

Complete genome sequence of *Bacteroides salanitronis* type strain (BL78^T)

Sabine Gronow¹, Brittany Held^{2,3}, Susan Lucas², Alla Lapidus², Tijana Glavina Del Rio², Matt Nolan², Hope Tice², Shweta Deshpande², Jan-Fang Cheng², Sam Pitluck², Konstantinos Liolios², Ioanna Pagani², Natalia Ivanova², Konstantinos Mavromatis², Amrita Pati², Roxane Tapia^{2,3}, Cliff Han^{2,3}, Lynne Goodwin^{2,3}, Amy Chen⁴, Krishna Palaniappan⁴, Miriam Land^{2,5}, Loren Hauser^{2,5}, Yun-Juan Chang^{2,5}, Cynthia D. Jeffries^{2,5}, Evelyne-Marie Brambilla¹, Manfred Rohde⁶, Markus Göker¹, John C. Detter^{2,3}, Tanja Woyke², James Bristow², Victor Markowitz⁴, Philip Hugenholtz^{2,8}, Nikos C. Kyrpides², Hans-Peter Klenk¹, and Jonathan A. Eisen^{2,7*}

¹ DSMZ - German Collection of Microorganisms and Cell Cultures GmbH, Braunschweig, Germany

² DOE Joint Genome Institute, Walnut Creek, California, USA

³ Los Alamos National Laboratory, Bioscience Division, Los Alamos, New Mexico, USA

⁴ Biological Data Management and Technology Center, Lawrence Berkeley National Laboratory, Berkeley, California, USA

⁵ Lawrence Livermore National Laboratory, Livermore, California, USA

⁶ HZI – Helmholtz Centre for Infection Research, Braunschweig, Germany

⁷ University of California Davis Genome Center, Davis, California, USA

⁸ Australian Centre for Ecogenomics, School of Chemistry and Molecular Biosciences, The University of Queensland, Brisbane, Australia

*Corresponding author: Jonathan A. Eisen

Keywords: strictly anaerobic, non-motile, rod-shaped, Gram-negative, mesophilic, cecum, poultry, chemoorganotrophic, *Bacteroidaceae*, GEBA

Bacteroides salanitronis Lan et al. 2006 is a species of the genus *Bacteroides*, which belongs to the family *Bacteroidaceae*. The species is of interest because it was isolated from the gut of a chicken and the growing awareness that the anaerobic microflora of the cecum is of benefit for the host and may impact poultry farming. The 4,308,663 bp long genome consists of a 4.24 Mbp chromosome and three plasmids (6 kbp, 19 kbp, 40 kbp) containing 3,737 protein-coding and 101 RNA genes and is a part of the *Genomic Encyclopedia of Bacteria and Archaea* project.

Introduction

Strain BL78^T (= DSM 18170 = CCUG 54637 = JCM 13657) is the type strain of *Bacteroides salanitronis* which belongs to the large genus *Bacteroides* [1,2]. Currently, there are 88 species placed in the genus *Bacteroides*. The species epithet is derived from the name of Joseph P. Salanitro, an American microbiologist. *B. salanitronis* strain BL78^T was isolated among other *Bacteroides* strains from the cecum of a healthy chicken. No other strain belonging to the same species has been identified [2]. Many *Bacteroides* species are common inhabitants of the intestine where they help to degrade complex molecules such as polysaccharides or transform steroids [3,4]. They also play a role as beneficent protectors of the gut against pathogenic microorganisms [5]. Here we present a sum-

mary classification and a set of features for *B. salanitronis* BL78^T, together with the description of the complete genomic sequencing and annotation.

Classification and features

A representative genomic 16S rRNA sequence of strain BL78^T was compared using NCBI BLAST 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 [6] and the relative frequencies, weighted by BLAST scores, of taxa and keywords (reduced to their stem [7]) were determined. The single most frequent genus was *Bacteroides* (100.0%) (1 hit in total). Regarding the single hit to sequences from members of the species, the average identity within

HSPs was 99.7%, whereas the average coverage by HSPs was 96.2%. No hits to sequences with (other) species names were found. The highest-scoring environmental sequence was DQ456041 ('pre-adolescent turkey cecum clone CFT112F11'), which showed an identity of 96.8% and an HSP coverage of 63.9%. The five most frequent keywords within the labels of environmental samples which yielded hits were 'fecal' (9.3%), 'microbiota' (7.5%), 'human' (7.1%), 'antibiot, effect, gut, pervas' (7.1%) and 'anim, beef, cattl, coli, escherichia, feedlot, habitat, synecolog' (2.2%) (249 hits in total).

Figure 1 shows the phylogenetic neighborhood of *B. salanitronis* in a 16S rRNA based tree. The sequences of the six 16S rRNA gene copies in the genome differ from each other by up to 26 nucleotides, and differ by up to 26 nucleotides from the previously published 16S rRNA sequence (AB253731).

The cells of *B. salanitronis* are generally rod-shaped (0.4-0.7 × 0.8-5.6 μm) with rounded ends (Figure 2). The cells are usually arranged singly or in pairs [2]. *B. salanitronis* is a Gram-negative, non-spore-forming bacterium (Table 1) that is described as non-motile, with only five genes associated with motility having been found in the genome (see below).

The temperature optimum for strain BL78^T is 37°C. *B. salanitronis* is a strictly anaerobic chemoorganotroph and is able to ferment glucose, mannose, sucrose, maltose, arabinose, cellobiose, lactose, xylose and raffinose [2]. The organism hydrolyzes esculin but does not liquefy gelatin, and neither reduces nitrate nor produces indole from tryptophan [2]. *B. salanitronis* does not utilize trehalose, glycerol, mannitol, sorbitol or melezitose; rhamnose and salicin are fermented weakly [2]. Growth is possible in the presence of bile [2]. Major fermentation products from broth (1% peptone, 1% yeast extract, and 1% glucose each (w/v)) are acetic acid and succinic acid, whereas isovaleric acid is produced in small amounts [2]. *B. salanitronis* shows activity for alkaline phosphatase, α- and β-galactosidases, α- and β-glucosidases, α-arabinosidase, leucyl glycine arylamidase, alanine arylamidase and glutamyl glutamic acid arylamidase but no activity for urease, catalase, glutamic acid decarboxylase, arginine dihydrolase, β-galactosidase 6-phosphate, β-glucuronidase, N-acetyl-β-glucosaminidase, α-fucosidase and arginine, proline, leucine, phenylalanine, pyroglutamic acid, tyrosine, glycine, histidine and serine arylamidase [2].

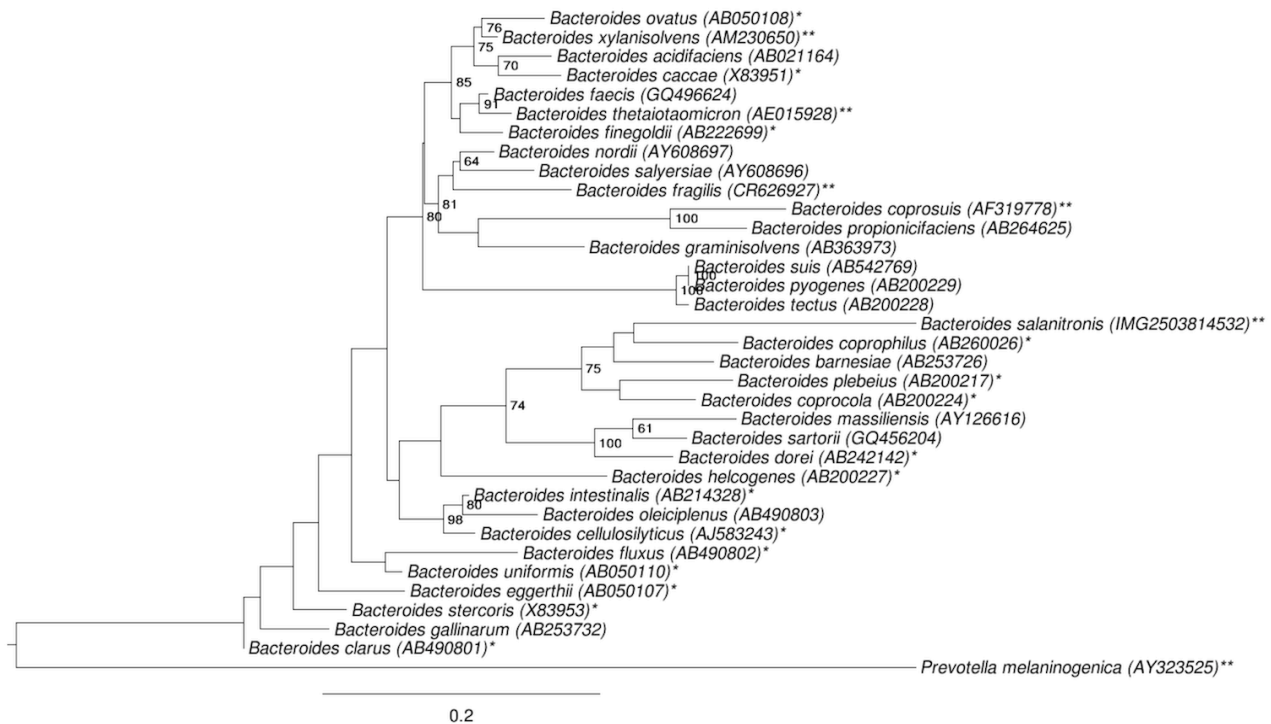


Figure 1. Phylogenetic tree highlighting the position of *B. salanitronis* relative to a selection of other type strains within the genus. The tree was inferred from 1,412 aligned characters [8,9] of the 16S rRNA gene sequence under the maximum likelihood criterion [10] and rooted in accordance with the current taxonomy. The branches are scaled in terms of the expected number of substitutions per site. Numbers to the right of bifurcations are support values from 1,000 bootstrap replicates [11] if larger than 60%. Lineages with type strain genome sequencing projects registered in GOLD [12] but unpublished are labeled with one asterisk, published genomes with two asterisks [13-15].

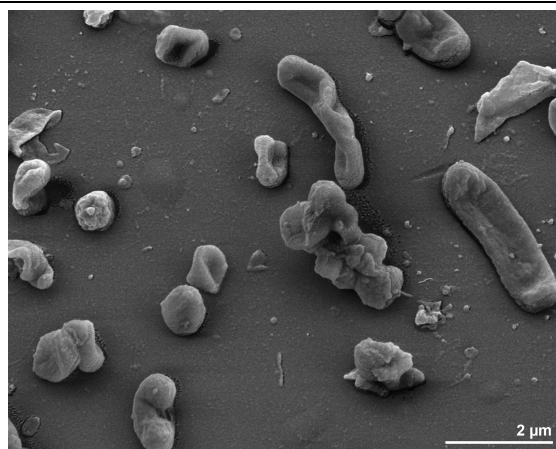


Figure 2. Scanning electron micrograph of *B. salanitronis* BL78^T

Table 1. Classification and general features of *B. salanitronis* BL78^T according to the MIGS recommendations [16].

MIGS ID	Property	Term	Evidence code
		Domain <i>Bacteria</i>	TAS [17]
		Phylum ' <i>Bacteroidetes</i> '	TAS [18]
		Class ' <i>Bacteroidia</i> '	TAS [19]
	Current classification	Order ' <i>Bacteroidales</i> '	TAS [20]
		Family <i>Bacteroidaceae</i>	TAS [21,22]
		Genus <i>Bacteroides</i>	TAS [21,23-26]
		Species <i>Bacteroides salanitronis</i>	TAS [2]
		Type strain BL78	TAS [2]
	Gram stain	negative	TAS [2]
	Cell shape	rod-shaped	TAS [2]
	Motility	non-motile	TAS [2]
	Sporulation	none	TAS [2]
	Temperature range	mesophile	TAS [2]
	Optimum temperature	37°C	TAS [2]
	Salinity	normal	NAS
MIGS-22	Oxygen requirement	strictly anaerobic	TAS [2]
	Carbon source	carbohydrates	TAS [2]
	Energy metabolism	chemoorganotroph	TAS [2]
MIGS-6	Habitat	chicken	TAS [2]
MIGS-15	Biotic relationship	free-living	NAS
MIGS-14	Pathogenicity	none	NAS
	Biosafety level	1	TAS [27]
	Isolation	chicken cecum	TAS [2]
MIGS-4	Geographic location	Japan	TAS [2]
MIGS-5	Sample collection time	November 2005	IDA
MIGS-4.1	Latitude	not reported	
MIGS-4.2	Longitude	not reported	
MIGS-4.3	Depth	not reported	
MIGS-4.4	Altitude	not reported	

Evidence codes - IDA: Inferred from Direct Assay (first time in publication); 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 [28]. If the evidence code is IDA, then the property was directly observed by one of the authors or an expert mentioned in the acknowledgements.

Chemotaxonomy

B. salanitronis strain BL78^T contains menaquinones MK-11 and MK-12 as principal respiratory quinones (43% each), small amounts of MK-10 (5%) and MK-13 (7%) are found as minor components [2]. The major fatty acids found were *anteiso*-C_{15:0} (32%), *iso*-C_{15:0} (14%), 3-hydroxy C_{16:0} (12%) and 3-hydroxy *iso*-C_{17:0} (10%). Fatty acids C_{14:0} (4%), C_{15:0} (2%), C_{16:0} (8%), C_{18:1} (2%), C_{18:2} (2%) and *iso*-C_{14:0} (2%) were found in minor amounts [2].

Genome sequencing and annotation

Genome project history

This organism was selected for sequencing on the basis of its phylogenetic position [29], and is part of the *Genomic Encyclopedia of Bacteria and Archaea* project [30]. The genome project is deposited in the Genomes On Line Database [31] 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	Three genomic libraries: one 454 pyrosequence standard library, one 454 PE library (7 kb insert size), one Illumina library
MIGS-29	Sequencing platforms	Illumina GAii, 454 GS FLX Titanium
MIGS-31.2	Sequencing coverage	283.0 × Illumina; 37.7 × pyrosequence
MIGS-30	Assemblers	Newbler version 2.3-PreRelease-09-14-2009-bin, Velvet, phrap version SPS 4.24
MIGS-32	Gene calling method	Prodigal 1.4, GenePRIMP
	INSDC ID	CP002530 (chromosome) CP002531 (plasmid 1) CP002532 (plasmid 2) CP002533 (plasmid 3)
	Genbank Date of Release	February 28, 2011
	GOLD ID	Gc001665
	NCBI project ID	40066
	Database: IMG-GEBA	2503754023
MIGS-13	Source material identifier	DSM 18170
	Project relevance	Tree of Life, GEBA

Growth conditions and DNA isolation

B. salanitronis BL78^T, DSM 18170, was grown anaerobically in DSMZ medium 104 (Peptone-Yeast extract-Glucose broth) [32] at 37°C. DNA was isolated from 0.5-1 g of cell paste using MasterPure Gram-positive DNA purification kit (Epicentre MGP04100) following the standard protocol as recommended by the manufacturer, adding 20 µL lysozyme (100mg/µl), and 10 µL mutanolysin, achromopeptidase, and lysostaphine, each, for 40 min lysis at 37°C followed by one hour incubation on ice. DNA is available through the DNA Bank Network [33].

Genome sequencing and assembly

The genome was sequenced using a combination of Illumina and 454 sequencing platforms. All general aspects of library construction and se-

quencing can be found at the JGI website [34]. Pyrosequencing reads were assembled using the Newbler assembler version 2.3-PreRelease-09-14-2009-bin (Roche). The initial Newbler assembly consisting of 100 contigs in two scaffolds was converted into a phrap assembly [35] by making fake reads from the consensus, to collect the read pairs in the 454 paired-end library. Illumina GAii sequencing data (920.8 Mb) was assembled with Velvet, version 0.7.63 [36] 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 109.0 Mb of 454 standard 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 [35] was used for sequence assem-

bly 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 [34], Dupfinisher [37], or sequencing cloned bridging PCR fragments with subcloning or transposon bombing (Epicentre Biotechnologies, Madison, WI). Gaps between contigs were closed by editing in Consed, by PCR and by Bubble PCR primer walks (J.-F.Chang, unpublished). A total of 193 additional reactions and four 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 [38]. The error rate of the completed genome sequence is less than 1 in 100,000. Together, the combination of the Illumina and 454 sequencing platforms provided $320.7 \times$ coverage of the genome. The final assembly contained 393,135 pyrosequence and 25,576,764 Illumina reads.

Genome annotation

Genes were identified using Prodigal [39] as part of the Oak Ridge National Laboratory genome an-

notation pipeline, followed by a round of manual curation using the JGI GenePRIMP pipeline [40]. The predicted CDSs were translated and used to search the National Center for Biotechnology Information (NCBI) nonredundant 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 [41].

Genome properties

The genome consists of a 4,242,803 bp long chromosome with a G+C content of 47%, as well as three plasmids of 6,277 bp, 18,280 bp and 40,303 bp length (Table 3 and Figure 3). Of the 3,838 genes predicted, 3,737 were protein-coding genes, and 101 RNAs; 96 pseudogenes were also identified. The majority of the protein-coding genes (57.3%) were assigned with 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.

Table 3. Genome Statistics

Attribute	Value	% of Total
Genome size (bp)	4,308,663	100.00%
DNA coding region (bp)	3,759,354	87.25%
DNA G+C content (bp)	2,003,128	46.49%
Number of replicons	4	
Extrachromosomal elements	3	
Total genes	3,838	100.00%
RNA genes	101	2.63%
rRNA operons	6	
Protein-coding genes	3,737	97.37%
Pseudo genes	96	2.50%
Genes with function prediction	2,200	57.32%
Genes in paralog clusters	876	22.82%
Genes assigned to COGs	2,013	52.45%
Genes assigned Pfam domains	2,269	59.12%
Genes with signal peptides	918	23.92%
Genes with transmembrane helices	794	20.69%
CRISPR repeats	0	

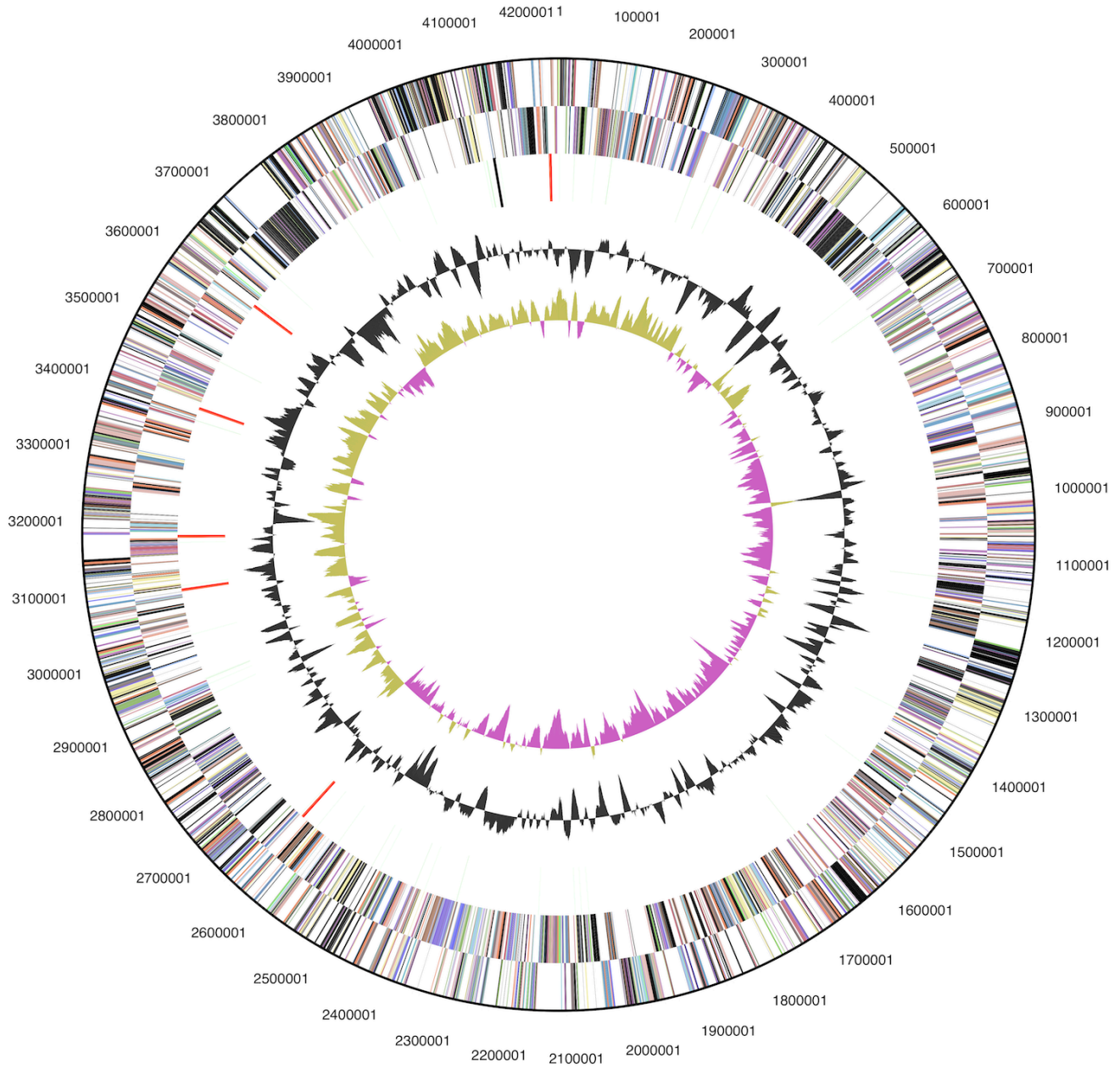


Figure 3. Graphical circular map of the chromosome (plasmid maps not shown). From outside to the center: Genes on forward strand (color by COG categories), Genes on reverse strand (color by COG categories), RNA genes (tRNAs green, rRNAs red, other RNAs black), GC content, GC skew.

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

Code	value	%age	Description
J	147	6.8	Translation, ribosomal structure and biogenesis
A	0	0.0	RNA processing and modification
K	143	6.6	Transcription
L	194	9.0	Replication, recombination and repair
B	0	0.0	Chromatin structure and dynamics
D	31	1.4	Cell cycle control, cell division, chromosome partitioning
Y	0	0.0	Nuclear structure
V	63	2.9	Defense mechanisms
T	85	3.9	Signal transduction mechanisms
M	193	8.9	Cell wall/membrane/envelope biogenesis
N	5	0.2	Cell motility
Z	0	0.0	Cytoskeleton
W	0	0.0	Extracellular structures
U	61	2.8	Intracellular trafficking, secretion, and vesicular transport
O	61	2.8	Posttranslational modification, protein turnover, chaperones
C	105	4.9	Energy production and conversion
G	174	8.0	Carbohydrate transport and metabolism
E	134	6.2	Amino acid transport and metabolism
F	68	3.1	Nucleotide transport and metabolism
H	98	4.5	Coenzyme transport and metabolism
I	62	2.9	Lipid transport and metabolism
P	104	4.8	Inorganic ion transport and metabolism
Q	29	1.3	Secondary metabolites biosynthesis, transport and catabolism
R	285	13.2	General function prediction only
S	125	5.8	Function unknown
-	1,825	47.6	Not in COGs

Acknowledgements

We would like to gratefully acknowledge the help of Sabine Welnitz (DSMZ) for growing cultures of *B. salanitronis*. This work was performed under the auspices of the US Department of Energy 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, UT-Battelle and Oak Ridge National Laboratory under contract DE-AC05-00OR22725, as well as German Research Foundation (DFG) INST 599/1-2.

References

- Garrity G. NamesforLife. BrowserTool takes expertise out of the database and puts it right in the browser. *Microbiol Today* 2010; 7:1.
- Lan PTN, Sakamoto M, Sakata S, Benno Y. *Bacteroides barnesiae* sp. nov., *Bacteroides salanitronis* sp. nov. and *Bacteroides gallinarum* sp. nov., isolated from chicken caecum. *Int J Syst Evol Microbiol* 2006; 56:2853-2859. [PubMed doi:10.1099/ijs.0.64517-0](#)
- Comstock LE. Importance of glycans to the host-bacteroides mutualism in the mammalian intestine. *Cell Host Microbe* 2009; 5:522-526. [PubMed doi:10.1016/j.chom.2009.05.010](#)

4. Bäckhed F, Ley RE, Sonnenburg JL, Peterson DA, Gordon JI. Host-bacterial mutualism in the human intestine. *Science* 2005; **307**:1915-1920. [PubMed doi:10.1126/science.1104816](#)
5. Hentges DJ. Role of the intestinal flora in host defense against infection. In *Human Intestinal Microflora in Health and Disease* 1983; pp. 311–331. Edited by D. J. Hentges. New York: Academic Press.
6. 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](#)
7. Porter MF. An algorithm for suffix stripping. *Program: electronic library and information systems* 1980; **14**:130-137.
8. 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](#)
9. Castresana J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol* 2000; **17**:540-552. [PubMed](#)
10. 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](#)
11. 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](#)
12. 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](#)
13. Cerdeño-Tárraga AM, Patrick S, Crossman LC, Blakely G, Abratt V, Lennard N, Poxton I, Duerden B, Harris B, Quail MA, et al. Extensive DNA inversions in the *B. fragilis* genome control variable gene expression. *Science* 2005; **307**:1463-1465. [PubMed doi:10.1126/science.1107008](#)
14. Xu J, Bjursell MK, Himrod J, Deng S, Carmichael LK, Chiang HC, Hooper LV, Gordon JI. A genomic view of the human *Bacteroides thetaiotaomicron* symbiosis. *Science* 2003; **299**:2074-2076. [PubMed doi:10.1126/science.1080029](#)
15. Land M, Held B, Gronow S, Abt B, Lucas S, Glavina Del Rio T, Nolan M, Tice H, Cheng JF, Pitluck S, et al. Non-contiguous finished genome sequence of *Bacteroides coprosuis* type strain (PC 139^T). *Stand Genomic Sci* 2011; **4**
16. 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](#)
17. 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](#)
18. 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.
19. Ludwig W, Euzéby 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. Taxonomic Outline 2008
20. Garrity GM, Holt JG. 2001. Taxonomic outline of the *Archaea* and *Bacteria*, p. 155-166. In G. M. Garrity, D. R. Boone, and R. W. Castenholz (ed.), *Bergey's Manual of Systematic Bacteriology*, 2nd ed, vol. 1. Springer, New York.
21. Skerman VBD, McGowan V, Sneath PHA. Approved Lists of Bacterial Names. *Int J Syst Bacteriol* 1980; **30**:225-420. [doi:10.1099/00207713-30-1-225](#)
22. Pribram E. Klassifikation der Schizomyceten. *Klassifikation der Schizomyceten (Bakterien)*, Franz Deuticke, Leipzig, 1933, p. 1-143.
23. Castellani A, Chalmers AJ. Genus *Bacteroides* Castellani and Chalmers, 1918. *Manual of Tropical Medicine*, Third Edition, Williams, Wood and Co., New York, 1919, p. 959-960.
24. Holdeman LV, Moore WEC. Genus I. *Bacteroides* Castellani and Chalmers 1919, 959. In: Buchanan RE, Gibbons NE (eds), *Bergey's Manual of Determinative Bacteriology*, Eighth Edition, The Wil-

- liams and Wilkins Co., Baltimore, 1974, p. 385-404.
25. Cato EP, Kelley RW, Moore WEC, Holdeman LV. *Bacteroides zoogloiformans*, Weinberg, Nativelle, and Prévot 1937) corrig. comb. nov.: emended description. *Int J Syst Bacteriol* 1982; **32**:271-274. [doi:10.1099/00207713-32-3-271](https://doi.org/10.1099/00207713-32-3-271)
 26. Shah HN, Collins MD. Proposal to restrict the genus *Bacteroides* (Castellani and Chalmers) to *Bacteroides fragilis* and closely related species. *Int J Syst Bacteriol* 1989; **39**:85-87. [doi:10.1099/00207713-39-1-85](https://doi.org/10.1099/00207713-39-1-85)
 27. Classification of bacteria and archaea in risk groups. [http://www.baua.de TRBA 466](http://www.baua.de/TRBA_466).
 28. 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](https://pubmed.ncbi.nlm.nih.gov/103875556/)
 29. 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](https://pubmed.ncbi.nlm.nih.gov/1016/j.syapm.2010.03.003)
 30. 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](https://pubmed.ncbi.nlm.nih.gov/1038/nature08656)
 31. 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](https://pubmed.ncbi.nlm.nih.gov/101093/bioinformatics/btp393)
 32. List of growth media used at DSMZ: http://www.dsmz.de/microorganisms/media_list.php
 33. Gemeinholzer B, Dröge G, Zetzsche H, Haszprunar 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](https://doi.org/10.1089/bio.2010.0029)
 34. The DOE Joint Genome Institute. <http://www.jgi.doe.gov/>
 35. Phrap and Phred for Windows, MacOS, Linux, and Unix. www.phrap.com
 36. 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](https://pubmed.ncbi.nlm.nih.gov/101101/gr.074492.107)
 37. Han C, Chain P. 2006. Finishing repeat regions automatically with Dupfinisher. In: Proceeding of the 2006 international conference on bioinformatics & computational biology. Arabina HR, Valafar H (eds), CSREA Press. June 26-29, 2006: 141-146.
 38. 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.
 39. 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](https://pubmed.ncbi.nlm.nih.gov/101186/1471-2105-11-119)
 40. Pati A, Ivanova NN, Mikhailova N, Ovchinnikova G, Hooper SD, Lykidis A, Kyrpides NC. GenePRIMP: a gene prediction improvement pipeline for prokaryotic genomes. *Nat Methods* 2010; **7**:455-457. [PubMed doi:10.1038/nmeth.1457](https://pubmed.ncbi.nlm.nih.gov/101038/nmeth.1457)
 41. 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](https://pubmed.ncbi.nlm.nih.gov/101093/bioinformatics/btp393)