# Complete genome sequence of the termite hindgut bacterium *Spirochaeta coccoides* type strain (SPN1$^T$), reclassification in the genus *Sphaerochaeta* as *Sphaerochaeta coccoides* comb. nov. and emendations of the family *Spirochaetaceae* and the genus *Sphaerochaeta*

Birte Abt[3], Cliff Han[1,2], Carmen Scheuner[3], Megan Lu[1,2], Alla Lapidus[1], Matt Nolan[1], Susan Lucas[1], Nancy Hammon[1], Shweta Deshpande[1], Jan-Fang Cheng[1], Roxane Tapia[1,2], Lynne Goodwin[1,2], Sam Pitluck[1], Konstantinos Liolios[1], Ioanna Pagani[1], Natalia Ivanova[1], Konstantinos Mavromatis[1], Natalia Mikhailova[1], Marcel Huntemann[1], Amrita Pati[1], Amy Chen[4], Krishna Palaniappan[4], Miriam Land[1,5], Loren Hauser[1,5], Evelyne-Marie Brambilla[3], Manfred Rohde[6], Stefan Spring[3], Sabine Gronow[3], Markus Göker[3], Tanja Woyke[1], James Bristow[1], Jonathan A. Eisen[1,7], Victor Markowitz[4], Philip Hugenholtz[1,8], Nikos C. Kyrpides[1], Hans-Peter Klenk[3]*, and John C. Detter[2]

[1] DOE Joint Genome Institute, Walnut Creek, California, USA
[2] Los Alamos National Laboratory, Bioscience Division, Los Alamos, New Mexico, USA
[3] DSMZ - German Collection of Microorganisms and Cell Cultures GmbH, Braunschweig, Germany
[4] Biological Data Management and Technology Center, Lawrence Berkeley National Laboratory, Berkeley, California, USA
[5] Oak Ridge National Laboratory, Oak Ridge, Tennessee, 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: Hans-Peter Klenk

## Keywords
obligately anaerobic, non-motile, termite hindgut, Gram-negative, di- and oligosaccharide-degrading, mesophilic, chemoorganotrophic, *Spirochaetaceae*, *Sphaerochaeta*, GEBA.

## Abstract
*Spirochaeta coccoides* Dröge *et al.* 2006 is a member of the genus *Spirochaeta* Ehrenberg 1835, one of the oldest named genera within the Bacteria. *S. coccoides* is an obligately anaerobic, Gram-negative, nonmotile, spherical bacterium that was isolated from the hindgut contents of the termite *Neotermes castaneus*. The species is of interest because it may play an important role in the digestion of breakdown products from cellulose and hemicellulose in the termite gut. Here we provide a taxonomic re-evaluation for strain SPN1$^T$ (= DSM 17374) and based on physiological and genomic characteristics we propose its reclassification as a novel species in the genus *Sphaerochaeta*, a recently published sister group of the *Spirochaeta*. The 2,227,296 bp long genome of strain SPN1$^T$ with its 1,866 protein-coding and 58 RNA genes is a part of the ***G**enomic **E**ncyclopedia of **B**acteria and **A**rchaea* project.

## Introduction
Strain SPN1$^T$ (= DSM 17374 = ATCC BAA-1237) is the type strain of *Spirochaeta coccoides* and was isolated from the hindgut contents of the lower dry-wood termite *Neotermes*

*castaneus* [1,2]. The genus *Spirochaeta* currently consists of 19 validly named species [49]. The genus name was derived from the latinized Greek words *speira*, 'a coil' and *chaitê,* 'hair', yielding the Neo-Latin 'Spirochaeta', the coiled hair [49]. The species epithet was derived from the neo-Greek words *coccos*, 'a berry' and *eidos,* meaning 'shape', yielding the Neo-Latin word *coccoides*, meaning berry-shaped [1]. Based on the nucleotide sequence of 16S rRNA gene strain SPN1$^T$ was assigned to the genus *Spirochaeta*, although its coccoid, non-motile cells differ from the morphology of all known validly named spirochetes [1]. Recently, the two spherical isolates *Spirochaeta* sp. Buddy and *Spirochaeta* sp. Grapes were validly named as *Sphaerochaeta globosa* and *Sphaerochaeta pleomorpha*, respectively [45]. Based on their unique morphology and the phylogenetic distance to their closest relatives Ritalahti *et al*. classified *Spirochaeta* sp. Buddy and *Spirochaeta* sp. Grapes into the novel genus *Sphaerochaeta*. On the basis of its morphological, physiological and genomic characteristics, indicating that *S. coccoides* is more closely related to *Sphaerochaeta* than to the remaining *Spirochaeta* species, we propose the reclassification of *S. coccoides* SPN1$^T$ into the genus *Sphaerochaeta*. Here we thus present a summary classification and a set of features for *S. coccoides* SPN1$^T$, a description of the complete genome sequencing and annotation, and a proposal to reclassify *S. coccoides* as a member of the genus *Sphaerochaeta* as *Sphaerochaeta coccoides* comb. nov..

## Classification and features

A representative genomic 16S rRNA sequence of strain SPN1$^T$ was compared using NCBI BLAST [4,5] 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 of taxa and keywords (reduced to their stem [8]) were determined, weighted by BLAST scores. The most frequently occurring genera were *Spirochaeta* (57.6%), *Sphaerochaeta* (39.7%) and *Cytophaga* (2.7%) (22 hits in total). Regarding the six hits to sequences from other members of the genus, the average identity within HSPs was 90.2%, whereas the average coverage by HSPs was 30.9%. Among all other species, the one yielding the highest score was *Spirochaeta bajacaliforniensis* (AJ698859), which corresponded to an identity of 90.3% and an HSP coverage of 32.6%. (Note that the Greengenes database uses the INSDC (= EMBL/NCBI/DDBJ) annotation, which is not an authoritative source for nomenclature or classification.) The highest-scoring environmental sequence was AY570600 ('biodegraded Canadian oil reservoir clone PL-16B9'), which showed an identity of 91.0% and an HSP coverage of 85.9%. The most frequently occurring keywords within the labels of all environmental samples which yielded hits were 'microbi' (6.5%), 'mat' (4.5%), 'hypersalin' (3.1%), 'termit' (2.8%) and 'hindgut' (2.6%) (228 hits in total). Environmental samples which yielded hits of a higher score than the highest scoring species were not found. The keywords are partially in agreement with the known environmental preferences of *S. coccoides* SPN1$^T$, but the results also indicate that the species itself is rarely found in environmental probes.

Figure 1 shows the phylogenetic neighborhood of *S. coccoides* in a 16S rRNA based tree. The sequences of the three 16S rRNA gene copies in the genome differ from each other by up to two nucleotides, and differ by up to two nucleotides from the previously published 16S rRNA sequence (AJ698092).

**Figure 1.** Phylogenetic tree highlighting the position of *S. coccoides* relative to the other type strains within the family *Spirochaetaceae*. The tree was inferred from 1,360 aligned characters [8,9] of the 16S rRNA gene sequence under the maximum likelihood criterion [10]. Rooting was done initially using the midpoint method [3] and then checked for its agreement with the current classification (Table 1). The branches are scaled in terms of the expected number of substitutions per site. Numbers adjacent to the branches are support values from 500 ML bootstrap replicates [11] (left) and from 1,000 maximum parsimony bootstrap replicates [37] (right) if larger than 60% if. Lineages with type strain genome sequencing projects registered in GOLD [12] are labeled with one asterisk, those also listed as 'Complete and Published' with two asterisks (see [38-41], CP002696 for *Treponema brennaborense,* CP002903 for *S. thermophila*, and CP002868 for *S. caldaria*). Also finished but second asterisk missing are *S. africana*, *S. pleomorpha* CP003155 and *S. globosa* CP002541.

In contrast to all other validly described spirochete species (except for those meanwhile placed in the novel genus *Sphaerochaeta* [45]) the cells of *S. coccoides* SPN1$^T$ are cocci (0.5 to 2.0 µm diameter) which are surrounded by an outer envelope. In the early growth phase cell aggregates are formed [1]. *S. coccoides* is a Gram-negative, non-motile and strictly anaerobic bacterium (Table 1). Strain SPN1$^T$ showed no catalase activity [1], although a gene probably coding a catalase (Spico_0266) was identified in the genome. The optimal growth temperature of strain SPN1$^T$ is 30°C, with no growth observed above 40°C or below 15°C [1]. The pH range for growth is 5.5-9.5, with an optimum at pH 7.4 [1]. Maltose is fermented to ethanol, formate, and acetate as the main fermentation products. There is no utilization of glucose, galactose, lactate, pyruvate, amino acids, and polysaccharides, but the organism is able to grow with yeast extract as the sole carbon and energy source [1]. A minimum yeast concentration of 0.2% was required for growing [1]. Activities of β-D-glucosidase, α-D-glucosidase, α-D-galactosidase, α-L-arabinosidase, β-D-fucosidase, and β-D-xylosidase are

exhibited [1]. These enzymatic activities seemed to be cell-bound, as no glycolytic activity was found in the supernatant of the culture [1].

**Table 1.** Classification and general features of *S. coccoides* SPN1$^T$ in accordance with the MIGS recommendations [19] and the NamesforLife database [42].

| MIGS ID | Property | Term | Evidence code |
|---|---|---|---|
| | Current classification | Domain *Bacteria* | TAS [20] |
| | | Phylum *Spirochaetae* | TAS [21] |
| | | Class *Spirochaetes* | TAS [22] |
| | | Order *Spirochaetales* | TAS [17,18] |
| | | Family *Spirochaetaceae* | TAS [16,17] |
| | | Genus *Spirochaeta* | TAS [13-15,17] |
| | | Species *Spirochaeta coccoides* | TAS [1,2] |
| | | Type strain SPN1 | TAS [1,2] |
| | Gram stain | negative | TAS [1] |
| | Cell shape | coccoid | TAS [1] |
| | Motility | nonmotile | TAS [1] |
| | Sporulation | none | TAS [1] |
| | Temperature range | mesophile | TAS [1] |
| | Optimum temperature | 30°C | TAS [1] |
| | Salinity | not reported | |
| MIGS-22 | Oxygen requirement | obligately anaerobic | TAS [1] |
| | Carbon source | pentoses (arabinose, xylose), oligosaccharides (maltose, cellobiose, maltotriose, maltotetraose), yeast extract | TAS [1] |
| | Energy metabolism | chemoorganotroph | TAS [1] |
| MIGS-6 | Habitat | digestive tract of lower dry-wood termites | TAS [1] |
| MIGS-15 | Biotic relationship | host associated commensal | TAS [1] |
| MIGS-14 | Pathogenicity | none | TAS [1] |
| | Biosafety level | 1 | TAS [23] |
| | Isolation | hindgut of *Neotermes castaneus* | TAS [1] |
| MIGS-4 | Geographic location | not reported | |
| MIGS-5 | Sample collection time | 2005 or before | TAS [1] |
| 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 - 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 [24].

# Genome sequencing and annotation

### Genome project history

This organism was selected for sequencing on the basis of its phylogenetic position [25], and is part of the ***G**enomic **E**ncyclopedia of **B**acteria and **A**rchaea* project [26]. The genome project is deposited in the Genomes On Line Database [12] 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 (8.9 kb insert size), one Illumina library |
| MIGS-29 | Sequencing platforms | Illumina GAii, 454 GS FLX Titanium |
| MIGS-31.2 | Sequencing coverage | 960.0 x Illumina; 40.0 x pyrosequence |
| MIGS-30 | Assemblers | Newbler version 2.3, Velvet version 0.7.63, phrap version SPS - 4.24 |
| MIGS-32 | Gene calling method | Prodigal 1.4, GenePRIMP |
| | INSDC ID | CP002659 |
| | Genbank Date of Release | April 27, 2011 |
| | GOLD ID | Gc01739 |
| | NCBI project ID | 48121 |
| | Database: IMG-GEBA | 2503904012 |
| MIGS-13 | Source material identifier | DSM 17374 |
| | Project relevance | Tree of Life, GEBA |

**Growth conditions and DNA isolation**

*S. coccoides* strain SPN1$^T$, DSM 17374, was grown anaerobically in DSMZ medium 1204 (*Spirochaeta coccoides* medium) [27] at 30°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 with modification st/DL for cell lysis as described in Wu *et al*. 2009 [26]. DNA is available through the DNA Bank Network [28].

**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 [35]. Pyrosequencing reads were assembled using the Newbler assembler (Roche). The initial Newbler assembly consisting of 97 contigs in one scaffold was converted into a phrap [36] assembly by making fake reads from the consensus, to collect the read pairs in the 454 paired end library. Illumina GAii sequencing data (2,245.3 Mb) was assembled with Velvet [29] and the consensus sequences were shredded into 2.0 kb overlapped fake reads and assembled together with the 454 data. The 454 draft assembly was based on 142.5 Mb 454 draft data and all of the 454 paired end data. Newbler parameters are -consed -a 50 -l 350 -g -m -ml 20. The Phred/Phrap/Consed software package [36] was used for sequence assembly and quality assessment in the subsequent finishing process. After the shotgun stage, reads were assembled with parallel phrap (High Performance Software, LLC). Possible mis-assemblies were corrected with gapResolution [35], Dupfinisher [30], or sequencing cloned bridging PCR fragments with subcloning. Gaps between contigs were closed by editing in Consed, by PCR and by Bubble PCR primer walks (J.-F. Chang, unpublished). A total of 308 additional reactions 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 [31]. 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 1,000.0 x coverage of the genome. The final assembly contained 137,682 pyrosequence and 58,694,953 Illumina reads.

**Genome annotation**

Genes were identified using Prodigal [32] as part of the Oak Ridge National Laboratory genome annotation pipeline, followed by a round of manual curation using the JGI GenePRIMP pipeline [33]. 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 [34].

# Genome properties

The genome consists of a 2,227,296 bp long chromosome with a G+C content of 50.6% (Table 3 and Figure 2). Of the 1,924 genes predicted, 1,866 were protein-coding genes, and 58 RNAs; 44 pseudogenes were also identified. The majority of the protein-coding genes (74.6%) 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) | 2,227,296 | 100.00% |
| DNA coding region (bp) | 2,003,786 | 89.96% |
| DNA G+C content (bp) | 1,126,077 | 50.56% |
| Number of replicons | 1 | |
| Extrachromosomal elements | 0 | |
| Total genes | 1,924 | 100.00% |
| RNA genes | 58 | 3.01% |
| rRNA operons | 3 | |
| Protein-coding genes | 1,866 | 96.99% |
| Pseudo genes | 44 | 2.29% |
| Genes with function prediction | 1,434 | 74.53% |
| Genes in paralog clusters | 733 | 38.10% |
| Genes assigned to COGs | 1,528 | 76.72% |
| Genes assigned Pfam domains | 1,518 | 78.90% |
| Genes with signal peptides | 314 | 16.32% |
| Genes with transmembrane helices | 524 | 27.23% |
| CRISPR repeats | 4 | |

**Figure 2.** Graphical map of the chromosome. 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 | COG counts and percentage of protein-coding genes | | Description |
|---|---|---|---|
| | Genome | | |
| | value | % of total | |
| J | 143 | 8.5 | Translation, ribosomal structure and biogenesis |
| A | 0 | 0.0 | RNA processing and modification |
| K | 118 | 7.0 | Transcription |
| L | 99 | 5.9 | Replication, recombination and repair |
| B | 0 | 0.0 | Chromatin structure and dynamics |
| D | 58 | 3.5 | Cell cycle control, cell division, chromosome partitioning |
| Y | 0 | 0.0 | Nuclear structure |

| | | | |
|---|---|---|---|
| V | 25 | 1.5 | Defense mechanisms |
| T | 59 | 3.5 | Signal transduction mechanisms |
| M | 46 | 2.7 | Cell wall/membrane/envelope biogenesis |
| N | 2 | 0.1 | Cell motility |
| Z | 39 | 2.4 | Cytoskeleton |
| W | 0 | 0.0 | Extracellular structures |
| U | 19 | 1.1 | Intracellular trafficking, secretion, and vesicular transport |
| O | 54 | 3.2 | Posttranslational modification, protein turnover, chaperones |
| C | 77 | 4.6 | Energy production and conversion |
| G | 260 | 15.5 | Carbohydrate transport and metabolism |
| E | 160 | 9.5 | Amino acid transport and metabolism |
| F | 58 | 3.6 | Nucleotide transport and metabolism |
| H | 42 | 2.5 | Coenzyme transport and metabolism |
| I | 44 | 2.6 | Lipid transport and metabolism |
| P | 58 | 3.6 | Inorganic ion transport and metabolism |
| Q | 15 | 0.9 | Secondary metabolites biosynthesis, transport and catabolism |
| R | 193 | 11.5 | General function prediction only |
| S | 109 | 6.5 | Function unknown |
| - | 396 | 20.6 | Not in COGs |

## Insights from the genome sequence, and taxonomic conclusions for *S. coccoides*

**Taxonomic interpretation for *S. coccoides* and neighboring species in the family *Spirochaetaceae* according to 16S rRNA data**

Based on its 16S rRNA sequence strain SPN1$^T$ was classified into the genus *Spirochaeta* [1], although it lacks the typical spiral morphology and is not motile. SPN1$^T$ showed highest similarity in 16S rRNA gene sequences to two at that time not validly published spherical isolates *Spirochaeta* sp. strain Buddy and *Spirochaeta* sp. strain Grapes [1]. Recently these isolates were classified into the new genus *Sphaerochaeta*, and validly published as *Sphaerochaeta globosa* and *Sphaerochaeta pleomorpha*, respectively [45].

The phylogenetic tree shown in Figure 1 demonstrates that the current classification of the group suffers from a non-homogenous location of species featured as members of the genus *Spirochaeta*. Not only is *Borrelia* placed within *Spirochaeta* (but without much branch support), but also appears *S. coccoides* as the sister group of *Sphaerochaeta* with maximum support. Support for a placement of *S. caldaria*, *S. stenostrepta* and *S. zuelzerae* more close to *Treponema* than to the other *Spirochaeta* species is also high and could only be considered a matter of rooting for the former two species (but note that the rooting is confirmed by a phylogenomic analysis described below).

To measure phylogenetic conflict caused by the taxonomic classification in detail, we conducted both unconstrained heuristic searches for the best tree under the maximum likelihood (ML) [10] and maximum parsimony (MP) criterion [37] as well as searches constrained for the monophyly of all genera (for details of the data matrix see the caption of figure 1). Our own re-implementation of CopyCat [47] in conjunction with AxPcoords and AxParafit [46] was used to determine those leaves (species) whose placement significantly deviated between the constrained and the unconstrained tree. AxParafit was applied to the ML trees with 1,000 rounds of random permutations of the associations. The ParaFit test was

originally introduced for comparing host and parasite phylogenies, but can be applied to the comparison of all kinds of trees. In contrast to other measures for the comparison of trees, it includes a statistical test for whether individual leaves significantly contribute to the agreement between two trees (a p value indicates how likely it is that this contribution is no more than random). All other leaves apparently cause more conflict than agreement.
The best-known ML tree had a log likelihood of -16,001.40, whereas the best tree found under the constraint had a log likelihood of -16,322.98. The constrained tree was significantly worse than the globally best one in the Shimodaira-Hasegawa test as implemented in RAxML [10] ($\alpha = 0.01$). The best-known MP trees had a score of 3,105, whereas the best constrained trees found had a score of 3,260 and were significantly worse in the Kishino-Hasegawa test as implemented in PAUP* [37] ($\alpha < 0.0001$). (See, e.g. chapter 21 in [48] for an in-depth description of such paired-site tests.) Accordingly, the current classification of the family as used by [49] and [50] is in significant conflict with the 16S rRNA data. Table 5 shows the ParaFit test results obtained by comparing the unconstrained tree and the one obtained with the genus-based constraint. The largest conflict is caused by the main *Spirochaeta aurantia*, probably because of its placement close to *Borrelia*, followed by *Sphaerochaeta* and then by the other members of the main *Spirochaeta* group.

To assess whether placing *S. coccoides* in *Sphaerochaeta* and the other three *Spirochaeta* species that cause conflict in *Treponema* would solve the problem, an according second constraint was created and used in phylogenetic analysis. The resulting ML tree had a log likelihood of -16,025.93 and was significantly worse than the best-known ML tree only for $\alpha = 0.05$. The MP trees inferred under the second constraint had a score of 3,123 and were not significantly worse than the best-known MP trees. Table 5 also shows the ParaFit test results obtained by comparing the unconstrained tree and the one obtained with the second constraint. Apparently the conflict is largely resolved; the only remaining p value above 0.05 is the one for *S. thermophilus*, which is nevertheless only slightly insignificant at this level (0.0539).

**Table 5.** Result (p values) from the test of individual links with ParaFit for the species with an insignificant result ($\alpha = 0.05$) in the first approach. The comparison was done between an unconstrained ML tree and the first, genus-based constraint (second column) or the second constraint, based on a revised classification of the group (third column). Note that with a single exception the phylogenetic conflict was resolved by assigning *S. coccoides* to *Sphaerochaeta* and three other *Spirochaeta* species to *Treponema*.

| species | p value, constraint 1 | p value, constraint 2 |
|---|---|---|
| *Spirochaeta aurantia* (M57740) | 0.2882 | 0.0038 |
| *Sphaerochaeta globosa* (AF357916) | 0.2844 | 0.0230 |
| *Sphaerochaeta pleomorpha* (AF357917) | 0.2754 | 0.0201 |
| *Spirochaeta cellobiosiphila* (EU448140) | 0.2076 | 0.0080 |
| *Spirochaeta americana* (AF373921) | 0.2001 | 0.0149 |
| *Spirochaeta alkalica* (X93927) | 0.1905 | 0.0145 |
| *Spirochaeta asiatica* (X93926) | 0.1830 | 0.0280 |
| *Spirochaeta halophila* (M88722) | 0.1806 | 0.0124 |
| *Spirochaeta bajacaliforniensis* (AJ698859) | 0.1765 | 0.0490 |
| *Spirochaeta dissipatitropha* (AY995150) | 0.1749 | 0.0278 |
| *Spirochaeta africana* (X93928) | 0.1656 | 0.0241 |
| *Spirochaeta isovalerica* (M88720) | 0.1654 | 0.0039 |
| *Spirochaeta smaragdinae* (U80597) | 0.1592 | 0.0454 |
| *Spirochaeta thermophila* (FR749903) | 0.1384 | 0.0539 |
| *Spirochaeta litoralis* (FR733665) | 0.1327 | 0.0025 |

| | | |
|---|---|---|
| *Spirochaeta coccoides* (IMG2503956950) | 0.0863 | 0.0217 |
| *Spirochaeta perfilievii* (AY337318) | 0.0716 | 0.0010 |

**Phylogenomic analyses**

For comparative analysis the genome sequences of *Sphaerochaeta globosa* (GenBank CP002541) and *Sphaerochaeta pleomorpha* (CP003155) [45], as well as the sequences of *S. smaragdinae* (GenBank CP002659) were used.

The genomes of the sequenced *Spirochaeta* and *Sphaerochaeta* species differ significantly in their size. The genome of *S. coccoides* (2.2 Mb, 1,866 protein-coding genes, G+C content 51 mol%) is the smallest in size. The genomes of *Sphaerochaeta pleomorpha* (3.6 Mb, 3,216 protein coding genes, G+C content 46 mol%), and *Sphaerochaeta globosa* (3.3 Mb, 3,057 protein-coding genes, G+C content 49 mol%) are bigger in size and the genome of *S. smaragdinae* counts 4.7 Mb with 4,306 protein-coding genes and a G+C content of 49 mol%.

An estimate of the overall similarity between *S. coccoides,* with both *Sphaerochaeta* species and *S. smaragdinae* was generated with the GGDC-Genome-to-Genome Distance Calculator [43,44]. This system calculates the distances by comparing the genomes to obtain HSPs (high-scoring segment pairs) and interfering distances from the set of formulas (1, HSP length / total length; 2, identities / HSP length; 3, identities / total length). Table 6 shows the results of the pairwise comparison.

The comparison of *S. coccoides* with both *Sphaerochaeta* species revealed the highest scores using the GGDC. The comparison of *S. coccoides* with *Sphaerochaeta globosa* and *Sphaerochaeta pleomorpha* revealed that 4.5% and 3.9% of the average of genome length are covered with HSPs. The identity within the HSPs was 83.2% and 83.3%, respectively, whereas the identity over the whole genome was 3.7% and 3.3%, respectively. Lower similarity scores were observed in the comparison of *S. coccoides* with *S. smaragdinae*: only 1.2% of the average of both genome lengths are covered with HSPs. The identity within these HSPs was 84.6%, whereas the identity over the whole genome was only 1.0%.

As expected, those distances relating HSP coverage (formula 1) and number of identical base pairs within HSPs to total genome length (formula 3) are higher between the *S. coccoides* and the *Sphaerochaeta* species than between *S. coccoides* and *S. smaragdinae*. That the distances relating the number of identical base pairs to total HSP length (formula 2) behave differently indicates that the genomic similarities between *S. coccoides* and *S. smaragdinae* are strongly restricted to more conserved sequences, a kind of saturation phenomenon [44].

**Table 6.** Pairwise comparison of *S. coccoides* with both *Sphaerochaeta* species and *Spirochaeta smaragdinae* using the GGDC-Genome-to-Genome Distance Calculator.

| | | **HSP length / total length [%]** | **identities / HSP length [%]** | **identities / total length [%]** |
|---|---|---|---|---|
| *Spirochaeta coccoides* | *Sphaerochaeta globosa* | 4.5 | 83.2 | 3.7 |
| *Spirochaeta coccoides* | *Sphaerochaeta pleomorpha* | 3.9 | 83.3 | 3.3 |
| *Spirochaeta coccoides* | *Spirochaeta smaragdinae* | 1.2 | 84.6 | 1.0 |
| *Sphaerochaeta globosa* | *Sphaerochaeta pleomorpha* | 14.2 | 82.0 | 11.7 |
| *Sphaerochaeta globosa* | *Spirochaeta smaragdinae* | 1.3 | 84.6 | 1.1 |

For conducting phylogenomic analyses of the group, amino-acid sequences from 16 *Spirochaetaceae* and outgroup (other *Spirochaeta* families) completed type-strain genomes were retrieved from INSDC and investigated as described in [51]. Parallel genome-against-genome protein BLAST was performed with NCBI BLAST version 2.2.17 (ftp://ftp.ncbi.nih.gov/blast/execuTables/release/2.2.17) using soft masking instead of complexity filtering. To determine orthologs, BLAST e-values were transformed using our own re-implementation of the OrthoMCL algorithm [52] in conjunction with MCL version 08-312 (http://micans.org/mcl/) [53] with the OrthoMCL default parameters (an e-value threshold of $10^{-5}$ and 2.0 as inflation parameter). OrthoMCL clusters containing inparalogs were reduced by selecting the most 'central' of several sequences from the same genome, that is, the sequence with the highest sum of within-cluster BLAST scores. These reduced clusters were aligned using MUSCLE version 3.7 under default settings [54]. The program scan_orphanerrs from the RASCAL package version 1.3.4 [55] was applied to detect orphan sequences (overall poorly aligned genes) within the alignments. After removal of orphan sequences (if present), poorly aligned columns and divergent regions were eliminated with GBLOCKS version 0.91b [8] using a minimum block length of two amino acids and allowing gap positions in all sequences. Filtered OrthoMCL cluster alignments comprising at least four sequences were concatenated to form a supermatrix for phylogenetic analysis, which was then cleaned from relatively uninformative genes using MARE [58] under default values (except that deleting taxa was disallowed). Maximum-likelihood trees were inferred from the supermatrix with RAxML [10] version 7.28 in conjunction with rapid bootstrapping and the bootstopping criterion [11] with subsequent search for the best tree. The best amino acid substitution model was determined beforehand by comparing the resulting log likelihoods on maximum-parsimony starting tree. Maximum-parsimony tree search (with 100 random sequence addition replicates and TBR branch swapping, saving no more than ten best trees per replicate) and bootstrapping (1,000 rounds, each with five tree-search replicates) was conducted with PAUP* version 4b10 [37].
In addition to the supermatrix analysis, homologous sequences were determined using our own re-implementation of the TribeMCL algorithm [56] combined with MCL (see above), applying an e-value threshold of $10^{-5}$ and an inflation parameter of 2.0. A gene-content (presence/absence) matrix was constructed, representing the occurrence of a gene of one genome within a clusters of homologs. Phylogenetic inference was done with the BINGAMMA model in RAxML and under maximum parsimony with PAUP*, other settings being as described above.
The supermatrix comprised 2,408 genes and 696,696 characters before, 522 genes and 140,413 characters after cleaning with MARE. The selected model was PROTGAMMALGF; the resulting tree had a log likelihood of -2,172,190.75 and is shown in Figure 3. The best maximum-parsimony tree found had a length of 346,334 steps (not counting uninformative characters) and was topologically identical. The gene-content matrix comprised 11,131 characters and yielded best tree with a log likelihood of -61,799.49 and a parsimony score of 10,229, respectively. Bootstrapping support values from all four methods applied are shown in Figure 3 if larger then 60%.

The sister-group relationship of *S. coccoides* and *Sphaerochaeta* was unanimously supported by all methods, much like the placement of *S. caldaria* within *Treponema*. The trees differed however, regarding the support for the placement of *Borrelia* as sister group to all other ingroup taxa. For this reason, we assessed via long-branch extraction [57] whether this positioning could be caused by long-branch attraction [48] between *Borrelia* and the outgroup. Removal of *Borrelia* and subsequent phylogenetic inference yielded a maximum-parsimony tree with the same topology that would have been obtained by pruning *Borrelia* from the tree depicted in Figure 3. Removal of the outgroup from the alignment, however,

yielded a maximum-parsimony tree in which *Borrelia* was placed as sister group of *S. thermophila*, supporting the long-branch attraction hypothesis (data not shown).

The phylogenomic analysis thus confirms the 16S rRNA tree (Figure 1) regarding the paraphyly of *Spirochaeta*. A first step to resolve this taxonomic problem is to move *S. coccoides* to the genus *Sphaerochaeta*. Given that *S. caldaria* and some other species are situated within *Treponema*, and that *Borrelia* probably is placed within the remaining *Spirochaeta* species, further taxonomic changed are probably necessary in the future.

**Figure 3.** Phylogenetic tree inferred from completely sequenced genomes of the *Spirochaeta* type strains. The tree was inferred from 140,413 aligned amino acid characters under the maximum likelihood (ML) criterion and rooted with *Leptospira*. The branches are scaled in terms of the expected number of substitutions per site. Numbers above the branches are bootstrapping support values (if larger than 60%) from (i) maximum-likelihood supermatrix analysis; (ii) maximum-parsimony supermatrix analysis; (iii) maximum-likelihood gene-content analysis; (iv) maximum-parsimony gene-content analysis. Note that the placement of *Borrelia* is probably caused by long-branch attraction. For further details see the text.

**Phenotypic data and taxonomic interpretation**

Table 7 gives an overview of some morphological and physiological features of *S. coccoides* compared with the genus descriptions of *Sphaerochaeta* and *Spirochaeta*. The coccoid cell morphology, the cell size, the lack of motility as well as the products of fermentation support the need to reclassify *S. coccoides* into the genus *Sphaerochaeta*. *S. coccoides* is so close to the original description of the genus *Sphaerochaeta* that only its reported GC content needs to be adapted.

**Table 7**. Typical features of reference taxa.

| | ***Spirochaeta coccoides* [1]** | **Genus *Sphaerochaeta* [45]** | **Genus *Spirochaeta* [14]** |
|---|---|---|---|
| Cell shape | coccoid, spherical, not spiral | coccoid, spherical, pleomorphic, not spiral | helicale, spiral, spherical bodies under unfavourable growth conditions |

| | | | |
|---|---|---|---|
| Cell size | 0.5-2.0 µm | 0.4-2.5 µm | 0.2-0.75 by 5-250 µm |
| Motility | non-motile | non-motile | motile |
| Flagellation | no flagella | no flagella | 2 periplasmic flagella (exception: *S. plicatilis*, with many flagella) |
| T-optimum | 30 °C | mesophilic | 25-68 °C |
| pH-optimum | 7.4 | neutrophilic | |
| Oxygen requirement | anaerob | anaerob | obligately anaerob or facultatively anaerob |
| Fermentation products | acetate, ethanol, formate | acetate, ethanol, formate | acetate, ethanol, $CO_2$, $H_2$ |
| G+C content | 56.6-57.4 mol% [1]<br>51 mol%, this study | 45-48 mol% | 51-65 mol% [14]<br>44-65 mol% [13] |

On the basis of the above mentioned physiological and phylogenetic characteristics of strain SPN1$^T$, its reclassification into the genus *Sphaerochaeta* is proposed. The inclusion of *Sphaerochaeta* in *Spirochaetaceae* also makes an emendation of the family necessary, as its previous description excludes features specifically found in *Sphaerochaeta*.

## Emended description of the family *Spirochaetaceae* Swellengrebel 1907

The description of the family *Spirochaetaceae* is given by Swellengrebel 1907 [16,17]. Some species form coccoid cells, have no flagella and are not motile. Some do not have L-ornithine in the peptidoglycan.

## Emended description of the genus *Sphaerochaeta*

The description of the genus *Sphaerochaeta* is as that given by Ritalahti *et al*. 2012 [45], with the following modification. DNA G+C content is 45-51 mol%.

## Description of *Sphaerochaeta coccoides* (Dröge *et al.* 2006) comb. nov.

Basonym: *Spirochaeta coccoides* Dröge *et al*. 2006.
The characteristics of the species are given in the species description by Dröge *et al*. 2006 [1].
The type strain is SPN1$^T$ (= DSM 17374 = ATCC BAA-1237).

## Acknowledgements

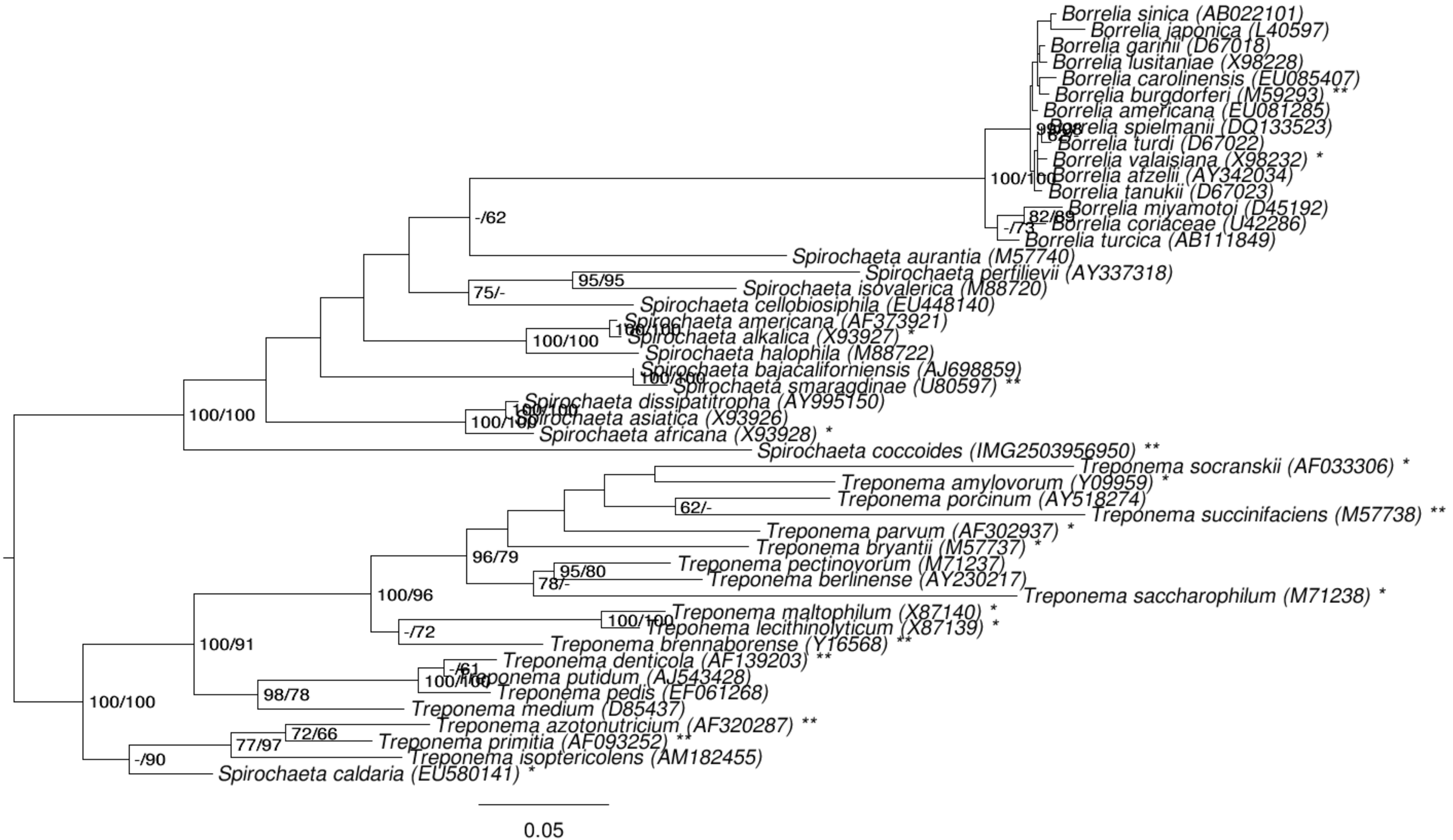
We would like to gratefully acknowledge the help of Sabine Welnitz (DSMZ) for growing *S. coccoides* cultures. This work was performed under the auspices of the US Department of Energy's Office of Science, Biological and Environmental Research Program, and by the University of California, Lawrence Berkeley 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, UT-Battelle and Oak Ridge National Laboratory under contract DE-AC05-00OR22725, as well as German Research Foundation (DFG) INST 599/1-2.

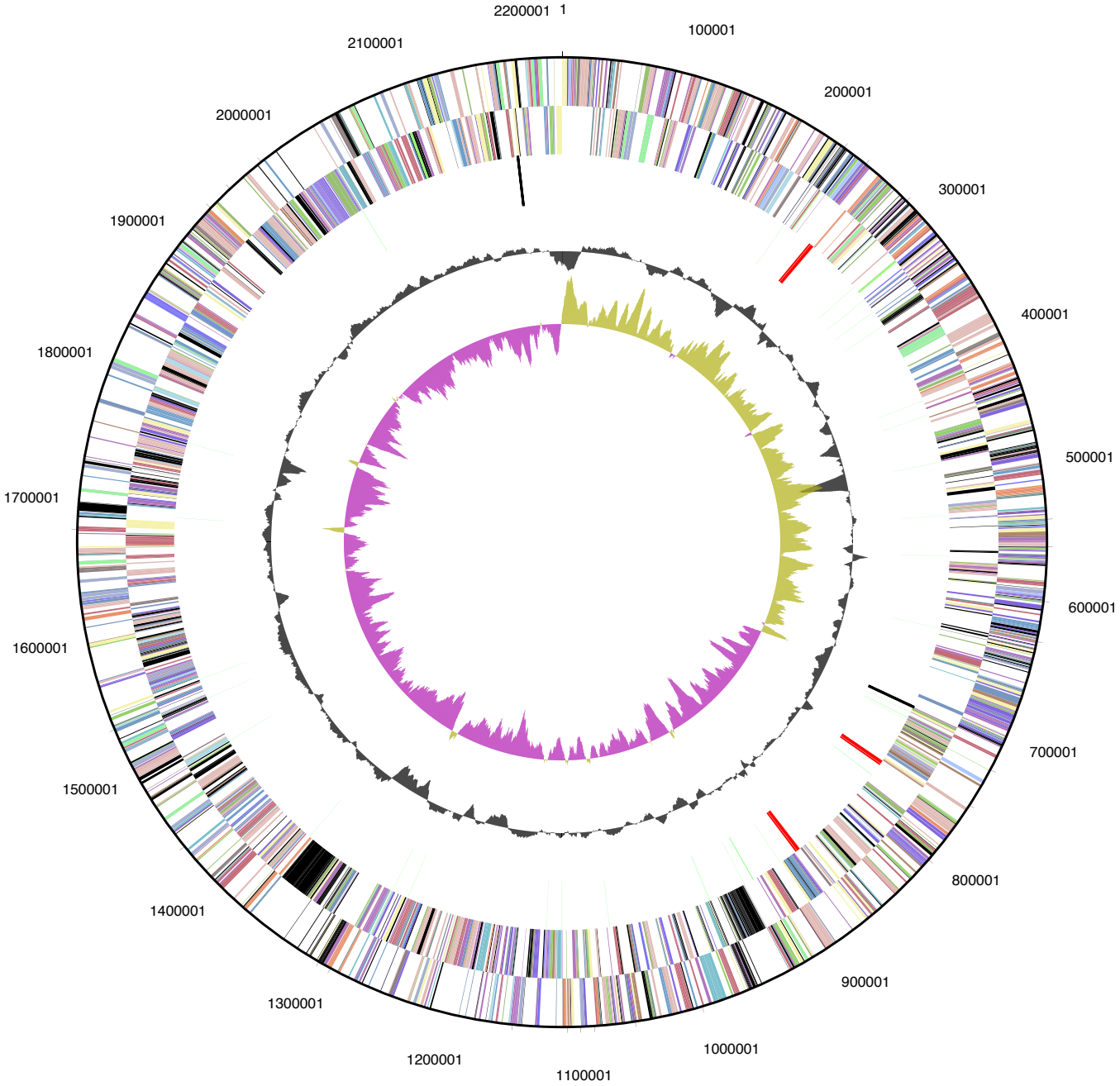
## References

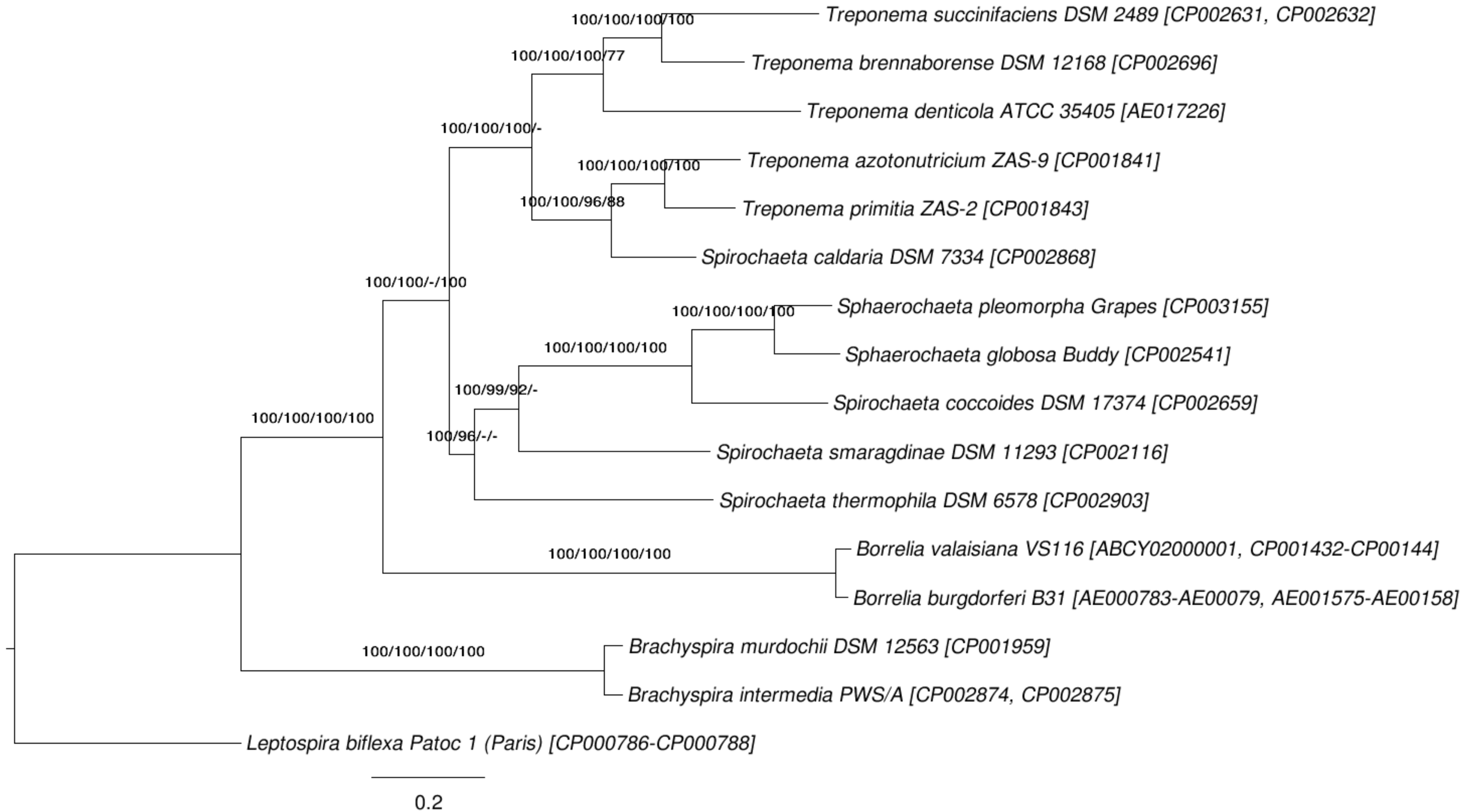
1. Dröge S, Fröhlich J, Radek R, König H. *Spirochaeta coccoides* sp. nov., a novel coccoid spirochete from the hindgut of the termite *Neotermes castaneus*. *Appl Envir Microbiol* 2006; **72**:392-397.
2. Validation List No 110. *Int J Syst Evol Microbiol* 2006; **56**:1459-1460.
3. Hess PN, De Moraes Russo CA. An empirical test of the midpoint rooting method. *Biol J Linn Soc* 2007; **92**:669-674.
4. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol* 1990; **215**: 403-410.
5. Korf I, Yandell M, Bedell J. BLAST, O'Reilly, Sebastopol, 2003
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 Env Microbiol* 2006; **72**:5069-5072.
7. Porter MF. An algorithm for suffix stripping. *Program: electronic library and information systems* 1980; **14**:130-137.
8. Castresana J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol* 2000; **17**:540-552.
9. Lee C, Grasso C, Sharlow MF. Multiple sequence alignment using partial order graphs. *Bioinformatics* 2002; **18**:452-464.
10. Stamatakis A, Hoover P, Rougemont J. A rapid bootstrap algorithm for the RAxML Web servers. *Syst Biol* 2008; **57**:758-771.
11. Pattengale ND, Alipour M, Bininda-Emonds ORP, Moret BME, Stamatakis A. How many bootstrap replicates are necessary? *Lect Notes Comp Sci* 2009; **5541**:184-200.
12. Pagani I, Liolios K, Jansson J, Chen IM, Smirnova T, Nosrat B, Markowitz VM, Kyrpides NC. The Genomes OnLine Database (GOLD) v.4: status of genomic and metagenomic projects and their associated metadata. *Nucleic Acids Res* 2012; **40**:D571-D579.
13. Pikuta EV, Hoover RB, Bej AK, Marsic D, Whitman WB, Krader P. *Spirochaeta dissipatitropha* sp. nov., an alkaliphilic, obligately anaerobic bacterium, and emended description of the genus *Spirochaeta* Ehrenberg 1835. *Int J Syst Evol Microbiol* 2009; **59**:1798-1804.
14. Canale-Parola E. Genus I. *Spirochaeta* Ehrenberg 1835, 313. *In:* Buchanan RE, Gibbons NE (*eds*), Bergey's Manual of Determinative Bacteriology, Eighth Edition, The Williams and Wilkins Co., Baltimore, 1974, p. 168-171.
15. Ehrenberg CG. Dritter Beitrag zur Erkenntniss grosser Organisation in der Richtung des kleinsten Raumes. Abhandlungen der Preussischen Akademie der Wissenschaften (Berlin), 1835, p. 143-336.
16. Swellengrebel NH. Sur la cytologie comparée des spirochètes et des spirilles. *Ann Inst Pasteur* (Paris) 1907; **21**:562-586.
17. Skerman VBD, McGowan V, Sneath PHA. Approved Lists of Bacterial Names. *Int J Syst Bacteriol* 1980; **30**:225-420.
18. Buchanan RE. Studies in the nomenclature and classification of bacteria. II. The primary subdivisions of the *Schizomycetes*. *J Bacteriol* 1917; **2**:155-164.
19. 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 Biotech* 2008; **26**:541-547.
20. 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.

21. Garrity G, Holt JG. Phylum B17 *Spirochaetes* phy. nov. Garrity and Holt. *In:* Garrity GM, Boone DR, Castenholz RW *(eds*), Bergey's Manual of Systematic Bacteriology, Second Edition, Volume 1, Springer, New York, 2001, p. 138.
22. Ludwig W, Euzeby J, Whitman WG. Draft taxonomic outline of the *Bacteroidetes, Planctomycetes, Chlamydiae, Spirochaetes, Fibrobacteres, Fusobacteria, Acidobacteria, Verrucomicrobia, Dictyoglomi*, and *Gemmatimonadetes*. http://www.bergeys.org/outlines/Bergeys_Vol_4_Outline.pdf. Taxonomic Outline 2008.
23. BAuA 2010, Classification of bacteria and archaea in risk groups. http://www.baua.de TRBA 466, p. 206.
24. 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.
25. Klenk HP, Göker M. En route to a genome-based classification of *Archaea* and *Bacteria*? *Syst Appl Microbiol* 2010; **33**:175-182.
26. 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.
27. List of growth media used at DSMZ: http://www.dsmz.de/catalogues/catalogue-microorganisms/culture-technology/list-of-media-for-microorganisms.html
28. 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. *Biopreserv Biobank* 2011; **9**:51-55.
29. Zerbino DR, Birney E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res* 2008; **18**:821-829.
30. Han C, Chain P. Finishing repeat regions automatically with Dupfinisher. *In:* Proceeding of the 2006 international conference on bioinformatics & computational biology. Arabnia HR, Valafar H (*eds*), CSREA Press. June 26-29, 2006: 141-146.
31. 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.
32. Hyatt D, Chen G-L, LoCascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 2010; **11**:119.
33. 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.
34. Markowitz VM, Ivanova NN, Chen I-MA, Chu K, Kyrpides NC. IMG ER: a system for microbial genome annotation expert review and curation. *Bioinformatics* 2009; **25:**2271-2278.
35. JGI website. http://www.jgi.doe.gov/
36. The Phred/Phrap/Consed software package. http://www.phrap.com.
37. Swofford DL. PAUP*: Phylogenetic Analysis Using Parsimony (*and Other Methods), Version 4.0 b10. Sinauer Associates, Sunderland, 2002.
38. Seshadri R, Myers GS, Tettelin H, Eisen JA, Heidelberg JF, Dodson RJ, Davidsen TM, DeBoy RT, Fouts DE, Haft DH, *et al*. Comparison of the genome of the oral pathogen *Treponema denticola* with other spirochete genomes. *Proc Natl Acad Sci USA* 2004; **101**:5646-5651.
39. Han C, Gronow S, Teshima H, Lapidus A, Nolan M, Lucas S, Hammon N, Deshpande S, Cheng JF, Zeytun A, *et al*. Complete genome sequence of *Treponema succinifaciens* type strain (6091$^{T}$). *Stand Genomic Sci* 2011; **4**:361-370.

40. Rosenthal AZ, Matson EG, Eldar A, Leadbetter JR. RNA-seq reveals cooperative metabolic interactions between two termite-gut spirochaete species in co-culture. *ISME J* 2011; *in press*
41. Mavromatis K, Yasawong M, Chertkov O, Lapidus A, Lucas S, Nolan M, Glavina Del Rio T, Tice H, Cheng JF, Pitluck S, *et al*. Complete genome sequence of *Spirochaeta smaragdinae* type strain (SEBR4228$^T$). *Stand Genomic Sci* 2010; **3**:136-144.
42. Garrity G. NamesforLife. BrowserTool takes expertise out of the database and puts it right in the browser. *Microbiol Today* 2010; **37**:9.
43. Auch AF, Von Jan M, Klenk HP, Göker M. Digital DNA-DNA hybridization for microbial species delineation by means of genome-to-genome sequence comparison. *Stand Genomic Sci* 2010; **2**:117-134.
44. Auch AF, Klenk HP, Göker M. Standard operating procedure for calculating genome-to-genome distances based on high-scoring segment pairs. *Stand Genomic Sci* 2010; **2**:142-148.
45. Ritalahti KM, Justicia-Leon SD, Cusick KD, Ramos-Hernandez N, Rubin M, Dornbush J, Löffler FE. *Sphaerochaeta globosa* gen. nov., sp. nov. and *Sphaerochaeta pleomorpha* sp. nov., free-living, spherical spirochaetes. *Int J Syst Evol Microbiol* 2012; **62**:210-216.
46. Stamatakis A, Auch AF, Meier-Kolthoff J, Göker M. AxPcoords & parallel AxParafit: statistical co-phylogenetic analyses on thousands of taxa. *BMC Bioinformatics* 2007; **8**:405.
47. Meier-Kolthoff JP, Auch AF, Huson DH, Göker M. COPYCAT: Co-phylogenetic Analysis tool. *Bioinformatics* 2007; **23**: 898-900.
48. Felsenstein J. Inferring phylogenies. Sinauer Associates Inc., Sunderland, Massachusetts 2004.
49. Euzéby JP. List of bacterial names with standing in nomenclature: A folder available on the Internet. *Int J Syst Bacteriol* 1997; **47**:590-592.
50. Yarza P, Ludwig W, Euzéby J, Amann R, Schleifer KH, Glöckner FO, Rosselló-Móra R. Update of the All-Species Living Tree Project based on 16S and 23S rRNA sequence analyses. *Syst Appl Microbiol* 2010; **33**:291-299.
51. Anderson I, Scheuner C, Göker M, Mavromatis K, Hooper SD, Porat I, Klenk H-P, Ivanova N, Kyrpides N. Novel Insights into the Diversity of Catabolic Metabolism from Ten Haloarchaeal Genomes. *PLoS ONE* 2011; **6**:e20237.
52. Li L, Stoeckert CJ, Roos DS. OrthoMCL: Identification of Ortholog Groups for Eukaryotic Genomes. *Genome Research* 2003; **13**:2178-2189.
53. van Dongen S. Graph Clustering by Flow Simulation. PhD Thesis, University of Utrecht, The Netherlands, 2000 (http://www.library.uu.nl/digiarchief/dip/diss/1895620/inhoud.htm).
54. Edgar RC MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* 2004; **32**:1792-1797.
55. Thompson JD, Thierry J-C, Poch O. RASCAL: rapid scanning and correction of multiple sequence alignments. *Bioinformatics* 2003; **19**:1155-1161.
56. Enright AJ, Van Dongen S, Ouzounis CA. An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Research* 2002; **30**:1575-1584.
57. Siddall ME, Whiting MF. Long-Branch Abstractions. *Cladistics* 1999; **15**:9-24.
58. Meusemann K, von Reumont BM, Simon S, Roeding F, Strauss S, *et al*. A Phylogenomic Approach to Resolve the Arthropod Tree of Life. *Mol Biol Evol* 2010; **27**:2451-2464.


Borrelia sinica (AB022101)
Borrelia japonica (L40597)
Borrelia garinii (D67018)
Borrelia lusitaniae (X98228)
Borrelia carolinensis (EU085407)
Borrelia burgdorferi (M59293) **
Borrelia americana (EU081285)
Borrelia spielmanii (DQ133523)
Borrelia turdi (D67022)
Borrelia valaisiana (X98232) *
Borrelia afzelii (AY342034)
Borrelia tanukii (D67023)
Borrelia miyamotoi (D45192)
Borrelia coriaceae (U42286)
Borrelia turcica (AB111849)
100/100
82/89
-/73
-/62
Spirochaeta aurantia (M57740)
Spirochaeta perfilievii (AY337318)
95/95
Spirochaeta isovalerica (M88720)
75/-
Spirochaeta cellobiosiphila (EU448140)
Spirochaeta americana (AF373921)
Spirochaeta alkalica (X93927) *
100/100
Spirochaeta halophila (M88722)
Spirochaeta bajacaliforniensis (AJ698859)
Spirochaeta smaragdinae (U80597) **
Spirochaeta dissipatitropha (AY995150)
Spirochaeta asiatica (X93926)
Spirochaeta africana (X93928) *
100/100
Spirochaeta coccoides (IMG2503956950) **
Treponema socranskii (AF033306) *
Treponema amylovorum (Y09959) *
Treponema porcinum (AY518274)
62/-
Treponema succinifaciens (M57738) **
Treponema parvum (AF302937) *
Treponema bryantii (M57737) *
96/79
Treponema pectinovorum (M71237)
95/80
Treponema berlinense (AY230217)
78/-
Treponema saccharophilum (M71238) *
100/96
Treponema maltophilum (X87140) *
Treponema lecithinolyticum (X87139) *
-/72
Treponema brennaborense (Y16568) **
100/91
Treponema denticola (AF139203) **
-/61
Treponema putidum (AJ543428)
Treponema pedis (EF061268)
98/78
Treponema medium (D85437)
100/100
Treponema azotonutricium (AF320287) **
72/66
Treponema primitia (AF093252) **
77/97
Treponema isoptericolens (AM182455)
-/90
Spirochaeta caldaria (EU580141) *
0.05


2200001
1
100001
200001
300001
400001
500001
600001
700001
800001
900001
1000001
1100001
1200001
1300001
1400001
1500001
1600001
1700001
1800001
1900001
2000001
2100001


Treponema succinifaciens DSM 2489 [CP002631, CP002632]
Treponema brennaborense DSM 12168 [CP002696]
Treponema denticola ATCC 35405 [AE017226]
Treponema azotonutricium ZAS-9 [CP001841]
Treponema primitia ZAS-2 [CP001843]
Spirochaeta caldaria DSM 7334 [CP002868]
Sphaerochaeta pleomorpha Grapes [CP003155]
Sphaerochaeta globosa Buddy [CP002541]
Spirochaeta coccoides DSM 17374 [CP002659]
Spirochaeta smaragdinae DSM 11293 [CP002116]
Spirochaeta thermophila DSM 6578 [CP002903]
Borrelia valaisiana VS116 [ABCY02000001, CP001432-CP00144]
Borrelia burgdorferi B31 [AE000783-AE00079, AE001575-AE00158]
Brachyspira murdochii DSM 12563 [CP001959]
Brachyspira intermedia PWS/A [CP002874, CP002875]
Leptospira biflexa Patoc 1 (Paris) [CP000786-CP000788]
100/100/100/100
100/100/100/77
100/100/100/-
100/100/100/100
100/100/96/88
100/100/-/100
100/100/100/100
100/100/100/100
100/99/92/-
100/100/100/100
100/96/-/-
100/100/100/100
100/100/100/100
0.2

## DISCLAIMER

This document was prepared as an account of work sponsored by the United States Government. While this document is believed to contain correct information, neither the United States Government nor any agency thereof, nor the Regents of the University of California, nor any of their employees, makes any warranty, express or implied, or assumes any legal responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by its trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or the Regents of the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof or the Regents of the University of California.