Complete genome sequence of *Arcanobacterium haemolyticum* type strain (11018^T)

Montri Yasawong[1], Hazuki Teshima[2,3], Alla Lapidus[2], Matt Nolan[2], Susan Lucas[2], Tijana Glavina Del Rio[2], Hope Tice[2], Jan-Fang Cheng[2], David Bruce[2,3], Chris Detter[2,3], Roxanne Tapia[2,3], Cliff Han[2,3], Lynne Goodwin[2,3], Sam Pitluck[2], Konstantinos Liolios[2], Natalia Ivanova[2], Konstantinos Mavromatis[2], Natalia Mikhailova[2], Amrita Pati[2], Amy Chen[4], Krishna Palaniappan[4], Miriam Land[2,5], Loren Hauser[2,5], Yun-Juan Chang[2,5], Cynthia D. Jeffries[2,5], Manfred Rohde[1], Johannes Sikorski[6], Rüdiger Pukall[6], Markus Göker[6], Tanja Woyke[2], James Bristow[2], Jonathan A. Eisen[2,7], Victor Markowitz[4], Philip Hugenholtz[2], Nikos C. Kyrpides[2], and Hans-Peter Klenk[6]*

[1] HZI – Helmholtz Centre for Infection Research, 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] Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA
[6] DSMZ - German Collection of Microorganisms and Cell Cultures GmbH, Braunschweig, Germany
[7] University of California Davis Genome Center, Davis, California, USA

*Corresponding author: Hans-Peter Klenk

Arcanobacterium haemolyticum (ex MacLean *et al.* 1946) Collins *et al.* 1983 is the type species of the genus *Arcanobacterium*, which belongs to the family *Actinomycetaceae*. The strain is of interest because it is an obligate parasite of the pharynx of humans and farm animal; occasionally, it causes pharyngeal or skin lesions. It is a Gram-positive, nonmotile and non-sporulating bacterium. The strain described in this study was isolated from infections amongst American soldiers of certain islands of the North and West Pacific. This is the first completed sequence of a member of the genus *Arcanobacterium* and the ninth type strain genome from the family *Actinomycetaceae*. The 1,986,154 bp long genome with its 1,821 protein-coding and 64 RNA genes is a part of the **G**enomic **E**ncyclopedia of **B**acteria and **A**rchaea project.

Introduction

Strain 11018^T (= DSM 20595 = CCM 5947 = ATCC 9345 = NBRC 15585) is the type strain of the species *A. haemolyticum*, which is the type species of its genus *Arcanobacterium* [1]. *Arcanobacterium* is one of six genera in the family *Actinomycetaceae* [2-4]. The genus currently consists of nine validly described species. The strain was first described in 1946 by MacLean as '*Corynebacterium haemolyticum*' [5]. Based on chemical features and the presence of unique phenotypic characteristics, the strain was subsequently transferred to the new genus *Arcanobacterium* as *A. haemolyticum* [1]

and emended by Lehnen *et al.* in 2006 [6]. The generic name drives from the Latin word '*arcanus*', meaning 'secretive' and the Latin word '*bacterium*', a small rod, meaning 'secretive bacterium' [1]. The species epithet is derived from the Latin word '*haema*' meaning 'blood' and the Neo-Latin word '*lyticus*' meaning 'able to loose or able to dissolve' referring to blood-dissolving or hemolytic when the cells grow on blood agar [1]. There are many medical case reports that *A. haemolyticum* is occasionally isolated in patients with brain abscess [7-9], cellulitis [10,11], endocarditis [12],

meningitis [13], peritonitis [14], post-traumatic ankle joint infection [15], septic arthritis [16], septicemia [17], sinusitis [11], soft tissue infections [18], venous ulcer infection [19], vertebral osteomyelitis [20] and wound infection [21,22]. Only rarely are cases reported in animals, where pathogenicity of *A. haemolyticum* has not been well documented [23-25]. Here we present a summary classification and a set of features for *A. haemolyticum* strain 11018[T], together with the description of the complete genomic sequencing and annotation.

Classification and features

Strain 11018[T] is an obligate parasite of the pharynx of human and farm animals; occasionally it causes pharyngeal or skin lesions [26]. The strain was isolated from infections in American soldiers [5]. The 16S rRNA gene sequence of strain 11018[T] (AJ234059) is 99% identical to six culturable strains that were reported in GenBank (status July 2010). Five strains were isolated from infected horses [23]. Another culturable strain, Tr2-2X-1 (FJ477385), was isolated from gasoline contaminated soil. The 16S rRNA gene of strain 11018[T] shares 93.3-97.9% sequence identity with the sequences of the type strains from the other members of the genus *Arcanobacterium* [27]. The next closest relative outside of the genus *Arcanobacterium* is *Dermacoccus barathri* MT2.1[T] (92.3% se-

quence similarity) [27]. No phylotypes from environmental screening or metagenomic surveys could be linked to *A. haemolyticum* or even the genus *Arcanobacterium*, indicating a rare occurrence of these species in the habitats screened thus far (as of July 2010). A representative genomic 16S rRNA sequence of *A. haemolyticum* 11018[T] was compared using BLAST with the most resent release of the Greengenes database [28] and the relative frequencies of taxa and keywords, weighted by BLAST scores, were determined. The five most frequent genera were *Arcanobacterium* (42.4%), *Dermacoccus* (12.6%), *Actinomyces* (10.8%), *Terrabacter* (9.9%) and *Sanguibacter* (5.7%). The five most frequent keywords within the labels of environmental samples were 'skin' (6.6%), 'human' (5.0%), 'feedlot' (4.6%), 'elbow' (3.4%) and 'microbiota' (3.3%). The BLAST keywords analysis supports the biological insights into *A. haemolyticum* strain 11018[T] as described above.

Figure 1 shows the phylogenetic neighborhood of *A. haemolyticum* strain 11018[T] in a 16S rRNA based tree. The sequences of the four 16S rRNA gene copies in the genome differ from each other by up to two nucleotides, and differ by up to five nucleotides from the previously published sequence generated from CIP 103370 (AJ234059) which contains one ambiguous base call.
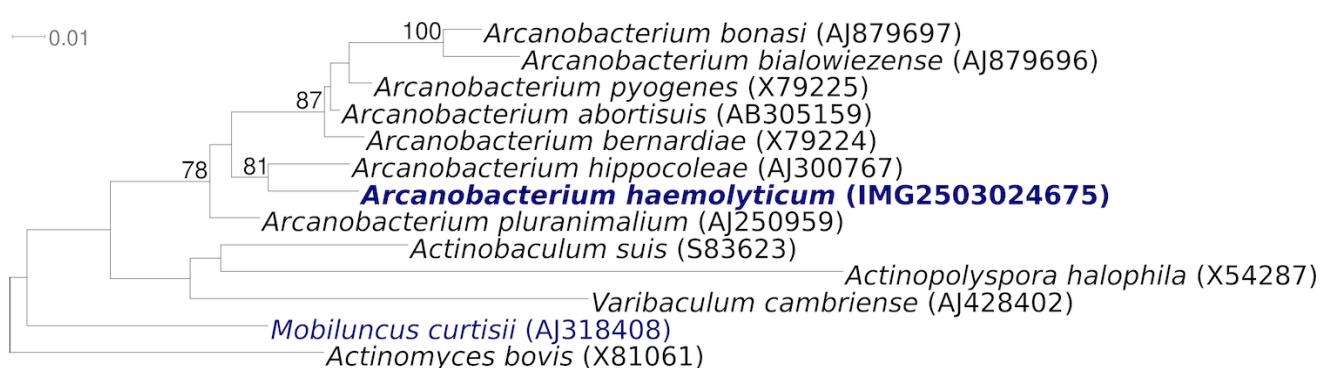

Figure 1. Phylogenetic tree highlighting the position of *A. haemolyticum* strain 11018[T] relative to the type strains of the other species within the genus *Arcanobacterium* and to the type strains of the other genera within the family *Actinomycetaceae*. The trees were inferred from 1,388 aligned characters [29,30] of the 16S rRNA gene sequence under the maximum likelihood criterion [31] and rooted in accordance with the current taxonomy. 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 [32] if larger than 60%. Lineages with type strain genome sequencing projects registered in GOLD [33] are shown in blue, published genomes in bold.

The cells of strain 11018[T] are slender or irregular rods (0.3-0.8 × 1.0-5.0 μm) [Table 1 and Figure 2]. The cells are Gram-positive, nonmotile, not acid-

fast and without endospores [1]. In young cultures, cells may show clubbed ends sometimes arranged in V formation, but there are no fila-

ments. In older cultures, cells segment into short, irregular rods and cocci [1]. Strain 11018[T] is facultatively anaerobic. The cells grow slowly on nutrient agar, but grow better on horse blood agar, giving small, convex, translucent colonies surrounded by a zone of complete hemolysis after two days at 37°C [1]. The selective medium for this strain was developed by Coman [39] and contains 5% sheep blood and 3.5% of NaCl. Cell growth is enhanced by the addition of CO_2 [1]. The optimum growth temperature is 37°C [1,26]. Cells do not withstand heating at 60°C for 15 min [1,5]. Strain 11018[T] is chemoorganotrophic and requires nutritionally rich media for growth [1,26]. The fermentative metabolism of this strain produces acid but does not produce gas from glucose and several other carbohydrates on which growth occurs [1,26]. Acid production is mainly acetic, lactic and succinic acids [1,26]. Catalase, nitrate reduction and gelatine hydrolysis reactions are negative [6]. Strain 11018[T] produces *N*-acetyl-*β*-galactosidase, alkaline phosphatase, extracellular DNase, *β*-galactosidase, *α*-glucosidase and pyrazinamidase. It does not produce acid phosphatase, *α*-chymotrypsin, cystine arylamidase, esterase (C4), esterase lipase (C8), *α*-fucosidase, *α*-

galactosidase, *β*-glucosidase, *β*-glucuronidase, leucine arylamidase, lipase (C14), *α*-mannosidase, naphthol-AS-BI-phosphohydrolase, trypsin, valine arylamidase and urease [1,6]. Strain 11018[T] is not able to ferment adonitol, L-arabitol, erythritol, D-fructose, glycerol, glycogen, D-mannitol and D-xylose. It is resistant to oxytetracycline (30μg per disc) but susceptible to nalidixic acid (30μg per disc), sulfamethoxazole trimethoprim (25μg per disc), amikacin (10μg per disc) or cefoxitin (30μg per disc) [1,42].

Chemotaxonomy

Strain 11018[T] possesses peptidoglycan type A5α based on L-Lys-L-Lys-D-Glu (unpublished, Norbert Weiss [43]). The predominant menaquinone is MK-9(H_4) (85%) complemented by 15% MK-8(H_4) [6]. The major cellular fatty acids when grown on blood agar at 35°C are straight-chain unsaturated acids $C_{18:1}$ $\omega 9c$ (37.0%), and saturated acids $C_{18:0}$ (24.7%), $C_{16:0}$ (22.5%) [6], which is similar to the cellular fatty acids spectrum reported from cells grown on sheep blood agar [31]: $C_{18:1}$ *cis*9 (29%), $C_{16:0}$ (23%), $C_{18:2}$ (18%), $C_{18:0}$ (17%), $C_{10:0}$ (3%) and $C_{14:0}$ (2%).
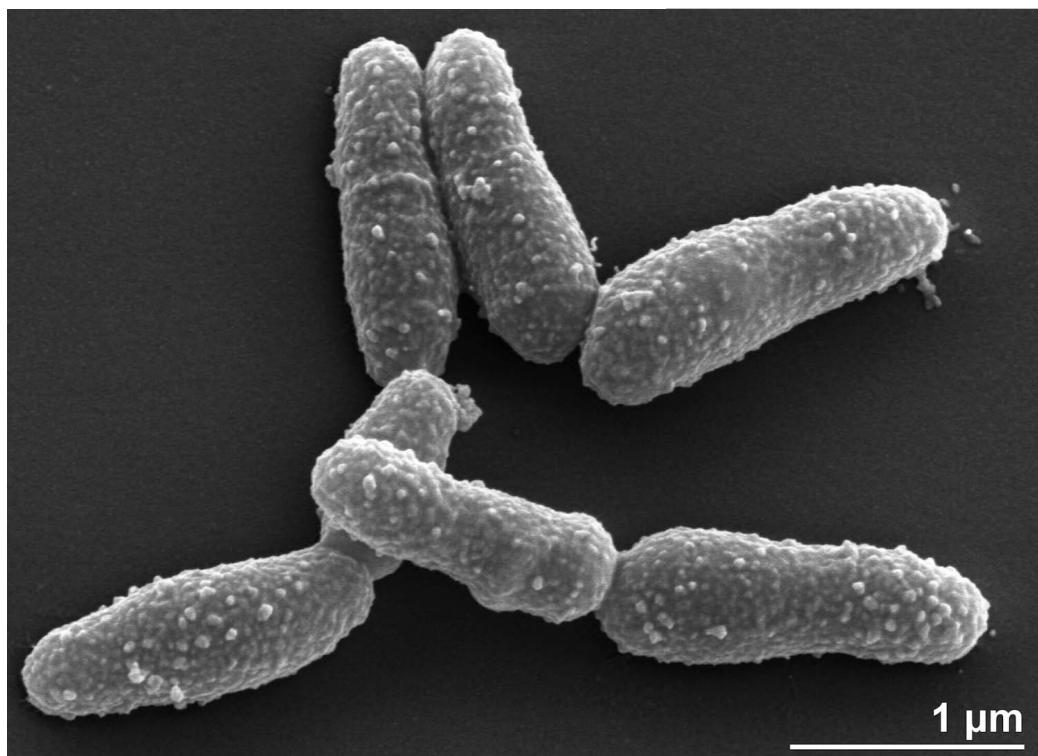

Figure 2. Scanning electron micrograph of *A. haemolyticum* strain 11018[T]

Table 1. Classification and general features of *A. haemolyticum* strain 11018[T] according to the MIGS recommendations [34].

MIGS ID	Property	Term	Evidence code
		Domain *Bacteria*	TAS [35]
		Phylum *Actinobacteria*	TAS [36]
		Class *Actinobacteria*	TAS [3]
		Subclass *Actinobacteridae*	TAS [3,4]
	Current classification	Order *Actinomycetales*	TAS [2-5,37]
		Suborder *Actinomycineae*	TAS [3,4]
		Family *Actinomycetaceae*	TAS [2-5,37]
		Genus *Arcanobacterium*	TAS [1,6,38]
		Species *Arcanobacterium haemolyticum*	TAS [1,5,38]
		Type strain 11018	TAS [1]
	Gram stain	positive	TAS [1]
	Cell shape	slender, irregular rods (0.3-0.8 ×1.0-5.0 µm)	TAS [1]
	Motility	none	TAS [1]
	Sporulation	none	TAS [1]
	Temperature range	not reported	
	Optimum temperature	37°C	TAS [1]
	Salinity	3.5%	TAS [39]
MIGS-22	Oxygen requirement	facultatively anaerobic	TAS [1]
	Carbon source	carbohydrates	TAS [1,5,6]
	Energy source	chemoorganotroph	TAS [26]
MIGS-6	Habitat	pharynx of humans and farm animals	TAS [26]
MIGS-15	Biotic relationship	obligate parasite	TAS [26]
MIGS-14	Pathogenicity	pharyngeal or skin lesions	TAS [26]
	Biosafety level	2	TAS [40]
	Isolation	infections amongst American soldiers	TAS [5]
MIGS-4	Geographic location	North and West Pacific	TAS [5]
MIGS-5	Sample collection time	1946 or before	TAS [1,5]
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 [41]. If the evidence code is IDA, then the property was directly observed 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 [44], and is part of the *Genomic Encyclopedia of Bacteria and Archaea* project [45]. The genome project is deposited in the Genome OnLine Database [33] 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: 454 pyrosequence standard, PE (12.5 kb insert size) libraries and one Illumina standard library
MIGS-29	Sequencing platforms	454 GS FLX Titanium, Illumina GAii
MIGS-31.2	Sequencing coverage	83.8 × pyrosequence, 36.8 x Illumina
MIGS-30	Assemblers	Newbler version 2.0.0-PostRelease-11/04/2008, phrap, Velvet
MIGS-32	Gene calling method	Prodigal 1.4, GenePRIMP
	INSDC ID	CP002045
	Genbank Date of Release	June 4, 2010
	GOLD ID	Gc01291
	NCBI project ID	37925
	Database: IMG-GEBA	646564505
MIGS-13	Source material identifier	DSM 20595
	Project relevance	Tree of Life, GEBA

Growth conditions and DNA isolation

A. haemolyticum strain 11018T, DSM 20595, was grown anaerobically in DSMZ medium 104 (PYG modified medium) [46] at 37°C. DNA was isolated from 1-1.5 g of cell paste using MasterPure Gram Positive DNA Purification Kit (Epicentre MGP04100), with a modified protocol for cell lysis, st/LALM, as described in Wu *et al.* [45].

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. Pyrosequencing reads were assembled using the Newbler assembler version 2.0.0-PostRelease-11/04/2008 (Roche). The initial Newbler assembly consisted of 116 contigs in 28 scaffolds and was converted into a phrap assembly by making fake reads from the consensus, collecting the read pairs in the 454 paired end library. Illumina GAii sequencing data was assembled with Velvet [47] and the consensus sequences were shredded into 1.5 kb overlapped fake reads and assembled together with the 454 data. Draft assemblies were based on 166.4 Mb 454 draft 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 was used for sequence assembly and quality assessment in the following finishing process. After the shotgun stage, reads were assembled with parallel phrap (High Performance Software, LLC). Possible mis-assemblies were corrected with gapResolu-tion, Dupfinisher [48], or sequencing cloned bridging PCR fragments with subcloning or transposon bombing (Epicentre Biotechnologies, Madison, WI) [49]. Gaps between contigs were closed by editing in Consed, by PCR and by Bubble PCR primer walks (J.-F. Chang, unpublished). A total of 140 additional reactions were necessary to close gaps and to raise the quality of the finished sequence. Illumina reads were also used to improve the final consensus quality using an in-house developed tool - the Polisher [50]. 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 120.6 ×coverage of the genome. The final assembly contains 2.03 million Illumina reads and 0.52 million pyrosequencing reads.

Genome annotation

Genes were identified using Prodigal [51] as part of the Oak Ridge National Laboratory genome annotation pipeline, followed by a round of manual curation using the JGI GenePRIMP pipeline [52]. 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 - Expert Review (IMG-ER) platform [53].

Table 3. Genome Statistics

Attribute	Value	% of Total
Genome size (bp)	1,986,154	100.00%
DNA coding region (bp)	1,744,192	87.82%
DNA G+C content (bp)	1,055,308	53.13%
Number of replicons	1	
Extrachromosomal elements	0	
Total genes	1,885	100.00%
RNA genes	64	3.40%
rRNA operons	4	
Protein-coding genes	1,821	96.60%
Pseudo genes	90	4.77%
Genes with function prediction	1,292	68.54%
Genes in paralog clusters	154	8.17%
Genes assigned to COGs	1,308	69.39%
Genes assigned Pfam domains	1,402	74.38%
Genes with signal peptides	391	20.74%
Genes with transmembrane helices	492	26.10%
CRISPR repeats	1	
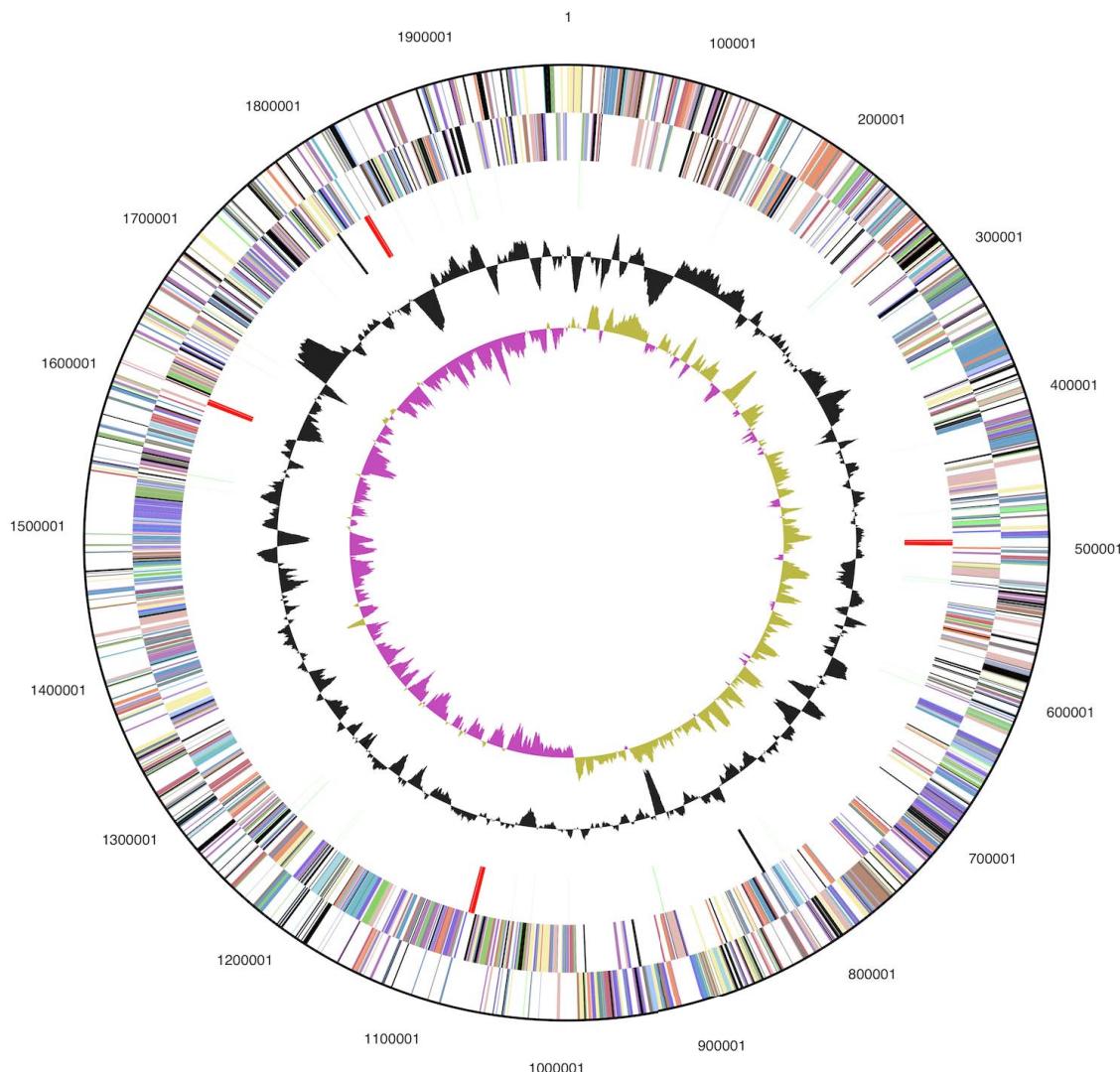
Figure 3. Graphical circular map of the genome. From outside to the center: Genes on forward strand (color by COG categories), Genes on reverse strand (color by COG categories), RNA genes (tRNAs green, rRNAs red, other RNAs black), GC content, GC skew.

Genome properties

The genome consists of a 1,986,154 bp long chromosome with a 53.1% GC content (Table 3 and Figure 3). Of the 1,885 genes predicted, 1,821 were protein-coding genes, and 64 RNAs; 90 pseudogenes were also identified. The majority of the protein-coding genes (68.5%) 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 4. Number of genes associated with the general COG functional categories

Code	Value	%age	Description
J	136	9.7	Translation, ribosomal structure and biogenesis
A	1	0.1	RNA processing and modification
K	99	7.1	Transcription
L	119	8.5	Replication, recombination and repair
B	0	0.0	Chromatin structure and dynamics
D	21	1.5	Cell cycle control, cell division, chromosome partitioning
Y	0	0.0	Nuclear structure
V	36	2.6	Defense mechanisms
T	51	3.6	Signal transduction mechanisms
M	75	5.4	Cell wall/membrane/envelope biogenesis
N	0	0.0	Cell motility
Z	0	0.0	Cytoskeleton
W	0	0.0	Extracellular structures
U	27	1.9	Intracellular trafficking and secretion, and vesicular transport
O	56	4.0	Posttranslational modification, protein turnover, chaperones
C	86	6.1	Energy production and conversion
G	125	8.9	Carbohydrate transport and metabolism
E	77	5.5	Amino acid transport and metabolism
F	58	4.1	Nucleotide transport and metabolism
H	56	4.0	Coenzyme transport and metabolism
I	34	2.4	Lipid transport and metabolism
P	93	6.6	Inorganic ion transport and metabolism
Q	12	0.9	Secondary metabolites biosynthesis, transport and catabolism
R	152	10.9	General function prediction only
S	87	6.2	Function unknown
-	577	30.6	Not in COGs

Acknowledgements

References

1. Collins MD, Jones D, Schofield GM. Reclassification of '*Corynebacterium haemolyticum*' (MacLean, Liebow & Rosenberg) in the genus *Arcanobacterium* gen.nov. as *Arcanobacterium haemolyticum* nom.rev., comb.nov. *J Gen Microbiol* 1982; **128**:1279-1281. PubMed

2. Buchanan RE. Studies in the Nomenclature and Classification of the *Bacteria*: VIII. The Subgroups and Genera of the *Actinomycetales*. *J Bacteriol* 1918; **3**:403-406. PubMed

3. Stackebrandt E, Rainey FA, Ward-Rainey NL. Proposal for a new hierarchic classification system, *Actinobacteria* classis nov. *Int J Syst Bacteriol* 1997; **47**:479-491. doi:10.1099/00207713-47-2-479

4. Zhi XY, Li WJ, 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 emended descriptions of the existing higher taxa. *Int J Syst Evol Microbiol* 2009; **59**:589-608. PubMed doi:10.1099/ijs.0.65780-0

5. MacLean PD, Liebow AA, Rosenberg AA. A hemolytic *Corynebacterium* resembling *Corynebacterium ovis* and *Corynebacterium pyrogenes* in man. *J Infect Dis* 1946; **79**:69-90.

6. Lehnen A, Busse H-J, Frölich K, Krasinska M, Kämpfer P, Speck S. *Arcanobacterium bialowiezense* sp. nov. and *Arcanobacterium bonasi* sp. nov., isolated from the prepuce of European bison bulls (Bison bonasus) suffering from balanoposthitis, and emended description of the genus *Arcanobacterium* Collins et al. 1983. *Int J Syst Evol Microbiol* 2006; **56**:861-866. PubMed doi:10.1099/ijs.0.63923-0

7. Altmann G, Bogokovsky B. Brain abscess due to *Corynebacterium haemolyticum. Lancet* 1973; **301**:378-379. doi:10.1016/S0140-6736(73)90177-3

8. Vargas J, Hernandez M, Silvestri C, Jimenez O, Guevara N, Carballo M, Rojas N, Riera J, Alayo E, Fernandez M, *et al*. Brain abscess due to *Arcanobacterium haemolyticum* after dental extraction. *Clin Infect Dis* 2006; **42**:1810-1811. PubMed doi:10.1086/504436

9. Washington JA, Martin WJ, Spiekerman RE. Brain abscess with *Corynebacterium hemolyticum*: report of a case. *Am J Clin Pathol* 1971; **56**:212-215. PubMed

10. Dobinsky S, Noesselt T, Rucker A, Maerker J, Mack D. Three cases of *Arcanobacterium haemolyticum* associated with abscess formation and cellulitis. *Eur J Clin Microbiol Infect Dis* 1999; **18**:804-806. PubMed doi:10.1007/s100960050404

11. Limjoco-Antonio AD, Janda WM, Schreckenberger PC. *Arcanobacterium haemolyticum* sinusitis and orbital cellulitis. *Pediatr Infect Dis J* 2003; **22**:465-467. PubMed doi:10.1097/00006454-200305000-00018

12. Worthington MG, Daly BD, Smith FE. *Corynebacterium hemolyticum* endocarditis on a native valve. *South Med J* 1985; **78**:1261-1262. PubMed

13. Minárik T, Sufliarsky J, Trupl J, Krcmery V, Jr. *Arcanobacterium haemolyticum* invasive infections, including meningitis in cancer patients. *BMC. J Infect* 1997; **34**:91. PubMed doi:10.1016/S0163-4453(97)80023-0

14. Farmer AD, Bruckner Holt CE, Le Roux G, Butterworth JR. Spontaneous bacterial peritonitis due to *Arcanobacterium haemolyticum. BMC. J Infect* 2007; **54**:516. PubMed doi:10.1016/j.jinf.2006.09.013

15. Hoosen AA, Rasool MN, Roux L. Posttraumatic ankle joint infection with *Arcanobacterium haemolyticum*: a case report. *J Infect Dis* 1990; **162**:780-781. PubMed

16. Goyal R, Singh NP, Mathur M. Septic arthritis due to *Arcanobacterium haemolyticum. Indian J Med Microbiol* 2005; **23**:63-65. PubMed doi:10.4103/0255-0857.13879

17. Ben-Yaacob D, Waron M, Boldur I, Gil I, Sompolinsky D. Septicemia due to *Corynebacterium haemolyticum. Isr J Med Sci* 1984; **20**:431-433. PubMed

18. Tan TY, Ng SY, Thomas H, Chan BK. *Arcanobacterium haemolyticum* bacteraemia and soft-tissue infections: case report and review of the literature. *J Infect* 2006; **53**:e69-e74. PubMed doi:10.1016/j.jinf.2005.10.008

19. Pânzaru C, Taranu T. Venous ulcer infection caused by *Arcanobacterium haemolyticum. Roum Arch Microbiol Immunol* 2001; **60**:323-327. PubMed

20. Ceilley RI. Foot ulceration and vertebral osteomyelitis with *Corynebacterium haemolyticum. Arch Dermatol* 1977; **113**:646-647. PubMed doi:10.1001/archderm.113.5.646

21. Barker KF, Renton NE, Lee PY, James DH. *Arcanobacterium haemolyticum* wound infection. *J Infect* 1992; **24**:214-215. PubMed doi:10.1016/0163-4453(92)93110-C

22. Ritter E, Kaschner A, Becker C, Becker-Boost E, Wirsing von Konig CH, Finger H. Isolation of *Arcanobacterium haemolyticum* from an infected foot wound. *Eur J Clin Microbiol Infect Dis* 1993; **12**:473-474. PubMed doi:10.1007/BF01967447

23. Hassan AA, Ulbegi-Mohyla H, Kanbar T, Alber J, Lammler C, Abdulmawjood A, Zschock M, Weiss R. Phenotypic and genotypic characterization of *Arcanobacterium haemolyticum* isolates from infections of horses. *J Clin Microbiol* 2009; **47**:124-128. PubMed doi:10.1128/JCM.01933-08

24. Richardson A, Smith PJ. Herd fertility and *Corynebacterium haemolyticum* in bovine semen. *Vet Rec* 1968; **83**:156-157. PubMed

25. Roberts RJ. Isolation of *Corynebacterium haemolyticum* from a case of ovine pneumonia. *Vet Rec* 1969; **84**:490. PubMed

26. Holt JG, Krieg NR, Sneath PHA, Staley JT, Williams ST. 1994. Bergey's manual of determinative bacteriology, 9th ed. Williams & Wilkins, Baltimore.

27. Chun J, Lee JH, Jung Y, Kim M, Kim S, Kim BK, Lim YW. EzTaxon: a web-based tool for the identification of prokaryotes based on 16S ribosomal RNA gene sequences. *Int J Syst Evol Microbiol* 2007; **57**:2259-2261. PubMed doi:10.1099/ijs.0.64915-0

28. 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

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

30. 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

31. 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

32. 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

33. 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

34. 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

35. 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

36. 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.

37. 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

38. Validation List no. 10. Validation of the publication of new names and new combinations previously effectively published outside the IJSB. *Int J Syst Bacteriol* 1983; **33**:438-440. doi:10.1099/00207713-33-2-438

39. Coman G, Panzaru C, Dahorea C. The isolation of *Arcanobacterium haemolyticum* from the pharyngeal exudate of children. *Bacteriol Virusol Parazitol Epidemiol* 1996; **41**:141-144. PubMed

40. Classification of bacteria and archaea in risk groups. TRBA 466.

41. 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

42. Schofield GM, Schaal KP. A numerical taxonomic study of members of the *Actinomycetaceae* and

related taxa. *J Gen Microbiol* 1981; **127**:237-259. PubMed

43. DSMZ. 2001. Catalogue of Strains, 7th ed. German Collection of Microorganisms and Cell Cultures, Braunschweig.

44. Klenk H-P, 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

45. 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

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

47. 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

48. Cliff S. Han, Patrick Chain. 2006. Finishing repeat regions automatically with Dupfinisher. *In:* Proceeding of the 2006 international conference on bioinformatics & computational biology. Hamid R Arabnia & Homayoun Valafar (eds), CSREA Press. June 26-29, 2006: 141-146.

49. Sims D, Brettin T, Detter JC, Han C, Lapidus A, Copeland A, Glavina Del Rio T, Nolan M, Chen F, Lucas S, *et al*. Complete genome sequence of *Kytococcus sedentarius* type strain (541^T). *Stand Genomic Sci* 2009; **1**:12-20. doi:10.4056/sigs.761

50. 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.

51. 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

52. 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

53. 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