Title

Complete genome sequence of the halophilic bacterium Spirochaeta africana type strain (Z-7692T) from the alkaline Lake Magadi in the East African Rift

Journal

Environmental Microbiome, 8(2)

ISSN

1944-3277

Authors

Liolos, Konstantinos

Abt, Birte

Scheuner, Carmen

et al.

Publication Date

2013-03-01

DOI

10.4056/sigs.3607108

Peer reviewed

Complete genome sequence of the halophilic bacterium *Spirochaeta africana* type strain (Z-7692^T) from the alkaline Lake Magadi in the East African Rift

Konstantinos Liolos[1†], Birte Abt[2†], Carmen Scheuner[2], Hazuki Teshima[3], Brittany Held[3], Alla Lapidus[1], Matt Nolan[1], Susan Lucas[1], Shweta Deshpande[1], Jan-Fang Cheng[1], Roxanne Tapia[1,3], Lynne A. Goodwin[1,3], Sam Pitluck[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], Manfred Rohde[6], Brian J. Tindall[2], John C. Detter[3], Markus Göker[2], James Bristow[1], Jonathan A. Eisen[1,7], Victor Markowitz[4], Philip Hugenholtz[1,8], Tanja Woyke[1], Hans-Peter Klenk[2]*, and Nikos C. Kyrpides[1]

[1] DOE Joint Genome Institute, Walnut Creek, California, USA
[2] Leibniz Institute DSMZ - German Collection of Microorganisms and Cell Cultures, Braunschweig, Germany
[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] 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

[†]Authors contributed equally

Keywords: anaerobic, aerotolerant, mesophilic, halophilic, spiral-shaped, motile, periplasmic flagella, Gram-negative, chemoorganotrophic, *Spirochaetaceae*, GEBA.

Spirochaeta africana Zhilina *et al*. 1996 is an anaerobic, aerotolerant, spiral-shaped bacterium that is motile *via* periplasmic flagella. The type strain of the species, Z-7692^T, was isolated in 1993 or earlier from a bacterial bloom in the brine under the trona layer in a shallow lagoon of the alkaline equatorial Lake Magadi in Kenya. Here we describe the features of this organism, together with the complete genome sequence, and annotation. Considering the pending reclassification of *S. caldaria* to the genus *Treponema*, *S. africana* is only the second 'true' member of the genus *Spirochaeta* with a genome-sequenced type strain to be published. The 3,285,855 bp long genome of strain Z-7692^T with its 2,817 protein-coding and 57 RNA genes is a part of the ***G**enomic **E**ncyclopedia of **B**acteria and **A**rchaea* project.

Introduction

Strain Z-7692^T (= DSM 8902 = ATCC 700263) is the type strain of the species *Spirochaeta africana* [1]. The genus *Spirochaeta* currently consists of 18 validly named species [2]. The genus name was derived from the latinized Greek words 'speira' meaning 'a coil' and 'chaitê' meaning 'hair', yielding the Neo-Latin word 'Spirochaeta', a 'coiled hair' [2]. The species epithet was derived from the Latin word 'africana', of African continent, found in the African alkaline Lake Magadi [1]. Here we present a summary classification and a set of features for *S. africana* strain Z-7692^T, together with the description of the complete genome sequencing and annotation.

Classification and features

16S rRNA analysis

A representative genomic 16S rRNA sequence of strain Z-7692^T was compared using NCBI BLAST [3,4] 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 [5] and the relative frequencies of taxa and keywords (reduced to their stem

[6]) were determined, weighted by BLAST scores. The most frequently occurring genera were *Spirochaeta* (91.1%), *Treponema* (5.8%) and *Cytophaga* (3.1%) (29 hits in total). Regarding the two hits to sequences from members of the species, the average identity within HSPs was 99.6%, whereas the average coverage by HSPs was 99.0%. Regarding the 19 hits to sequences from other members of the genus, the average identity within HSPs was 89.1%, whereas the average coverage by HSPs was 78.9%. Among all other species, the one yielding the highest score was *Spirochaeta asiatica* (NR_026300), which corresponded to an identity of 96.6% and an HSP coverage of 98.8%. (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 AF454308 (Greengenes short name 'spirochete clone ML320J-13'), which showed an identity of 90.6% and an HSP coverage of 99.3%. The most frequently occurring keywords within the labels of all environmental samples which yielded hits were 'microbi' (10.5%), 'mat' (8.8%), 'hypersalin' (6.3%), 'new' (4.2%) and 'world' (4.1%) (221 hits in total). Environmental samples which yielded hits of a higher score than the highest scoring species were not found, indicating that this species is rarely found in environmental sequencing.

Figure 1 shows the phylogenetic neighborhood of *S. africana* in a 16S rRNA based tree. The sequences of the three identical 16S rRNA gene copies in the genome differ by two nucleotides from the previously published 16S rRNA sequence (X93928).

Morphology and physiology

Cells of strain Z-7692T are 0.25 to 0.3 µm in diameter and 15 to 30 µm (occasionally 7 to 40 µm) in length and form regular, stable primary coils [1] (Figure 2); spherical bodies were seen in stationary-phase cultures (not visible in Figure 2). The cells are motile by periplasmic flagella [1] (not visible in Figure 2). The cell mass is orange [1]. *S. africana* is a Gram-negative, anaerobic, aerotolerant, mesophilic microorganism (Table 1) with an optimal growth temperature between 30°C and 37°C, and no growth observed above 47°C [1]. The optimum pH is 8.8-9.8, no growth is observed at pH 8 or pH 10.8 [1]. *S. africana* is halophilic and does not grows at NaCl concentrations below 3% or above 10% (wt/vol) [1].

S. africana utilizes mainly mono- and disaccharides as carbon and energy sources. Amino acids cannot be fermented. Glucose is fermented to acetate, ethanol and H_2 as the main fermentation products, with a minor amount of lactate produced in stationary phase [1]. Strain Z-7692T is able to ferment fructose, maltose, trehalose, saccharose, cellobiose, glucose, glycogen, starch. Poor growth was observed with mannose and or xylose, no growth with galactose, N-acetylglucosamin or ribose. A supplement of vitamins is required [1].

Chemotaxonomy

Major components detected in the fatty acid analysis are the fatty acids $C_{14:0}$ (6.6%), $C_{16:1cis9}$ (6.3%), $C_{16:0}$ (19.0%), $C_{18:1cis-9}$ (1.4%), summed feature 10 ($C_{18:1cis11/trans9/trans6}$ and/or an unknown fatty acid with an equivalent chain length of 17.834) (34.9%), $C_{18:0}$ (1.8%), $C_{20:1cis13/trans11}$ (2.4%), as well as dimethyl acetals (DMA)/aldehydes (ALDE) probably derived from plasmalogens, $C_{14:0}$ DMA (5.0%), $C_{16:0}$ ALDE (3.8%), $C_{16:1cis-9}$ DMA (1.1%), $C_{16:0}$ DMA (15.3%), $C_{18:1cis11}$ DMA (0.8%) [35]. No data are available on polar lipid, quinone or other cell wall/envelope components that may be taxonomically significant

Taxonomic perspective

The data presented in Figure 1, based on an evaluation of the 16S rRNA gene sequence data provide an interesting insight into the nomenclature and classification of members of the genus *Spirochaeta*. In determining which species currently placed in this genus should remain members of this genus it is important to note that the primary criterion is which species group with the type strain of the type species of the genus *Spirochaeta*. It should be noted that the type species of this genus is *Spirochaeta plicatilis* and only a description serves as the type since no type strain appears to be available. This makes it difficult to determine which species represented by living type strains belong within the genus *Spirochaeta*. This is important because the monophyletic group delineated by the majority of the members of the genus *Spirochaeta* and members of the genus *Borrelia* does not split into two monophyletic groups corresponding with the members of the genus *Spirochaeta* and *Borrelia*, but causes the members of the genus *Spirochaeta* to appear to be paraphyletic. If one of the goals of modern taxonomy is to classify species in a single

genus only if the members of the genus constitute a monophyletic group, then there are three possible solutions. The first is that all members of the genus *Borrelia* should be transferred to the genus *Spirochaeta*, although this is also complicated by the fact that a type strain for the type species of the genus *Borrelia*, *Borrelia anserine* has never been designated. The second alternative would be to create a number of genera based on monophyletic groups to be found within the current analysis of members of the genus *Spirochaeta*. The third alternative would be to accept the *status quo* whereby members of the genus *Spirochaeta* appear to constitute a paraphyetic group. However, a key factor in attempting to undertake such a reclassification would be the absence of type strains of the type species of the genera *Spirochaeta* and *Borrelia*. There are already indications that the evolutionary group constituting members of the genera *Spirochaeta* and *Borrelia* show an interesting degree of diversity at the level of morphology, physiology and the genome.

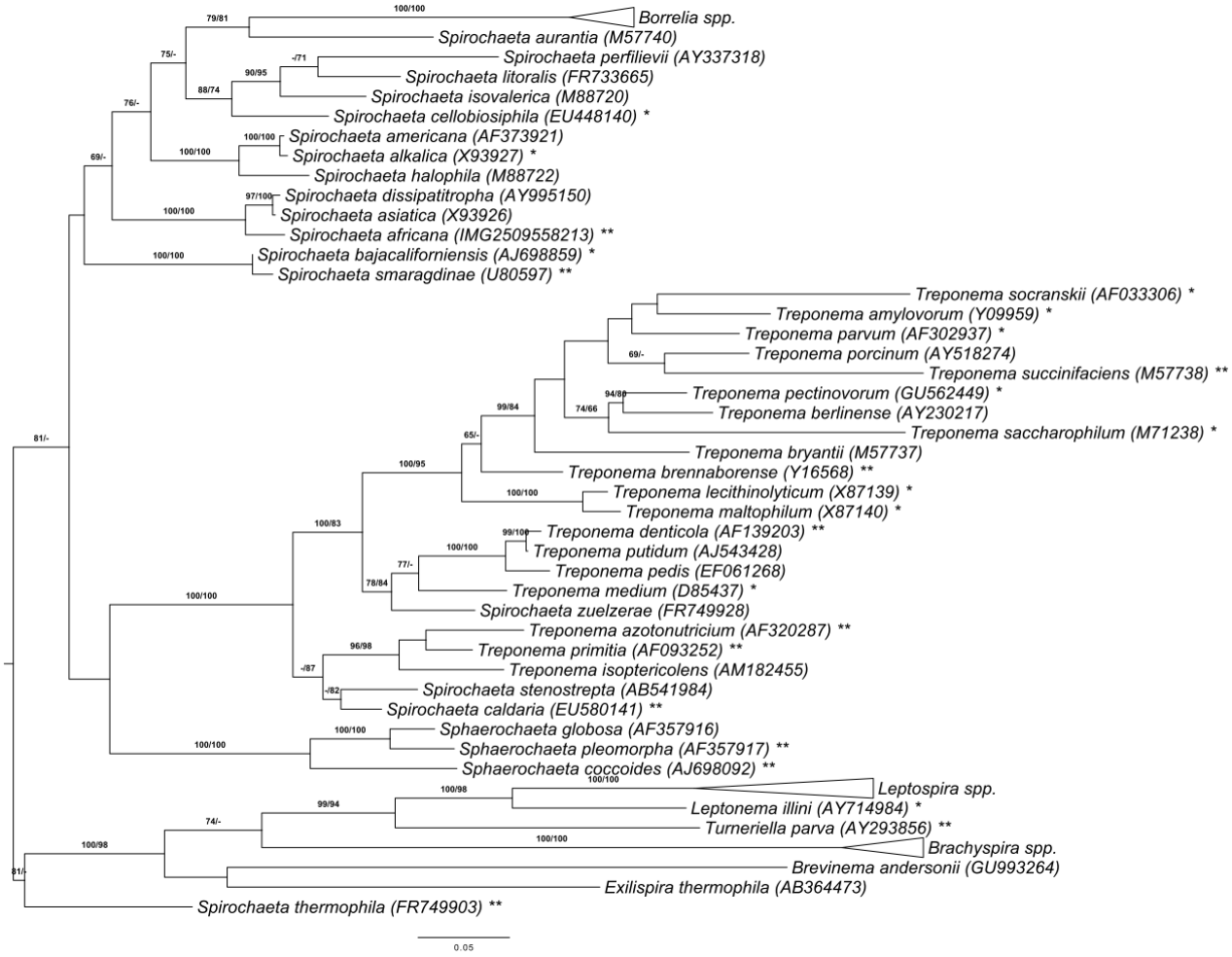

Figure 1. Phylogenetic tree highlighting the position of *S. africana* relative to the type strains of the other species within the phylum *'Spirochaetes'*. The tree was inferred from 1,332 aligned characters [7,8] of the 16S rRNA gene sequence under the maximum likelihood (ML) criterion [9]. Rooting was done initially using the midpoint method [10] 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 350 ML bootstrap replicates [11] (left) and from 1,000 maximum-parsimony bootstrap replicates [12] (right) if larger than 60%. Lineages with type strain genome sequencing projects registered in GOLD [13] are labeled with one asterisk, those also listed as 'Complete and Published' with two asterisks (see [14-20] and CP003155 for *Sphaerochaeta pleomorpha*, CP002903 for *Sphaerochaeta thermophila*, CP002696 for *Treponema brennaborense*, CP001841 for *T. azotonutricium* and CP001843 for *T. primitia*. Note: *Spirochaeta caldaria*, *S. stenostrepta and S. zuelzerae* were effectively renamed to *T. caldaria, T. stenostrepta and T. zuelzerae* in [15], however, the names have not yet been validily published.

Table 1. Classification and general features of *S. africana* Z-7692^T according to the MIGS recommendations [21] and the NamesforLife database [22].

MIGS ID	Property	Term	Evidence code
	Current classification	Domain *Bacteria*	TAS [23]
		Phylum *Spirochaetae*	TAS [24,25]
		Class *Spirochaetes*	TAS [25,26]
		Order *Spirochaetales*	TAS [27,28]
		Family *Spirochaetaceae*	TAS [27,29]
		Genus *Spirochaeta*	TAS [27,30-32]
		Species *Spirochaeta africana*	TAS [1]
		Type strain Z-7692	TAS [1]
	Gram stain	negative	TAS [1]
	Cell shape	spiral shaped	TAS [1]
	Motility	motile	TAS [1]
	Sporulation	none	TAS [1]
	Temperature range	mesophile	TAS [1]
	Optimum temperature	30 - 37°C	TAS [1]
	Salinity	halophile	TAS [1]
MIGS-22	Oxygen requirement	anaerobic, aerotolerant	TAS [1]
	Carbon source	saccharolytic, utilize carbohydrates	TAS [1]
	Energy metabolism	chemoorganotroph	TAS [1]
MIGS-6	Habitat	alkaline salt lakes, fresh water	TAS [1]
MIGS-15	Biotic relationship	free living	TAS [1]
MIGS-14	Pathogenicity	none	TAS [1]
	Biosafety level	1	TAS [33]
	Isolation	bacterial bloom in the brine under trona from alkaline lake	TAS [1]
MIGS-4	Geographic location	Lake Magadi (Kenya)	TAS [1]
MIGS-5	Sample collection time	1993 or before	NAS
MIGS-4.1	Latitude	-1.945	NAS
MIGS-4.2	Longitude	36.253	NAS
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). Evidence codes are from the Gene Ontology project [34].

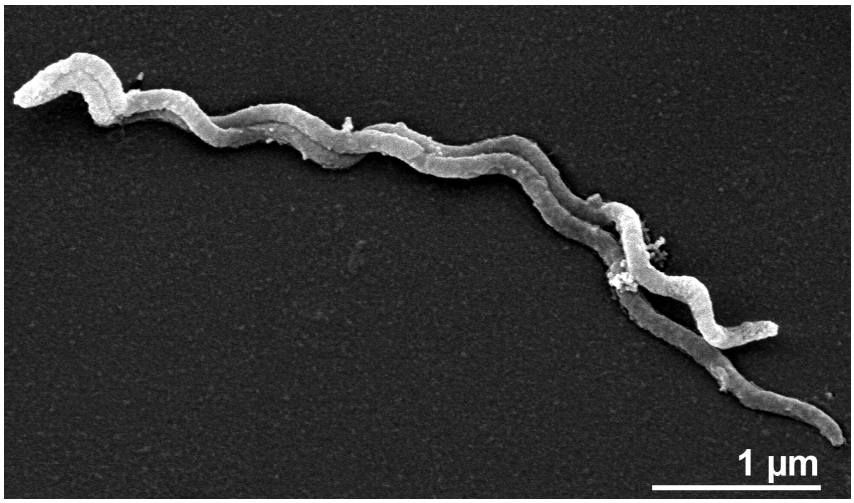

Figure 2. Scanning electron micrograph of *S. africana* strain Z-7692^T

Genome sequencing and annotation

Genome project history

This organism was selected for sequencing on the basis of its phylogenetic position [36,37], and is part of the ***G**enomic **E**ncyclopedia of **B**acteria and **A**rchaea* project [38]. The genome project is deposited in the Genomes On Line Database [13] and the complete genome sequence is deposited in GenBank. Sequencing, finishing and annotation were performed by the DOE Joint Genome Institute (JGI) using state of the art sequencing technology [39]. A summary of the project information is shown in Table 2.

Growth conditions and DNA isolation

S. africana strain Z-7692^T, DSM 8902, was grown anaerobically in DSMZ medium 700 (Alkaliphilic *Spirochaea* medium) [40] at 37°C. DNA was isolated from 0.5-1 g of cell paste using MasterPure Gram-positive DNA purification kit (Epicentre MGP04100) following the standard protocol as recommended by the manufacturer with modification st/LALM for cell lysis as described in Wu *et al.* 2009 [41]. DNA is available through the DNA Bank Network [42].

Table 2. Genome sequencing project information

MIGS ID	Property	Term
MIGS-31	Finishing quality	Finished
MIGS-28	Libraries used	Four genomic libraries: one 454 pyrosequence standard library, two 454 PE libraries (4 kb and 6 kb insert size), one Illumina library
MIGS-29	Sequencing platforms	Illumina GAii, 454 GS FLX Titanium
MIGS-31.2	Sequencing coverage	123.6 × Illumina; 23.4 × pyrosequence
MIGS-30	Assemblers	Newbler version 2.3-PreRelease-6/30/2009, Velvet 1.0.13, phrap version SPS - 4.244
MIGS-32	Gene calling method	Prodigal 1.4, GenePRIMP
	INSDC ID	CP003282
	GenBank Date of Release	April 2, 2012
	GOLD ID	Gc02193
	NCBI project ID	52939
	Database: IMG	2509276057
MIGS-13	Source material identifier	DSM 8902
	Project relevance	Tree of Life, GEBA

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 [43]. Pyrosequencing reads were assembled using the Newbler assembler (Roche). The initial Newbler assembly consisting of 511 contigs in one scaffold was converted into a phrap [44] assembly by making fake reads from the consensus, to collect the read pairs in the 454 paired end library. Illumina GAii sequencing data (459.3 Mb) was assembled with Velvet [45] and the consensus sequences were shredded into 1.5 kb overlapped fake reads and assembled together with the 454 data. The 454 draft assembly was based on 234.5 Mb 454 draft data and all of the 454 paired end data. Newbler parameters are -consed -a 50 -l 350 -g -m -ml 21. The Phred/Phrap/Consed software package [44] 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 [43], Dupfinisher [46], 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 132 additional reactions were necessary to close some gaps and to raise the quality of the final contigs. Illumina reads were also used to correct potential base errors and increase consensus quality using a software Polisher developed at JGI [47]. The error rate of the final genome sequence is less than 1 in 100,000. Together, the combination of the Illumina and 454 sequencing platforms provided 480.9 x coverage of the genome. The final assembly contained 509,107 pyrosequence and 12,708,968 Illumina reads.

Genome annotation

Genes were identified using Prodigal [48] as part of the DOE-JGI genome annotation pipeline [20], followed by a round of manual curation using the JGI GenePRIMP pipeline [49]. 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 [50].

Genome properties

The genome consists of a 3,285,855 bp long chromosome with a G+C content of 57.8% (Table 3 and Figure 3). Of the 2,874 genes predicted, 2,817 were protein-coding genes, and 57 RNAs; 35 pseudogenes were also identified. The majority of the protein-coding genes (74.2%) were assigned a putative function while the remaining ones were annotated as hypothetical proteins. The distribution of genes into COGs functional categories is presented in Table 4.

Table 3. Genome Statistics

Attribute	Value	% of Total
Genome size (bp)	3,285,855	100.00%
DNA coding region (bp)	3,080,373	93.75%
DNA G+C content (bp)	1,898,112	57.77%
Number of replicons	1	
Extrachromosomal elements	0	
Total genes	2,874	100.00%
RNA genes	57	1.98%
rRNA operons	3	
Protein-coding genes	2,817	98.02%
Pseudo genes	35	1.22%
Genes with function prediction	2,133	74.22%
Genes in paralog clusters	1,205	41.93%
Genes assigned to COGs	2,153	74.91%
Genes assigned Pfam domains	2,235	77.77
Genes with signal peptides	247	8.59
Genes with transmembrane helices	847	29.47%
CRISPR repeats	1	

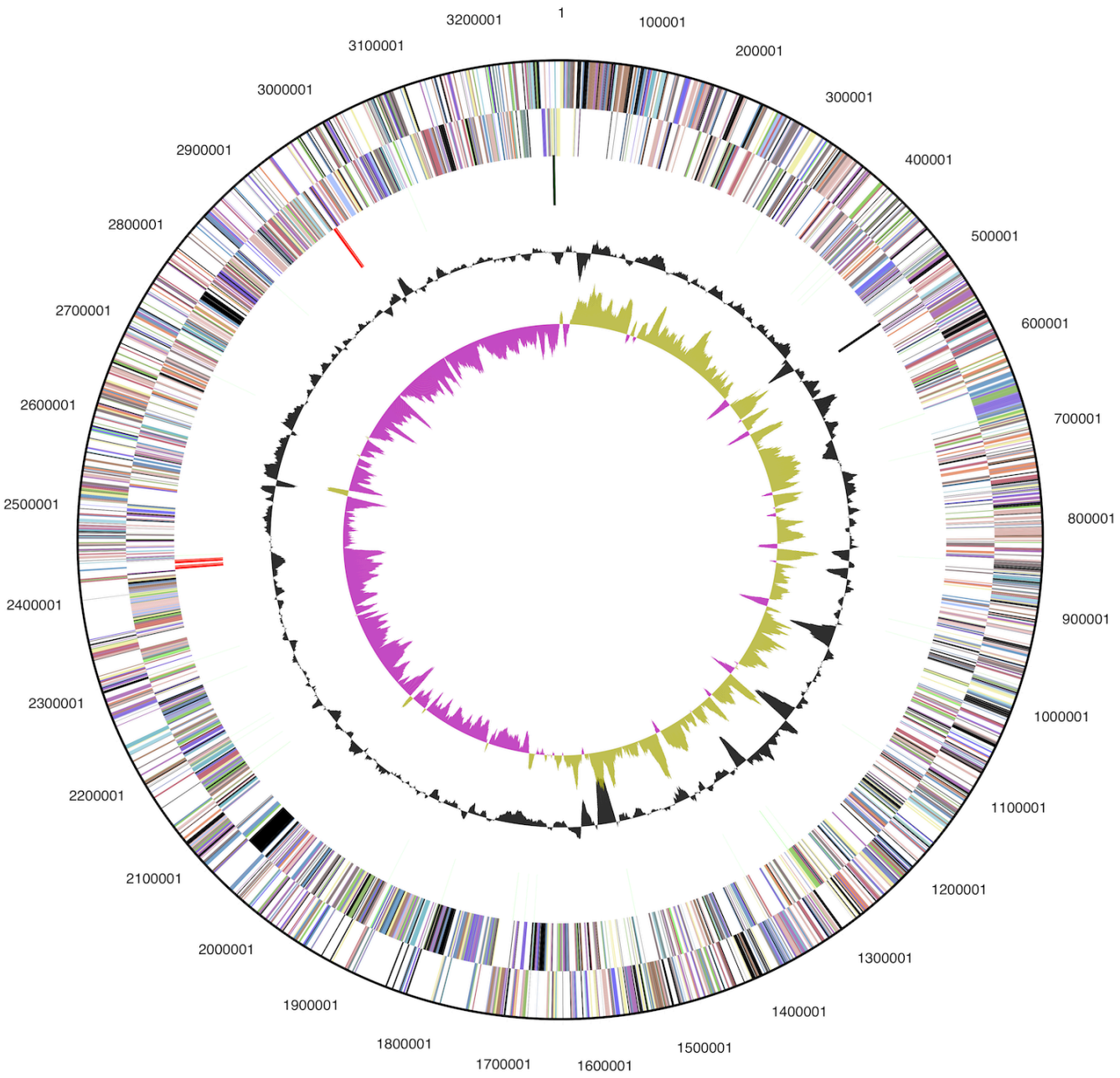

Figure 3. Graphical map of the chromosome. From outside to the center: Genes on forward strand (color by COG categories), genes on reverse strand (color by COG categories), RNA genes (tRNAs green, rRNAs red, other RNAs black), GC content (black), GC skew (purple/olive).

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

Code	Value	%age	Description
J	151	6.4	Translation, ribosomal structure and biogenesis
A	0	0.0	RNA processing and modification
K	137	5.8	Transcription
L	138	5.8	Replication, recombination and repair
B	0	0.0	Chromatin structure and dynamics
D	24	1.0	Cell cycle control, cell division, chromosome partitioning
Y	0	0.0	Nuclear structure
V	49	2.1	Defense mechanisms
T	239	10.1	Signal transduction mechanisms
M	149	6.3	Cell wall/membrane/envelope biogenesis
N	93	3.9	Cell motility
Z	0	0.0	Cytoskeleton
W	1	0.0	Extracellular structures
U	60	2.5	Intracellular trafficking, secretion, and vesicular transport
O	100	4.2	Posttranslational modification, protein turnover, chaperones
C	109	4.6	Energy production and conversion
G	185	7.8	Carbohydrate transport and metabolism
E	179	7.5	Amino acid transport and metabolism
F	62	2.6	Nucleotide transport and metabolism
H	68	2,9	Coenzyme transport and metabolism
I	57	2.4	Lipid transport and metabolism
P	103	4.3	Inorganic ion transport and metabolism
Q	23	1.0	Secondary metabolites biosynthesis, transport and catabolism
R	267	11.3	General function prediction only
S	181	7.6	Function unknown
-	721	25.1	Not in COGs

Insights from the genome sequence

Phylogenomic analyses

According to the results from 16S rRNA gene analysis (Figure 1), for a comparative analysis the genome sequences of *S. africana* (GenBank ID CP003282), *S. alkalica* (GenBank ID PRJNA169743), *S. caldaria* (CP002868) and *S. smaragdinae* (CP002116) were used. The genomes of *S. caldaria* (3.2 Mb, 2,928 protein-coding genes), *S. africana* (3.3 Mb, 2,874 protein-coding genes) and *S. alkalica* (3.4 Mb, 2,938 protein-coding genes) have a similar size, whereas the genome of *S. smaragdinae* (4.7 Mb, 4,363 protein-coding gene) is significantly larger in size. *S. caldaria* and *S. smaragdinae* have similar G+C contents, 46% and 49%, respectively. The G+C contents of *S. alkalica* and *S. africana* are significantly higher, 61% and 58%, respectively.

An estimate of the overall similarity between the genomes of *S. africana*, and those of the other *Spirochaeta* species was generated with the GGDC-Genome-to-Genome Distance Calculator [51,52]. 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 5 shows the results of the pairwise comparison.

The comparison of *S. africana* with *S. alkalica* reached the highest scores using the GGDC, 5.2% of the average of genome length are covered with HSPs. The identity within the HSPs was 86.4%, whereas the identity over the whole genome was 4.5%. Lower similarity scores were observed in the comparison of *S. africana* with *S. caldaria* and with *S. smaragdinae* only 1.62% and 1.64%, respectively, of the average of both genome lengths are covered with HSPs. The identity within these HSPs was 84.5% and 83.5%, respectively, whereas the identity over the whole genome was only 1.4% in both comparisons. *S. alkalica* shows the highest GGDC scores with *S. smaragdinae*: 2.5% of the average of genome length are covered with HSPs and the identity within the HSPs was 87.7%, whereas the identity over the whole genome was 2.2% [51].

Table 5. Pairwise comparison of *S. africana* with *S. alkalica*, *S. caldaria*, and *S. smaragdinae*, using the GGDC-Genome-to-Genome Distance Calculator.

		HSP length / total length [%]	identities / HSP length [%]	identities / total length [%]
S. africana	*S. alkalica*	5.21	86.44	4.51
S. africana	*S. caldaria*	1.62	84.50	1.37
S. africana	*S. smaragdinae*	1.64	83.52	1.37
S. smaragdinae	*S. alkalica*	2.51	87.71	2.20
S. smaragdinae	*S. caldaria*	1.52	83.91	1.28
S. caldaria	*S. alkalica*	2.08	88.57	1.85

Acknowledgements

We would like to gratefully acknowledge the help of Helga Pomrenke for growing *S. africana* cultures and Evelyne-Marie Brambilla for DNA extraction and quality control (both at DSMZ). This work was performed under the auspices of the US Department of Energy's Office of Science, Biological and Environmental Research Program, and by the University of California, Lawrence Berkeley National Laboratory under contract No. DE-AC02-05CH11231, Lawrence Livermore National Laboratory under Contract No. DE-AC52-07NA27344, and Los Alamos National Laboratory under contract No. DE-AC02-06NA25396, UT-Battelle and Oak Ridge National Laboratory under contract DE-AC05-00OR22725, as well as German Research Foundation (DFG) INST 599/1-2.

Note

IJSEM will validate the names *T. caldarium*, *T. stenostreptum* and *T. zuelzerae* in Validation List 153 (September 2013)

References

1. Zhilina TN, Zavarzin GA, Rainey F, Kevbrin VV, Kostrikina NA, Lysenko AM. *Spirochaeta alkalica* sp. nov., *Spirochaeta africana* sp. nov., and *Spirochaeta asiatica* sp. nov., alkaliphilic anaerobes from the continental soda lakes in Central Asia and the East African Rift. *Int J Syst Bacteriol* 1996; **46**:305-312. PubMed http://dx.doi.org/10.1099/00207713-46-1-305

2. Euzéby JP. List of bacterial names with standing in nomenclature: A folder available on the Internet. *Int J Syst Bacteriol* 1997; **47**:590-592. PubMed http://dx.doi.org/10.1099/00207713-47-2-590

3. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol* 1990; **215**:403-410. PubMed

4. Korf I, Yandell M, Bedell J. BLAST, O'Reilly, Sebastopol, 2003.

5. 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 compat-

ible with ARB. *Appl Environ Microbiol* 2006; **72**:5069-5072. PubMed http://dx.doi.org/10.1128/AEM.03006-05

6. Porter MF. An algorithm for suffix stripping. *Program: electronic library and information systems* 1980; **14**:130-137.

7. Lee C, Grasso C, Sharlow MF. Multiple sequence alignment using partial order graphs. *Bioinformatics* 2002; **18**:452-464. PubMed http://dx.doi.org/10.1093/bioinformatics/18.3.452

8. Castresana J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol* 2000; **17**:540-552. PubMed http://dx.doi.org/10.1093/oxfordjournals.molbev.a026334

9. Stamatakis A, Hoover P, Rougemont J. A rapid bootstrap algorithm for the RAxML web-servers. *Syst Biol* 2008; **57**:758-771. PubMed http://dx.doi.org/10.1080/10635150802429642

10. Hess PN, De Moraes Russo CA. An empirical test of the midpoint rooting method. *Biol J Linn Soc Lond* 2007; **92**:669-674. http://dx.doi.org/10.1111/j.1095-8312.2007.00864.x

11. Pattengale ND, Alipour M, Bininda-Emonds ORP, Moret BME, Stamatakis A. How Many Bootstrap Replicates Are Necessary? *Lect Notes Comput Sci* 2009; **5541**:184-200. http://dx.doi.org/10.1007/978-3-642-02008-7_13

12. Swofford DL. PAUP*: Phylogenetic Analysis Using Parsimony (*and Other Methods), Version 4.0 b10. Sinauer Associates, Sunderland, 2002.

13. 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. PubMed http://dx.doi.org/10.1093/nar/gkr1100

14. Abt B, Han C, Scheuner C, Lu M, Lapidus A, Nolan M, Lucas S, Hammon N, Deshpande S, Cheng JF, *et al*. 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*. *Stand Genomic Sci* 2012; **6**:194-209. PubMed http://dx.doi.org/10.4056/sigs.2796069

15. Abt B, Göker M, Scheuner C, Han C, Lu M, Misra M, Lapidus A, Nolan M, Lucas S, Hammon N, *et al*. Genome sequence of the thermophilic freshwater bacterium *Spirochaeta caldaria* type strain (H1^T), reclassification of *Spirochaeta caldaria* and *Spirochaeta stenostrepta* in the genus *Treponema* as *Treponema caldaria* comb. nov. and *Treponema stenostrepta* comb. nov., revival of the name *Treponema zuelzerae* comb. nov., and emendation of the genus *Treponema*. *Stand Genomic Sci* 2012;8:88-105. http://dx.doi.org/10.4056/sigs.3096473

16. Mavromatis K, Yasawong M, Chertkov O, Lapidus A, Lucas S, Nolan M, Rio TGD, Tice H, Cheng JF, Pitluck S, *et al*. Complete genome sequence of *Spirochaeta smaragdinae* type strain (SEBR 4228^T). *Stand Genomic Sci* 2010; **3**:136-144. PubMed

17. 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. PubMed http://dx.doi.org/10.4056/sigs.1984594

18. Pati A, Sikorski J, Gronow S, Lapidus A, Copeland A, Tio TGD, Nolan M, Lucas S, Chen F, Tice H, *et al*. Complete genome sequence of *Brachyspira murdochii* type strain (56-150^T). *Stand Genomic Sci* 2010; **2**:260-269. PubMed http://dx.doi.org/10.4056/sigs.831993

19. 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. PubMed http://dx.doi.org/10.1073/pnas.0307639101

20. Mavromatis K, Ivanova NN, Chen IM, Szeto E, Markowitz VM, Kyrpides NC. The DOE-JGI Standard operating procedure for the annotations of microbial genomes. *Stand Genomic Sci* 2009; **1**:63-67. PubMed http://dx.doi.org/10.4056/sigs.632

21. Field D, Garrity GM, 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 http://dx.doi.org/10.1038/nbt1360

22. Garrity GM. NamesforLife. BrowserTool takes expertise out of the database and puts it right in the browser. *Microbiol Today* 2010; **37**:9.

23. Woese CR, Kandler O, Wheelis ML. Towards a natural system of organisms. Proposal for the domains Archaea and Bacteria. *Proc Natl Acad Sci USA* 1990; **87**:4576-4579. PubMed http://dx.doi.org/10.1073/pnas.87.12.4576

24. 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.

25. Judicial Commission of the International Committee on Systematics of Prokaryotes. The nomenclatural types of the orders *Acholeplasmatales, Halanaerobiales, Halobacteriales, Methanobacteriales, Methanococcales, Methanomicrobiales, Planctomycetales, Prochlorales, Sulfolobales, Thermococcales, Thermoproteales* and *Verrucomicrobiales* are the genera *Acholeplasma, Halanaerobium, Halobacterium, Methanobacterium, Methanococcus, Methanomicrobium, Planctomyces, Prochloron, Sulfolobus, Thermococcus, Thermoproteus* and *Verrucomicrobium*, respectively. Opinion 79. *Int J Syst Evol Microbiol* 2005; **55**:517-518. PubMed http://dx.doi.org/10.1099/ijs.0.63548-0

26. 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.

27. Skerman VBD, McGowan V, Sneath PHA. Approved Lists of Bacterial Names. *Int J Syst Bacteriol* 1980; **30**:225-420. http://dx.doi.org/10.1099/00207713-30-1-225

28. Buchanan RE. Studies in the nomenclature and classification of bacteria. II. The primary subdivisions of the *Schizomycetes*. *J Bacteriol* 1917; **2**:155-164. PubMed

29. Swellengrebel NH. Sur la cytologie comparée des spirochètes et des spirilles. [Paris]. *Ann Inst Pasteur (Paris)* 1907; **21**:562-586.

30. 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. PubMed

31. 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.

32. 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.

33. BAuA. 2010, Classification of bacteria and archaea in risk groups. http://www.baua.de TRBA 466, p. 206.

34. 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 http://dx.doi.org/10.1038/75556

35. 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 Bacteriol* 2009; **59**:1798-1804. PubMed

36. Klenk HP, Göker M. *En route* to a genome-based classification of *Archaea* and *Bacteria*? *Syst Appl Microbiol* 2010; **33**:175-182. PubMed http://dx.doi.org/10.1016/j.syapm.2010.03.003

37. Göker M, Klenk HP. Phylogeny-driven target selection of large scale genome-sequencing (and other) projects. *Stand Genomic Sci* 2013; (accepted for publication).

38. 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 http://dx.doi.org/10.1038/nature08656

39. Mavromatis K, Land ML, Brettin TS, Quest DJ, Copeland A, Clum A, Goodwin L, Woyke T, Lapidus A, Klenk HP, *et al*. The fast changing landscape of sequencing technologies and their impact on microbial genome assemblies and annotation. *PLoS ONE* 2012; **7**:e48837. PubMed http://dx.doi.org/10.1371/journal.pone.0048837

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

41. 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 http://dx.doi.org/10.1038/nature08656

42. 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. http://dx.doi.org/10.1089/bio.2010.0029

43. JGI website. http://www.jgi.doe.gov/.

44. The Phred/Phrap/Consed software package. http://www.phrap.com.

45. Zerbino DR, Birney E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res* 2008; **18**:821-829. PubMed http://dx.doi.org/10.1101/gr.074492.107

46. 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.

47. 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.

48. Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal Prokaryotic Dynamic Programming Genefinding Algorithm. *BMC Bioinformatics* 2010; **11**:119. PubMed http://dx.doi.org/10.1186/1471-2105-11-119

49. Pati A, Ivanova N, Mikhailova N, Ovchinikova G, Hooper SD, Lykidis A, Kyrpides NC. GenePRIMP: A Gene Prediction Improvement Pipeline for microbial genomes. *Nat Methods* 2010; **7**:455-457. PubMed http://dx.doi.org/10.1038/nmeth.1457

50. 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 http://dx.doi.org/10.1093/bioinformatics/btp393

51. 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. PubMed http://dx.doi.org/10.4056/sigs.531120

52. 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. PubMed http://dx.doi.org/10.4056/sigs.541628