

## Complete genome sequence of *Atopobium parvulum* type strain (IPP 1246<sup>T</sup>)

Alex Copeland<sup>1</sup>, Johannes Sikorski<sup>2</sup>, Alla Lapidus<sup>1</sup>, Matt Nolan<sup>1</sup>, Tijana Glavina Del Rio<sup>1</sup>, Susan Lucas<sup>1</sup>, Feng Chen<sup>1</sup>, Hope Tice<sup>1</sup>, Sam Pitluck<sup>1</sup>, Jan-Fang Cheng<sup>1</sup>, Rüdiger Pukall<sup>2</sup>, Olga Chertkov<sup>1,3</sup>, Thomas Brettin<sup>1,3</sup>, Cliff Han<sup>1,3</sup>, John C. Detter<sup>1,3</sup>, Cheryl Kuske<sup>1,3</sup>, David Bruce<sup>1,3</sup>, Lynne Goodwin<sup>1,3</sup>, Natalia Ivanova<sup>1</sup>, Konstantinos Mavromatis<sup>1</sup>, Natalia Mikhailova<sup>1</sup>, Amy Chen<sup>4</sup>, Krishna Palaniappan<sup>4</sup>, Patrick Chain<sup>1,5</sup>, Manfred Rohde<sup>6</sup>, Markus Göker<sup>2</sup>, Jim Bristow<sup>1</sup>, Jonathan A. Eisen<sup>1,7</sup>, Victor Markowitz<sup>4</sup>, Philip Hugenholtz<sup>1</sup>, Nikos C. Kyrpides<sup>1</sup>, Hans-Peter Klenk<sup>2\*</sup>, and John C. Detter<sup>1,3</sup>

<sup>1</sup> DOE Joint Genome Institute, Walnut Creek, California, USA

<sup>2</sup> DSMZ - German Collection of Microorganisms and Cell Cultures GmbH, Braunschweig, Germany

<sup>3</sup> Los Alamos National Laboratory, Bioscience Division, Los Alamos, New Mexico, USA

<sup>4</sup> Biological Data Management and Technology Center, Lawrence Berkeley National Laboratory, Berkeley, California, USA

<sup>5</sup> Lawrence Livermore National Laboratory, Livermore, California, USA

<sup>6</sup> HZI - Helmholtz Centre for Infection Research, Braunschweig, Germany

<sup>7</sup> University of California Davis Genome Center, Davis, California, USA

\*Corresponding author: Hans-Peter Klenk

**Keywords:** halitosis, obligately anaerobic, human respiratory tract, risk group 2, malodor, *Coriobacteriaceae*

---

*Atopobium parvulum* (Weinberg *et al.* 1937) Collins and Wallbanks 1993 comb. nov. is the type strain of the species and belongs to the genomically yet unstudied *Atopobium/Olsenella* branch of the family *Coriobacteriaceae*. The species *A. parvulum* is of interest because its members are frequently isolated from the human oral cavity and are found to be associated with halitosis (oral malodor) but not with periodontitis. Here we describe the features of this organism, together with the complete genome sequence, and annotation. This is the first complete genome sequence of the genus *Atopobium*, and the 1,543,805 bp long single replicon genome with its 1369 protein-coding and 49 RNA genes is part of the *Genomic Encyclopedia of Bacteria and Archaea* project.

---

### Introduction

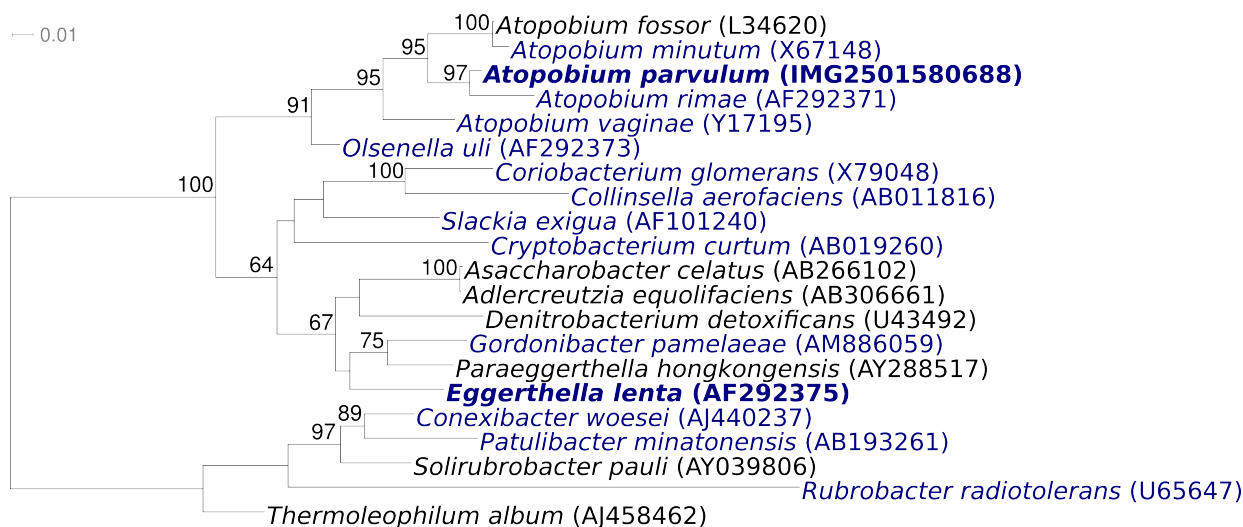
Strain IPP 1246<sup>T</sup> (= DSM 20469 = ATCC 33793 = JCM 10300) is the type strain of the species *Atopobium parvulum* and was first described by Weinberg *et al.* 1937 as '*Streptococcus parvulus*' (basonym) [1]. In 1992 it was reclassified as *A. parvulum* [2]. *A. parvulum* is of high interest because it has frequently been isolated from the human oral cavity, especially from the tongue dorsum, where it has been associated with patients suffering from halitosis (oral malodor) [3,4]. In general, the malodorous compounds are volatile sulfur compounds, with the most frequent ones being hydrogen sulfide, methyl mercaptan, and dimethyl sulfide, which are produced by bacterial metabolism of

the sulfur containing amino acids cysteine and methionine [3,4]. However, for *A. parvulum* itself, the production of these substances has not yet been studied. *A. parvulum* has not been found to be significantly associated with chronic periodontitis, though a participation in periodontitis can not be fully excluded [5]. Nevertheless, *A. parvulum* has been associated with odontogenic infections, *e.g.* dental implants, but also with the saliva of healthy subjects [6]. Here we present a summary classification and a set of features for *A. parvulum* IPP 1246<sup>T</sup> together with the description of the complete genomic sequencing and annotation.

## Classification and features

Phylotypes with significant 16S sequence similarity to strain IPP 1246<sup>T</sup> were observed from intubated patients (EF510777) and from metagenomic human skin surveys (GQ081350) [7]. No significant similarities were found in human gut metagenomes (highest similarity is 92%, BABE01011651) [8], or in marine metagenomes (87%, AACY020304192) [9] (status June 2009).

Figure 1 shows the phylogenetic neighborhood of *A. parvulum* strain IPP P1246<sup>T</sup> in a 16S rRNA based tree. The sequence of the sole copy of the 16S rRNA gene in the genome is identical with the previously published sequence generated from ATCC 22793 (AF292372), but differs by four nucleotides from the sequence used for the last taxonomic emendation (X67150) [2].



**Figure 1.** Phylogenetic tree of *A. parvulum* strain IPP 1246<sup>T</sup>, all other type strains of the genus *Atopobium* and the type strains of all other genera within the *Coriobacteriaceae*, inferred from 1345 aligned characters [10,11] of the 16S rRNA gene sequence under the maximum likelihood criterion [12]. The tree was rooted with the type strains of the genera within the subclass *Rubrobacteridae*. The branches are scaled in terms of the expected number of substitutions per site. Numbers above branches are support values from 1000 bootstrap replicates if larger than 60%. Lineages with type strain genome sequencing projects registered in GOLD [13] are shown in blue, published genomes in bold, including two of which are reported in this issue of *SIGS* [14,15]

The cells are cocci (approximately 0.3 to 0.6  $\mu\text{m}$  in diameter) that occur singly, in pairs, in clumps, and in short chains, occasionally with central swelling [16,17] (Table 1 and Figure 2). The strains are non-motile and obligate anaerobic. Interestingly, growth is substantially stimulated by 0.02% (vol/vol) Tween 80 and by 10% (vol/vol) rabbit serum added to culture media [16]. Strain IPP 1246<sup>T</sup> is susceptible to chloramphenicol (12  $\mu\text{g/ml}$ ), clindamycin (1.6  $\mu\text{g/ml}$ ), erythromycin (3  $\mu\text{g/ml}$ ), penicillin G (2 U/ml), and tetracycline (6  $\mu\text{g/ml}$ ) [17].

Strain IPP 1126<sup>T</sup> produces acid (final pH < 4.7) from cellobiose, esculin, fructose, galactose, glucose, inulin, lactose, maltose, mannose, salicin,

sucrose, and trehalose; erythritol and xylose were weakly fermented; no acid was produced from amygdalin, arabinose, glycerol, glycogen, inositol, mannitol, melezitose, melibiose, pectin, raffinose, rhamnose, ribose, sorbitol, or starch. Esculin was hydrolyzed; neither starch nor hippurate was hydrolyzed. Nitrate was not reduced. Indole was not formed. A solid acid curd formed in milk; neither milk, gelatin, nor meat was digested. Neither catalase, urease, deoxyribonuclease, lecithinase, nor lipase was detected [17]. Other enzyme activities are positive for acid phosphatase, alanine arylamidase, arginine arylamidase,  $\beta$ -galactosidase, leucine arylamidase, pyroglutamic acid arylamidase, glycine arylamidase, tyrosine arylamidase,

but negative for arginine dihydrolase, histidine arylamidase, proline arylamidase, serine arylamidase, as determined using the API system [24].

### Chemotaxonomy

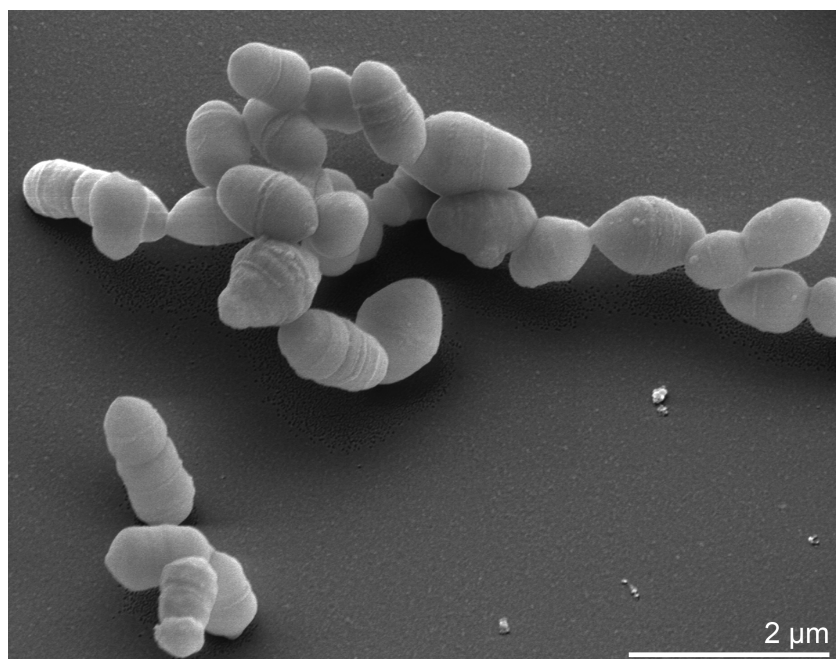
The chemotaxonomy of *A. parvulum* IPP 1246<sup>T</sup> is unfortunately hardly studied. There are no data known on the polar lipids. The strain possesses cell-wall peptidoglycan of type A4 $\alpha$ , L-Lys-D-Asp

(type A11.31 according to the DSMZ catalogue of strains) [25]. The major cellular fatty acids (FAME: fatty acid methyl ester; DMA: dimethylacetyl) are C<sub>18:1</sub> *cis*-9 (38.2%, FAME), C<sub>18:1</sub> *cis*-9 (24.1%, DMA), C<sub>16:1</sub> *cis*-9 (5.0%, FAME), C<sub>17:1</sub> *cis*-8 (5.0%, FAME), C<sub>18:1</sub> c11/t9/t6 (5.0%, FAME), C<sub>18:1</sub> *cis*-11 (3.9%, DMA), C<sub>14:0</sub> (3.4%, FAME), C<sub>10:0</sub> (3.0%, FAME) [16].

**Table 1.** Classification and general features of *A. parvulum* IPP 1146<sup>T</sup> according to the MIGS recommendations [18].

MIGS ID	Property	Term	Evidence code
		Domain <i>Bacteria</i>	TAS [19]
		Phylum <i>Actinobacteria</i>	TAS [20]
		Class <i>Actinobacteria</i>	TAS [20]
		Subclass <i>Coriobacteridae</i>	TAS [21]
	Current classification	Order <i>Coriobacteriales</i>	TAS [21]
		Suborder " <i>Coriobacterineae</i> "	TAS [21]
		Family <i>Coriobacteriaceae</i>	TAS [21]
		Genus <i>Atopobium</i>	TAS [2]
		Species <i>Atopobium parvulum</i>	TAS [2]
		Type strain IPP 1246	
	Gram stain	positive	TAS [16]
	Cell shape	small cocci that occasionally appear to be elliptical	TAS [16]
	Motility	nonmotile	TAS [17]
	Sporulation	nonsporulating	TAS [16]
	Temperature range	25°C–45°C	TAS [17]
	Optimum temperature	37°C–45°C	TAS [17]
	Salinity	less than 6.5% NaCl	TAS [17]
MIGS-22	Oxygen requirement	obligate anaerobic	TAS [17]
	Carbon source	acid production from cellobiose, esculin, fructose, galactose, glucose, inulin, lactose, maltose, mannose, salicin, sucrose, and trehalose	TAS [17]
	Energy source	carbohydrates	TAS [17]
MIGS-6	Habitat	human respiratory tract.	TAS [1,17]
MIGS-15	Biotic relationship	free living	NAS
MIGS-14	Pathogenicity	associated with halitosis and human oral infections	TAS [3,4,6]
	Biosafety level	2	TAS [22]
	Isolation	unknown for this specific strain, but Weinberg <i>et al</i> reported that the principal habitat was the respiratory tract.	TAS [1,17]
MIGS-4	Geographic location	unknown, probably France	TAS [1,17]
MIGS-5	Sample collection time	before 1937	TAS [1,17]
MIGS-4.1	Latitude – Longitude	unknown	
MIGS-4.2			
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 [23]. If the evidence code is IDA the property was directly observed for a living isolate by one of the authors or another expert mentioned in the acknowledgements.



**Figure 2.** Scanning electron micrograph of *A. parvulum* IPP 1246<sup>T</sup>

## Genome sequencing and annotation

### Genome project history

This organism was selected for sequencing on the basis of each phylogenetic position, and is part of the *Genomic Encyclopedia of Bacteria and Archaea* project. The genome project is deposited in the Genome OnLine Database [13] 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.

### Growth conditions and DNA isolation

*A. parvulum* strain IPP 1246<sup>T</sup>, DSM 20469, was grown anaerobically in DSMZ medium 104 (modified PYG; Medium [26]) at 37°C. DNA was isolated from 0.5-1 g of cell paste using the JGI CTAP procedure with a modified protocol for cell lysis as described in Wu *et al.* [27].

**Table 2.** Genome sequencing project information

MIGS ID	Property	Term
MIGS-31	Finishing quality	Finished
MIGS-28	Libraries used	Two Sanger libraries: 8kb pMCL200 and fosmid pcc1Fos One 454 pyrosequence standard library
MIGS-29	Sequencing platforms	ABI3730, 454 GS FLX
MIGS-31.2	Sequencing coverage	7.8× Sanger; 43.4× pyrosequence
MIGS-30	Assemblers	Newbler, phrap
MIGS-32	Gene calling method	Prodigal, GenePRIMP
	Genbank ID	CP001721
	Genbank Date of Release	September 9, 2009
	GOLD ID	Gc01099
	NCBI project ID	29401
	Database: IMG-GEBA	2501533209
MIGS-13	Source material identifier	DSM 20469
	Project relevance	Tree of Life, GEBA

### 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 on the JGI website. 454 Pyrosequencing reads were assembled using the Newbler assembler version 1.1.02.15 (Roche). Large Newbler contigs were broken into 1,716 overlapping fragments of 1000bp and entered into assembly as pseudo-reads. The sequences were assigned quality scores based on Newbler consensus q-scores with modifications to account for overlap redundancy and to adjust inflated q-scores. A hybrid 454/Sanger assembly was made using the parallel phrap assembler (High Performance Software, LLC). Possible mis-assemblies were corrected with Dupfinisher [28] or transposon bombing of bridging clones (Epicentre Biotechnologies, Madison, WI). Gaps between contigs were closed by editing in Consed, custom primer walk or PCR amplification. A total of 125 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 51.2 x coverage of the genome. The

final assembly contains 12,842 Sanger and 359,479 pyrosequence reads.

### Genome annotation

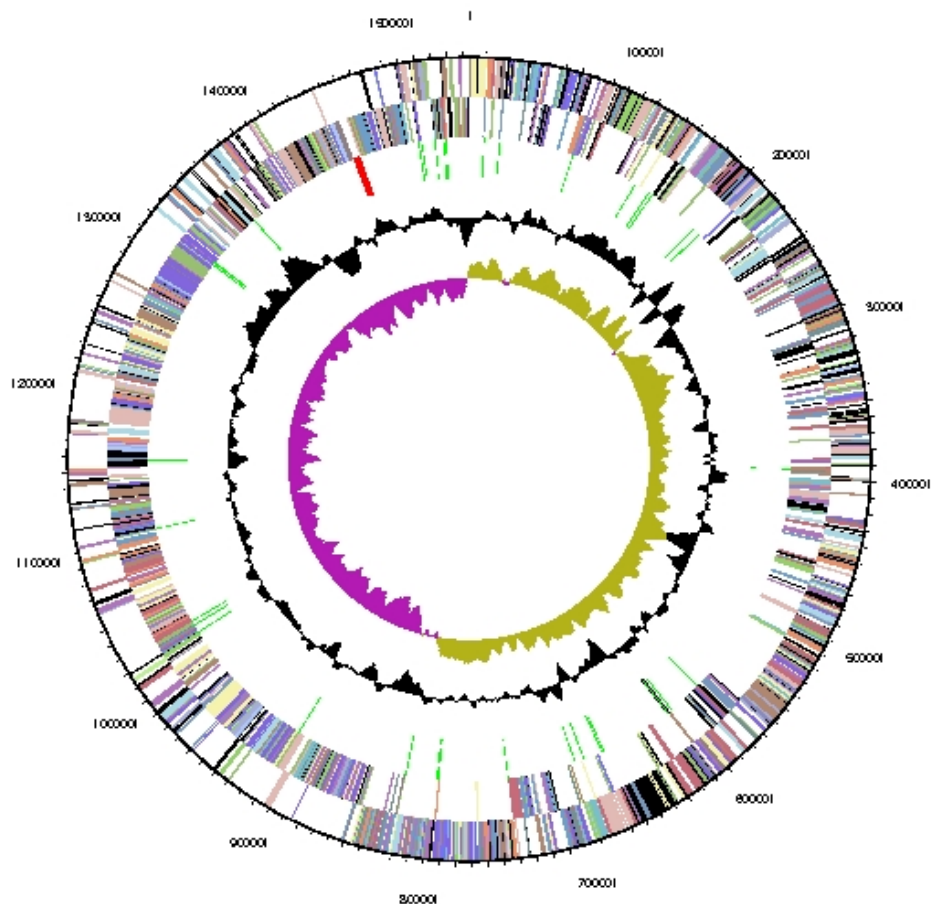
Genes were identified using Prodigal [29] as part of the Oak Ridge National Laboratory genome annotation pipeline, followed by a round of manual curation using the JGI GenePRIMP pipeline [30]. The predicted CDSs were translated and used to search the National Center for Biotechnology Information (NCBI) nonredundant database, UniProt, TIGRFam, Pfam, PRIAM, KEGG, COG, and InterPro databases. Additional gene prediction analysis and functional annotation were performed within the Integrated Microbial Genomes Expert Review (IMG-ER) platform [31].

### Genome properties

The genome is 1,543,805 bp long and comprises one main circular chromosome with a 45.7% GC content (Table 3 and Figure 3). Of the 1419 genes predicted, 1369 were protein coding genes, and 50 RNAs. Sixteen pseudogenes were also identified. The majority of the genes (74.5%) 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 3.** Genome Statistics

Attribute	Value	% of Total
Genome size (bp)	1,543,805	100.00%
DNA Coding region (bp)	1,396,223	90.44%
DNA G+C content (bp)	705,312	45.69%
Number of replicons	1	
Extrachromosomal elements	0	
Total genes	1419	100.00%
RNA genes	49	3.52%
rRNA operons	1	
Protein-coding genes	1369	96.48%
Pseudo genes	16	1.13%
Genes with function prediction	1059	74.63%
Genes in paralog clusters	69	4.86%
Genes assigned to COGs	1096	77.24%
Genes assigned Pfam domains	1084	76.39%
Genes with signal peptides	240	16.91%
Genes with transmembrane helices	339	23.89%
CRISPR repeats	0	



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

Code	Value	% age	Description
J	128	9.3	Translation, ribosomal structure and biogenesis
A	0	0.0	RNA processing and modification
K	85	6.2	Transcription
L	72	5.3	Replication, recombination and repair
B	1	0.1	Chromatin structure and dynamics
D	18	1.3	Cell cycle control, mitosis and meiosis
Y	0	0.0	Nuclear structure
V	42	3.1	Defense mechanisms
T	46	3.4	Signal transduction mechanisms
M	70	5.1	Cell wall/membrane biogenesis
N	1	0.1	Cell motility
Z	0	0.0	Cytoskeleton
W	0	0.0	Extracellular structures
U	20	1.5	Intracellular trafficking and secretion
O	44	3.2	Posttranslational modification, protein turnover, chaperones
C	44	3.2	Energy production and conversion
G	115	8.4	Carbohydrate transport and metabolism

**Table 4.** Number of genes associated with the general COG functional categories (cont.)

Code	Value	% age	Description
E	105	7.7	Amino acid transport and metabolism
F	53	3.9	Nucleotide transport and metabolism
H	37	2.7	Coenzyme transport and metabolism
I	23	1.7	Lipid transport and metabolism
P	59	4.3	Inorganic ion transport and metabolism
Q	11	0.8	Secondary metabolites biosynthesis, transport and catabolism
R	125	9.1	General function prediction only
S	90	6.6	Function unknown
-	273	19.9	Not in COGs

## Acknowledgements

We would like to gratefully acknowledge the help of Katja Steenblock for growing *A. parvulum* 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 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

- Weinberg M, Nativelle R, Prévot AR. Les microbes anaérobies. Masson and Co., Paris. 1937.
- Collins MD, Wallbanks S. Comparative sequence analyses of the 16S rRNA genes of *Lactobacillus minutus*, *Lactobacillus rimae* and *Streptococcus parvulus*: proposal for the creation of a new genus *Atopobium*. *FEMS Microbiol Lett* 1992; **74**: 235-240. [PubMed doi:10.1111/j.1574-6968.1992.tb05372.x](#)
- Kazor CE, Mitchell PM, Lee AM, Stokes LN, Loesche WJ, Dewhirst FE, Paster BJ. Diversity of bacterial populations on the tongue dorsa of patients with halitosis and healthy patients. *J Clin Microbiol* 2003; **41**:558-563.
- Riggio MP, Lennon A, Rolph HJ, Hodge PJ, Donaldson A, Maxwell AJ, Bagg J. Molecular identification of bacteria on the tongue dorsum of subjects with and without halitosis. *Oral Dis* 2008; **14**: 251-258. [PubMed doi:10.1111/j.1601-0825.2007.01371.x](#)
- Kumar PS, Griffen AL, Barton JA, Paster BJ, Moeschberger ML, Leys EJ. New bacterial species associated with chronic periodontitis. *J Dent Res* 2003; **82**: 338-344. [PubMed doi:10.1177/154405910308200503](#)
- Downes J, Munson MA, Spratt DA, Konen E, Tarkka E, Jousimies-Somer H, Wade WG. Characterisation of *Eubacterium*-like strains isolated from oral infections. *J Med Microbiol* 2001; **50**: 947-951. [PubMed](#)
- Grice EA, Kong HH, Conlan S, Deming CB, Davis J, Young AC, NISC Comparative Sequencing Program, Bouffard GG, Blakesley RW, Murray RR, et al. Topographical and temporal diversity of the human skin microbiome. *Science* 2009; **324**: 1190-1192. [PubMed](#)
- Kurokawa K, Itoh T, Kuwahara T, Oshima K, Toh H, Toyoda A, Takami H, Morita H, Sharma VK, Srivastava TP, et al. Comparative metagenomics revealed commonly enriched gene sets in human gut microbiomes. *DNA Res* 2007; **14**: 169-181. [PubMed doi:10.1093/dnares/dsm018](#)
- Venter JC, Remington K, Heidelberg JF, Halpern AL, Rusch D, Eisen JA, Wu D, Paulsen J, Nelson KE, Nelson W, et al. Environmental genome shotgun sequencing of the Sargasso Sea. *Science* 2004; **304**: 66-74. [PubMed doi:10.1126/science.1093857](#)
- Castresana J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol* 2000; **17**: 540-552. [PubMed](#)
- 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](#)

12. 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](https://pubmed.ncbi.nlm.nih.gov/doi/10.1080/10635150802429642)
13. Liolios K, Mavromatis K, Tavernarakis N, Kyrpides NC. The Genomes On Line Database (GOLD) in 2007: status of genomic and metagenomic projects and their associated metadata. *Nucleic Acids Res* 2008; **36**: D475-D479. [PubMed doi:10.1093/nar/gkm884](https://pubmed.ncbi.nlm.nih.gov/doi/10.1093/nar/gkm884)
14. Mavrommatis K, Pukall R, Rohde C, Chen F, Sims D, Brettin T, Kuske C, Detter JC, Han C, Lapidus A, et al. Complete genome sequence of *Cryptobacterium curtum* type strain (12-3T). *Stand Genomic Sci* 2009; **1**: 93-101. [doi:10.4056/sigs.15195](https://doi.org/10.4056/sigs.15195)-
15. Saunders E, Pukall R, Abt B, Lapidus A, Galvina Del Rio T, Copeland A, Tice H, Cheng JF, Lucas S, Chen F, et al. Complete genome sequence of *Eggerthella lenta* type strain (IPP VPI 0255T). *Stand Genomic Sci* 2009; **1**: 174-182. [doi:10.4056/sigs.33592](https://doi.org/10.4056/sigs.33592)
16. Olsen I, Johnson JL, Moore LVM, Moore WEC. *Lactobacillus uli* sp. nov. and *Lactobacillus rimae* sp. nov. from the human gingival crevice and emended descriptions of *Lactobacillus minutus* and *Streptococcus parvulus*. *Int J Syst Bacteriol* 1991; **41**: 261-266. [PubMed](https://pubmed.ncbi.nlm.nih.gov/doi/10.1093/ijb/41.3.261)
17. Cato EP. Transfer of *Peptostreptococcus parvulus* (Weinberg, Nativelle, and Prévot 1937) Smith 1957 to the Genus *Streptococcus*: *Streptococcus parvulus* (Weinberg, Nativelle, and Prévot 1937) comb. nov., nom. rev., emend. *Int J Syst Bacteriol* 1983; **33**: 82-84.
18. 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](https://pubmed.ncbi.nlm.nih.gov/doi/10.1038/nbt1360)
19. 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](https://pubmed.ncbi.nlm.nih.gov/doi/10.1073/pnas.87.12.4576)
20. Garrity GM, Holt J. In: Garrity GM, Boone DR, Castenholz RW, eds. Taxonomic Outline of the *Archaea* and *Bacteria*. *Bergey's Manual of Systematic Bacteriology*, 2nd Ed. Vol 1 The *Archaea*, Deeply Branching and Phototrophic *Bacteria*. 2001 pp. 155-166
21. 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.
22. Anonymous. Biological Agents. Technical rules for biological agents www.baua.de TRBA 466.
23. 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. [PubMed doi:10.1038/75556](https://pubmed.ncbi.nlm.nih.gov/doi/10.1038/75556)
24. Rodriguez Jovita MR, Collins MD, Sjoden B, Falsen E. Characterization of a novel *Atopobium* isolate from the human vagina: description of *Atopobium vaginae* sp. nov. *Int J Syst Bacteriol* 1999; **49**: 1573-1576. [PubMed](https://pubmed.ncbi.nlm.nih.gov/doi/10.1093/ijb/49.10.1573)
25. Weiss N. Cell wall structure of anaerobic cocci. *Rev Inst Pasteur Lyon* 1981; **14**: 53-59.
26. List of media used at DSMZ for cell growth: [http://www.dsmz.de/microorganisms/media\\_list.php](http://www.dsmz.de/microorganisms/media_list.php)
27. Wu M, Hugenholtz P, Mavromatis K, Pukall R, Dalin E, Ivanova N, Kunin V, Goodwin L, Wu M, Tindall BJ, et al. A phylogeny-driven genomic encyclopedia of *Bacteria* and *Archaea*. *Nature* 2009; *(In press)*
28. 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 of *Kytococcus sedentarius* type strain (541<sup>T</sup>). *Stand Genomic Sci* 2009; **1**: 12-20. [doi:10.4056/sigs.761](https://doi.org/10.4056/sigs.761)
29. Anonymous. Prodigal Prokaryotic Dynamic Programming Genefinding Algorithm. Oak Ridge National Laboratory and University of Tennessee 2009 <http://compbio.ornl.gov/prodigal>
30. Pati A, Ivanova N, Mikhailova N, Ovchinnikova G, Hooper SD, Lykidis A, Kyrpides NC. GenePRIMP: A Gene Prediction Improvement Pipeline for microbial genomes. (Submitted) 2009.
31. Markowitz VM, Mavromatis K, Ivanova NN, Chen IMA, Chu K, Kyrpides NC. Expert Review of Functional Annotations for Microbial Genomes. *Bioinformatics* 2009; **25**: 2271-2278. [PubMed doi:10.1093/bioinformatics/btp393](https://pubmed.ncbi.nlm.nih.gov/doi/10.1093/bioinformatics/btp393)