Complete genome sequence of *Beutenbergia cavernae* type strain (HKI 0122^T)

Miriam Land[1,2], Rüdiger Pukall[3], Birte Abt[3], Markus Göker[3], Manfred Rohde[4], Tijana Glavina Del Rio[1], Hope Tice[1], Alex Copeland[1], Jan-Fang Cheng[1], Susan Lucas[1], Feng Chen[1], Matt Nolan[1], David Bruce[1,5], Lynne Goodwin[1,5], Sam Pitluck[1], Natalia Ivanova[1], Konstantinos Mavromatis[1] , Galina Ovchinnikova [1], Amrita Pati[1], Amy Chen[6], Krishna Palaniappan[6], Loren Hauser[1,2], Yun-Juan Chang[1,2], Cynthia C. Jefferies[1,2], Elizabeth Saunders[5], Thomas Brettin[1,5], John C. Detter[1,5], Cliff Han[1,5], Patrick Chain[1,7], James Bristow[1], Jonathan A. Eisen[1,8], Victor Markowitz[6], Philip Hugenholtz[1], Nikos C. Kyrpides[1], Hans-Peter Klenk[3], and Alla Lapidus[1*]

[1] DOE Joint Genome Institute, Walnut Creek, California, USA

[2] Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA

[3] DSMZ - German Collection of Microorganisms and Cell Cultures GmbH, Braunschweig, Germany

[4] HZI - Helmholtz Centre for Infection Research, Braunschweig, Germany

[5] Los Alamos National Laboratory, Bioscience Division, Los Alamos, New Mexico USA

[6] Biological Data Management and Technology Center, Lawrence Berkeley National Laboratory, Berkeley, California, USA

[7] Lawrence Livermore National Laboratory, Livermore, California, USA

[8] University of California Davis Genome Center, Davis, California, USA

[*]Corresponding author: Alla Lapidus

Keywords: mesophile, non-pathogenic, aerobic and microaerophilic, rod-coccus growth cycle, MK-8(H_4), actinomycete, *Micrococcineae*

Beutenbergia cavernae (Groth *et al.* 1999) is the type species of the genus and is of phylogenetic interest because of its isolated location in the actinobacterial suborder *Micrococcineae*. *B. cavernae* HKI 0122^T is a Gram-positive, non-motile, non-spore-forming bacterium isolated from a cave in Guangxi (China). *B. cavernae* grows best under aerobic conditions and shows a rod-coccus growth cycle. Its cell wall peptidoglycan contains the diagnostic L-lysine ← L-glutamate interpeptide bridge. 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 poorly populated micrococcineal family *Beutenbergiaceae*, and this 4,669,183 bp long single replicon genome with its 4225 protein-coding and 53 RNA genes is part of the ***G**enomic **E**ncyclopedia of **B**acteria and **A**rchaea* project.

Introduction

Beutenbergia cavernae strain HKI 0122^T (DSM 12333 = ATCC BAA-8 = JCM 11478) is the type strain of the species, which represents the type species of the genus *Beutenbergia*, the type genus of the family *Beutenbergiaceae* [1]. *B. cavernae* was described by Groth et al. 1999 as Gram-positive, non-motile and non-spore-forming [1]. The organism is of significant interest for its position in the tree of life within the small (2 type strains) family *Beutenbergiaceae* Zhi, et al., 2009 emend. Schumann et al. 2009 in the actinobacterial suborder *Micrococcineae* [2], which in addition to the genus *Beutenbergia* contains only the genus *Salana* [3,4] (Figure 1), also otherwise stated in a

recent overview on the class *Actinobacteria* [2]. Here we present a summary classification and a set of features for *B. cavernae* strain HKI 0122^T (Table 1), together with the description of the complete genome sequencing and annotation.

In addition to strain HKI 0122^T, only one other strain (HKI 0132) was isolated from the soil sample collected in the Reed Flute Cave near Guilin, Guangxi, China. HKI 0132 was also classified in the species *B. cavernae* [1]. No closely related isolates and uncultivated clones with more than 97% 16S rRNA gene sequence identity are recorded in the microbiological literature, nor can any phylotype from environmental samples or genomic surveys be directly linked to *B. cavernae.*

B. cavernae cells vary in shape and colonies grown on rich medium vary in color from cream to bright yellow. In young cultures, cells are irregular rods arranged in palisades, clusters or in pairs at an angle to give V-formations [1]. Cells in stationary cultures are predominantly coccoid, occurring singly, in pairs, irregular clusters and short chains. During growth in complex media a rod-coccus growth cycle was observed [1]. *B. cavernae* grows well under aerobic and microaerophilic conditions, but not under anaerobic conditions [1]. The optimal growth temperature is 28°C [1].

B. cavernae is able to degrade casein, esculin, gelatin and potato starch. Acids are produced from L-arabinose, D-cellobiose, dextrin, D-fructose, D-galactose, D-glucose, glycerol, inulin, maltose, D-mannose, D-raffinose, L-rhamnose, D-ribose, salicin, sucrose, starch, trehalose and D-xylose. There is no acid production from D-glucitol, lactose and D-mannitol. Nitrate is reduced to nitrite, H_2S is produced [1].

Classification and features

Figure 1. shows the phylogenetic neighborhood of *B. cavernae* strain HKI 0122^T in a 16S rRNA based tree. Analysis of the two identical 16S rRNA gene sequences in the genome of strain HKI differed by four nucleotides from the previously published 16S rRNA sequence generated from DSM 12333 (Y18378). The slight differences between the genome data and the reported 16S rRNA gene sequence is most likely due to sequencing errors in the previously reported sequence data .

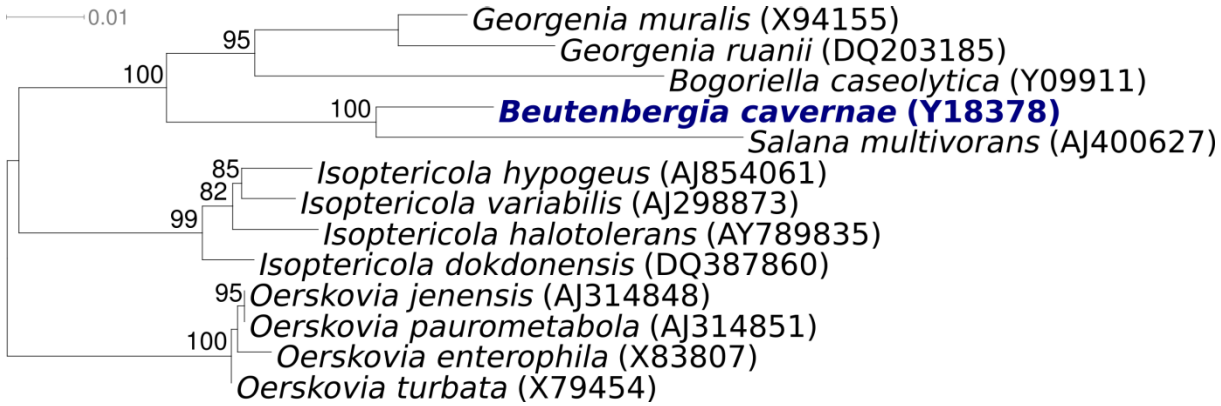

Figure 1. Phylogenetic tree of *B. cavernae* HKI 0122^T and all type strains of the genus *Beutenbergia*, inferred from 1,411 aligned characters [5, 6] of the 16S rRNA sequence under the maximum likelihood criterion [7]. The tree was rooted with species from the genera *Isoptericola* and *Oerskovia*, both also members of the actinobacterial suborder *Micrococcineae*. The branches are scaled in terms of the expected number of substitutions per site. Numbers above branches are support values from 1,000 bootstrap replicates if larger than 60%. Strains with a genome-sequencing project registered in GOLD [8] are printed in blue; published genomes in bold.

Chemotaxonomy

The peptidoglycan of *B. cavernae* HKI 0122^T contains D- and L-alanine, D- and L-glutamic acid and L-lysine, with the latter widely distributed among actinobacteria [1]. The strain possesses a type A4⟨ peptidoglycan with a diagnostic L-Lys←L-Glu interpeptide bridge, type A11.54 according to DSMZ. Glucose, mannose and galactose are the cell wall sugars [1]. The fatty acid profile of strain *B. cavernae* HKI 0122^T is dominated by 13-methyl tetradecanoic (iso-$C_{15:0}$; 43.7%) and 12-methyl tetradecanoic (anteiso-$C_{15:0}$; 34.6%) saturated, branched chain acids. Other predominantly saturated fatty acids play a minor role in the cellular fatty acid composition of the strain: iso-$C_{14:0}$ (0.9%), $C_{14:0}$ (1.9%); $C_{15:0}$ (0.9%) iso$C_{16:0}$ (2.3%), $C_{16:0}$ (6.8%), iso$C_{17:0}$ (3.1%), anteiso-$C_{17:0}$ (4.9%), und $C_{18:1}$ (0.9%) [1]. Mycolic acids are not present [1]. MK-8(H_4) is the major menaquinone, complemented by minor amounts of MK-8(H_2), MK-8 and MK-9(H_4) [1]. The combination of the L-Lys←L-Glu interpeptide bridge and MK-8(H_4) as the dominating menaquinone is shared with the organisms from the neighboring genera *Bogoriella* and *Georgenia*. The polar lipids of strain HKI 0122^T consist of phosphatidylinositol and diphosphatidylglycerol together with three yet unidentified phospholipids [1].

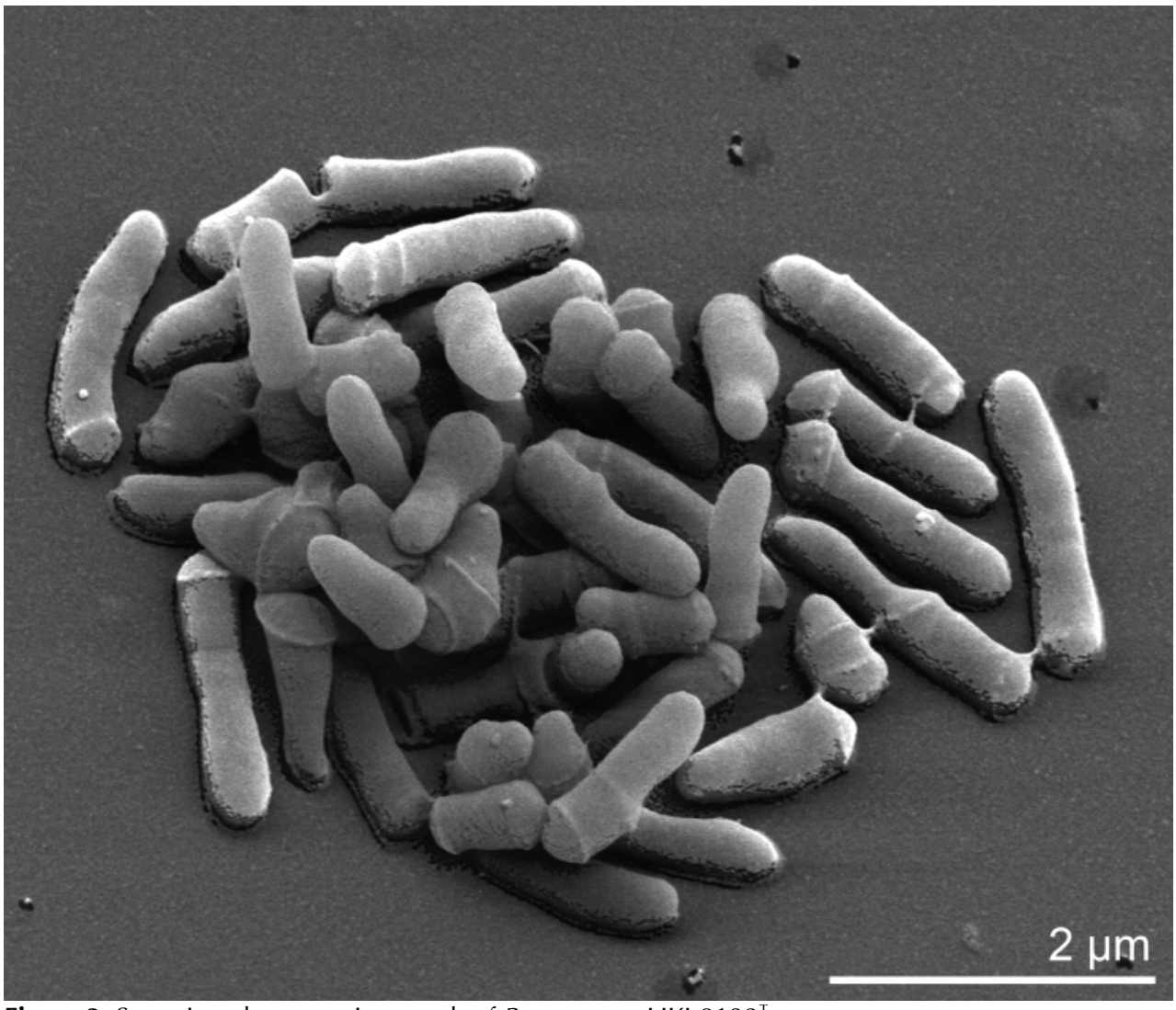

Figure 2. Scanning electron micrograph of *B. cavernae* HKI 0122^T

Table 1. Classification and general features of *B. cavernae* HKI 0122[T] based on the MIGS recommendations [9]

MIGS ID	Property	Term	Evidence code
	Current classification	Domain *Bacteria*	
		Phylum *Actinobacteria*	
		Class *Actinobacteria*	TAS [10]
		Order *Actinomycetales*	TAS [10]
		Suborder *Micrococcineae*	TAS [2]
		Family *Beutenbergiaceae*	TAS [2]
		Genus *Beutenbergia*	TAS [1]
		Species *Beutenbergia cavernae*	TAS [1]
		Type strain HKI 0122	
	Gram stain	positive	TAS [1]
	Cell shape	varies; rod-coccus growth cycle	TAS [1]
	Motility	nonmotile	TAS [1]
	Sporulation	non-sporulating	TAS [1]
	Temperature range	mesophile	TAS [1]
	Optimum temperature	28°C	TAS [1]
	Salinity	tolerance of 2-4% (w/v) NaCl	TAS [1]
MIGS-22	Oxygen requirement	aerobic and microaerobic, no growth under anaerobic conditions	TAS [1]
	Carbon source	glucose, maltose, mannose, cellobiose	TAS [1]
	Energy source	unknown	
MIGS-6	Habitat	cave (soil)	TAS [1]
MIGS-15	Biotic relationship		
MIGS-14	Pathogenicity	none	NAS
	Biosafety level	1	TAS [11]
	Isolation	cave, soil between rocks	TAS [1]
MIGS-4	Geographic location	Guangxi, China	TAS [1]
MIGS-5	Sample collection time	about 1999	TAS [1]
MIGS-4.1 MIGS-4.2	Longitude - Latitude	110.263306 - 25.307878	TAS [1]
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 the Gene Ontology project [12]. If the evidence code is IDA the property was directly observed for a live isolate by one of the authors or an expert mentioned in the acknowledgements.

Genome sequencing and annotation

Genome project history

This organism was selected for sequencing on the basis of its phylogenetic position, and is part of the ***G**enomic **E**ncyclopedia of **B**acteria and **A**rchaea* project. The genome project is deposited in the Genomes OnLine Database [8] and the complete genome sequence in GenBank (CP001618). 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: two Sanger libraries - 8 kb pMCL200 and fosmid pcc1Fos - and one 454 pyrosequence standard library
MIGS-29	Sequencing platforms	ABI3730, 454 GS FLX
MIGS-31.2	Sequencing coverage	8.56x Sanger; 10.86x pyrosequence
MIGS-30	Assemblers	Newbler version 1.1.02.15, phrap
MIGS-32	Gene calling method	Prodigal
	INSDC / Genbank ID	CP001618
	Genbank Date of Release	07-MAY-2009
	GOLD ID	Gc01025
	NCBI project ID	20827
	Database: IMG-GEBA	2501416922
MIGS-13	Source material identifier	DSM 12333
	Project relevance	Tree of Life, GEBA

Growth conditions and DNA isolation

B. cavernae HKI 0122^T, DSM 12333, was grown in DSMZ medium 736 (Rich Medium) [13] at 28°C. DNA was isolated from 0.5-1 g of cell paste using Qiagen Genomic 500 DNA Kit (Qiagen, Hilden, Germany) with a modification of the standard protocol for cell lysis in first freezing for 20 min. (-70°C), then heating 5 min. (98°C), and cooling 15 min to 37°C; adding 1.5 ml lysozyme (standard: 0.3 ml, only), 1.0 ml achromopeptidase, 0.12 ml lysostaphine, 0.12 ml mutanolysine, 1.5 ml proteinase K (standard: 0.5 ml, only), followed by overnight incubation at 35°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 the JGI website. 454 Pyrosequencing reads were assembled using the Newbler assembler version 1.1.02.15 (Roche). Large Newbler contigs were broken into 5,256 overlapping fragments of 1,000 bp and entered into the 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 [14]. Gaps between contigs were closed by editing in Consed, custom primer walking or PCR amplification. A total of 1,627 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 19.42x coverage of the genome.

Genome annotation

Genes were identified using Prodigal [15] as part of the Oak Ridge National Laboratory genome annotation pipeline, followed by a round of manual curation using the JGIGenePRIMP pipeline [16]. 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 [17].

Genome properties

The genome is 4,669,183 bp long and comprises one main circular chromosome with a 73.1% GC content. (Table 3 and Figure 3). Of the 4,278 genes predicted, 4,225 were protein coding genes, and 53 RNAs. Twenty eight pseudogenes were also identified. The majority of the genes

(74.3%) 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.

Table 3. Genome Statistics

Attribute	Value	% of Total
Genome size (bp)	4,669,183	
DNA Coding region (bp)	4,347,731	93.12%
DNA G+C content (bp)	3,413,947	73.12%
Number of replicons	1	
Extrachromosomal elements	0	
Total genes	4278	100.00%
RNA genes	53	1.24%
rRNA operons	2	
Protein-coding genes	4225	98.76%
Pseudo genes	28	0.65%
Genes with function prediction	3183	74.40%
Genes in paralog clusters	689	16.11%
Genes assigned to COGs	3109	72.67%
Genes assigned Pfam domains	3246	75.88%
Genes with signal peptides	1034	24.17%
Genes with transmembrane helices	1135	26.53%
CRISPR repeats	1	

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

Code	Value	%	Description
J	169	4	Translation, ribosomal structure and biogenesis
A	4	0.1	RNA processing and modification
K	384	9.1	Transcription
L	122	2.9	Replication, recombination and repair
B	1	0	Chromatin structure and dynamics
D	25	0.6	Cell cycle control, mitosis and meiosis
Y	0	0	Nuclear structure
V	95	2.3	Defense mechanisms
T	138	3.3	Signal transduction mechanisms
M	166	3.9	Cell wall/membrane biogenesis
N	1	0	Cell motility
Z	0	0	Cytoskeleton
W	0	0	Extracellular structures
U	27	0.6	Intracellular trafficking and secretion
O	89	2.1	Posttranslational modification, protein turnover, chaperones

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

Code	Value	%	Description
J	169	4	Translation, ribosomal structure and biogenesis
G	546	12.9	Carbohydrate transport and metabolism
E	264	6.3	Amino acid transport and metabolism
F	92	2.2	Nucleotide transport and metabolism
H	129	3.1	Coenzyme transport and metabolism
I	101	2.4	Lipid transport and metabolism
P	183	4.3	Inorganic ion transport and metabolism
Q	62	1.5	Secondary metabolites biosynthesis, transport and catabolism
R	433	10.3	General function prediction only
S	249	5.9	Function unknown
-	1116	26.4	Not in COGs

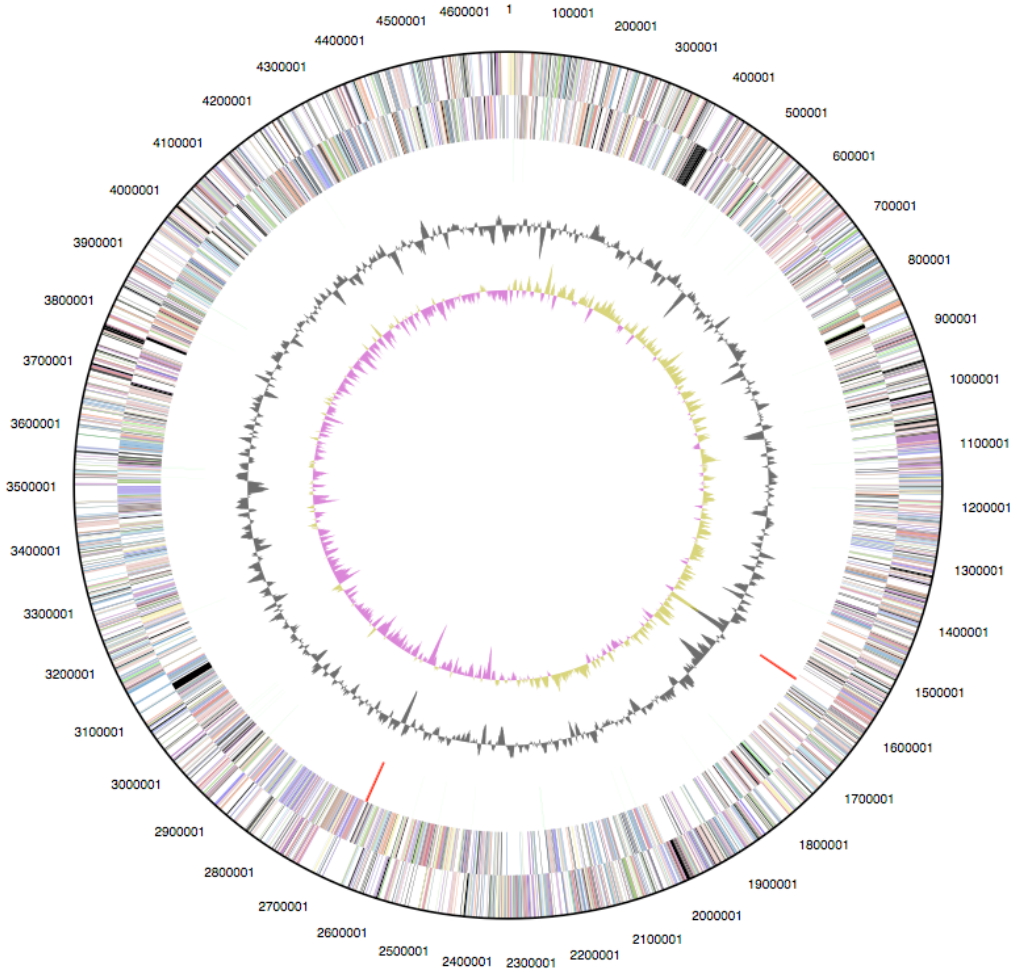

Figure 3. Graphical circular map of the genome. From outside to the center: Genes on forward strand (color by COG categories), Genes on reverse strand (color by COG categories), RNA genes (tRNAs green, rRNAs red, other RNAs black), GC content, GC skew

Acknowledgements

We would like to gratefully acknowledge the help of Katja Steenblock for growing *B. cavernae* 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, as well as German Research Foundation (DFG) INST 599/1-1.

References

1. Groth I, Schumann P, Schütze B, Augsten K, Kramer I, Stackebrandt E. *Beutenbergia cavernae*, gen. nov., sp. nov., a L-lysine-containing actinomycete isolated from a cave. *Int J Syst Bacteriol* 1999; **49**:1733-40. PMID:10555355 doi:10.1099/00207713-49-4-1733
2. Zhi X-Y, Li W-J, Stackebrandt E. An update of the structure and 16S rRNA gene sequence-based definition of higher ranks of the class *Actinobacteria*, with the proposal of two new suborders and four new families and embedded descriptions of the existing higher taxa. *Int J Syst Evol Micobiol* 2009; **59**:589-608. PMID:19244447 doi:10.1099/ijs.0.65780-0
3. von Wintzingerode F, Göbel UB, Siddiqui RA, Rösick U, Schumann P, Frühling A, Rohde M, Pukall R, Stackebrandt E. *Salana multivorans* gen. nov., sp. nov., a novel actinobacterium isolated from an anaerobic bioreactor and capable of selenate reduction. *Int J Syst Evol Bacteriol* 2001; **51**:1653-61. PMID:11594592
4. Schumann, P, Kampfer, P, Busse, HJ, Evtushenko, LI Proposed minimal standards for describing new genera and species of the suborder *Micrococcineae*. *Int J Syst Evol Microbiol*. 2009 published ahead of print PMID: 19542112 doi:10.1099/ijs.0.12971-0
5. Lee C, Grasso C, Sharlow MF. Multiple sequence alignment using partial order graphs. *Bioinformatics* 2002; **18**:452-64. PMID: 11934745 doi:10.1093/bioinformatics/18.3.452
6. Castresana J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol* 2000; **17**:540-52. PMID:10742046
7. Stamatakis A, Hoover P, Rougemont J. A rapid bootstrap algorithm for the RAxML web-servers. *Syst Biol* 2008; **57**:758-71. PMID:18853362 doi:10.1080/10635150802429642
8. 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. PMID:17981842 doi:10.1093/nar/gkm884
9. Field D, Garrity G, Gray T, Morrison N, Selengut J, Sterk P, Tatusova T, Thomson N, Allen MJ, Angiuoli SV. Towards a richer description of our complete collection of genomes and metagenomes: the "Minimum Information about a Genome Sequence" (MIGS) specification. *Nature Biotechnology* 2008; **26**:541-7. PMID:18464787 doi:10.1038/nbt1360
10. Stackebrandt E, Rainey FA, Ward-Rainey NL. Proposal for a new hierarchic classification system, *Actinobacteria* classis nov. Int J Syst Bacteriol 1997; **47**:479-91. doi:10.1099/00207713-47-2-479
11. Anonymous. Biological Agents: Technical rules for biological agents. <http://www.baua.de>.
12. 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. The Gene Ontology Consortium. *Nat Genet* 2000; **25**:25-29. PMID:10802651 doi:10.1038/75556
13. Yamada K, Komagata K. Taxonomic studies on coryneform bacteria IV. Morphological, cultural, biochemical, and physiological characteristics. *J Gen Appl Microbiol* 1997; **18**:399-416.
14. Sims D, Brettin T, Detter JC, Han C, Lapidus A, et al. Complete genome of Kytococcus sedentarius type strain (strain 541T). *SIGS*, 2009 reviewed. doi: 10.4056/sigs.761
15. Anonymous. Prodigal Prokaryotic Dynamic Programming Genefinding Algorithm. Oak Ridge National Laboratory and University of Tennessee 2009 <compbio.ornl.gov/prodigal/>
16. Pati. GenePRIMP: A Gene Prediction Improvement Pipeline for microbial genomes. (Submitted).
17. Markowitz V, Mavromatis K, Ivanova N, Chen I-M, Chu K, Palaniappan K, Szeto E, Anderson I, Lykidis A, Kyrpides N. Expert Review of Functional Annotations for Microbial Genomes. (Submited 2009).