

The Chimeric Genome of *Sphaerochaeta*: Nonspiral Spirochetes That Break with the Prevalent Dogma in Spirochete Biology

A. Caro-Quintero,^a K. M. Ritalahti,^{b,c} K. D. Cusick,^b F. E. Löffler,^{b,c,d} and K. T. Konstantinidis^{a,e}

School of Biology, Georgia Institute of Technology, Atlanta, Georgia, USA^a; Department of Microbiology, University of Tennessee, Knoxville, Tennessee, USA^b; Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA^c; Department of Civil and Environmental Engineering, University of Tennessee, Knoxville, Tennessee, USA^d; and School of Civil and Environmental Engineering, Georgia Institute of Technology, Atlanta, Georgia, USA^e

ABSTRACT *Spirochaetes* is one of a few bacterial phyla that are characterized by a unifying diagnostic feature, namely, the helical morphology and motility conferred by axial periplasmic flagella. Their unique morphology and mode of propulsion also represent major pathogenicity factors of clinical spirochetes. Here we describe the genome sequences of two coccoid isolates of the recently described genus *Sphaerochaeta* which are members of the phylum *Spirochaetes* based on 16S rRNA gene and whole-genome phylogenies. Interestingly, the *Sphaerochaeta* genomes completely lack the motility and associated signal transduction genes present in all sequenced spirochete genomes. Additional analyses revealed that the lack of flagella is associated with a unique, nonrigid cell wall structure hallmarked by a lack of transpeptidase and transglycosylase genes, which is also unprecedented in spirochetes. The *Sphaerochaeta* genomes are highly enriched in fermentation and carbohydrate metabolism genes relative to other spirochetes, indicating a fermentative lifestyle. Remarkably, most of the enriched genes appear to have been acquired from nonspirochetes, particularly clostridia, in several massive horizontal gene transfer events (>40% of the total number of genes in each genome). Such a high level of direct interphylum genetic exchange is extremely rare among mesophilic organisms and has important implications for the assembly of the prokaryotic tree of life.

IMPORTANCE Spiral shape and motility historically have been the unifying hallmarks of the phylum *Spirochaetes*. These features also represent important virulence factors of highly invasive pathogenic spirochetes such as the causative agents of syphilis and Lyme disease. Through the integration of genome sequencing, microscopy, and physiological studies, we conclusively show that the spiral morphology and motility of spirochetes are not universal morphological properties. In particular, we found that the genomes of the members of the recently described genus *Sphaerochaeta* lack the genes encoding the characteristic flagellar apparatus and, in contrast to most other spirochetes, have acquired many metabolic and fermentation genes from clostridia. These findings have major implications for the isolation and study of spirochetes, the diagnosis of spirochete-caused diseases, and the reconstruction of the evolutionary history of this important bacterial phylum. The *Sphaerochaeta* sp. genomes offer new avenues to link ecophysiology with the functionality and evolution of the spirochete flagellar apparatus.

Received 25 January 2012 Accepted 18 April 2012 Published 15 May 2012

Citation Caro-Quintero A, Ritalahti KM, Cusick KD, Löffler FE, Konstantinidis KT. 2012. The chimeric genome of *Sphaerochaeta*: nonspiral spirochetes that break with the prevalent dogma in spirochete biology. *mBio* 3(3):e00025-12. doi:10.1128/mBio.00025-12.

Editor David Relman, VA Palo Alto Health Care System

Copyright © 2012 Caro-Quintero et al. This is an open-access article distributed under the terms of the Creative Commons Attribution-Noncommercial-Share Alike 3.0 Unported License, which permits unrestricted noncommercial use, distribution, and reproduction in any medium, provided the original author and source are credited.

Address correspondence to Konstantinos (Kostas) Konstantinidis, kostas@ce.gatech.edu.

Spirochaetes is a diverse, deeply branching phylum of Gram-negative bacteria. Members of this phylum share distinctive morphological features, i.e., a spiral shape and axial, periplasmic flagella (1, 2). These traits enable propulsion through highly viscous media and thus are directly associated with the ecological niches spirochetes occupy. For instance, motility mediated by axial flagella represents a major pathogenicity factor that allows strains of the *Treponema*, *Borrelia*, and *Leptospira* genera to invade and colonize host tissues, resulting in important diseases such as Lyme disease and syphilis. Several studies have shown that disruption of the flagellar genes or the chemotaxis genes that control the periplasmic flagella attenuates the pathogenic potential of spirochetes (3–5).

The focus on clinical isolates has biased our understanding of the ecology, physiology, and diversity of the phylum *Spirochaetes*. Indeed, free-living, nonpathogenic spirochetes are greatly under-

represented in culture collections, while culture-independent studies have revealed that spirochetes are ubiquitous in anoxic environments, implying that they are key players in anaerobic food webs (6–9). Consistent with the latter findings, studies of members of the genus *Sphaerochaeta* have demonstrated that environmental isolates possess physiological properties distinct from those of their pathogenic relatives; e.g., they encode a diverse set of saccharolytic enzymes (7), while other members of the genus are alkaliphiles (10) and thermophiles (11). More recently, screening of environmental samples revealed a novel genus of free-living spirochetes, *Sphaerochaeta* (12). Phylogenetic analysis of 16S rRNA genes identified this group as a member of the phylum *Spirochaetes*, most closely related to the genus *Spirochaeta*. Interestingly, *Sphaerochaeta pleomorpha* strain Grapes and *Sphaerochaeta globosa* strain Buddy are nonmotile and show the same spherical morphology during laboratory cultivation (12). How-

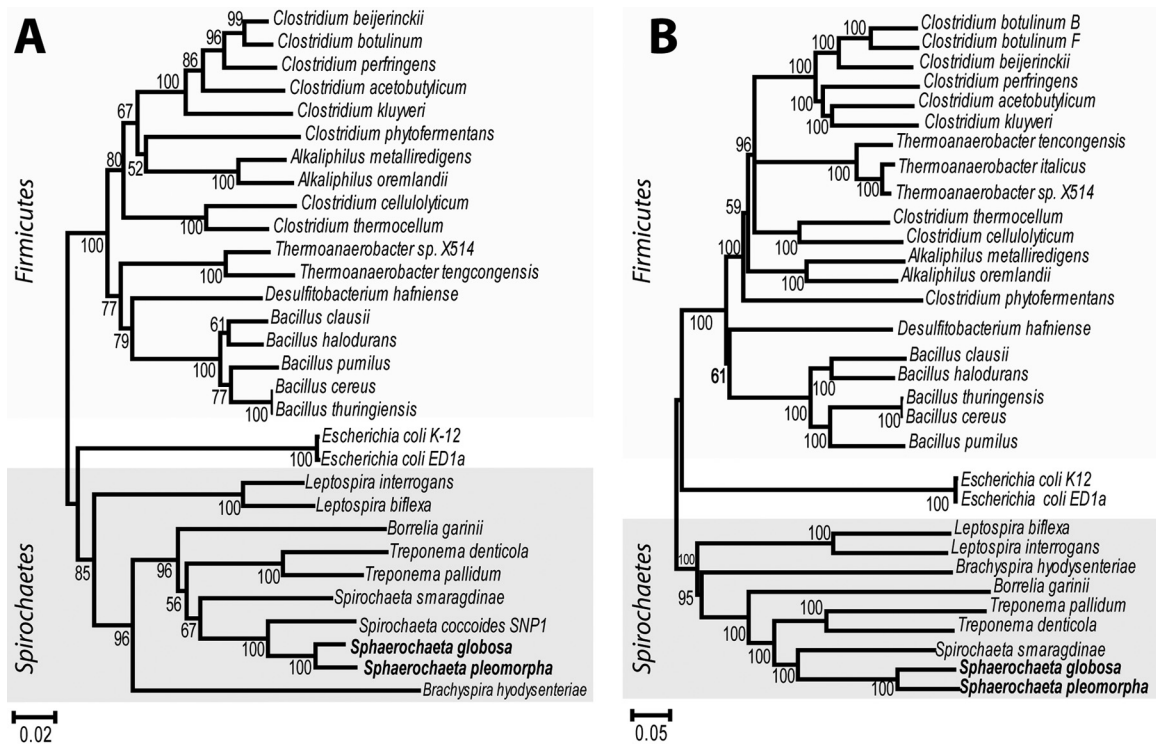


FIG 1 Phylogenetic affiliation of *S. globosa* and *S. pleomorpha*. Neighbor-joining phylogenetic trees of *Sphaerochaeta* and selected bacterial species based on 16S rRNA gene sequences (A) and the concatenated alignment of 43 single-copy informational gene sequences (B) are shown. Values at the nodes represent bootstrap support from 1,000 replicates. The scale bar represents the number of nucleotide (A) or amino acid (B) substitutions per site.

ever, it remains unclear whether this unusual morphology and the lack of motility represent a distinct stage of the cell cycle and/or responses to culture conditions or if these distinguishing features have a genetic basis.

To elucidate the metabolic properties and evolutionary history of environmental, nonpathogenic spirochetes and to provide insights into the unusual morphological features of *Sphaerochaeta*, we sequenced the genomes of strain Grapes and strain Buddy, which represent the type strains of *S. pleomorpha* and *S. globosa*, respectively. Our analyses of the two complete genome sequences suggest that the members of the genus *Sphaerochaeta* are unique spirochetes that completely lack the genes for the motility apparatus and have acquired nearly half of their genomes from Gram-positive bacteria, an extremely rare event among mesophilic organisms.

RESULTS

Phylogenetic affiliation. The *S. pleomorpha* strain Grapes and *S. globosa* strain Buddy genomes contain about 3,200 and 3,000 putative protein coding sequences and have average G+C contents of 46 and 49% and sizes of 3.5 and 3.2 Mbp, respectively (see Table S1 in the supplemental material). The two genomes share about 1,850 orthologous genes (i.e., 57 to 61% of the total number of genes in the genome, depending on the reference genome), and these genes show, on average, 65% amino acid identity. Therefore, the two genomes represent two divergent species of the genus *Sphaerochaeta* according to current taxonomic standards (13).

Phylogenetic analysis of the concatenated alignment of 43 highly conserved, single-copy informational genes (see Table S2 in the supplemental material), which showed no obvious horizontal

gene transfer (HGT) signal when their individual trees were assessed against the 16S rRNA gene tree, corroborated previous 16S rRNA gene-based findings (12). The genus *Sphaerochaeta* represents a distinct lineage of the phylum *Spirochaetes* most closely related to members of the genus *Spirochaeta*, e.g., *Spirochaeta coccoides* and *Spirochaeta smaragdinae* (Fig. 1). The average amino acid identity between *S. smaragdinae* and *S. pleomorpha* or *S. globosa* was 46% (based on 900 shared orthologous genes). This level of genomic relatedness is typically observed between organisms of different families, if not orders (14); hence, *Sphaerochaeta* and *Spirochaeta* represent distantly related genera of the phylum *Spirochaetes*. Other spirochetal genomes had fewer orthologous genes in common with *Sphaerochaeta* (i.e., 300 to 500), and these genes showed lower levels of amino acid identity than those of *S. smaragdinae* (e.g., 30 to 45%). No obvious inter- or intraphylum HGT of any of the 43 informational genes was observed when the phylogenetic analysis was expanded to include selected genomes of *Proteobacteria* and Gram-positive bacteria (see below).

Motility and chemotaxis. Typical spirochetal flagella are composed of about 30 different proteins (15), and about a dozen additional regulatory and sensory proteins have been demonstrated to interact directly with flagellar proteins, such as the methyl-accepting chemotaxis proteins encoded by the *che* operon (1). To determine whether or not the *Sphaerochaeta* genomes possess motility genes, we queried the protein sequences of the *Treponema pallidum* flagellar and chemotaxis genes against the *S. pleomorpha* and *S. globosa* genome sequences (tBLASTn). Although the *T. pallidum* sequences had clear orthologs in all available spirochetal genomes, none of the motility or chemotaxis genes were present in

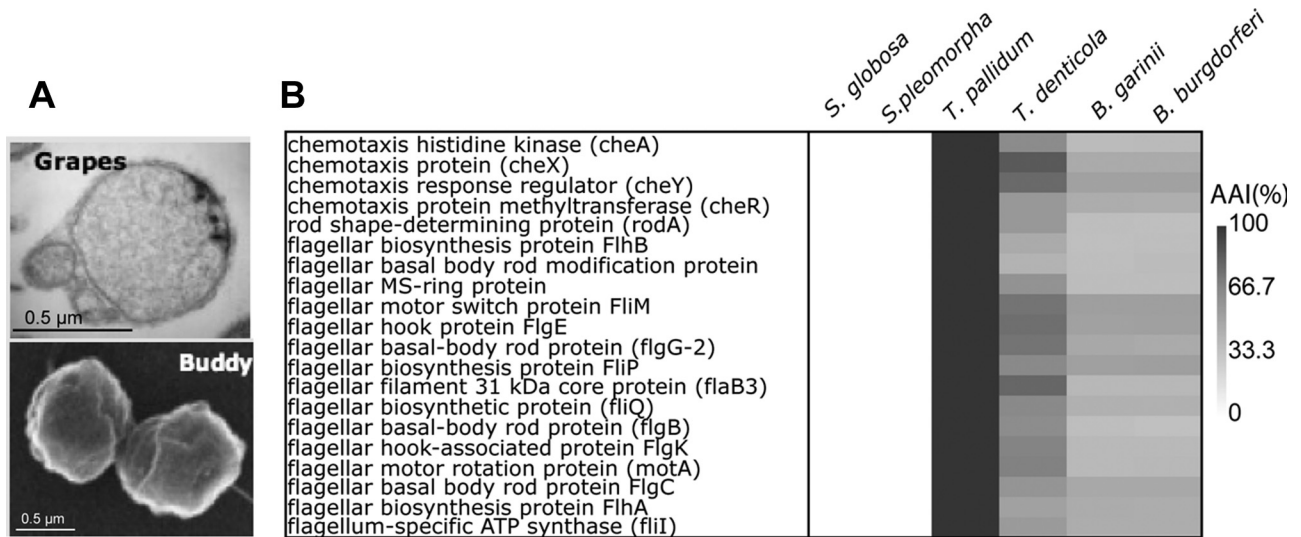


FIG 2 Absence of flagellar and chemotaxis genes from *Sphaerochaeta* genomes. (A) Transmission electron micrograph showing the nonspiral shape of *S. globosa* strain Buddy and *S. pleomorpha* strain Grapes cells. (B) Heat map showing the presence or absence and the level of amino acid identity (see scale) of *T. pallidum* chemotaxis, flagellar assembly, and locomotion gene homologs in selected spirochetal genomes.

the *S. pleomorpha* or *S. globosa* genome (Fig. 2B). Incomplete sequencing, assembly errors, or low sequence similarity did not present a plausible explanation for these results since the flagellar genes are typically located in three distinct, large gene clusters, each 20 to 30 kbp long, and it is not likely that such clusters were missed in genome sequencing and annotation. Consistent with these interpretations, all of the informational genes encoding ribosomal proteins and RNA and DNA polymerases were recovered in the assembled genome sequences. These results were consistent with previous microscopic observations and corroborated the finding that the spherical morphology characteristic of *Sphaerochaeta* is related to the absence of axial flagella (12).

A unique cell wall structure. Our analyses revealed additional features of *Sphaerochaeta* that are unusual among spirochetes and Gram-negative bacteria in general and are probably linked to the lack of axial flagella. Both *Sphaerochaeta* genomes contain all of the genes required for peptidoglycan biosynthesis, and electron microscopy verified the presence of a cell wall in growing cells (12); however, the genomes lack genes for penicillin-binding proteins (PBP). PBP catalyze the formation of linear glycan chains (transglycosylation) during cell elongation and the transpeptidation of murein glycan chains (see Table S3 in the supplemental material), which confers rigidity on the cell wall (16, 17). Consequently, *Sphaerochaeta* spp. are resistant to β -lactam antibiotics (ampicillin at up to 250 μ g/ml, which was the highest concentration tested). In Gram-negative bacteria without antibiotic resistance mechanisms, including clinical spirochetes, β -lactam antibiotics block PBP functionality, resulting in cell lysis. Often, β -lactam-treated, cell wall-deficient cells can be maintained in isotonic growth media as so-called L forms with characteristic spherical morphologies (18–20). While *Sphaerochaeta* sp. cells occur in spherical morphologies (Fig. 2A), they possess a cell wall, grow in defined hypertonic and hypotonic media without the addition of osmotic stabilizers (12), and are not L forms. It is conceivable that a rigid cell wall is required for anchoring of the axial flagella. Thus, the absence of both axial flagella and PBP genes

presumably explains the atypical spirochete morphology of the members of the genus *Sphaerochaeta*. The loss of the flagella and PBP genes likely occurred in the ancestor of *Sphaerochaeta*, since both members of the genus lack these genes.

Extensive gene acquisition from Gram-positive bacteria. Searching of all *Sphaerochaeta* protein sequences against the nonredundant (NR) protein database of GenBank revealed that ~700 of the protein-encoding genes had best matches to genes of members of the order *Clostridiales* (phylum *Firmicutes*), ~700 had best matches to genes of members of the phylum *Spirochaetes*, and ~100 had best matches to genes of members of the class *Bacilli* (see Fig. S1 in the supplemental material). Consistent with the best-match results, *S. pleomorpha* and *S. globosa* exclusively shared more unique genes with clostridia than with other members of the phylum *Spirochaetes* (~110 versus ~70 genes, respectively). Both species exclusively had a substantial number of unique genes in common with *Bacilli* (phylum *Firmicutes*, 25 genes) and *Escherichia* (*Gammaproteobacteria*, 60 and 10 genes for *S. pleomorpha* and *S. globosa*, respectively) (Fig. 3B). Functional analysis based on the COG database showed that the spirochete-like genes of *Sphaerochaeta* were associated mostly with informational categories, e.g., transcription and translation, whereas the clostridium-like genes were highly enriched in metabolic functions, e.g., carbohydrate and amino acid metabolism and transport (see Fig. S1 and S2 in the supplemental material). Several of the carbohydrate and amino acid metabolism genes, such as the multidomain glutamate synthase (SpiBuddy_0108-0113) and genes related to polysaccharide biosynthesis (SpiBuddy_0254-0259), were found in large gene clusters, indicating that their acquisition likely occurred in single HGT events. Interestingly, many of the clostridium-like genes had high sequence identity to their clostridial homologs (>70% amino acid identity), even though these genes did not encode informational proteins (e.g., ribosomal proteins and RNA/DNA polymerases). While informational genes tend to show high levels of sequence conservation, much lower sequence conservation was expected for (vertically inherited)

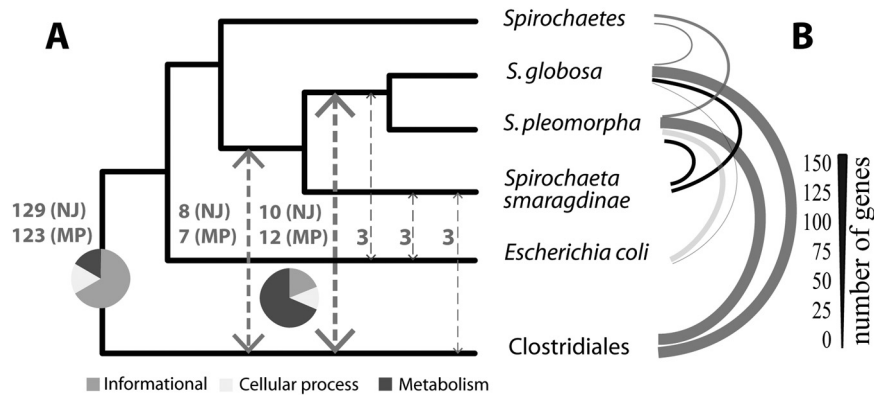


FIG 3 HGT between *Sphaerochaeta* spp. and *Clostridiales*. The cladogram depicts the 16S rRNA gene phylogeny. Arrows connecting branches represent cases of HGT (A); the values next to the arrows indicate the numbers of genes exchanged (out of a total of 178 genes examined). Pie charts show the distribution of the genes in major COG functional categories (the key at the bottom shows the category designations by color). Orthologous genes shared exclusively by *Sphaerochaeta* and other taxa are graphically represented by arcs in panel B. The thickness of each arc is proportional to the number of genes shared (see scale bar).

metabolic genes shared across phyla, revealing that some of the genetic exchange events between *Sphaerochaeta* and *Clostridiales* occurred recently relative to the divergence of the *Spirochaetes* and *Firmicutes* phyla.

Homology-based (best-hit) bioinformatic analyses are inherently prone to artifacts, including uneven numbers of representative genomes in the database, disparate G+C contents, different rates of evolution, multidomain proteins, and gene loss (21, 22). To provide further insights into the genome fluidity of *Sphaerochaeta* and the interphylum HGT events, we performed a detailed phylogenetic analysis of 223 orthologous proteins that had at least one homologous sequence in each of the taxa evaluated (i.e., *Sphaerochaeta* spp., *S. smaragdinae*, other members of the phylum *Spirochaetes*, *Escherichia coli*, and *Clostridiales*). We evaluated genetic exchange events based on embedded quartet decomposition analysis (23) by using both the maximum-parsimony (MP) and neighbor-joining (NJ) methods and 178 trees with at least 50% bootstrap support in all branches. The gene set contributing to the trees was biased toward informational functions; hence, it was not surprising that the most frequent topology obtained (123 trees [MP] and 129 trees [NJ]) was congruent with the 16S rRNA gene-based topology, denoting no interphylum genetic exchange. Nonetheless, the analysis also provided trees with topologies consistent with genetic exchange between *Clostridiales* and *Sphaerochaeta* and identified 19 (MP) and 18 (NJ) genes (i.e., ~10% of the total number of trees evaluated) that were most likely subject to interphylum HGT. This gene set was enriched in genes encoding metabolic functions, e.g., carbohydrate metabolism (Fig. 3A). About half of the 19 trees identified by MP analysis were consistent with genetic exchange between *Clostridiales* and the ancestor of both *S. smaragdinae* and *Sphaerochaeta*, while the other trees were consistent with exchange between the ancestor of *Clostridiales* and *Sphaerochaeta* (more recent events; Fig. 3). The phylogenetic distribution of the genes exchanged between *Clostridiales* and *Sphaerochaeta* in other spirochetes and Gram-positive bacteria (e.g., see Fig. S3 in the supplemental material) suggested that members of the order *Clostridiales* were predominantly the donors (>95% of the genes examined) in these genetic exchange events (unidirectional HGT). These findings corroborated those of the best-match analysis and collectively revealed that, with the exception of informational genes, interphylum HGT and gene loss

(e.g., flagellar genes) have shaped more than half of the *Sphaerochaeta* genomes through evolutionary time.

How unique is the case of *Sphaerochaeta*-*Clostridiales* gene transfer? We evaluated how frequently a high level of interphylum gene transfer such as that observed between *Clostridiales* and *Sphaerochaeta* genomes occurs within the prokaryotic domain. To this end, the ratio of the number of genes of a reference genome with best matches in a genome of a different phylum versus the number of genes of the reference genome with best matches to a genome of a member of the same phylum was determined. To account for differences in the coverage of phyla with sequenced representatives, the analysis was performed using three genomes at a time (two of the same phylum and one of a different phylum). Further, only genomes of the same phylum that showed genetic relatedness among them, measured by the genome-aggregate average amino acid identity, or gAAI (14), similar to that between *Sphaerochaeta* and selected *Spirochaetes* genomes, i.e., *Leptospira* (48% gAAI) and *Treponema* (52% gAAI) genomes, were compared. This strategy sidesteps the limitation that the number of genes common to any two genomes depends on the genetic relatedness between the genomes (see Fig. S4 in the supplemental material) (24) and thus can affect estimates of the number of best-match genes and HGT. The sets compared represented 12 different bacterial and 3 archaeal phyla and 308 and 249 different genomes (150,022 and 86,516 unique 3-genome sets) for the 48% and 52% gAAI set comparisons, respectively. The analysis revealed that the extent of genetic exchange between *Sphaerochaeta* and *Clostridiales* is highly uncommon relative to that which occurs among other genomes, i.e., the upper 99.74th and 99.99th percentiles for the 48% and 52% gAAI sets, respectively. Similar results were obtained when all of the genes in the genome or only the genes common to the three genomes, which were enriched in conserved housekeeping functions, were evaluated (Fig. 4). Most of the clostridium-like genes in *Sphaerochaeta* genomes had best matches within a phylogenetically narrow group of clostridia that included fermenters such as *Clostridium saccharolyticum* and *Clostridium phytofermentans*, which are associated with anaerobic organic matter decomposition (25), and species such as *Eubacterium rectale* (26) and *Butyrivibrio proteoclasticus* (27), which are associated with the animal gut.

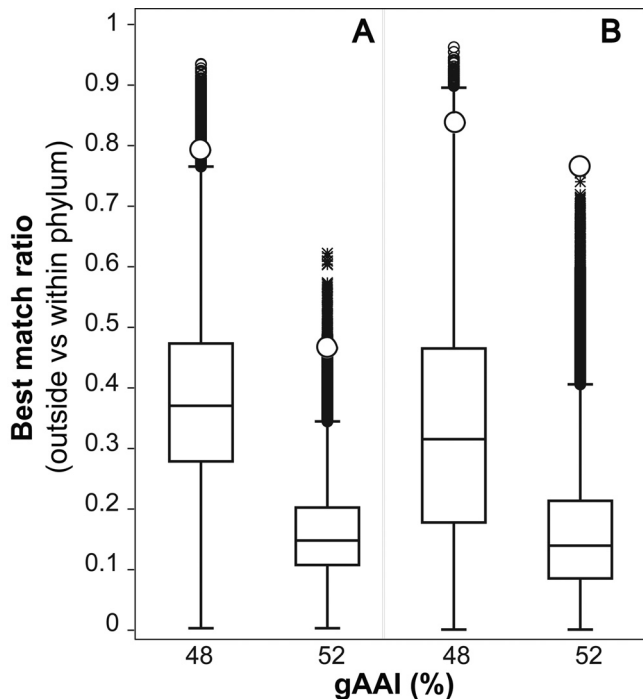


FIG 4 Comparisons of the extents of interphylum HGT. The ratio of the number of genes of a reference genome with best BLASTP matches in a genome of a different phylum relative to a genome of the same phylum as the reference genome was determined in three-genome comparisons (sets) as described in the text. The graph shows the distribution of the ratios for 150,022 and 86,516 comparisons that included genomes of the same phylum showing ~48% and ~52% gAAI, respectively; the distributions were based on all of the genes common to the three genomes in a comparison (A) and all of the genes in the reference genome (B). Horizontal bars represent the median, the upper and lower box boundaries represent the upper and lower quartiles, and the upper and lower whiskers represent the 99th percentile. Open circles represent the values for the *Sphaerochaeta-Clostridiales* case.

Metabolic properties of *Sphaerochaeta*. Metabolic genome reconstruction revealed that most of the central metabolic pathways were common to *S. pleomorpha* and *S. globosa* (Fig. 5). The complete glycolytic and pentose phosphate pathways were present in both genomes. Only a few genes for the tricarboxylic acid (TCA) cycle, such as those for citrate lyase, 2-oxoglutarate oxidoreductase, and succinate dehydrogenase, were found, suggesting an incomplete TCA cycle. A recent study of *Synechococcus* sp. strain PCC 7002, a photosynthetic cyanobacterium, identified missing cyanobacterial TCA cycle functions among the uncharacterized genes of this genome. Two proteins, encoded by the SynPCC7002_A2770 and SynPCC7002_A2771 genes, were reported to carry out the (previously) missing functions of 2-oxoglutarate decarboxylase and succinic semialdehyde dehydrogenase, respectively (28). Searches for homology between these two genes and the *Sphaerochaeta* genomes detected only one homolog, that of SynPCC7002_A2770, with 56% amino acid identity. These results indicate that the missing functions of the TCA cycle in *Sphaerochaeta* might be found among the uncharacterized genes of the genome.

Another important feature of the two genomes was the absence of key components of respiratory electron transport chains such as c-type cytochromes and the ubiquinol-cytochrome *c* reductase

(cytochrome *bc*₁ complex), corroborating physiological test findings that *Sphaerochaeta* spp. do not respire. Instead, cellular energy conservation (ATP, reducing power) in *Sphaerochaeta* relies on fermentation, a feature common to several other spirochetes lacking respiratory functions, including members of the *Spirochaeta*, *Borrelia*, and *Treponema* genera (29). In *Sphaerochaeta*, homofermentation of lactate and mixed-acid fermentation appear to be the dominant fermentation pathways, producing lactate, acetate, formate, ethanol, H₂, and CO₂, consistent with physiological observations. A few genes possibly related to alternative means of energy generation were also present in the genomes and included the *rnf* and *nqr* redox complexes. The *rnf* and *nqr* complexes export protons and/or ions (e.g., Na⁺) by coupling the flow of electrons from a reduced ferredoxin to NAD⁺ (30). This transmembrane potential can be used by V-type ATPases (e.g., SpiGrapes_0737-0742) for ATP synthesis or energize ion-dependent transporters for the uptake of sugars or amino acids.

The two *Sphaerochaeta* genomes also encode an assortment of transport proteins for the uptake and utilization of oligo- and monosaccharides. Genes involved in carbohydrate metabolism and amino acid transport and metabolism are also overrepresented relative to those in other spirochete genomes. In contrast, genes involved in signal transduction, intracellular trafficking, motility, posttranscriptional modification, and cell wall and membrane biogenesis are underrepresented in *Sphaerochaeta* genomes (see Fig. S2 in the supplemental material). Consistent with an anaerobic lifestyle (6, 9), several genes related to oxidative stress and protection from reactive oxygen species were found in the *Sphaerochaeta* genomes. Genes encoding alkyl hydroperoxide reductase, superoxide dismutase, manganese superoxide dismutase, glutaredoxin, peroxidase, and catalase indicate that *Sphaerochaeta* spp. are adapted to environments with oxidative stress fluctuations. The genome analysis provided no evidence for the formation of selenocysteine.

Each *Sphaerochaeta* genome contains about 850 species-specific genes (~25% of the genome), the majority of which have unknown or poorly characterized functions (see Fig. S5 in the supplemental material). Nevertheless, our analyses identified a few genes or pathways that can functionally differentiate the two *Sphaerochaeta* species and might have implications for the habitat distribution of each species. For example, *S. pleomorpha*-specific genes were enriched in sugar metabolism and energy production functions, including genes for trehalose and maltose utilization and the complete (TCA cycle-independent) fermentation pathway for citrate utilization (31) (green genes in Fig. 5). Further, the genome of *S. pleomorpha* uniquely contains several genes involved in cell wall and capsule formation, such as those for phosphoheptose isomerase (capsular heptose biosynthesis) and anhydro-*N*-acetylmuramic acid kinase (peptidoglycan recycling) (32). These findings revealed that *S. pleomorpha* has a potential for capsule formation and can use a wider range of carbohydrates than *S. globosa*, which are both consistent with previously reported experimental observations (12). Almost all of the *S. globosa*-specific genes have unknown or poorly characterized functions.

Bioinformatic predictions in deeply branching organisms. *Sphaerochaeta* spp. probably represent a new family or even an order within the phylum *Spirochaetes* based on their divergent genomes and unique morphological and phylogenetic features. Bioinformatic functional predictions, particularly for such deeply branching organisms, are often limited by weak sequence similar-

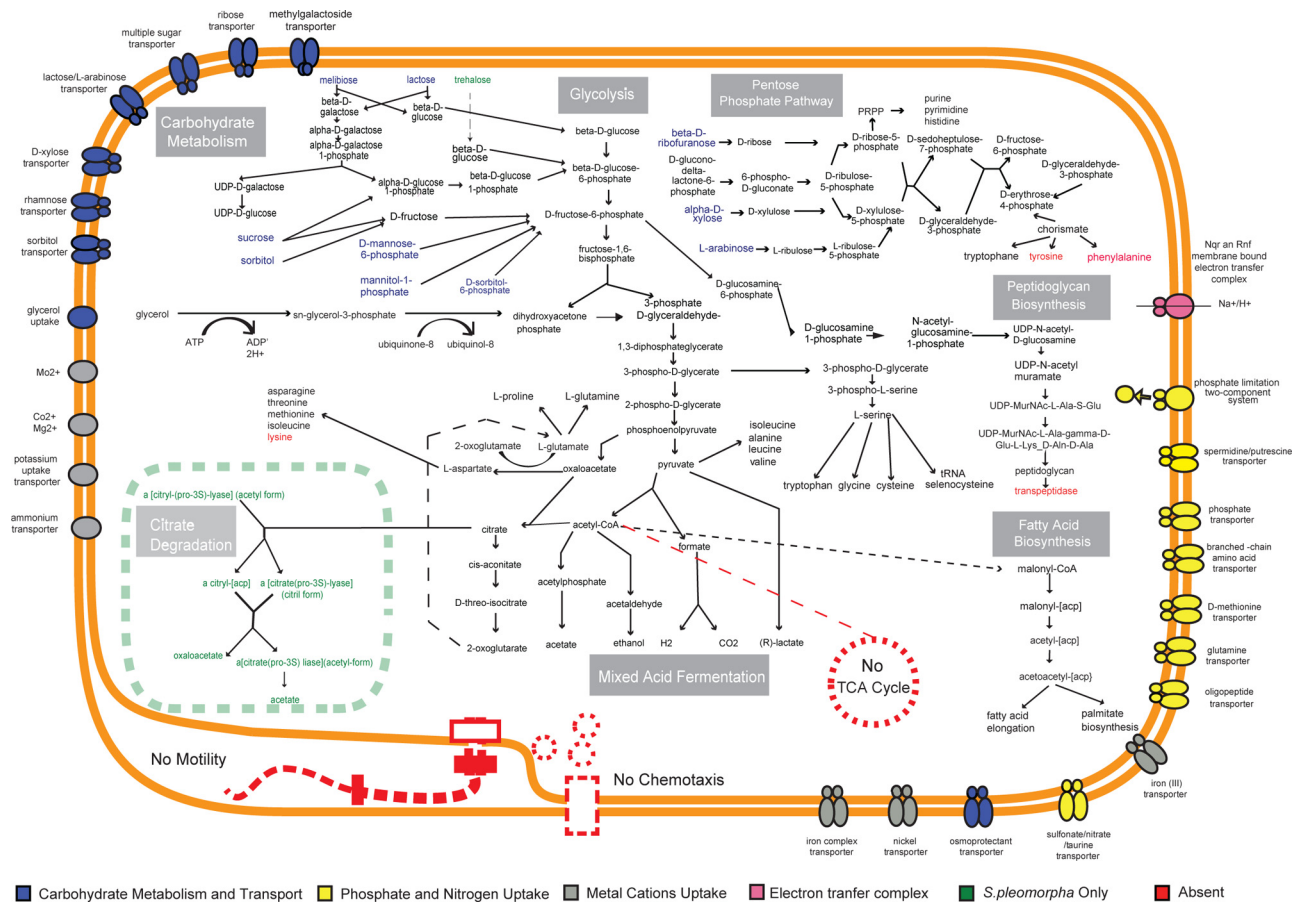


FIG 5 Overview of the metabolic pathways encoded by the *S. pleomorpha* and *S. globosa* genomes. Shown are the primary energy generation pathways, diversity of carbohydrate metabolism pathways, biosynthesis genes for amino acids and fatty acids, and cell wall features encoded by both genomes. Pathways not found in the genomes, such as those encoding flagellar and two-component signal transduction systems related to motility, are in red. The substrates and pathways found exclusively in *S. pleomorpha* are in green. Transporters related to carbohydrate metabolism (in blue), metal ion transport and metabolism (in gray), and phosphate and nitrogen uptake (in yellow) are also shown.

ity and/or uncertainty about the actual function of homologous genes or pathways. Nonetheless, bioinformatic analysis remains a powerful tool for hypothesis generation, as well as for understanding of the phenotypic differences among organisms. For the genus *Sphaerochaeta*, experimental evidence confirmed all of our bioinformatic predictions. For instance, we have confirmed experimentally (12) the predictions regarding the resistance of *Sphaerochaeta* to β -lactam antibiotics (based on the lack of PBP), utilization of various oligo- and monosaccharides, an unusual cell wall structure, absence of motility, and tolerance to oxygen. These results revealed that bioinformatic-analysis-based inferences about the metabolism and physiology of deep-branching organisms such as those in the genus *Sphaerochaeta* can be robust and reliable.

***Sphaerochaeta* and reductive dechlorination.** *Sphaerochaeta* spp. commonly co-occur with obligate organohalide respirers of the genus *Dehalococcoides* (9, 12). The reasons for this association are unclear, but it may have important practical implications for the bioremediation of chloro-organic pollutants. The *Sphaerochaeta* genomes have provided some clues and led to new hypotheses with respect to the potential interactions between free-living, nonmotile *Sphaerochaeta* spp. and *Dehalococcoides* dechlorinators. For instance, it was previously hypothesized that *Sphaero-*

chaeta may provide a corrinoid to dechlorinators, an essential cofactor for reductive dechlorination activity (33). However, genome analyses revealed that *Sphaerochaeta* genomes encode only the cobalamin salvage pathway, which is not in agreement with the corrinoid hypothesis. Alternative intriguing hypotheses include the possibility that the fermentation carried out by *Sphaerochaeta* provides essential substrates (e.g., acetate and H_2) to *Dehalococcoides* or that *Sphaerochaeta* spp. help to protect highly redox-sensitive *Dehalococcoides* cells from oxidants (i.e., oxygen) (34).

DISCUSSION

Genomic analyses revealed the absence of motility genes, the underrepresentation of sensing/regulatory genes (Fig. 2; see Fig. S2 in the supplemental material), and the unusual lack of transpeptidase and transglycosylase genes involved in cell wall formation and explained the resistance of *Sphaerochaeta* spp. to β -lactam antibiotics and their unusual cell morphology. These findings demonstrate that a spiral shape and motility are not attributes shared by all of the members of the phylum *Spirochaetes*, breaking with the prevalent dogma in spirochete biology that “spirochetes are one of the few major bacterial groups whose natural phylogenetic relationships are evident at the level of phenotypic charac-

teristics" (35). The reasons for the loss of motility genes in the members of the genus *Sphaerochaeta* are not clear, but the lack of transpeptidase activity (i.e., loss of cell wall rigidity) may have been associated with the loss of axial flagella. Cell wall rigidity is presumably necessary for anchoring of the two ends of the axial flagellum; hence, permanent loss of cell wall rigidity is likely detrimental to the proper functioning of an axial flagellum. It is also possible that habitats such as anoxic sediments enriched in organic matter and/or characterized by a constant influx of nutrients do not select for motility (36, 37) and favor the loss of genes encoding the motility apparatus; *Sphaerochaeta* spp. were obtained from such habitats (12).

Their unusual nonrigid cell wall structure likely imposes additional challenges to the maintenance of cell integrity by *Sphaerochaeta* organisms. A cellular adaptation to maintain membrane integrity, possibly accounting for the lack of a rigid cell wall, is the tight regulation of intracellular osmotic potential. Several genes encoding the biosynthesis of osmoregulating periplasmic glucans, osmoprotectant ABC transporters, an uptake system for betaine and choline, and potassium homeostasis were found in the genomes of *S. globosa* and *S. pleomorpha*, suggesting fine-tuned responses to osmotic stressors. The importance of these findings for explaining *Sphaerochaeta* sp. survival and ecological success in the environment remains to be experimentally verified.

The loss of motility genes imposes new challenges for the identification of nonmotile spirochetes in environmental or clinical samples. Free-living spirochetes are isolated routinely by selective enrichment for spiral motility, using specialized filters and/or solidified media, and by taking advantage of the unique spiral morphology, mode of propulsion, and natural rifampin resistance of spirochetes (38). Therefore, traditional isolation methods have failed to recognize and have likely underestimated the abundance and distribution of nonmotile spirochetes. New isolation procedures should be adopted to expand our understanding of the ecology and diversity of this clinically and environmentally important bacterial phylum. The genome sequences reported here will greatly assist such efforts; for instance, they have revealed that *Sphaerochaeta* spp. are naturally resistant to β -lactam antibiotics. The *Sphaerochaeta* genomes also provide a long-needed negative control (i.e., lack of axial flagella) to launch new investigations into the flagellum-mediated infection process of spirochetes causing life-threatening diseases. Further, the recently determined genome sequence of *Spirochaeta coccoides* (accession number CP002659) also lacks the flagellum, chemotaxis, and PBP genes and is more closely related to *Sphaerochaeta* than to other members of the genus *Spirochaeta* (e.g., *S. smaragdinae*). These findings indicate that, to date, nonspiral cell morphology is phylogenetically restricted to the closely related genera *Spirochaeta* and *Sphaerochaeta* within the phylum *Spirochaetes* and that *S. coccoides* may justifiably be considered a member of the genus *Sphaerochaeta*.

Our analyses revealed that more than 10% of the core genes and presumably more than 50% of the auxiliary and secondary metabolism genes of *Sphaerochaeta* were acquired from Gram-positive members of the phylum *Firmicutes*. The extensive unidirectional HGT (i.e., *Clostridiales* to *Sphaerochaeta*) implied that the two taxa (or their ancestors) have an ecological niche(s) and/or physiological properties in common. Consistent with these interpretations, ecological overlap between *Clostridiales* and both host-associated and free-living spirochetes was observed previ-

ously. For instance, several genes related to carbohydrate metabolism in *Brachyspira hyodysenteriae*, an anaerobic, commensal spirochete, appear to have been acquired from co-occurring members of the genera *Escherichia* and *Clostridium* in the porcine large intestine (29). Among free-living spirochetes, ecological overlap is likely to occur within anaerobic food webs where spirochetes and clostridia coexist (36, 39). For example, the biomass yields of and rates of cellulose degradation by *Clostridium thermo-cellum* increase when it is grown in coculture with *Spirochaeta caldaria* (40). In agreement with these studies, the genes transferred between *Sphaerochaeta* and *Clostridiales* were heavily biased toward carbohydrate uptake and fermentative metabolism functions. A more comprehensive phylogenetic analysis that included 35 spirochetal and clostridial genomes (see Table S1 in the supplemental material) indicated that *Sphaerochaeta* acquired several, but not all, of its clostridium-like genes from the ancestor of the anaerobic cellulolytic bacterium *C. phytofermentans* (see Fig. S3 in the supplemental material), which was also consistent with the BLASTP-based results of the three-genome comparisons.

Such a high level of interphylum genetic exchange is extremely rare among mesophilic organisms like *Sphaerochaeta* (Fig. 4) (41). This level of HGT has been reported previously only for thermophilic *Thermotoga* spp. (i.e., organisms living under extreme environmental selection pressures) (42). On the other hand, we did not observe HGT that affected informational proteins such as ribosomal proteins and DNA/RNA polymerases, suggesting that the reconstruction of spirochetal phylogenetic relationships, and in general, the construction of the bacterial tree of life, can be attained even in cases of extensive genetic exchange of metabolic genes (for a contrasting opinion, see reference 43). In the case of *Sphaerochaeta*, the massive HGT was apparently favored by an ecological niche overlap with *Clostridiales* and/or strong functional interactions within anoxic environments. The altered, nonrigid cell wall structure of *Sphaerochaeta* might have played a role in the high level of genetic exchange observed, e.g., by facilitating DNA transfer across the cell wall, although experimental evidence for this hypothesis is lacking. These findings highlight the importance of both ecology and environment in determining the rates and magnitudes of HGT. The acquisition of quantitative insights into the role of the environment and shared ecological niches in HGT will lead to a more educated assembly of the prokaryotic tree of life based on measurable and quantifiable properties.

MATERIALS AND METHODS

Organisms used in this study. The genome sequence of each *Sphaerochaeta* species used in this study is shown in Table S1 in the supplemental material. The accession numbers of the genomes are CP003155 (*S. pleomorpha*) and CP002541 (*S. globosa*). Details regarding the conditions used to isolate type species are available elsewhere (12).

Sequence analysis and metabolic reconstruction. Orthologous proteins between *Sphaerochaeta* and selected publicly available genomes were identified by a reciprocal best-match approach and a minimum cutoff for a match of 70% coverage of the query sequence and 30% amino acid identity, as described previously (44). For phylogenetic analysis, sequence alignments were constructed using the ClustalW software (45) and trees were built using the NJ algorithm as implemented in the MEGA4 package (46). Central metabolic pathways were reconstructed using Pathway Tools version 14 (47). The annotation files required as the input to Pathway Tools were prepared from the consensus results of two approaches. First, amino acid sequences of predicted proteins were annotated based on their best BLAST matches against the NR (48), KEGG (49), and COG (50)

databases. Second, the whole-genome sequences were submitted to the RAST annotation pipeline (51) to ensure that the previous approach did not miss any important genes and to assign protein sequences to functions and enzymatic reactions (EC numbers). The results of both approaches were used to extract gene names and EC numbers. Disagreements between the two approaches were resolved by manual curation.

HGT analysis. For best-match analysis, strain Buddy protein sequences were searched for using BLASTP against two databases, (i) all completed prokaryotic genomes available in January 2011 ($n = 1,445$) and (ii) the NR database (release 178). The best match for each query sequence, with better than 70% coverage of the length of the query protein and 30% amino acid identity, was identified, and the taxonomic affiliation of the genome containing the best match was extracted from the NCBI taxonomy browser. HGT events were identified as follows. Orthologous protein sequences present in at least one representative genome from the five groups used (i.e., *Sphaerochaeta*, *S. smaragdinae*, other spirochetes, *Clostridiales*, and *E. coli*) were identified and aligned as described above. Phylogenetic trees for each alignment were built in PHYLIP v3.6 (J. Felsenstein, University of Washington, Seattle, WA [<http://evolution.genetics.washington.edu/phylip.html>]) by using both the MP and NJ algorithms and bootstrapped 100 times using Seqboot. The topology of the resulting consensus tree was compared to the 16S rRNA gene-based tree topology, and conflicting nodes between the two trees which also had bootstrap support higher than 50 were identified as cases of HGT.

To evaluate how unique the case of interphylum gene transfer between and *Sphaerochaeta* is, the following approach was used. All of the available completed bacterial and archaeal genomes (as of January 2011, $n = 1,445$) that showed genetic relatedness among them similar to the relatedness among the *Sphaerochaeta* genomes (i.e., $65\% \pm 0.5\%$ gAAI) were assigned to the same group. All protein-coding genes common to the genomes of different groups were subsequently determined by using the BLASTP algorithm as described above. The BLASTP results were analyzed by using sets of three genomes at a time, each genome representing one of three distinct groups, (i) a reference group, (ii) a group from the same phylum as the reference group, or (iii) a group from another phylum. The ratio of the number of genes of the reference genome with best matches in the genome of the different phylum versus the number of genes in the reference genome with best matches to the genome of the same phylum was determined for each set and plotted against the gAAI between the reference genome and the genome of the same phylum (Fig. 4). Groups of genomes with fewer than 40 genes in common were removed from further analysis to reduce noisy results from very distantly related or small genomes.

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at <http://mbio.asm.org/lookup/suppl/doi:10.1128/mBio.00025-12/-/DCSupplemental>.

Figure S1, DOCX file, 0.1 MB.
 Figure S2, DOCX file, 0.2 MB.
 Figure S3, DOCX file, 0.2 MB.
 Figure S4, DOCX file, 0.1 MB.
 Figure S5, DOCX file, 0.1 MB.
 Table S1, DOCX file, 0.2 MB.
 Table S2, DOCX file, 0.6 MB.
 Table S3, DOCX file, 0.1 MB.

ACKNOWLEDGMENTS

We thank the DOE Joint Genome Institute, particularly the people associated with the Community Sequencing Program, for sequencing and assembling the *Sphaerochaeta* genomes.

This work was supported by the National Science Foundation under grant 0919251.

REFERENCES

- Charon NW, Goldstein SF. 2002. Genetics of motility and chemotaxis of a fascinating group of bacteria: the Spirochetes. *Annu. Rev. Genet.* 36: 47–73.
- Paster BJ, Dewhirst FE. 2000. Phylogenetic foundation of Spirochetes. *J. Mol. Microbiol. Biotechnol.* 2:341–344.
- Lux R, Miller JN, Park NH, Shi W. 2001. Motility and chemotaxis in tissue penetration of oral epithelial cell layers by *Treponema denticola*. *Infect. Immun.* 69:6276–6283.
- Rosey EL, Kennedy MJ, Yancey RJ, Jr. 1996. Dual *flaA1 flaB1* mutant of *Serpulina hyodysenteriae* expressing periplasmic flagella is severely attenuated in a murine model of swine dysentery. *Infect. Immun.* 64: 4154–4162.
- Sadziene A, Thomas DD, Bundoc VG, Holt SC, Barbour AG. 1991. A flagella-less mutant of *Borrelia burgdorferi*. Structural, molecular, and *in vitro* functional characterization. *J. Clin. Invest.* 88:82–92.
- Franzmann PD, Dobson SJ. 1992. Cell wall-less, free-living spirochetes in Antarctica. *FEMS Microbiol. Lett.* 76:289–292.
- Leschine S, Paster B, Canale-Parola E. 2006. Free-living saccharolytic spirochetes: the genus “*Spirochaeta*,” p 195–210. *In The Prokaryotes*. Springer Verlag, New York, NY.
- Magot M, et al. 1997. *Spirochaeta smaragdinae* sp. nov., a new mesophilic strictly anaerobic spirochete from an oil field. *FEMS Microbiol. Lett.* 155: 185–191.
- Ritalahti KM, Löffler FE. 2004. Populations implicated in anaerobic reductive dechlorination of 1,2-dichloropropane in highly enriched bacterial communities. *Appl. Environ. Microbiol.* 70:4088–4095.
- Zhilina TN, et al. 1996. *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.* 46:305–312.
- Janssen PH, Morgan HW. 1992. Glucose catabolism by *Spirochaeta thermophila* RI 19.B1. *J. Bacteriol.* 174:2449–2453.
- Ritalahti KM, et al. 2012. Isolation of *Sphaerochaeta* (gen. nov.), free-living, spherical spirochetes, and characterization of *Sphaerochaeta pleomorpha* (sp. nov.) and *Sphaerochaeta globosa* (sp. nov.). *Int. J. Syst. Evol. Microbiol.* 62:210–216.
- Goris J, et al. 2007. DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. *Int. J. Syst. Evol. Microbiol.* 57:81–91.
- Konstantinidis KT, Tiedje JM. 2005. Towards a genome-based taxonomy for prokaryotes. *J. Bacteriol.* 187:6258–6264.
- Liu R, Ochman H. 2007. Stepwise formation of the bacterial flagellar system. *Proc. Natl. Acad. Sci. U. S. A.* 104:7116–7121.
- Mattei PJ, Neves D, Dessen A. 2010. Bridging cell wall biosynthesis and bacterial morphogenesis. *Curr. Opin. Struct. Biol.* 20:749–755.
- Sauvage E, Kerff F, Terrak M, Ayala JA, Charlier P. 2008. The penicillin-binding proteins: structure and role in peptidoglycan biosynthesis. *FEMS Microbiol. Rev.* 32:234–258.
- Allan EJ, Hoischen C, Gumpert J. 2009. Bacterial L-forms. *Adv. Appl. Microbiol.* 68:1–39.
- Mursic VP, et al. 1996. Formation and cultivation of *Borrelia burgdorferi* spheroplast-L-form variants. *Infection* 24:218–226.
- Wise EM, Jr, Park JT. 1965. Penicillin: its basic site of action as an inhibitor of a peptide cross-linking reaction in cell wall mucopeptide synthesis. *Proc. Natl. Acad. Sci. U. S. A.* 54:75–81.
- Eisen JA. 2000. Horizontal gene transfer among microbial genomes: new insights from complete genome analysis. *Curr. Opin. Genet. Dev.* 10: 606–611.
- Koski LB, Golding GB. 2001. The closest BLAST hit is often not the nearest neighbor. *J. Mol. Evol.* 52:540–542.
- Zhaxybayeva O, Gogarten JP, Charlebois RL, Doolittle WF, Papke RT. 2006. Phylogenetic analyses of cyanobacterial genomes: quantification of horizontal gene transfer events. *Genome Res.* 16:1099–1108.
- Konstantinidis KT, Tiedje JM. 2005. Genomic insights that advance the species definition for prokaryotes. *Proc. Natl. Acad. Sci. U. S. A.* 102: 2567–2572.
- Warnick TA, Methé BA, Leschine SB. 2002. *Clostridium phytofermentans* sp. nov., a cellulolytic mesophile from forest soil. *Int. J. Syst. Evol. Microbiol.* 52:1155–1160.
- Mahowald MA, et al. 2009. Characterizing a model human gut microbiota composed of members of its two dominant bacterial phyla. *Proc. Natl. Acad. Sci. U. S. A.* 106:5859–5864.
- Moon CD, et al. 2008. Reclassification of *Clostridium proteoclasticum* as *Butyrivibrio proteoclasticus* comb. nov., a butyrate-producing ruminal bacterium. *Int. J. Syst. Evol. Microbiol.* 58:2041–2045.

28. Zhang S, Bryant DA. 2011. The tricarboxylic acid cycle in cyanobacteria. *Science* 334:1551–1553.
29. Bellgard MI, et al. 2009. Genome sequence of the pathogenic intestinal spirochete *Brachyspira hyodysenteriae* reveals adaptations to its lifestyle in the porcine large intestine. *PLoS One* 4:e4641.
30. Biegel E, Schmidt S, González JM, Müller V. 2011. Biochemistry, evolution and physiological function of the Rnf complex, a novel ion-motive electron transport complex in prokaryotes. *Cell. Mol. Life Sci.* 68: 613–634.
31. Bott M. 1997. Anaerobic citrate metabolism and its regulation in enterobacteria. *Arch. Microbiol.* 167:78–88.
32. Uehara T, et al. 2005. Recycling of the anhydro-*N*-acetylmuramic acid derived from cell wall murein involves a two-step conversion to *N*-acetylglucosamine-phosphate. *J. Bacteriol.* 187:3643–3649.
33. He J, Holmes VF, Lee PK, Alvarez-Cohen L. 2007. Influence of vitamin B₁₂ and cocultures on the growth of *Dehalococcoides* isolates in defined medium. *Appl. Environ. Microbiol.* 73:2847–2853.
34. Morris JJ, Kirkegaard R, Szul MJ, Johnson ZI, Zinser ER. 2008. Facilitation of robust growth of *Prochlorococcus* colonies and dilute liquid cultures by “helper” heterotrophic bacteria. *Appl. Environ. Microbiol.* 74: 4530–4534.
35. Paster BJ, et al. 1991. Phylogenetic analysis of the spirochetes. *J. Bacteriol.* 173:6101–6109.
36. Canale-Parola E. 1978. Motility and chemotaxis of spirochetes. *Annu. Rev. Microbiol.* 32:69–99.
37. Kimsey RB, Spielman A. 1990. Motility of lyme disease spirochetes in fluids as viscous as the extracellular matrix. *J. Infect. Dis.* 162:1205–1208.
38. Breznak JA, Canale-Parola E. 1975. Morphology and physiology of *Spirochaeta aurantia* strains isolated from aquatic habitats. *Arch. Microbiol.* 105:1–12.
39. Harwood CS, Canale-Parola E. 1984. Ecology of spirochetes. *Annu. Rev. Microbiol.* 38:161–192.
40. Leschine SB. 1995. Cellulose degradation in anaerobic environments. *Annu. Rev. Microbiol.* 49:399–426.
41. Beiko RG, Harlow TJ, Ragan MA. 2005. Highways of gene sharing in prokaryotes. *Proc. Natl. Acad. Sci. U. S. A.* 102:14332–14337.
42. Zhaxybayeva O, et al. 2009. On the chimeric nature, thermophilic origin, and phylogenetic placement of the *Thermotogales*. *Proc. Natl. Acad. Sci. U. S. A.* 106:5865–5870.
43. Doolittle WF, Baptiste E. 2007. Pattern pluralism and the tree of life hypothesis. *Proc. Natl. Acad. Sci. U. S. A.* 104:2043–2049.
44. Konstantinidis KT, et al. 2009. Comparative systems biology across an evolutionary gradient within the *Shewanella* genus. *Proc. Natl. Acad. Sci. U. S. A.* 106:15909–15914.
45. Thompson JD, Gibson TJ, Higgins DG. 2002. Multiple sequence alignment using ClustalW and ClustalX. *Curr. Protoc. Bioinformatics Chapter 2:Unit 2.3.*
46. Tamura K, Dudley J, Nei M, Kumar S. 2007. MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* 24:1596–1599.
47. Paley SM, Karp PD. 2006. The pathway tools cellular overview diagram and omics viewer. *Nucleic Acids Res.* 34:3771–3778.
48. Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Wheeler DL. 2007. GenBank. *Nucleic Acids Res.* 35:D21–D25.
49. Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M. 2007. KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res.* 35:W182–W185.
50. Tatusov R, et al. 2003. The COG database: an updated version includes eukaryotes. *BMC Bioinform.* 4:41.
51. Aziz RK, et al. 2008. The RAST server: rapid annotations using subsystems technology. *BMC Genomics* 9:75.