (62.9%) and *Fluviicola* (37.1%) (3 hits in total). Among all other species, the one yielding the highest score was *'Brumimicrobium mesophilum'* (DQ660382), which corresponded to an identity of 92.1% and an HSP coverage of 58.0%. (Note that the Greengenes database uses the INSDC (= EMBL/NCBI/DDBJ) annotation, which is not an authoritative source for nomenclature or classification.) The most frequently occurring keywords within the labels of all environmental samples which yielded hits were 'lake' (9.1%), 'tin' (3.4%), 'microbi' (2.5%), 'depth' (2.0%) and 'tract' (1.7%) (247 hits in total). The most frequently occurring keywords within those labels of environmental samples which yielded hits of a higher score than the highest scoring species were 'lake' (9.2%), 'tin' (3.8%), 'microbi' (2.3%), 'depth' (2.0%) and 'tract' (1.8%) (169 hits in total). The most frequent key-

word 'lake' may reflect the freshwater origin of strain RW262^T, whereas the keywords 'tin' and 'depth' may allude to some until now unrecognized ecological features of *F. taffensis.*

Figure 1 shows the phylogenetic neighborhood of *F. taffensis* in a 16S rRNA based tree. The sequences of the two identical 16S rRNA gene copies in the genome differ by two nucleotides from the previously published 16S rRNA sequence (AF493694), which contains one ambiguous base call.

Strain RW262^T is strictly aerobic, Gram-negative, motile by gliding and flexirubin-pigmented [1]. Cells are flexible rods with rounded ends (Figure 2), 0.4-0.5 μm in diameter and 1.5-5.7 μm in length, with rare longer filaments of up to 51 μm in length [1]. Growth occurs at 4ºC and 20ºC, but not in the presence of Na$^+$ ions [1]. Growth of strain RW262^T at 4 ºC is only weak, so that *F. taffensis* should not be considered to be psychrotolerant like the other members of the family [1,2]. Strain RW262^T is capable of DNA hydrolysis [1], is catalase positive but oxidase negative, able to catalyze the hydrolysis of arginine, aesculin or starch, whereas it weakly hydrolyzes gelatine [1]. It is negative for nitrate and nitrite reduction; indole production; β-galactosidase, urease and xylanase activity; hydrolysis of agar, arginine, aesculin and starch; and acid production from carbohydrates [1]. The strain is not able to utilize glucose, arabinose, mannose, mannitol, *N*-acetylglucosamine, maltose, gluconate, caprate, adipate, malate, citrate or phenyl acetate [1]. However, within the genome are several genes for utilization of complex organic carbon compounds. The strain is resistant to chloramphenicol (10 μg), streptomycin (10 μg), and kanamycin (30 μg) but susceptible to penicillin G (10 units), ampicillin (10 μg), rifampicin (5 μg) and tetracycline (10 μg) [1].

Chemotaxonomy

The predominant cellular acid of strain RW262^T was the branched-chain saturated fatty acid *iso*-$C_{15:0}$ (44.2%) [1]. Unsaturated branched-chain fatty acids, straight-chain saturated and mono-unsaturated fatty acids occur only in lower amounts: $C_{14:0}$ (3.2%), $C_{15:0}$ (7.5%), $C_{16:0}$ (3.0%), *iso*-$C_{15:1\ \omega10c}$ (11.8%), *iso*-$C_{16:1\ \omega12c}$ (4.9%). Lipopolysaccharide hydroxy fatty acids constitute 20.4% of the total cellular fatty acids, mainly composed of *iso*-$C_{17:0\ 3-OH}$ (12.3%), iso-$C_{15:0\ 3-OH}$ (4.2%) and iso-$C_{15:0\ 2-OH}$ (3.5%) [1].
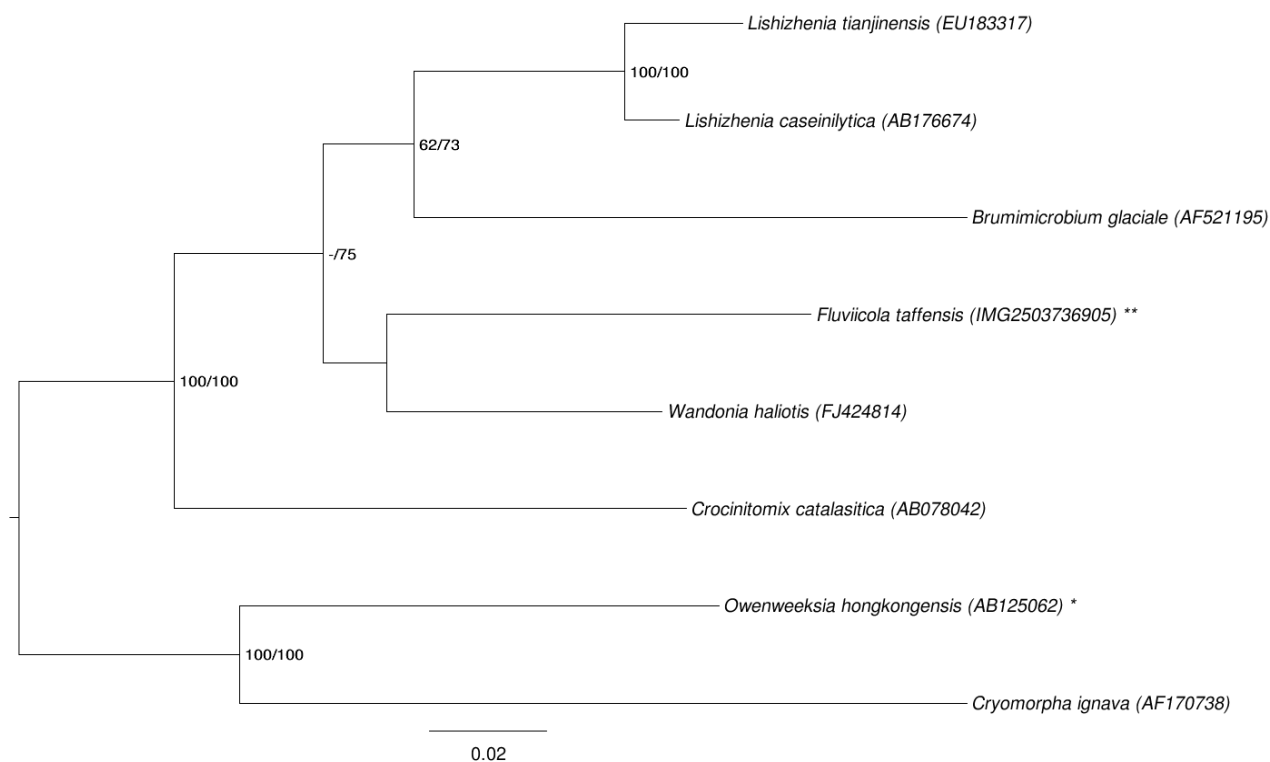
Figure 1. Phylogenetic tree highlighting the position of *F. taffensis* relative to the type strains of the other species within the family *Cryomorphaceae*. The tree was inferred from 1,429 aligned characters [13,14] of the 16S rRNA gene sequence under the maximum likelihood (ML) criterion [15]. Rooting was done initially using the midpoint method [16] and then checked for its agreement with the current classification (Table 1). The branches are scaled in terms of the expected number of substitutions per site. Numbers adjacent to the branches are support values from 300 ML bootstrap replicates [17] (left) and from 1,000 maximum parsimony bootstrap replicates [18] (right) if larger than 60%. Lineages with type strain genome sequencing projects registered in GOLD [19] are labeled with one asterisk, those also listed as 'Complete and Published' (as well as the target genome) with two asterisks.
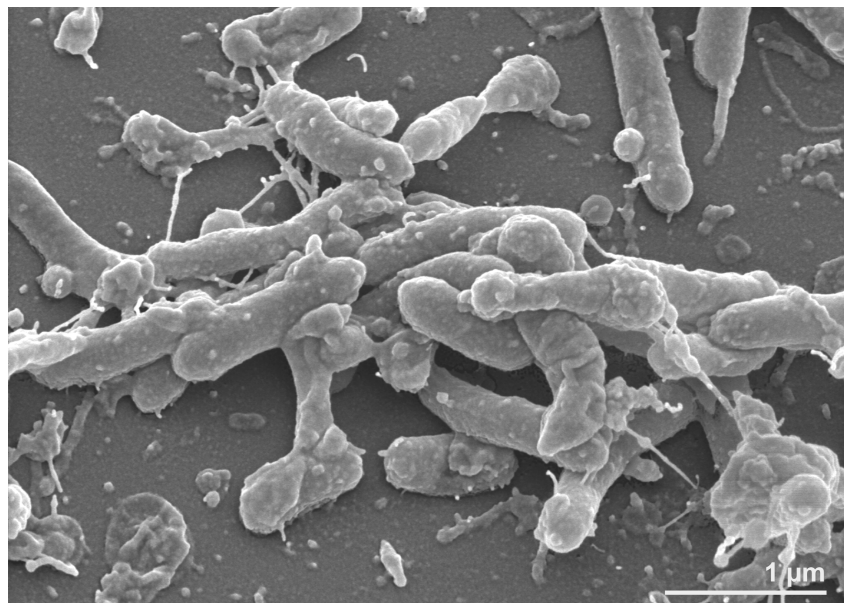

Figure 2. Scanning electron micrograph of *F. taffensis* RW262[T]

Table 1. Classification and general features of *F. taffensis* RW262[T] according to the MIGS recommendations [20] and the NamesforLife database [21].

MIGS ID	Property	Term	Evidence code
		Domain *Bacteria*	TAS [22]
		Phylum *Bacteroidetes*	TAS [23]
		Class "*Flavobacteria*"	TAS [24]
	Current classification	Order "*Flavobacteriales*"	TAS [25]
		Family *Cryomorphaceae*	TAS [2]
		Genus *Fluviicola*	TAS [1]
		Species *Flaviicola taffensis*	TAS [1]
		Type strain RW262	TAS [1]
	Gram stain	negative	TAS [1]
	Cell shape	rod-shaped	TAS [1]
	Motility	by gliding	TAS [1]
	Sporulation	none	TAS [1]
	Temperature range	4°C-25°C	TAS [1]
	Optimum temperature	20°C	TAS [1]
	Salinity	obligate 0%	TAS [1]
MIGS-22	Oxygen requirement	strict aerobe	TAS [1]
	Carbon source	probably amino acids; unable to use carbohydrates	NAS
	Energy metabolism	chemoorganotroph	TAS [1]
MIGS-6	Habitat	fresh water	TAS [1]
MIGS-15	Biotic relationship	free-living	NAS
MIGS-14	Pathogenicity	none	NAS
	Biosafety level	1	TAS [26]
	Isolation	fresh river water	TAS [1]
MIGS-4	Geographic location	River Taff near Cardiff, UK	TAS [1]
MIGS-5	Sample collection time	January 2000	TAS [1]
MIGS-4.1	Latitude	51.85	TAS [1]
MIGS-4.2	Longitude	-2.32	TAS [1]
MIGS-4.3	Depth	not reported	
MIGS-4.4	Altitude	sea level	NAS

Evidence codes - TAS: Traceable Author Statement (i.e., a direct report exists in the literature); NAS: Non-traceable Author Statement (i.e., not directly observed for the living, isolated sample, but based on a generally accepted property for the species, or anecdotal evidence). These evidence codes are from of the Gene Ontology project [27].

Genome sequencing and annotation
Genome project history

This organism was selected for sequencing on the basis of its phylogenetic position [28], and is part of the *Genomic Encyclopedia of Bacteria and Archaea* project [29]. The genome project is deposited in the Genome On Line Database [19] and the complete genome sequence is deposited in GenBank. 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	Libraries used	Tree genomic libraries: one 454 pyrosequence standard library, one 454 PE library (11 kb insert size), one Illumina library
MIGS-29	Sequencing platforms	Illumina GAii, 454 GS FLX Titanium
MIGS-31.2	Sequencing coverage	351.0 × Illumina; 23.0 × pyrosequence
MIGS-30	Assemblers	Newbler version 2.3, Velvet, phrap version SPS – 4.24
MIGS-32	Gene calling method	Prodigal 1.4, GenePRIMP
	INSDC ID	CP002542
	Genbank Date of Release	April 1, 2011
	GOLD ID	Gc01706
	NCBI project ID	47603
	Database: IMG-GEBA	2503707007
MIGS-13	Source material identifier	DSM 16823
	Project relevance	Tree of Life, GEBA

Growth conditions and DNA isolation

F. taffensis RW262[T], DSM 16823, was grown in DSMZ medium 948 (Oxoid nutrient medium) [30] at 28°C. DNA was isolated from 0.5-1 g of cell paste using JetFlex Genomic DNA Purification kit (GENOMED 600100) following the standard protocol as recommended by the manufacturer, but with additional 20 µl proteinase K incubation (one hour) at 58° for improved cell lysis. DNA is available through the DNA Bank Network [31].

Genome sequencing and assembly

The genome was sequenced using a combination of Illumina and 454 sequencing platforms. All general aspects of library construction and sequencing can be found at the JGI website [32]. Pyrosequencing reads were assembled using the Newbler assembler (Roche). The initial Newbler assembly consisting of 51 contigs in one scaffold was converted into a phrap [33] assembly by making fake reads from the consensus, to collect the read pairs in the 454 paired end library. Illumina GAii sequencing data (801.4 Mb) was assembled with Velvet [34] and the consensus sequences were shredded into 1.5 kb overlapped fake reads and assembled together with the 454 data. The 454 draft assembly was based on 164.9 Mb 454 draft data and all of the 454 paired end data. Newbler parameters are -consed -a 50 -l 350 -g -m -ml 20. The Phred/Phrap/Consed software package [33] was used for sequence assembly and quality assessment in the subsequent finishing process. After the shotgun stage, reads were assembled with parallel phrap (High Performance Software, LLC). Possible mis-assemblies were corrected with gapResolution [32], Dupfinisher [35], or sequencing clones bridging PCR fragments with subcloning. Gaps between contigs were closed by editing in Consed, by PCR and by Bubble PCR primer walks (J.-F. Chang, unpublished). A total of 161 additional reactions and shatter libraries were necessary to close gaps and to raise the quality of the finished sequence. Illumina reads were also used to correct potential base errors and increase consensus quality using a software Polisher developed at JGI [36]. The error rate of the completed genome sequence was less than 1 in 100,000. Together, the combination of the Illumina and 454 sequencing platforms provided 374.0 × coverage of the genome. The final assembly contained 232,904 pyrosequence and 44,902,395 Illumina reads.

Genome annotation

Genes were identified using Prodigal [37] as part of the Oak Ridge National Laboratory genome annotation pipeline, followed by a round of manual curation using the JGI GenePRIMP pipeline [38]. The predicted CDSs were translated and used to search the National Center for Biotechnology Information (NCBI) non-redundant database, UniProt, TIGR-Fam, Pfam, PRIAM, KEGG, COG, and InterPro databases. Additional gene prediction analysis and functional annotation was performed within the Integrated Microbial Genomes - Expert Review (IMG-ER) platform [39].

Genome properties

The genome consists of a 4,633,577 bp long chromosome with a G+C content of 36.5% (Figure 3 and Table 3). Of the 4,131 genes predicted, 4,082 were protein-coding genes, and 49 RNAs; 49 pseudogenes were also identified. The majority of the protein-coding genes (55.0%) were assigned a putative function while the remaining ones were annotated as hypothetical proteins. The distribution of genes into COGs functional categories is presented in Table 4.
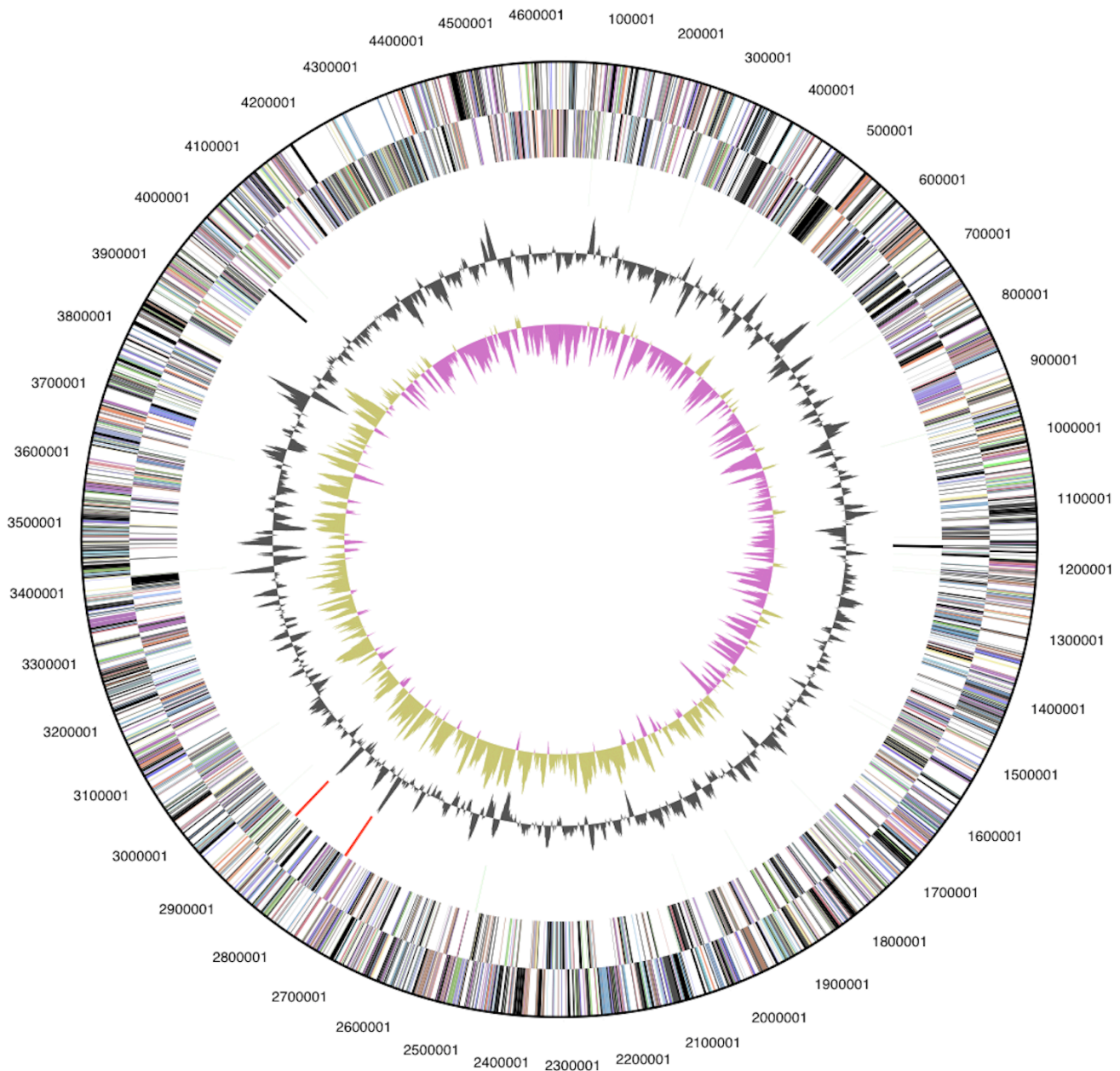

Figure 3. Graphical circular map of the genome. From outside to the center: Genes on forward strand (color by COG categories), Genes on reverse strand (color by COG categories), RNA genes (tRNAs green, rRNAs red, other RNAs black), GC content, GC skew.

Table 3. Genome Statistics

Attribute	Value	% of Total
Genome size (bp)	4,633,577	100.00%
DNA coding region (bp)	4,192,830	90.49%
DNA G+C content (bp)	1,691,009	36.49%
Number of replicons	1	
Extrachromosomal elements	0	
Total genes	4,131	100.00%
RNA genes	49	1.19%
rRNA operons	2	
Protein-coding genes	4,082	98.81%
Pseudo genes	49	1.19%
Genes with function prediction	2,271	54.97%
Genes in paralog clusters	532	12.88%
Genes assigned to COGs	2,169	52.51%
Genes assigned Pfam domains	2,420	58.58%
Genes with signal peptides	1,331	32.22%
Genes with transmembrane helices	911	22.05%
CRISPR repeats	1	

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

Code	value	%age	Description
J	168	6.0	Translation, ribosomal structure and biogenesis
A	0	0.0	RNA processing and modification
K	212	8.8	Transcription
L	137	5.7	Replication, recombination and repair
B	1	0.0	Chromatin structure and dynamics
D	22	0.9	Cell cycle control, cell division, chromosome partitioning
Y	0	0.0	Nuclear structure
V	57	2.4	Defense mechanisms
T	183	7.6	Signal transduction mechanisms
M	222	9.2	Cell wall/membrane/envelope biogenesis
N	8	0.3	Cell motility
Z	0	0.0	Cytoskeleton
W	0	0.0	Extracellular structures
U	42	1.8	Intracellular trafficking, secretion, and vesicular transport
O	106	4.4	Posttranslational modification, protein turnover, chaperones
C	117	4.9	Energy production and conversion
G	76	3.2	Carbohydrate transport and metabolism
E	136	5.7	Amino acid transport and metabolism
F	63	2.6	Nucleotide transport and metabolism
H	114	4.7	Coenzyme transport and metabolism
I	101	4.2	Lipid transport and metabolism
P	116	4.8	Inorganic ion transport and metabolism
Q	45	1.9	Secondary metabolites biosynthesis, transport and catabolism
R	286	11.9	General function prediction only
S	194	8.1	Function unknown
-	1,962	47.5	Not in COGs

Acknowledgements

References

1. O'Sullivan LA, Rinna J, Humphreys G, Weightman AJ, Fry JC. *Fluviicola taffensis* gen. nov., sp. nov., a novel freshwater bacterium of the family *Cryomorphaceae* in the phylum '*Bacteroidetes*'. *Int J Syst Evol Microbiol* 2005; **55**:2189-2194. PubMed doi:10.1099/ijs.0.63736-0

2. Bowman JP, Mancuso Nichols C, Gibson JAE. *Algoriphagus ratkowskyi* gen. nov., sp. nov., *Brumimicrobium glaciale* gen. nov., sp. nov., *Cryomorpha ignava* gen. nov., sp. nov. and *Crocinitomix catalasitica* gen. nov., sp. nov., novel flavobacteria isolated from various polar habitats. *Int J Syst Evol Microbiol* 2003; **53**:1343-1355. PubMed doi:10.1099/ijs.0.02553-0

3. Euzéby JP. List of bacterial names with standing in nomenclature: A folder available on the internet. *Int J Syst Bacteriol* 1997; **47**:590-592. PubMed doi:10.1099/00207713-47-2-590

4. Kirchman DL. The ecology of *Cytophaga-Flavobacteria* in aquatic environments. *FEMS Microbiol Ecol* 2002; **39**:91-100. PubMed

5. Cottrell MT, Kirchman DL. Natural assemblages of marine proteobacteria and members of the *Cytophaga-Flavobacter* cluster consuming low - and high-molecular-weight dissolved organic matter. *Appl Environ Microbiol* 2000; **66**:1692-1697. PubMed doi:10.1128/AEM.66.4.1692-1697.2000

6. Lau KWK, Ng CYM, Ren J, Lau SCL, Qian PY, Wong PK, Lau TC and WU M. *Owenweeksia hongkongensis* gen. nov., sp. nov., a novel marine bacterium of the phylum '*Bacteroidetes*'. *Int J Syst Evol Microbiol* 2005; **55**:1051-1057. PubMed doi:10.1099/ijs.0.63155-0

7. Lee DH, Choi EK, Moon SR, Ahn S, Lee YS, Jung JS, Jeon CO, Whang KS, Kahng HY. *Wandonia haliotis* gen. nov., sp. nov., a marine bacterium of the family *Cryomorphaceae*, phylum *Bacteroidetes*. *Int J Syst Evol Microbiol* 2010; **60**:510-514. PubMed doi:10.1099/ijs.0.012674-0

8. Lau KWK, Ren J, Wai NLM, Qian PY, Wong PK, Wu M. *Lishizhenia caseinilytica* gen. nov., sp. nov., a marine bacterium of the phylum *Bacteroidetes*. *Int J Syst Evol Microbiol* 2006; **56**:2317-2322. PubMed doi:10.1099/ijs.0.64415-0

9. Bernardet JF, Nakagawa Y, Holmes B. Proposed minimal standards for describing new taxa of the family *Flavobacteriaceae* and emended description of the family. *Int J Syst Evol Microbiol* 2002; **52**:1049-1070. PubMed doi:10.1099/ijs.0.02136-0

10. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol* 1990; **215**:403-410. PubMed

11. DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K, Huber T, Dalevi D, Hu P, Andersen GL. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl Environ Microbiol* 2006; **72**:5069-5072. PubMed doi:10.1128/AEM.03006-05

12. Porter MF. An algorithm for suffix stripping. *Program: electronic library and information systems* 1980; **14**:130-137.

13. Lee C, Grasso C, Sharlow MF. Multiple sequence alignment using partial order graphs. *Bioinformatics* 2002; **18**:452-464. PubMed doi:10.1093/bioinformatics/18.3.452

14. Castresana J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol* 2000; **17**:540-552. PubMed

15. Stamatakis A, Hoover P, Rougemont J. A rapid bootstrap algorithm for the RAxML web-servers. *Syst Biol* 2008; **57**:758-771. PubMed doi:10.1080/10635150802429642

16. Hess PN, De Moraes Russo CA. An empirical test of the midpoint rooting method. *Biol J Linn Soc Lond* 2007; **92**:669-674. doi:10.1111/j.1095-8312.2007.00864.x

17. Pattengale ND, Alipour M, Bininda-Emonds ORP, Moret BME, Stamatakis A. How many bootstrap replicates are necessary? *Lect Notes Comput Sci*

2009; **5541**:184-200. doi:10.1007/978-3-642-02008-7_13

18. Swofford DL. PAUP*: Phylogenetic Analysis Using Parsimony (*and Other Methods), Version 4.0 b10. Sinauer Associates, Sunderland, 2002.

19. Liolios K, Chen IM, Mavromatis K, Tavernarakis N, Hugenholtz P, Markowitz VM, Kyrpides NC. The Genomes On Line Database (GOLD) in 2009: status of genomic and metagenomic projects and their associated metadata. *Nucleic Acids Res* 2010; **38**:D346-D354. PubMed doi:10.1093/nar/gkp848

20. Field D, Garrity G, Gray T, Morrison N, Selengut J, Sterk P, Tatusova T, Thomson N, Allen MJ, Angiuoli SV, *et al*. The minimum information about a genome sequence (MIGS) specification. *Nat Biotechnol* 2008; **26**:541-547. PubMed doi:10.1038/nbt1360

21. Garrity G. NamesforLife. BrowserTool takes expertise out of the database and puts it right in the browser. *Microbiol Today* 2010; **37**:9.

22. Woese CR, Kandler O, Wheelis ML. Towards a natural system of organisms: proposal for the domains *Archaea, Bacteria*, and *Eucarya. Proc Natl Acad Sci USA* 1990; **87**:4576-4579. PubMed doi:10.1073/pnas.87.12.4576

23. Garrity GM, Holt JG. The Road Map to the Manual. *In*: Garrity GM, Boone DR, Castenholz RW (*eds*), Bergey's Manual of Systematic Bacteriology, Second Edition, Volume 1, Springer, New York, 2001, p. 119-169.

24. Ludwig W, Euzeby J, Whitman WG. Draft taxonomic outline of the *Bacteroidetes, Planctomycetes, Chlamydiae, Spirochaetes, Fibrobacteres, Fusobacteria, Acidobacteria, Verrucomicrobia, Dictyoglomi*, and *Gemmatimonadetes*. http://www.bergeys.org/outlines/Bergeys_Vol_4_Outline.pdf

25. Garrity GM, Holt JG. 2001. Taxonomic outline of the *Archaea* and *Bacteria*, p. 155-166. *In*: Garrity GM, Boone DR, Castenholz RW (*ed*), Bergey's Manual of Systematic Bacteriology, 2nd ed, vol. 1. Springer, New York.

26. BAuA. 2010, Classification of bacteria and archaea in risk groups. http://www.baua.de TRBA 466, p. 89.

27. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, *et al*. Gene Ontology: tool for the unification of biology. *Nat Genet* 2000; **25**:25-29. PubMed doi:10.1038/75556

28. Klenk HP, Göker M. En route to a genome-based classification of *Archaea* and *Bacteria? Syst Appl Microbiol* 2010; **33**:175-182. PubMed doi:10.1016/j.syapm.2010.03.003

29. Wu D, Hugenholtz P, Mavromatis K, Pukall R, Dalin E, Ivanova NN, Kunin V, Goodwin L, Wu M, Tindall BJ, *et al*. A phylogeny-driven genomic encyclopaedia of *Bacteria* and *Archaea. Nature* 2009; **462**:1056-1060. PubMed doi:10.1038/nature08656

30. List of growth media used at DSMZ: http://www.dsmz.de/microorganisms/media_list.php

31. Gemeinholzer B, Dröge G, Zetzsche H, Hasprunar G, Klenk HP, Güntsch A, Berendsohn WG, Wägele JW. The DNA Bank Network: the start from a German initiative. *Biopreservation and Biobanking* 2011; **9**:51-55. doi:10.1089/bio.2010.0029

32. JGI website. http://www.jgi.doe.gov

33. The Phred/Phrap/Consed software package. http://www.phrap.com

34. Zerbino DR, Birney E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res* 2008; **18**:821-829. PubMed doi:10.1101/gr.074492.107

35. Han C, Chain P. Finishing repeat regions automatically with Dupfinisher. *In:* Proceeding of the 2006 international conference on bioinformatics & computational biology. Arabnia HR, Valafar H (*eds*), CSREA Press. June 26-29, 2006: 141-146.

36. Lapidus A, LaButti K, Foster B, Lowry S, Trong S, Goltsman E. POLISHER: An effective tool for using ultra short reads in microbial genome assembly and finishing. AGBT, Marco Island, FL, 2008.

37. Hyatt D, Chen GL, LoCascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 2010; **11**:119. PubMed doi:10.1186/1471-2105-11-119

38. Pati A, Ivanova NN, Mikhailova N, Ovchinnikova G, Hooper SD, Lykidis A, Kyrpides NC. Gene-PRIMP: a gene prediction improvement pipeline for prokaryotic genomes. *Nat Methods* 2010; **7**:455-457. PubMed doi:10.1038/nmeth.1457

39. Markowitz VM, Ivanova NN, Chen IMA, Chu K, Kyrpides NC. IMG ER: a system for microbial genome annotation expert review and curation. *Bioinformatics* 2009; **25**:2271-2278. PubMed doi:10.1093/bioinformatics/btp393

　　　　　　　　　　　　　　　　　　　　　　The Genomic Standards Consortium