

# A phylogenomic and molecular signature based approach for characterization of the phylum Spirochaetes and its major clades: proposal for a taxonomic revision of the phylum

# Radhey S. Gupta \*, Sharmeen Mahmood and Mobolaji Adeolu

Department of Biochemistry and Biomedical Sciences, McMaster University, Hamilton, ON, Canada

### Edited by:

Hiromi Nishida, Toyama Prefectural University, Japan

#### Reviewed by:

Viktoria Shcherbakova, Institute of Biochemistry and Physiology of Microorganisms, Russian Academy of Sciences, Russia David L. Bernick, University of California, Santa Cruz, USA

#### \*Correspondence:

Radhey S. Gupta, Department of Biochemistry and Biomedical Sciences, McMaster University, 1280 Main Street West, Hamilton, ON L8N 325, Canada e-mail: gupta@mcmaster.ca The Spirochaetes species cause many important diseases including syphilis and Lyme disease. Except for their containing a distinctive endoflagella, no other molecular or biochemical characteristics are presently known that are specific for either all Spirochaetes or its different families. We report detailed comparative and phylogenomic analyses of protein sequences from Spirochaetes genomes to understand their evolutionary relationships and to identify molecular signatures for this group. These studies have identified 38 conserved signature indels (CSIs) that are specific for either all members of the phylum Spirochaetes or its different main clades. Of these CSIs, a 3 aa insert in the FIgC protein is uniquely shared by all sequenced Spirochaetes providing a molecular marker for this phylum. Seven, six, and five CSIs in different proteins are specific for members of the families Spirochaetaceae, Brachyspiraceae, and Leptospiraceae, respectively. Of the 19 other identified CSIs, 3 are uniquely shared by members of the genera Sphaerochaeta, Spirochaeta, and Treponema, whereas 16 others are specific for the genus Borrelia. A monophyletic grouping of the genera Sphaerochaeta, Spirochaeta, and *Treponema* distinct from the genus *Borrelia* is also strongly supported by phylogenetic trees based upon concatenated sequences of 22 conserved proteins. The molecular markers described here provide novel and more definitive means for identification and demarcation of different main groups of Spirochaetes. To accommodate the extensive genetic diversity of the Spirochaetes as revealed by different CSIs and phylogenetic analyses, it is proposed that the four families of this phylum should be elevated to the order level taxonomic ranks (viz. Spirochaetales, Brevinematales ord. nov., Brachyspiriales ord. nov., and Leptospiriales ord. nov.). It is further proposed that the genera Borrelia and Cristispira be transferred to a new family Borreliaceae fam. nov. within the order Spirochaetales.

Keywords: Spirochaetes, Spirochaetes phylogeny and taxonomy, molecular signatures, *Spirochaetaceae, Borreliaceae, Brachyspiriales, Leptospiriales,* conserved signature indels

## **INTRODUCTION**

The phylum Spirochaetes consists of a large group of motile bacteria which are widespread in the environment and are highly prevalent disease causing agents (Seshadri et al., 2004; Paster, 2011a). The members of this phylum share a distinguishing morphological feature, the endoflagella, a special class of flagella that folds back into the cell and remains within the periplasm (Li et al., 2008). Most spirochetes have one or more of these structures protruding from either pole of the cell, forming an axial filament, which gives rise to the characteristic jerky, corkscrewlike motility of the members of the phylum (Li et al., 2008; Paster, 2011a).Currently, the phylum Spirochaetes consists of 15 genera which are highly divergent in terms of their lifestyle and other characteristics (Euzéby, 2013). They live in marine sediments, deep within soil, commensally in the gut of arthropods, including termites, as well as in vertebrates as obligate parasites. They can also be free-living or host-associated, pathogenic or non-pathogenic, and aerobic or anaerobic (Paster, 2011a). There is also enormous variability in the genome sizes and organization of Spirochaetes species **Table 1**. However, despite the diverse characteristics of its members, the phylum Spirochaetes is currently comprised of a single class, *Spirochaetia*, containing a single order, *Spirochaetales*, which is made up of four families (viz. *Spirochaetaceae*, *Brachyspiraceae*, *Leptospiraceae*, and *Brevinemataceae*) (Paster, 2011a; Euzéby, 2013).

There are four clinically important genera of the phylum Spirochaetes whose species are the causative agents of many globally prevalent illnesses, *Treponema*, *Borrelia*, *Leptospira*, and *Brachyspira* (Bellgard et al., 2009). Of these, *Treponema* and *Borrelia* are members of the family *Spirochaetaceae*, which also includes the genera *Clevelandina*, *Cristispira*, *Diplocalyx*, *Hollandina*, *Pillotina*, *Spirochaeta*, and *Sphaerochaeta* (Paster, 2011b; Euzéby, 2013). However, the genera *Clevelandina*, *Diplocalyx*, *Hollandina*, and *Pillotina* have yet to be isolated and

## Table 1 | Genome characteristics of the sequenced members of the phylum Spirochaetes.

Strain name	Accession number	Size (Mb)	GC %	Chromosomes	Plasmids	Genome source
Borrelia afzelii PKo	NC 017238	14	27 90	1	17	Casiens et al. 2011
Borrelia bissettii DN127	NC 015921	14	28.33	1	16	Schutzer et al. 2012
Borrelia burgdorferi $B31^{T}$	NC 001318	1.52	28.18	1	21	Zhong and Barbour, 2004
Borrelia crocidurae Achema	NC 017808	1.53	29.06	1	39	Elbir et al 2012
Borrelia duttonii Ly	NC_011229	1.50	28.02	1	16	Lescot et al. 2008
Borrelia garinii PBi	NC 006156	0.99	28.12	1	11	Glöckner et al. 2004
Borrelia bermsii DAH	NC 010673	0.93	29.81	1	2	Dai et al. 2006
Borrelia recurrentis A1	NC 011244	1 24	27.51	1	7	Unité des Bickettsies <sup>1</sup>
Borrelia sp. SV1	NZ_AB.IZ0000000	1.21	28.27	1	9	Casiens et al 2011
Borrelia spielmanii A14S	NZ_ABKB00000000	1 25	27.69	_	8	Schutzer et al. 2012
Borrelia turicatae 91E135	NC 008710	0.92	29.10	1	_	Bocky Mountain Laboratories <sup>2</sup>
Borrelia valaisiana VS116 <sup>T</sup>	N7 ABCY0000000	0.35	25.10	_	11	Schutzer et al. 2012
Brachyspira byodysenteriae $\Delta TCC 27164^{T}$	NZ_ABSY00000000	3.05	27.00	1	1	DOF-IGI3
Brachyspira intermedia $PW/S/\Delta^{T}$	NC 017243	3 31	27.00	1	1	Håfström et al. 2011
Brachyspira murdochii DSM 12563 <sup>T</sup>	NC 01/150	3.24	27.10	1	_	Pati et al. 2010
Brachyspira milocioli $PA3/6/78^{T}$	NC_014130	2.56	27.00	1	-	Lip of al. 2013
L optopoma illini DSM 21528 <sup>T</sup>	NZ AHKTOOOOOOO	2.50	Z7.30	I	-	
Leptonerna mini DSW 21526	NC_010942	2.06	20 00	-	1	Doc-501
	NC_009500	2.30	40.20	2	I	Pulach at al. 2006
		3.93	40.20	2	-	ICV4
Leptospira biodrilli $5599$		4.49	42.90	-	-	
		4.57	25.00	-	-	
		4.0	25.00	Z	-	
		4.4	35.90	-	-	1014
		4.48	44.70 25.00	-	-	
		4.21	35.90	-	-	
		4.19	38.00	-	_	
		3.88	41.80	-	_	Chou et al., 2012
		4.04	41.70	-	_	
	NZ_AFLV0000000	4.37	40.80	_	-	
	NC_015436	2.23	50.60	1	_	Abt et al., 2012
Sphaerochaeta globosa Buddy'	NC_015152	3.32	48.90	1	-	DOE-JGP
Sphaerochaeta pleomorpha Grapes'	NC_016633	3.59	46.20	1	-	DOE-JGP
Spirochaeta africana DSM 8902	NC_017098	3.29	57.80	1	-	DOE-JGP
Spirochaeta smaragdinae DSIM 11293	NC_014364	4.65	49.00	1	-	Mavromatis et al., 2010
Spirochaeta thermophila DSM 6578	NC_017583	2.56	60.90	1	-	DOE-JGP
Ireponema azotonutricium ZAS-9'	NC_015577	3.86	49.80	1	-	
Ireponema brennaborense DSM 12168'	NC_015500	3.06	51.50	1	-	DOE-JGI <sup>3</sup>
Treponema caldaria DSM 7334	NC_015732	3.24	45.60	1	-	Abt et al., 2013
Ireponema denticola AICC 35405'	NC_002967	2.84	37.90	1	1	Seshadrı et al., 2004
Ireponema pallidum Nichols	NC_000919	1.14	52.80	1	-	Fraser et al., 1997
Treponema paraluiscuniculi Cuniculi A	NC_015714	1.13	52.70	-	-	Smajs et al., 2011
Ireponema phagedenis F0421	NZ_AEFH00000000	2.83	40.10	-	-	WUGSC <sup>5</sup>
Treponema primitia ZAS-21	NC_015578	4.06	50.80	1	-	JCV <sup>4</sup>
Ireponema saccharophilum DSM 2985 <sup>1</sup>	NZ_AGRW00000000	3.45	53.20	-	-	DOE-JGI <sup>3</sup>
Ireponema sp. JC4	NZ_AJGU0000000	3.03	40.30	-	-	CSIRO®
Treponema succinifaciens DSM 2489	NC_015385	2.9	39.17	1	1	Han et al., 2011
Ireponema vincentii ATCC 35580	NZ_ACYH00000000	2.51	45.70	-	-	JCV <sup>4</sup>
Iurneriella parva DSM 215271	NC_018020	4.41	53.60	1	1	DOE-JGI3

Genomic information was collected from: http://www.ncbi.nlm.nih.gov/genomes/

<sup>1</sup>Unité des Rickettsies: Genome sequenced by Unité des Rickettsies at Center National de Référence.

<sup>2</sup> Rocky Mountain Laboratories: Genome sequenced by the Laboratory of Human Bacterial Pathogenesis at Rocky Mountain Laboratories.

<sup>3</sup>DOE-JGI: Genome sequenced by the United States Department of Energy Joint Genome Institute.

<sup>4</sup>JCV: Genome sequenced by the J. Craig Venter Institute.

<sup>5</sup>WUGSC: Genome sequenced by the Washington University Genome Sequencing Center.

<sup>6</sup>CSIRO: Genome sequenced by the Commonwealth Scientific and Industrial Research Organization.

grown in pure or mixed culture and their phylogeny is based largely on analyses of morphological characteristics (Bermudes et al., 1988). Treponema pallidum subspecies pallidum is the causative agent of syphilis, a sexually transmitted disease which affects at least 25 million adults worldwide (Gerbase et al., 1998). Other members of the genus Treponema are responsible for diseases such bejel, yaws, and pinta and play important role in periodontal diseases (Ellen and Galimanas, 2005; Visser and Ellen, 2011; Smajs et al., 2012). Members of the genus Borrelia, namely Borrelia burgdorferi s and Borrelia recurrentis, are important human pathogens that cause Lyme disease and relapsing fever, respectively (Dworkin et al., 2008; Nau et al., 2009; Cutler, 2010). Leptospira and Brachyspira, are members of the families Leptospiraceae and Brachyspiraceae, and causative agents of the diseases leptospirosis and intestinal spirochaetosis, respectively (Adler and de la Peña Moctezuma, 2010; Anthony et al., 2013; Euzéby, 2013).

Despite the importance of species of the phylum Spirochaetes in causing many important human diseases, the evolutionary relationship of species within this phylum remains poorly understood and no distinguishing molecular features are known that are specific for all members of the different families (Olsen et al., 2000; Paster and Dewhirst, 2000; Paster, 2011a). The availability of genome sequences provides a valuable resource to identify/discover novel molecular markers that are helpful in these regards and to gain insights into their evolutionary relationships. Genomes from 48 species covering the three main families of the phylum Spirochaetes are now available in the NCBI database (Table 1) (NCBI, 2013). The availability of genome sequences allows for the use of comparative genomic approaches to identify molecular markers that are specific for different bacterial taxa at various taxonomic levels. Using genomic sequences, one useful approach pioneered by our lab involves the discovery of Conserved Signature insertions/deletions (i.e., Indels) or CSIs present in protein sequences that are specific for different groups of organisms. Due to the specificity of these CSIs for particular groups/taxa of species, they provide valuable molecular markers of common evolutionary descent (i.e., synapomorphies) for identification and demarcation of different phylogenetic/taxonomic clades of organisms in molecular terms. Additionally, based upon the presence or absence of these CSIs in outgroup species, it is possible to infer whether the observed genetic change is an insert or a deletion and a rooted phylogenetic relationship among different groups can be derived (Baldauf and Palmer, 1993; Gupta, 1998; Griffiths and Gupta, 2004; Gao and Gupta, 2012a).

In this work, we report the results of comparative analyses on protein sequences for the phylum Spirochaetes to identify molecular markers (CSIs) that are specific for the species from the phylum and its subgroups, or those that provide information regarding interrelationships among them. These studies have led to identification of 38 CSIs providing novel molecular markers for the species from the phylum and clarifying their evolutionary relationships. Additionally, we have also constructed a phylogenetic tree for all genome sequenced members of the phylum Spirochaetes based upon concatenated sequences for 22 conserved proteins. The inferences from different identified CSIs are strongly supported by the branching pattern of species in the phylogenetic tree indicating that the identified CSIs provide reliable molecular markers for the indicated groups of Spirochaetes.

# METHODS

# PHYLOGENETIC SEQUENCE ANALYSIS

Phylogenetic analysis was performed on a concatenated sequence alignment of 22 highly conserved proteins (viz. UvrD, GyrA, GyrB, RpoB, RpoC, EF-G, EF-Tu, RecA, ArgRS, IleRS, ThrRS, TrpRS, SecY, DnaK, and ribosomal proteins L2, L5, S2, S3, and S9) which have been widely used for phylogenetic analysis (Harris et al., 2003; Gao and Gupta, 2012a). Sequences for these proteins were obtained from the NCBI database for representative strains of all the sequenced Spirochaetes species (Table 1) and Thermosvnechococcus elongatus and Nostoc flagelliforme which were used to root the tree. Multiple sequence alignments for these proteins were created using Clustal\_X 1.83 (Jeanmougin et al., 1998) and concatenated into a single alignment file. Poorly aligned regions from this alignment file were removed using Gblocks 0.91 b (Castresana, 2000). The resulting alignment, which contained 7411 aligned amino acids, was used for phylogenetic analysis. The maximum likelihood (ML) and neighbor joining (NJ) trees based on 100 bootstrap replicates of this alignment were constructed using MEGA 5.1 (Tamura et al., 2011) employing the Whelan and Goldman (Whelan and Goldman, 2001) and Jones-Taylor-Thornton (Jones et al., 1992) substitution models, respectively.

A 16S rRNA gene sequence tree was also created for 107 sequences that included representative species for all 11 cultured Spirochaetes genera. 16S rRNA gene sequences larger than 1200 bp were obtained for all type species classified under the phylum Spirochaetes in release 114 of the SILVA database (Quast et al., 2013). Information for these sequences is provided in Supplemental Table 1. A ML tree based on these sequences was created using 100 bootstrap replicates of the 16S rRNA sequence alignments in MEGA 5.1 (Tamura et al., 2011) employing the General Time-Reversible (Tavaré, 1986) substitution model.

## **IDENTIFICATION OF MOLECULAR MARKERS (CSIs)**

To identify CSIs that are commonly shared by different groups of Spirochaetes, BLASTp searches (Altschul et al., 1997) were performed on each protein in the genome of Treponema pallidum subspecies pallidum strain Nichols. These searches were performed using the default BLAST parameters against all available sequences in the GenBank non-redundant database. For those proteins for whom high scoring homologs (*E*-values  $< 1e^{-20}$ ) were present in other species from the phylum Spirochaetes and some other bacterial groups multiple sequence alignments were created using the Clustal\_X 1.83 program (Jeanmougin et al., 1998). These alignments were visually inspected for the presence of insertions or deletions that were flanked on both sides by at least 4-5 conserved amino acid residues in the neighboring 30-40 amino acids. Indels that were not flanked by conserved regions were not further considered, as they do not provide useful molecular markers (Gupta, 1998; Gao and Gupta, 2012a; Adeolu and Gupta, 2013). The specificity of potentially useful indels for members of the Spirochaetes was further evaluated by carrying out detailed Blastp searches on short sequence segments containing the indel and the flanking conserved regions (60–100 amino acids long). To ensure that the identified signatures are only present in the Spirochaetes homologs, a minimum of 250 blast hits with the highest similarity to the query sequence were examined for the presence or absence of these CSIs. In this work, we report the results of only those CSIs that are specific for different groups of Spirochaetes and where similar CSIs were not observed in any other bacteria in the top 250 blast hits. The sequence alignment files presented here contain sequence information for all sequenced genera within Spirochaetes. However, due to size restraints, different strains and/or species of the sequenced genera are not shown as they all exhibited similar patterns.

# RESULTS

# **GENOMIC CHARACTERISTICS OF THE SEQUENCED SPIROCHAETES**

There are currently 48 genome sequenced species of Spirochaetes. Table 1 lists some characteristics of representative strains for all Spirochaetes species that have been completely sequenced. The genome sizes of these species of Spirochaetes showed a large amount of variation, ranging from 0.92 to 4.7 Mb in length. The G + C content of these species also showed a large amount of variation, ranging from 25.8 to 60.9%. The members of the phylum Spirochaetes also exhibited a large amount of variation in genome structure. The genome structure of members of genus Borrelia is one of the most unique among prokaryotes (Chaconas, 2005; Chaconas and Kobryn, 2010). The Borrelia genome consists of 6-24 DNA segments, including a linear chromosome about 900 kb in length which is accompanied by multiple essential linear and circular plasmids ranging from 5 to 220 kb in length (Chaconas and Kobryn, 2010). Linear chromosomes and plasmids terminated by covalently closed hairpin telomers are particularly uncommon genomic features among prokaryotes and are only found in the genomes of the Borrelia species and the species Agrobacterium tumefaciens (Goodner et al., 2001; Kobryn, 2007; Chaconas and Kobryn, 2010). Members of the genus Leptospira also have an unusual genome structure consisting of two circular chromosomes, a big chromosome about 3.6-4.2 Mb in length and a smaller chromosome about 300 kb in length (Ren et al., 2003; Picardeau et al., 2008).

# PHYLOGENETIC ANALYSES OF THE SEQUENCED SPIROCHAETES

The branching order of species within the phylum Spirochaetes has primarily been determined using 16S rRNA sequence based phylogenetic trees (Paster and Dewhirst, 2000; Paster, 2011a). In these trees, the four families with the phylum branch into distinct monophyletic clades separated by long branches. However, the interrelationships of members of the family Spirochaetaceae are not reliably resolved (Paster, 2011b) (Figure 2). Phylogenetic trees derived from large numbers of conserved genes/proteins provide greater resolving power than those based on any single gene or protein (Rokas et al., 2003; Ciccarelli et al., 2006; Wu et al., 2009; Gao and Gupta, 2012a). In this study, we have constructed phylogenetic trees of the genome sequenced members of the phylum Spirochaetes listed in Table 1 using 22 conserved housekeeping and ribosomal proteins. The trees were constructed using both the NJ and ML methodologies and branching patterns generated by both methodologies were highly similar (Figure 1).

In the concatenated protein trees, which are rooted using the species T. elongatus and N. flagelliforme, the members of the three sequenced families of Spirochaetes (viz. Spirochaetaceae, Brachyspiraceae, and Leptospiraceae) formed three distinct monophyletic clades (Figure 1). Additionally, the branching order of members of the family Spirochaetaceae is well-resolved in the concatenated protein trees. Within the Spirochaetaceae clade, the genera Treponema, Spirochaeta, and Sphaerochaeta formed a well-supported monophyletic clade separated from the members of the genus Borrelia by a long branch. The Treponema, Spirochaeta, and Sphaerochaeta clade exhibited a large amount of diversity and consisted of a number of strongly supported subclades. Members of each of the sequenced genera within Spirochaetes formed monophyletic clusters with the exception of the genus Spirochaeta, where Spirochaeta smaragdinae branched with the genus Sphaerochaeta. Another Spirochaeta species, S. caldaria, which branched within the Treponema has recently been reclassified as Treponema caldaria (Abt et al., 2013). The remaining Spirochaeta (viz. S. thermophila and S. africana) branched deeply within the Treponema, Spirochaeta, and Sphaerochaeta clade (Figure 1). The monophyletic clade containing all the members of the genus Borrelia consisted of two highly distinct subclades, one containing Borrelia burgdorferi, and related species of Borrelia and the other containing Borrelia recurrentis related species.

The 16S rRNA tree shown in **Figure 2** includes all of the members included in the concatenated protein tree as well as other cultured members of the phylum Spirochaetes which have yet to be genome sequenced. The branching patterns in the 16S rRNA phylogenetic tree were similar to those observed in the concatenated protein tree; all families within the phylum branched distinctly. Within the cluster consisting of members of the family *Spirochaetaceae* the genera *Treponema*, *Sphaerochaeta*, and most members of the genus *Spirochaeta* formed a monophyletic clade. The genera *Borrelia* and *Cristispira* also formed a well-supported monophyletic clade that was distinct from the genera *Treponema*, *Spirochaeta*, and *Sphaerochaeta* within the *Spirochaetaceae* clade. The different sequenced members of the genus *Borrelia* also formed two distinct clusters in the 16S rRNA tree (**Figure 2**).

# CSI SPECIFIC FOR THE PHYLUM SPIROCHAETES

CSIs that are restricted to a group of related species are a novel class of molecular marker with high utility for evolutionary studies (Gupta, 1998; Rokas et al., 2003; Gupta, 2009; Gao and Gupta, 2012a). The co-occurrence of multiple CSIs in different species may be due to shared evolutionary history, convergent evolution, lateral gene transfer. However, the unique shared presence of multiple CSIs in a diverse range by a related group of species is most parsimoniously explained by the occurrence of the rare genetic changes that resulted in these CSIs in a common ancestor of the group, followed by vertical transmission of these CSIs to various descendant species (Gupta, 1998; Rokas and Holland, 2000; Gogarten et al., 2002; Gupta and Griffiths, 2002; Gao and Gupta, 2012a). Hence, these CSIs represent molecular synapomorphies of common evolutionary descent and they provide useful markers for identifying different groups of organisms in molecular terms and for understanding their interrelationships independently of



FIGURE 1 | A phylogenetic tree of genome sequenced members of the phylum Spirochaetes based on the concatenated amino acid sequences of 22 conserved proteins. The tree shown is a maximum-likelihood (ML) distance tree. Bootstrap values are shown at branch nodes for both maximum-likelihood and neighbor-joining tree construction methods as ML/NJ. The different sequenced families and two main clades of the family *Spirochaetaceae* supported by the tree are marked. The letter <sup>T</sup> refers to the type strain of the species.

phylogenetic trees (Gupta, 1998; Gupta and Griffiths, 2002; Gao and Gupta, 2012a,b). The CSI-based approach has recently been used to propose important taxonomic changes for a number of groups of bacteria (viz. Chloroflexi, *Coriobacateriia, Neisseriales*,

and *Bacillus*) at different taxonomic ranks (Gupta et al., 2012, 2013; Adeolu and Gupta, 2013; Bhandari et al., 2013). In the present work, we have completed comprehensive genomic analyses to identify CSIs that are primarily restricted to the phylum



Spirochaetes or its subgroups. Information regarding the species specificities of these CSIs and their evolutionary significances are discussed below.

Our analyses have identified 38 CSIs in diverse and important proteins that are specific for members of the Spirochaetes. One CSI has been identified that is specifically found in all of the sequenced members of the phylum Spirochaetes and not found in homologous proteins from any other bacterial species (in the top 250 Blast hits) (**Figure 3**). This CSI consists of a 3 amino acid (aa) insertion located in the flagellar basal-body rod protein FlgC, a component of the basal body which comprises a large portion of the flagella (Macnab, 2003). This CSI represents a unique molecular characteristic of the phylum Spirochaetes and may be related to

					92		122
	[	Treponema p	pallidum	291059787	YDPTHPDA	ILS	GPKAGYVEYPNVDIVTEMVD
		Treponema	vincentii	257457889	- E- D H-	- V-	NN
		Treponema d	denticola	42526722	к-	LKY	
		Treponema p	paraluiscuniculi	3387063			
		Treponema p	phagedenis	320535379	N	- K-	NNN
		Treponema p	primitia	333997909		- KA	QMV
		Treponema a	azotonutricium	333994363		- TT	RQM
		Treponema b	orennaborense	332297607	S	- Q-	NNN
		Treponema	vincentii	257457889	- E- D H-	- V -	NN
		Treponema s	succinifaciens	328947992	S	- Q-	NKN
		Treponema o	caldaria	339500288		- KT	RMN
		Spirochaeta	a africana	373485215		-R-	REMNE
		Spirochaeta	a thermophila	307718596		- K-	LKNN
		Spirochaeta	a smaragdinae	302338075		- QT	KMNE
		Borrelia bi	issettii	343127611	s	- G-	-DRKLNL-E
		Borrelia tu	uricatae	119953092		- K-	-ERKFNA-E
		Borrelia du	uttonii	203284214		- K-	-ELKFNV-E
Sp	irochaetes	Borrelia at	fzelii	111115118	s	-S-	-D-KLNL-E
	(48/48)	Borrelia he	ermsii	187918167		- R-	-ELKFNA-E
		Borrelia sp	pielmanii	224534284	S	-S-	-D-KLNL-E
		Borrelia bu	urgdorferi	1448943	AS	-S-	-D-KLNL-E
		Borrelia ga	arinii	51598553	s	- SF	-D-KLNL-E
		Borrelia va	alaisiana	224531802	S	- SF	-D-KLNL-E
		Leptospira	noguchii	359724699		-QI	KMNT-
		Leptospira	licerasiae	359687476		-QI	KMNT-
		Leptospira	santarosai	359684672		-QI	NK-LMNT-
		Leptospira	weilii	359729323	FE	- GR	-S-KFIILEE-L
		Leptospira	biflexa	183220163		- QT	KLINT-
		Leptospira	Borgpetersenii	116327145		-QI	KMNT-
		Leptospira	interrogans	294827625	S	-QI	KMNT-
		Leptonema :	illini	373872765	FG	-KI	QMNET-
		Turneriella	a parva	392405560	FS	VKV	KS
		Brachyspira	a murdochii	296127542	····S····	- KY	-EMNP
		Brachyspira	a pilosicoli	300870979	S	- KY	-EMNP
		Brachyspira	a hyodysenteriae	225619783	····S····	- KY	-EMNP
		Brachyspira	a intermedia	384208491	S	- KY	-EMNP
		Butyrivibri	io crossotus	260438810	····S····		DENKNNT-
		Clostridium	n difficile	255099413	S		KPNN-LVT-
		Syntrophomo	onas wolfei	114566383	SE-		DNENNEIN
		Oribacteri	um sinus	227873538	-N-N		DAQNLK-T
		Eubacterium	n yurii	306820111	G		DENKMELI
		Rose, intes	stinalis	240145509	S		DENTNTTN
		Lysinibaci	Llus	299534789			NADNMPLK
		Bacillus ce	ereus	218230982	H-		N-ERI-VTATN
	Other	Listeria se	eligeri	289433989	N-		NEN-AMTATN
	Dutier	Abiotrophia	a defectiva	229826300	····S····		NEDTNT-QT-
	Bacteria	Lactobacil	Lus ruminis	227528457	N		DANMNMAD
	(0/>250)	Thermotoga	neapolitana	222100200			UENRMNR
		Acido. caps	sulatum	225871925			D-QSEINP
		Desultarcul	Lus Daars11	302342345	HS		UADLVN
		nelicobacte	ar pytori	308064340	8		NAQANA-VA-
		Selenomonas	s noxia	292009/34	-E-G		NADKNA
		Brucella ce		200108120	SA-		D.D
		Conovibrat	s syringae	20009138	-E-N5-		DED AM ND
		Papagabara	lla marico	204041007	E		
	L	Persepnone.	lia marina	220800273	FS		DunAPLR

FIGURE 3 | A partial sequence alignment of the flagellar basal-body rod protein FIgC, showing a CSI (boxed) that is uniquely present in all members of the phylum Spirochaetes. Sequence information for only a limited number of species from the Spirochaetes and other bacteria is shown here, but unless otherwise indicated similar CSIs were detected in all members of the indicated group and not detected in any other bacterial species in the top 250 Blastp hits. The dashes (--) in the alignments indicate identity with the residue in the top sequence. GenBank identification (GI) numbers for each sequence are indicated in the second column. Sequence homologs for this protein were not identified from members of the genus *Sphaerochaeta*.

the characteristic flagellar morphology shared by members of the phylum.

## **CSIs THAT ARE SPECIFIC FOR DIFFERENT FAMILIES OF SPIROCHAETES**

Many of the CSIs identified by our analyses are specific for the different sequenced families within the phylum Spirochaetes (viz. *Spirochaetaceae*, *Brachyspiraceae*, and *Leptospiraceae*) allowing us to demarcate these families in clear molecular terms. Seven of the CSIs identified by our analyses are specific for the family *Spirochaetaceae*. One example of a CSI that is specific for the species from the family *Spirochaetaceae* is a 15 aa insertion in a highly conserved region of the protein phosphoribosylpyrophosphate synthetase, which is uniquely found in all members of the family *Spirochaetaceae* but not in any other sequenced bacterial groups (**Figure 4**). Sequence information for 6 other CSIs in diverse proteins (viz. Alanyl-tRNA synthetase,

			97		143
1	Borrelia sp. SV1	496158147	ILKTIRNKDIFIVQDVA	NAYEVEINSSEKII	MTVNDHIMNLMTTIDA
	Borrelia burgdorferi	218249888			
	Borrelia garinii	51598795		.1	
	Borrelia bissettii	343127845		·····S·····	
	Borrelia spielmanii	493478988	······	N+	*************
	Borrelia valaisiana	492960169	**************	-TN	*************
	Borrelia afzelii	384207032	***************	-TNV-	
	Borrelia crocidurae	386859775	•••••\$	-T-P-TV-NN-EVV	•••••L•••••V••
	Borrelia duttonii	203284450	·····\$	-T-P-TV-NN-EVV	•••••V••
	Borrelia recurrentis	203287984	S	- I - P - TV - NN - EVV	V
	Borrelia turicatae	119953329	•••••\$	- T - P - NV - NN VV	IL
	Borrelia hermsii	187918407	•••••S	- T - P - NV - NN VV	IL
Spirochastaceae	Spirochaeta thermophila	307719014	L-SSV-D-EVYE	-R-PQRFSDGKEYT	LSNFIM-LV
(20/20) X	Spirochaeta smaragdinae	302337090	L-SSVKGMSVIE	-H-PLAFYNL-ENVC	LSV-M-F
(30/30)	Sphaerochaeta globosa	325971852	DAV-GLRIY-LS	-SEP-KVAG-DEPVK	LSL-F-FN-
	Sphaerochaeta pleomorpha	374317254	DAV-GLRVY-LS	-MEP-KVAG-DEP-H	LSL-F-LN-
	Sphaerochaeta coccoides	330837008	M-APV-GLRLIHT	-ESPIKVAGLDEPQV	FSIL-F-FN-
	Treponema sp. JC4	496394911	L-EGVFE	- HEVLTL - GGKNKAV	LSV-S-LV
	Treponema succinitaciens	328948985	L-DCGVY-FE	-HEKIAL-DGKNVLS	LSV-S-LV
	Treponema denticola	488752940	L-E-V-GVY-FE	-KQ-11EGKNKEV	FSL-S-IV
	Treponema caldaria	339498915	ESGFE	- HHPLPF - DGKN-KV	LSLV-V
	Treponema brennaborense	332297752	-NECGVFE	-HQPIKL-GGKNLQV	LSSMLV
	Treponema saccharophilum	488/91143	ECV-GY-FE	-HEPIKM-GGKNEVV	FSV-T-LV
	Treponema vincentii	493197584	ECG-NVFE	-H-PLSV-EGKNOHV	FSL-SM-V
	Treponema azotonutricium	333994388	ESGYE	-H-P-SF-DGDLKKA	LSL-TMFV-V
	Treponema pallidum	15639286	-NECGVFE	- HQP - LV - NGKSKKI	FSV-M-IV
	Treponema phageden1s	488785632	-NECGVFE	- HQVL TV - QGKNKQR	FSL-S-IV
	Treponema primitia	333996912	ESGMY-FE	- HHPLNF - NGSLSKA	LSIL-SIFV-V
(	Brachyspira pilosicoli	404476574	-EE-ATY-I-PTG	R	PSS-ESW-E-FCV
1	Brachyspira murdochii	296125819	-EE-VTY-I-PTG	R	PSSSENW-E-YCI
Other	Brachyspira hyodysenteriae	225216214	-EE-VTYVI-PTG	R	PSSSESW-E-YCI
Spirachasta	Turneriella parva	392404244	-SENV-GS-V-VI-STS		NPANL-ELIV
Sphochaetes <	Leptonema illini	488860053	-NENGVPTS		APANL-E-LLI
(0/18)	Leptospira fainei	514358906	-EENV-GRLI-STS		APAL-E-LLIM
	Leptospira inadai	498101250	-EENV-GRLI-STS		APAL-E-LLIM
L. L.	Leptospira broomii	498256941	-DENV-GRLI-STS		APAL-E-LLIM
	Leptospira interrogans	463284931	-EDNV-G-EVSTS		APAL-E-ILIM
/	Bacteriovorax marinus	374290019	-E-NV-GA-V-VI-STS		SPV-E-LIM
1	Helicobacter pylori	385226974	-SESV-GPTC		- P NL-E-LVMV
	Hippea maritima	327399268	-DESV-GA-V-LI-SLS		SPNE-LV-L
	Syntrophobacter fumaroxidans	116748649	-GENV-GAVSG-		QPL-E-LVM
	Anaerococcus lactolyticus	490965715	-NESGVPTS		NPTNL-E-LIM
	Staphylococcus massiliensis	496848324	-EESGC-VI-PTS		NPL-E-LIM
	Streptococcus macacae	489169291	- EESV - GS Y - I - STS		NPLWE-LIMV
	Facklamia ignava	493751806	VAESGSHVYSTS		NPE-L-E-LIA
	Streptococcus thermophilus	386087062	- EESV - GD Y - I - STS		NPLWE-LIM
	Bacillus cereus	446029563	-EESGC-V1-STS		FPEE-LIM
	Lactobacillus malefermentans	498305941	-EESV-GD-V-LI-S-S		NPSML-EIM
Other	Leuconostoc lactis	497688032	-EESGDNV-VI-ST-		APNL-E-LIM
Ouler	Pediococcus claussenii	377809552	-EESGAEVI-SIS		DPISL-E-LIM
Bacteria <	Eremococcus coleocola	493462576	-EESGDH-YSTS		NPNL-EIC
(0/>250)	Halothiobacillus neapolitanus	261855768	ENV - GR - V - V PTC		DPTNEVMV
A RECEIVED AND	Acidithiobacillus caldus	491008878	ENV-GR-V-AI-PTC		APTL-E-L-L
	Selenomonas sputigena	330839472	-SESGI-PTS		QPNL-E-LIMT
	Centipeda periodontii	493349495	-EESGI-STS		FPNL-E-LILT
	Megamonas hypermegale	479205918	-SDSGVI-PTS		QPNL-E-LVMA
	Mitsuokella multacida	492431171	-SESGI-PTS		YPNL-E-LIMA
	Capsaspora owczarzaki	470296469	-KESV-DE-VYSGS		GD+NL-E+LIM-N-
	Spiroplasma chrysopicola	507379176	AVNSV-GY-I-STS		NPENL-E-LIA
	Mycoplasma iowae	490124334	SKE-VC-CLLI-STS		NPNL-E-LIG
	Cyanobium gracile	427701574	-QESGC-V-LI-PTC		APL-EIMV
	Microcystis aeruginosa	488876880	-QESGC-VYLI-PCC		NPL-E-LIM
	Chlamydomonas reinhardtii	159472214	VQESGC-V-LI-PTC		PPL-E-LI
A					

FIGURE 4 | A partial sequence alignment of the protein alanyl-tRNA synthetase showing a two amino acid insertion (boxed) identified in homologs from the family *Spirochaetaceae*, but not found in the

sequence homologs of any other sequenced bacteria. Sequence information for other *Spirochaetaceae* specific CSIs is presented in Supplemental Figures 3–6 and summarized in **Table 2**.

phosphoribosylpyrophosphate synthetase, preprotein translocase SecY, peptide chain release factor 2, DNA mismatch repair protein MutS, and DNA mismatch repair protein MutL) that are also specifically present in members of the family *Spirochaetaceae* is presented in Supplementary Figures 1–6 and some of their characteristics are summarized in **Table 2**.

Our analyses have also identified 6 CSIs in diverse proteins that are specifically found in members of the family Brachyspiraceae and absent in all other bacterial groups. One of these Brachyspiraceae-specific CSIs, a 1 aa insertion, is present in the flagellar hook-associated protein FlgK, a protein involved in flagellar hook morphogenesis (Figure 5A) (Homma et al., 1990). Another Brachyspiraceae-specific CSI, a 1 aa insertion, is found in a highly conserved region of DNA polymerase I (Figure 5B). These proteins represent highly conserved and essential components of members of the family Brachyspiraceae which contain conserved molecular changes not found in any other sequenced bacterial group. Sequence information for 4 other CSIs in three other proteins (viz. valyl-tRNA synthetase, ATP-dependent protease La, and glutamyl-tRNA amidotransferase subunit B) that are also specifically present in members of the family Brachyspiraceae is presented in Supplemental Figures 7-10 and some of their characteristics are summarized in Table 3.

We have also identified 5 CSIs that are uniquely present in members of the family *Leptospiraceae*. Two examples of such CSIs are shown in **Figure 6**. The first of these CSIs, an 8 aa insertion in the 50S ribosomal protein L14, is shown in **Figure 6A**, and the other CSI, a 4 aa insert in alanyl-tRNA synthetase, is shown in **Figure 6B**. Both of these CSIs are found in members of the the family *Leptospiraceae* and absent in every other sequenced bacterial group. Sequence information for 4 other CSIs in diverse proteins (viz. 30S Ribosomal protein S2, flagellar basal-body rod protein FlgG, and flagellar filament core protein FlaB) that are also specifically present in members of the family *Leptospiraceae* is presented in Supplemental Figures 11–14 and some of their characteristics are summarized in **Table 4**.

# CSIs DISTINGUISHING TWO CLADES WITHIN THE FAMILY Spirochaetaceae

In addition to the numerous CSIs identified in our analyses for the sequenced families within the phylum Spirochaetes, we have also identified a number of CSIs that elucidate the relationship of the genera within the family *Spirochaetaceae*. Three of the identified CSIs are uniquely shared by the genera *Treponema*, *Spirochaeta*,

and *Sphaerochaeta*. One example of a CSI specific to these three genera, a 1 aa deletion in the 30S ribosomal protein S13, a component of the protein translation complex, is shown in **Figure 7A**. Sequence information for 2 other CSIs specifically found in these three genera is provided in **Table 5** and Supplemental Figures 14, 15. An additional 16 CSIs were uniquely shared by members of the genus *Borrelia*. One example of a CSI consisting of a 6 aa insertion in the glycolysis related protein, phosphofructokinase, that is specific to the members of the genus *Borrelia* and information for them is presented in **Table 5** and Supplemental Figures 16–30.

# **DISCUSSION**

The phylum Spirochaetes is currently distinguished from other bacteria on the basis of both branching in 16S rRNA sequence based phylogenies and the presence of the endoflagella that characterizes the phylum (Paster, 2011a; Euzéby, 2013). Apart from the presence of endoflagella, no reliable morphological, biochemical, or molecular characteristics are known that are specifically shared by all members of the phylum. Additionally, the phylum contains four divergent lineages, contained within a single class/order, that are demarcated largely on the basis of 16S rRNA sequence based phylogenies (Paster, 2011a). In this work, we have utilized comparative genomic techniques to identify large numbers of novel molecular signatures (CSIs) that are distinctive characteristics of either all members of the phylum Spirochaetes or for its different subgroups at multiple phylogenetic levels and which can be used to demarcate these groups in more definitive molecular terms. A summary diagram depicting the species distribution of the identified CSIs is shown in Figure 8.

The phylum Spirochaetes is rare in having a defining morphological characteristic, the endoflagella, which correlates to the clustering of the members of the phylum in 16S rRNA phylogenetic trees (Ludwig and Klenk, 2001; Cavalier-Smith, 2002; Paster, 2011a). The endoflagella is a unique feature of the phylum and is thought to responsible for the great pathogenic and ecological diversity of its many members (Ren et al., 2003). Of the 38 CSIs we have identified in this study, one was uniquely shared by all 48 members of the phylum Spirochaetes and absent in every other sequenced group of bacteria. The identified CSI is located in the flagellar basal-body rod protein FlgC, a core component of the motor complex of the flagella (Macnab, 2003). This CSI provides a novel means to distinguish the members of the phylum from all

### Table 2 | Conserved signature Indels that are specific for members of the family Spirochaetaceae.

Protein name	Gene	GI	Figure	Indel	Indel
	name	number	number	size	position
Phosphoribosylpyrophosphate synthetase	prsA	496158147	Figure 4	15 aa ins	97–143
Alanyl-tRNA synthetase	alaS	386859446	Supplemental Figure 1	2 aa ins	277–306
Phosphoribosylpyrophosphate synthetase	prsA	387827445	Supplemental Figure 2	8 aa ins	256–297
Preprotein translocase	secY	15639201	Supplemental Figure 3	1 aa del	340–373
Peptide chain release factor 2	prfB	257457828	Supplemental Figure 4	1 aa del	137–176
DNA mismatch repair protein MutS	mutS	224532424	Supplemental Figure 5	2 aa del	720–751
DNA mismatch repair protein MutL	mutL	338706271	Supplemental Figure 6	4 aa del	494–520

P-dund	Brachyspira hyodysenteriae	225620569	GHNINNVATKGYSRORV	VMRTFD K PLEQPSLNRAERAGQIGQG
Brachyspiraceae	Brachyspira pilosicoli Brachyspira muchachii	300870953		·····
(4/4)	Brachyspira intermedia	343387999		SA
	Treponema brennaborense	332298485	S-AS-E	QL-Q IYR-D-ST
	Treponema phagedenis	320538492	VA-SS-PT	NLDA-EYR-D-STP
	Treponema succinifaciens	328948264	S-AD-E	IVKS-E - IYR-D-EMVP
	Treponema caldaria	339499734	LS-AS-E	EFSP-E -IYL-GE-TP
	Spirochaeta africana	373483402	LS-ASSE	ELQAYP A-DR-G-AP
	Leptospira licerasiae	359688911	S-ADN-H	-LQATYEHVP
Other	Leptospira weilii	359728975	S-ADN-H-A	T-TAMYDNLP
Chinashastas	Leptospira nogucnii	359684276	S-ADN-H-A	T-TAM-
Spirochaetes _	Leptonema illini	488861851	S-ADNEN-A	T-ESMYDAGP
(0/44)	Turneriella parva	392403086	S-ADN-N-AI	D-E-MHYESSGL
	Treponema pallidum	15639647	VS-SS-P	NLDS-E YR-D-SQLP
	Treponema vincentii	257457497	S-SS-ET	QLDSYEYR-D-STP
	Treponema denticola	42527853	S-SSTI	NLEA-E -IYR-D-TTP
	Borrelia valaisiana	224532252	LS-ATKP	T-K-TTYA-QKKQL
	Borrella spielmanii	224534287	LS-ATKP	1-1-11YA-QKKQL
	Borrella offolii	111115002	LS ATKP	
1	Oceanobacillus ihevensis	23099962	S-AN-PT	NLEAMN GFPYA-BPNMPM-T-
	Bacillus coahuilensis	205375049	A- AN- PQ-1	S-S-TT - FPTAGE-APQIPT-
	Paenibacillus polymyxa	310644334	A-AN-ETT-I	K-KASI -M-AYGISNVPL-T-
0.1	Brevibacillus brevis	226314925	S-AN-E	N-QATT G-PYVGMQASIEP-LL-T-
Other	Aminomonas paucivorans	310783099	A-ADVEI	EASSTFTD-G-A-PAIPT-
Bacteria -	Clostridium botulinum	168180854	A- AS- E	ELQ-TR -YCKM-S-AGPV-T-
(0/>250)	Desulfo. reducens	134300275	T- AN- P EA	-ISPST -WTR-DYLT-MTPL-T-
(0. 200)	Ammonifex degensii	260892170	AA-AS-PI	ELAASS AYTIVL-PAAPL-T-
	Clostpidium liumadeblid	114505761	A-AN-P-FAAP	51V-15 - PHA-MN-N-VL-T-
	Halothermothrix orenii	220932550	A	HT-T- YTV-GETSPAGV-T-
		220302000	CONTRACTOR CONTRACTOR	
			810	85
	Brachyspira murdochii	296127550	GYVRTMCGRIRDL	KTINSSNAMARNEAERMALNTLIQGSAA
Brachusniracana	Brachyspira hyodysenteriae	225621485	-F	
(A(A)	Brachyspira nilosicoli	300871901		·····vv·····
(4/4)	Brachyspira intermedia	343385040		
	Loptopomo illini	499957666	0.18	ID T REPKEC A T.V. PV TO
	Ceptonema IIIIni	400007000	Q-LSH-QV	CDI VIIA DEV EA I VIII DV OT
	Turneriella parva	392404191	AENLFR-P-	GDLKNA-HFV-EAL-IPV-SI
	Leptospira terpstrae	489061584	E-LLR-Y-	PDKHKSEA-K-V-I-SPTS-
	Leptospira yanagawae	505584257	E-LLR-Y-	PDHK-VSEG-K-V-I-SPTS-
	Leptospira wolbachii	505590055	E-LLR-Y-	PD-H-KHKSEA-K-V-I-SPTS-
	Leptospira meyeri	463323994	E-LLR-Y-	PD-H-KHKSEA-K-V-I-SPTS-
	Leptospira biflexa	183220086	E-LLR-Y-	PD-HHK-VSEG-K-V-I-SPTS-
	Leptospira interrogans	445566922	Q-LTR-PV	TDTHKS-KEA-K-I-I-SPTS-
	Treponema paraluiscuniculi	338706083	TSLAR-YI	RD-R-TLE-ARQS
	Treponema phagedenis	320535266	E-LMR-YI	YAA-KLEQAGV-IP
Other	Treponema pallidum	291059523	TSLAR-YI	RD-R-TLE-ARQS
Other	Treponema vincentii	257457843	- F-E-LMR-YI	RAR-QTQ-AAI-IP
Spirochaetes	Treponema denticola	42527145	E-LMR-Y-	PAK-KVEKAGI-VPT
(0/44)	Spirochaeta thermophila	307717956	T-LFK-P-	PY-T-R-KTQKTGI-VPG-
	Spirochaeta africana	383791874	-BST-LLE-ST	PA-H-B-KNEKSGI-VPT
	Spirochaeta smaradinae	302340150	-VSK-11 F-ST	PAB-KTEKAGT-VPT
	Sphaerochaeta alabasa	325970645	S-S-11-W/-TT	TEBS-VE-AKT-VV-
	Sphaenochaete plaenorthe	350351036		NE.H.P. COE.AK.O. TOV. V
	Regnolia opicimenti	004524260	OF THE DAY	E N VIE 64 TOT OT
	Dorreita spielmanii	224034002	SE-ILK-H-YI	N-TLE-SAISI-SI
	Borrella Dissettii	343127849	SE-ILK-R-YI	-EN-YLE-SAI-I-ST
	Borrelia valaisiana	224532146	SE-ISK-R-YI	-EN-YLE-SAI-I-SI
	Borrelia afzelii	216263573	SE-ILK-R-YI	-EN-YLE-SAI-I-SI
	Borrelia burgdorferi	195941682	SE-ILK-R-YI	-EN-YLE-SAI-I-SI
	Borrelia garinii	219685277	SE-ILK-R-YI	-EN-YLE-SAI-I-SA
[	Selenomonas noxía	292669881	T-LYR-E-	SAR-Y-Q-SFMPT
	Eubacterium rectale	291525363	TKFPI	PELF-Q-QFGV-M-SPT
1201	Coprococcus eutactus	163814036	KFPI	PELKF-Q-SFGV-M-SP
Other	Veillonella dispar	238019447	-KFQ-E-	PDR-FNR-SFT-MPT
Bacteria -	Roseburia intestinalis	291535124	TFR-PV	PELSF-Q-SFGV-M-SPT
(0/>250)	Flavobacteria bacterium	126663071	E-ISR-Y-	-DAIV-GGN-V-AP
(01-250)	Kordia algicida	163753855	E-VLR-Y-	-DRIV-GAN-V-AP
	Prevotella ruminicola	294675119	T-LFB-Y-	PDHTV-GFN-I-APT
	Bacteroides dorei	212692553	IE- IFK-Y-	PDR-SVV-GYN-I-AP
1			and the state of the	1999 - 1997 - 19
equence alignments	of (A) Flagellar hook-associ	i <b>ated</b> flagel	llar hook-associated p	orotein FlgK were not identifie

other bacteria in molecular terms and provides another delimiting marker for the group in addition to the endoflagella. While the role of this CSI in the function or morphology of the Spirochaetes flagella is currently unknown, the unique presence of this CSI in a flagellar protein in all members of the phylum Spirochaetes suggests that it may be related to the unique flagella ultrastructure of the phylum. Earlier work has established that the CSIs are pri-

marily located on surface loops of proteins which are important

## Table 3 | Conserved signature Indels that are specific for members of the family Brachyspiraceae.

Protein name	Gene	GI	Figure	Indel	Indel
	name	number	number	size	position
Flagellar hook-associated protein FlgK	flgK	225620569	Figure 5A	1 aa ins	62–104
DNA polymerase I	polA	296127550	Figure 5B	1 aa ins	810–852
ValyI-tRNA synthetase	valS	300871449	Supplemental Figure 7	1 aa ins	225–263
ValyI-tRNA synthetase	valS	300871449	Supplemental Figure 8	2 aa del	660–703
ATP-dependent protease La	lon	225620632	Supplemental Figure 9	1 aa ins	760–793
Glutamyl-tRNA amidotransferase subunit B	gatB	300871379	Supplemental Figure 10	1 aa ins	325–361

			36		73
	Leptospira interrogans	5163214	GDEIIVAVKDAQPAFGLKDS	TGKKVHNK	AVQRAVVVRT
	Leptospira borgpetersenii	116327234			*******
Leptospiraceae	Leptospira biflexa	183221337	YR-G	Q	
(12/12)	Leptospira noguchii	359723222			
(15/15)	Leptospira licerasiae	359687079	Y	•••••	
	Turneriella parva	392404093	I-VA-TY-I	AD-K-	QI
	Leptonema illini	373876420	I-VSV-EYN	R	K
	Treponema primitia	333998879	I-VL-TSTI-KG		T-EK
	Treponema brennaborense	332298903	I-VI-TSTI-KG		SIEKI
	Treponema denticola	488746686	I-VL-TSTI-KG		S-EKI
	Treponema azotonutricium	333996155	I-VL-TSAI-KG		EK
Other	Spirochaeta Africana	383789798	V-V		N-EI
Other	Spirochaeta thermonhila	307718216			T
Spirochaetes	Spirochaeta smaradinae	302337487			DesKestes
(0/35)	Sphaerochaeta coccoides	220827645	V V N NGAI KG		DIKII
	Sphaerochaeta coccordes	330037043			D WK I
	Sphaerochaeta preomorpha	374310502	V CO T II TOOLEKO		D-MK1
	Brachyspira intermedia	384210002	VCS-T-TI-TCSTERG		K-VK1
	Brachyspira murdochii	296126875	V-VCS-1-11-TCSTERG		K-VK1
	Brachyspira pilosicoli	300870637	V-VCS-1-11-1CS1EKG		K-VK1
	Niabeila soli	373233633	K-V-TTG-I-KG		AKI
	Niastella koreensis	361061937	K-V-TIG-I-KG		TKI
Other	Alistipes putredinis	167752395	K-VS-S-SGDV-KG		SK
Other	Dokdonia donghaensis	86132554	K-V-ST-NGQV-KG		ST
Bacteria _	Psychroflexus torquis	91216900	KVV-SE-T-NGNI-KG		ST
(0/>250)	Sphingo. spiritivorum	227538708	K-V-TS-L-SGNV-KG		S-SK
	Prevotella copri	281422241	V-VNVI-SSDKG		SK-LI
	Zunongwangia profunda	295135691	K-V-SE-T-NGNI-KG		ST
	Mari. ferrooxydans	114778814	V-VE-V-NGKV-KG		E
	Lentosnina weilii	350728223			
	Leptospira weilii Leptospira santarosai	359728223 359683791			-R
Leptospiraceae	Leptospira weilii Leptospira santarosai Leptospira borgpetersenii	359728223 359683791 116331894	V	D	-R
Leptospiraceae (12/13) —	Leptospira weilii Leptospira santarosai Leptospira borgpetersenii Leptospira biflexa	359728223 359683791 116331894 183220780	V P	D D T-G NNPN	-R 
Leptospiraceae (12/13) —	Leptospira weilii Leptospira santarosai Leptospira borgpetersenii Leptospira biflexa Leptospira licerasiae	359728223 359683791 116331894 183220780 359686873		D D T-G NNPN D-G VKYE	-R -RYYL
Leptospiraceae (12/13) —	Leptospira weilii Leptospira santarosai Leptospira borgpetersenii Leptospira biflexa Leptospira licerasiae Leotonema illini	359728223 359683791 116331894 183220780 359686873 488860306		D D T-G NNPN D-G VKYE PTG CGDP	-RYYL
Leptospiraceae (12/13) —	Leptospira weilii Leptospira santarosai Leptospira borgpetersenii Leptospira liflexa Leptospira licerasiae Leptonema illini Turneriella parva	359728223 359683791 116331894 183220780 359686873 488860306 392404112		D D T-G NNPN D-G VKYE PTG CGDP EQN	-RYYL 
Leptospiraceae (12/13) –	Leptospira weilii Leptospira santarosai Leptospira borgpetersenii Leptospira biflexa Leptospira licerasiae Leptonema illini Turneriella parva Brachvospira intermedia	359728223 359683791 116331894 183220780 359686873 488860306 392404112 343385516		D D T-G NNPN D-G VKYE PTG CGDP EQN KPD	-RYYL DTSYYL RAGSGML FVE-YLE
Leptospiraceae (12/13) —	Leptospira weilii Leptospira santarosai Leptospira borgpetersenii Leptospira biflexa Leptospira licerasiae Leptonema illini Turneriella parva Brachyspira intermedia Brachyspira hydysenteriae	359728223 359683791 116331894 183220780 359686873 48860306 392404112 343385516 225620397		D D T-G NNPN D-G VKYE PTG CGDP EQN KPD KPD	-RYYL DTSYYL RAGSGML -FVE-YLE -FVE-YLE
Leptospiraceae (12/13) —	Leptospira weilii Leptospira santarosai Leptospira borgpetersenii Leptospira licerasiae Leptonema illini Turneriella parva Brachyspira intermedia Brachyspira murdcohii	359728223 359683791 116331894 183220780 359686873 48860873 488860306 392404112 343385516 225620397 296126160		D D T-G NNPN D-G VKYE PTG CGDP EQN KPD KPD KPD	-R
Leptospiraceae (12/13) —	Leptospira weilii Leptospira borgpetersenii Leptospira borgpetersenii Leptospira biflexa Leptospira licerasiae Leptonema illini Turneriella parva Brachyspira intermedia Brachyspira intermedia Brachyspira murdochii Borrelia turicatae	359728223 359683791 116331894 183220780 359686873 488860306 392404112 343385516 225620397 296126160 119953017	V P	D D T-G NNPN D-G VKYE PTG CGDP EQN KPD KPD ECD	-R
Leptospiraceae (12/13) —	Leptospira weilii Leptospira borghetersenii Leptospira borghetersenii Leptospira biflexa Leptospira licerasiae Leptonema illini Turneriella parva Brachyspira intermedia Brachyspira hydysenteriae Brachyspira murdochii Borreila turicatae Borreila hermsii	359728223 359683791 116331894 183220780 359686873 48860306 392404112 343385516 225620397 296126160 119953017 187918093		D D T-G NNPN D-G VKYE PTG CGDP EQN KPD KPD KPD ECD RCD	-R
Leptospiraceae (12/13) —	Leptospira weilii Leptospira santarosai Leptospira borghetersenii Leptospira biflexa Leptospira licerasiae Leptonema illini Turneriella parva Brachyspira intermedia Brachyspira hyodysenteriae Brachyspira murdochii Borrelia turicatae Borrelia hermsii Borrelia burgdorferi	359728223 359683791 116331894 183220780 359686873 48860306 392404112 343385516 225620397 296126160 119953017 187918093 488734245		D D T-G NNPN D-G VKYE PTG CGDP EQN KPD KPD KPD ECD RCD DCN	-R
Leptospiraceae (12/13) –	Leptospira weilii Leptospira borgpetersenii Leptospira borgpetersenii Leptospira biflexa Leptospira licerasiae Leptonema illini Turneriella parva Brachyspira intermedia Brachyspira murdochii Borrelia turicatae Borrelia hermsii Borrelia dremsii Borrelia dremsii	359728223 359683791 116331894 183220780 359666673 48860306 392404112 343385516 225620397 296126160 119953017 187918093 488734245 384206719		D D D PTG CGDP EQN KPD KPD ECD RCD ECD RCD DCN NCN	-R
Leptospiraceae (12/13)	Leptospira weilii Leptospira santarosai Leptospira borgpetersenii Leptospira biflexa Leptospira licerasiae Leptonema illini Turneriella parva Brachyspira intermedia Brachyspira hydysenteriae Brachyspira murdochii Borreila turicatae Borreila hemsii Borreila burgdorferi Borreila sfzelii Spirochesta africana	359728223 359683791 116331894 183220780 359686873 488660306 392404112 343385516 225620397 296126160 119953017 187918093 488734245 384206719 383791404	V P P P -P -P -MHCKG -E-F-TD 	D            D            T-G         NNPN           D-G         VKYE           PTG         CGDP           EGN         KPD           ECD         RCD           DCN         NCN           SCR         SCR	-R
Leptospiraceae (12/13) – Other Spirochaetes –	Leptospira weilii Leptospira santarosai Leptospira borghetersenii Leptospira biflexa Leptospira licerasiae Leptonema illini Turneriella parva Brachyspira intermedia Brachyspira hyodysenteriae Brachyspira murdochii Borrelia turicatae Borrelia burgdorferi Borrelia burgdorferi Borrelia afzelii Spirochaeta africana Spirochaeta thermonbila	359728223 359683791 116331894 1183220780 359686873 488660306 392404112 34385516 225620397 226126160 119953017 187918093 488734245 384206719 383791404		D            D            T-G         NNPN           D-G         VKYE           PTG         GGDP           EQN         KPD           KPD         ECD           RCD         DCN           NCN         SCR           DCK         MCK	-R
Leptospiraceae (12/13) Other Spirochaetes (0/35)	Leptospira weilii Leptospira borgpetersenii Leptospira borgpetersenii Leptospira licerasiae Leptonema illini Turneriella parva Brachyspira intermedia Brachyspira murdochii Borrelia turicatae Borrelia hermsii Borrelia hermsii Borrelia hermsii Sorrelia afzelii Spirochaeta africana Spirochaeta africana Spinochaeta afnida	359728223 339663791 116331894 133220780 359666873 488860306 332404112 343385516 225620397 296126160 119953017 187918093 488734245 384206719 383791404 386347103 390837065		D            D            D            D         O           P            D         O           P         O           P         O           P         O           P         O           P         O	-R
Leptospiraceae (12/13) Other Spirochaetes (0/35)	Leptospira weilii Leptospira santarosai Leptospira borgpetersenii Leptospira biflexa Leptospira licerasiae Leptonema illini Turneriella parva Brachyspira intermedia Brachyspira hydysenteriae Brachyspira murdochii Borreila turicatae Borreila hemsii Borreila burgdorferi Borreila dirgdorferi Borreila afzelii Spirochaeta africana Spirochaeta africana Sphaerochaeta pleamorpha	359728223 359683791 116331894 183220780 359686873 488660306 392404112 343385516 225620397 296126160 119953017 187918093 488734245 384206719 383791404 386347103 330837065	V P P P P -P 	Image: Constraint of the second sec	-R
Leptospiraceae (12/13) Other Spirochaetes (0/35)	Leptospira weilii Leptospira santarosai Leptospira borghetersenii Leptospira biflexa Leptospira licerasiae Leptonema illini Turneriella parva Brachyspira intermedia Brachyspira hyodysenteriae Brachyspira murdochii Borrelia turicatae Borrelia turicatae Borrelia burgdorferi Borrelia burgdorferi Borrelia birgdorferi Borrelia afzelii Spirochaeta africana Spinochaeta thermophila Sphaerochaeta pleomorpha Trenonema denticala	359728223 359683791 116331894 1183220780 359686873 488660306 392404112 34385516 225620397 226126160 119953017 187918093 488734245 384206719 383791404 386347103 330837065 374317032 488766025		D            D            D            T-G         NNPN           D-G         VKYE           PTG         CGDP           EON         KPD           KPD         ECD           RCD         DCN           NCN         SCR           DCK         DSR           D         -	-R
Leptospiraceae (12/13) Other Spirochaetes (0/35)	Leptospira weilii Leptospira borgpetersenii Leptospira borgpetersenii Leptospira biflexa Leptospira licerasiae Leptonema illini Turneriella parva Brachyspira intermedia Brachyspira murdochii Borrelia turicatae Borrelia hermsii Borrelia hermsii Borrelia hermsii Borrelia fatelii Spirochaeta africana Spirochaeta africana Spinochaeta thermophila Sphaerochaeta pleomorpha Treponema bergnaburenge	359728223 339663791 116331894 183220780 339686873 48866036 332404112 343385516 225820397 226126160 119953017 1197318093 488734245 384206719 383791404 386347103 330837065 374317032	V           P           P           P           P           AHICKG           I-M-E           I-M-E           O           I-M-E           O           P	D            D            T         GNPN           D         CGDP           EAN         KPD           KPD         ECD           RCD         DCN           NCN         SCR           DCK         DSR           D        R           NCR         SCR	-R
Leptospiraceae (12/13) – Other Spirochaetes – (0/35)	Leptospira weilii Leptospira santarosai Leptospira borgetersenii Leptospira biflexa Leptospira licerasiae Leptonema illini Turneriella parva Brachyspira intermedia Brachyspira hydysenteriae Brachyspira murdochii Borreila turicatae Borreila hermsii Borreila burgdorferi Borreila dirgdorferi Borreila afzelii Spirochaeta africana Spirochaeta thermopila Sphaerochaeta pleomorpha Treponema denticola Treponema primitia	359728223 359683791 116331894 116321894 1183220780 359686873 488660306 392404112 433385516 225620397 296126160 119953017 187918093 488734245 384206719 383791404 386347103 386347103 236447103 236447103 236447103 248766025 374317032	V P P P P -P 	D            D            T-G         NNPN           D-G         VKYE           PTG         GGDP           EGN         KPD           KPD         ECD           RCD         DCN           NCN         SGR           DCK         DSR           D-Q        R           NCR         SGR	-R
Leptospiraceae (12/13) – Other Spirochaetes – (0/35)	Leptospira weilii Leptospira borgpetersenii Leptospira borgpetersenii Leptospira biflexa Leptospira licerasiae Leptonema illini Turneriella parva Brachyspira intermedia Brachyspira murdochii Borrelia turicatae Borrelia turicatae Borrelia hermsii Borrelia direlii Spirochaeta africana Spirochaeta africana Spirochaeta africana Spirochaeta africana Spirochaeta africana Spirochaeta africana Spirochaeta coccides Sphaerochaeta coccides Sphaerochaeta coccides Sphaerochaeta pleomorpha Treponema brennaborense Treponema primitia	359728223 339663731 116331894 133220780 359666673 488860306 332404112 245320580 225620397 296126160 119953017 187918093 488734245 384206719 383791404 386547103 38426719 383791404 386547103 230837065 374317032 488766025	V           P	Image: Constraint of the second sec	-R
Leptospiraceae (12/13) Other Spirochaetes (0/35)	Leptospira weilii Leptospira borgpetersenii Leptospira borgpetersenii Leptospira licerasiae Leptonema illini Turneriella parva Brachyspira hydysenteriae Brachyspira hydysenteriae Brachyspira hydysenteriae Brachyspira hydysenteriae Brachyspira hydysenteriae Borrelia turicatae Borrelia hermsii Borrelia hermsii Borrelia hermsii Borrelia hermsii Borrelia hermsii Borrelia hermsii Spirochaeta africana Spinochaeta tafricana Spinochaeta tafricana Sphaerochaeta pleomorpha Treponema denticola Treponema primatia Treponema primitia	359728223 359663791 116331894 183220780 359668673 488660306 392404112 343385516 225620397 296126160 119953017 187918093 488734245 384206719 383791404 386347103 330837065 374317032 488766025 332297434 333999524 189026240	V P P P P -P 	D            D            D            D         0           VKYE         PTG           CGDP         ECN           KPD         KPD           ECD         CCN           NCN         SCR           DCN         NCN           SCR         DCN           NCN         SCR           DCN         NCN           SCR         SCR           DCN         NCR           K-G         SCR	-R
Leptospiraceae (12/13) –	Leptospira weilii Leptospira santarosai Leptospira borgpetersenii Leptospira biflexa Leptospira licerasiae Leptonema illini Turneriella parva Brachyspira intermedia Brachyspira hydysenteriae Brachyspira murdochii Borrelia turicatae Borrelia hermsii Borrelia burgdorferi Borrelia burgdorferi Borrelia afzelii Spirochaeta africana Spirochaeta thermopila Sphaerochaeta pleomorpha Treponema denticola Treponema brennaborense Treponema pallidum Eubacteriaceae bacterium Ciactotidum eticklendii	359728223 359683791 116331894 116321894 1183220780 359686873 488660306 392404112 433385516 225620397 296126160 119953017 187918093 488734245 384206719 383791404 386347103 330837065 374317032 488766025 332297434 33399524 1189026240 363892665	V P P P P P -P 	D            D            T-G         NNPN           D-G         VKYE           PTG         GGDP           EGN         KPD           KPD         KPD           KPD         DON           NCR         SCR           DCK         DSR           D        R           NCR         SCR           DK         SCR           DCK         SCR           DCA        R           NCR         SCR           DCN	-R
Leptospiraceae (12/13) –	Leptospira weilii Leptospira borgpetersenii Leptospira borgpetersenii Leptospira biflexa Leptospira licerasiae Leptonema illini Turneriella parva Brachyspira intermedia Brachyspira murdochii Borrelia turicatae Borrelia turicatae Borrelia hermsii Borrelia dirzlii Spirochaeta africana Spirochaeta africana Spinochaeta africana Spinochaeta africana Spinochaeta africana Spinochaeta coccides Sphaerochaeta coccides Sphaerochaeta pleomorpha Treponema brennaborense Treponema primitia Treponema plildum Eubacteriaceae bacterium Clostridium sticklandii	359728223 339663731 116331894 133220780 359666873 488860306 332404112 245303516 125520397 296126160 119953017 187918093 488734245 384206719 383791404 386547103 380837065 374317032 488766025 374317032 488766025 332297434 33399524 183399524	V           P           P           P           P           F-TD           I-M-E-GG           I-M-E-GG           I-M-E-GG           P-DT.IFV-T-K-GSV           P-DT-IFV-T-K-CSV           P-DT-IFV-T-K-CSV           P-DT-IFV-T-K-CSV           P-DT-IFV-T-K-CSV           P-DT-IFV-T-K-CSV           P-DT-IFF-T-VPCSV           P-DT-IFF-T-VPCSV           P-D-DT-IFF-T-KP8C-           P-D-D-IFF-T-KRACSE           P-D-D-IFF-T-KRACSE           P-D-D-IFF-T-VPCSV           P-D-D-IFF-T-VPCSV           P-D-D-IFF-T-VPCSV           P-D-DT-IFF-T-VPCSV           P-D-T-IFF-T-VPCSV           P-D-T-IFF-T-VPCSV           P-D-IT-IF-T-KACSE           P-D-D-IFF-T-VPCSV	Image: Constraint of the second sec	-R
Leptospiraceae (12/13) Other Spirochaetes (0/35)	Leptospira weilii Leptospira borgpetersenii Leptospira borgpetersenii Leptospira biflexa Leptospira licerasiae Leptonema illini Turneriella parva Brachyspira intermedia Brachyspira murdochii Borrelia turicatae Borrelia hermsii Borrelia hermsii Borrelia hermsii Borrelia hermsii Borrelia hermsii Borrelia tafziii Spirochaeta africana Spinochaeta africana Spinochaeta termophila Sphaerochaeta pleomorpha Treponema denticola Treponema primitia Treponema primitia Treponema palldum Eubacteriaceae bacterium Clostridium sticklandii Peptostreptococcus stomatis	359728223 359663791 116331894 183220780 359666873 48866030 392404112 225620397 296126160 119953017 187918093 488734245 384206719 383791404 386347103 330837065 374317032 488766025 332297434 33399524 18926240 36389265 310658795 307243232	V P P P P P -P 	D            D            T-G         NNPN           D-G         VKYE           PTG         GGDP           EON         KPD           KPD         ECD           RCD         DCN           NCR         SCR           DD-0        R           NCR         K-G           SCR         DPN           DPN         DPN           DP0         DCN	-R
Leptospiraceae (12/13) –	Leptospira weilii Leptospira santarosai Leptospira borgpetersenii Leptospira biflexa Leptospira licerasiae Leptonema illini Turneriella parva Brachyspira intermedia Brachyspira hydysenteriae Brachyspira murdochii Borrelia turicatae Borrelia hermsii Borrelia burgdorferi Borrelia dirgdorferi Borrelia afzelii Spirochaeta africana Spirochaeta thermopila Sphaerochaeta thermopila Sphaerochaeta pleomorpha Treponema denticola Treponema brennaborense Treponema pallidum Eubacteriaceae bacterium Clostridium sticklandii Peptostreptococcus stomatis Filifactor alocis	359728223 359683791 116331894 116321894 1183220780 359686873 488660306 392404112 433385516 225620397 296126160 119953017 187918093 488734245 384206719 3837045 384206719 3837044 386547103 386547103 386547103 386547103 236647103 330837065 374317032 488766025 33297434 33399524 13905240 36389265 310658795 310658795 307242322 320120352	V           P	D            D            T-G         NNPN           D-G         VKYE           PT         GGDP           EGN         KPD           KPD         KCD           DCN         NCN           SCR         DCK           DSR         D           NCR         SCR           DCK         SSR           DCK         SSR           DCK         SCR           DCH         CCD           DPN         CCD           DEH         K-G           SCR         DPN           CED         DEH	-R
Leptospiraceae (12/13)	Leptospira weilii Leptospira borgpetersenii Leptospira borgpetersenii Leptospira licerasiae Leptonema illini Turneriella parva Brachyspira intermedia Brachyspira nurdochii Borrelia turicatae Borrelia hermsii Borrelia hermsii Borrelia dafzelii Spirochaeta africana Spirochaeta coccides Sphaerochaeta coccides Sphaerochaeta coccides Treponema brennaborense Treponema primitia Treponema pallidum Eubacteriaceae bacterium Clostridium sticklandii Peptostreptococcus stomatis Filifactor alocis	359728223 339663791 116331894 133220780 359686873 488860306 332404112 343385516 225620397 296126160 119953017 1197518093 488734245 384206719 383791404 3863746025 374317032 488766025 374317032 488766025 332297434 33399524 139026240 363892665 310658796 307243232 320120352 306820448	V           P           P           P           P           P           P           P           P           I-M-E           FF-TD           P	Image: Constraint of the second sec	-R
Conter Spirochaetes - (0/35)	Leptospira weilii Leptospira borgpetersenii Leptospira borgpetersenii Leptospira licerasiae Leptonei alieni Turneriella parva Brachyspira hydysenteriae Brachyspira hydysenteriae Brachyspira hydysenteriae Brachyspira hydysenteriae Brachyspira hydysenteriae Borrelia turicatae Borrelia burgdorferi Borrelia burgdorferi Borrelia dirzelii Spirochaeta africana Spinochaeta africana Spinochaeta africana Sphaerochaeta africana Treponema brennaborense Treponema primitia Treponema primitia Treponema pimitia Treponema pimitia Diostridium eticklandii Peptostreptococcus stomatis Filifactor alocis Eubacterium yunii Thermoanaerobacterium thermoza	359728223 359663791 116331894 183220780 359666873 488660306 392404112 343385516 225620397 296126160 119953017 187918093 488734245 384206719 383791404 386347103 30837065 374317032 488766025 332297434 333999524 189026240 363892645 31058795 307243232 320120352 306820448 304316715	V           P	Image: Constraint of the second sec	-R
Leptospiraceae (12/13) – Other Spirochaetes – (0/35)	Leptospira weilii Leptospira sontarosai Leptospira borgpetersenii Leptospira biflexa Leptospira licerasiae Leptomem illini Turneriella parva Brachyspira intermedia Brachyspira murdochii Borrelia turicatae Borrelia turicatae Borrelia hermsii Borrelia direlii Spirochaeta afzelii Spirochaeta afzelii Spirochaeta thermophila Sphaerochaeta coccides Sphaerochaeta coccides Sphaerochaeta pleomorpha Treponema primitia Treponema primitia Treponema pimitia Freponema pimitia Filifactor alocis Eubacterium yurii Thermoanaerobacter imathranii	359728223 359683791 116331894 116331894 1183220780 359686873 48860306 392404112 433385516 225620397 226126160 119953017 187918093 488734245 384206719 383791404 386547103 386547103 386547103 386547103 386547103 2374317032 488766025 332297434 33399524 139026240 363892655 310658795 3107243232 306820448 3007245232 306820448		Image: Constraint of the second sec	-R
Conter Spirochaetes (0/35)	Leptospira weilii Leptospira borgpetersenii Leptospira borgpetersenii Leptospira licerasiae Leptonema illini Turneriella parva Brachyspira intermedia Brachyspira nurdochii Borrelia turicatae Borrelia hermsii Borrelia hermsii Borrelia hermsii Borrelia fatelii Spirochaeta africana Spinochaeta africana Spinochaeta africana Spinochaeta thermophila Sphaerochaeta pleomorpha Treponema denticola Treponema primitia Treponema primitia Treponema pallidum Eubacteriaceae bacterium Clostridium sticklandii Peptostreptococcus stomatis Filifactor alocis Eubacteriaerobacter mathranii Cadicellulosiruptor lactoacet	359728223 339663791 116331894 183220780 339686873 48860306 332404112 343385516 225820397 226126160 119953017 1197318093 488734245 384206719 383791404 386347103 330837065 374317032 488766025 332297434 33399524 18926240 33399524 18926240 363892665 310658795 307243232 306820448 304316715 306820448 304316715	V           P           P           P           P           P           P           P           P           I-M-E           P           DT           P <td>D            D            T-G         NNPN           D-G         VKYE           PTG         GGDP           EON         KPD           ECD         RCD           DCN         NCN           SSGR         DCN           DCN        R           NCR         K-G           SSGR         DPN           DPD         CED           DPN         DPN           DPD         CED           KPT         KPT           KKPT         KPT</td> <td>-R</td>	D            D            T-G         NNPN           D-G         VKYE           PTG         GGDP           EON         KPD           ECD         RCD           DCN         NCN           SSGR         DCN           DCN        R           NCR         K-G           SSGR         DPN           DPD         CED           DPN         DPN           DPD         CED           KPT         KPT           KKPT         KPT	-R
Leptospiraceae (12/13) – Other Spirochaetes – (0/35) –	Leptospira weilii Leptospira borgpetersenii Leptospira borgpetersenii Leptospira licerasiae Leptonena illini Turneriella parva Brachyspira hydysenteriae Brachyspira hydysenteriae Brachyspira hydysenteriae Brachyspira hydysenteriae Brachyspira hydysenteriae Borrelia turicatae Borrelia burgdorferi Borrelia fizelii Spirochaeta africana Spirochaeta africana Sphaerochaeta africana Sphaerochaeta africana Treponema brennaborense Treponema primitia Treponema primitia Treponema primitia Filifactor alocis Eubacterium yurii Thermoanaerobacterium thermosa Thermoanaerobacterium thermosa	359728223 359663791 116331894 183220780 35966873 48860306 392404112 343385516 225620397 296126160 119953017 187918093 488734245 384206719 383791404 386347103 30837065 332297434 333999524 189026240 36389265 307243232 300724323 3007243232 300724323 300724323 300724323 3007243232 300724323 30072433 300724323 30072433 30072433 30072433 3007243 30	V           P	Image: Constraint of the second sec	-R
Leptospiraceae (12/13) - Other Spirochaetes - (0/35) -	Leptospira weilii Leptospira borgptersenii Leptospira borgptersenii Leptospira biflexa Leptospira licerasiae Leptonema illini Turneriella parva Brachyspira intermedia Brachyspira murdochii Borrelia turicatae Borrelia turicatae Borrelia hermsii Borrelia direlii Spirochaeta africana Spirochaeta africana Spirochaeta africana Spirochaeta africana Spirochaeta coccides Sphaerochaeta pleomorpha Treponema primitia Treponema primitia Treponema primitia Freponema primitia Filifactor alocis Eubacteriaceae bacterium Clostridium sticklandii Peptostreptococcus stomatis Filifactor alocis Eubacterium yurii Thermoanaerobacter mathranii Caldicellulosiruptor lactaaet Natranaerobius thermophilus Ethanoligenens harbinense	359728223 339663731 116331894 133220780 359666873 488860306 332404112 24520397 226520397 226126160 119953017 187918093 488734245 384206719 383791404 386547103 380837065 374317032 488766025 332297434 33399524 18392665 3107243232 307243232 307243232 306820448 304316715 297544523 34496005	V           P           P           P           P           P           P           P           P           I-M-E           F-TD           P<	Image: Constraint of the	-R
Conter Spirochaetes - (0/35)	Leptospira weilii Leptospira borgpetersenii Leptospira borgpetersenii Leptospira licerasiae Leptonema illini Turneriella parva Brachyspira intermedia Brachyspira nurdochii Borrelia turicatae Borrelia hermsii Borrelia hermsii Borrelia hermsii Borrelia fatelii Spirochaeta africana Spinochaeta africana Spinochaeta africana Spinochaeta thermophila Sphaerochaeta coccoides Sphaerochaeta pleomorpha Treponema denticola Treponema primitia Treponema pallidum Eubacteriaceae bacterium Clostridium sticklandii Peptostreptococcus stomatis Filifactor alocis Eubacterium yurii Thermoanaerobacter mathranii Caldicellulosiruptor lactoacet Natranaerobis thermophilus Ethanoligenens harbinenese Acetivibrio cellulolyticus	359728223 339663791 116331894 183220780 339686873 48860306 332404112 343385516 225820397 226126160 119953017 187918093 488734245 384206719 383791404 38637065 374317032 488766025 332297434 33399524 189026240 363892665 31058795 307243232 320120352 306820448 304316715 306820448 304316715 306820448 304316715 306820448 304316715 306820448 304316715 306820448 304316715 306820448 304316715 306820448 304316715 306820448 304316715 306820448 304316715 3068205 307555 307555 307555 307555 307555 3075555 3075555555555	V           P           P           P           P           P           FF-TD           I-M-E           FF-TD           I-M-E           P	D            D            T         NNPN           D         VKYE           PTO         GGDP           EON         KPD           KPD         CGDP           DCN         NCR           K         SCR           DD-        R           NCR         SCR           DDN         DON           DD-         OR           NCR         SCR           DDPD         DDD           CED         DPN           DPD         DEH           HKD         KPT           KPT         SPD           KPT         KPT	-R
Leptospiraceae (12/13)	Leptospira weilii Leptospira borgpetersenii Leptospira borgpetersenii Leptospira biflexa Leptospira licerasiae Leptonema illini Turneriella parva Brachyspira intermedia Brachyspira hydysenteriae Brachyspira murdochii Borrelia turicatae Borrelia turicatae Borrelia burgdorferi Borrelia direta africana Spirochaeta africana Spirochaeta africana Sphaerochaeta africana Sphaerochaeta africana Treponema brennaborense Treponema primitia Treponema primitia Treponema primitia Eubacteriaceae bacterium Clostridium sticklandii Peptostreptococcus stomatis Filifactor alocis Eubacterium yurii Thermoanaerobacterium thermosa Thermoanaerobacter mathranii Caldicelluosiruptor lactoacet Natranaerobius thermophilus Ethanoligenems harbinense Acetivibrio cellulolyticus	359728223 359663791 116331894 183220780 35966873 48860306 392404112 343385516 225620397 296126160 119953017 187918093 488734245 384206719 383791404 386347103 30837065 332297434 189026240 36389265 307243232 306820448 304316715 297544523 306820448 304316715 297544523 31439005 188586395 317133026	V           P	Image: Constraint of the second sec	-R

FIGURE 6 | Partial sequence alignments of (A) 50S Ribosomal protein L14 and (B) Alanyl-tRNA synthetase, showing two CSIs that are specific for the family *Leptospiraceae*, but not found in the sequence **homologs of any other sequenced bacteria.** Sequence information for other *Leptospiraceae* specific CSIs is presented in Supplemental Figures 11–13 and summarized in **Table 4**.

## Table 4 | Conserved signature Indels that are specific for members of the family Leptospiraceae.

Protein name	Gene	GI	Figure	Indel	Indel
	name	number	Number	Size	Position
50S Ribosomal protein L14	rplN	5163214	Figure 6A	8 aa ins	36–73
Alanyl-tRNA synthetase	alaS	45656657	Figure 6B	4 aa ins	165–211
30S Ribosomal protein S2	rpsB	116330588	Supplemental Figure 11	2 aa ins	108–141
Flagellar filament core protein FlaB Flagellar basal-body rod protein FlgG	flaB flgG	12657818 294828153	Supplemental Figure 12 Supplemental Figure 13	4 aa del 1 aa ins	130–168 80–123

			148	184
	Treponema pallidum	15639102	GGSLLGTSRGGG N RVVDIVDGIERLNLHILF	FIIGGDG
	Treponema paraluiscuniculi	338706086		
	Treponema phagedenis	320536657	·····	-V
	Treponema succinifaciens	328949033	TF D - TS S MGINMM	
	Treponema vincentii	257456783	TI TEN	-V
Trenonema	Treponema denticola	42525583	F TTE QM- INMV	•F
Spirochasta and	Treponema primitia	333999680	TFSA K E-EKAM-QMNV-	·T····
Sphoendeld, and	Treponema azotonutricium	333993411	TYSA K E-GKAQNM	·T····
Sphaerochaela	Treponema caldaria	339498823	TIA K Q-TEAN	·A
(18/18)	Spirochaeta thermophila	307717981	TIS E -TEELAL-QINV	·T····
	Spirochaeta africana	373484532	I E -TEETMN	·T····
	Spirochaeta smaragdinae	302339733	TVSA DGE-A-A INM	·T····
	Sphaerochaeta coccoides	330837727	TIS D QISESLM-VN	VAV
	Sphaerochaeta pleomorpha	359352086	TISA K Q-EESLM-IN)	/1/
	Sphaerochaeta globosa	325972723		
	Borrella sp. SVI	220001773		-N
	Borrelia bissettii	343128024	TT-ST KP-ETIW-TNWT	N
	Borrelia burgdorferi	218249692	TT.ST KP.FTIN.TNNT.	N
Other	Borrelia garinii	219684340	TISI KP-FTLM-INMT	N
Spirochaetes -	Borrelia valaisiana	224531593	TISI KP-ETLM-INMI	N
(0/30)	Leptonema illini	373874404	TISNQ DP-IMTLISHGVN	-AV
(0,50)	Leptospira santarosai	359683664	M-ASNQ SP-EMA-CLVLYGIK	-C
	Leptospira weilii	359725956	M-SSNQ SPIEMCLSFYGVKM-	·C
	Leptospira licerasiae	359688124	TI-ASNQ SPS-MRLSLYGVKM	-C
	Leptospira interrogans	24217172	M-SSNQ SPEEMSLSFYGVK	-C
	Desulfohalobium retbaense	258406234	TMSHQ SAEEALM-IN	-M
	Syntrophobacter fumaroxidans	116750380	TISPQ D-LEMTLD-M-IG	·T····
	Desulfomicrobium baculatum	256828343	TI-SSPQ APEEALM-ISV	-M
04	Lawsonia intracellularis	94986497	TISPQ QPEEASLVHH-IN	·V
Other	Sorangium cellulosum	162450529	TVNQ DPHQMTL-A-GINV	-VV
Bacteria -	Kineococcus radiotolerans	152968217	TIQQ DA-EVCLMGIS	-V
(0/>250)	Frankia alni	111223772	TISQQ DPGECLS-M-IN	-V
	Aeromicrobium marinum	311/4205/	TT EO DEAE CL MOTHU	·V
	Versusomicsobium spinosum	171013866	THSHO NTG-MTL-DHKVD	·v
	Victivallis vadensis	281358019	TVSRQ DE-TM-ETLK-M-IK	·C
			101 III III 101 GO ADHOIDIN	
-				
В			275	319
В	Borrelia garinii	219685531	275 LIPELDFDIEGPNGFLVHLERRLL EKESLE	319 EIPHAVILIAEGAGQ
В	Borrelia garinii Borrelia spielmanii	219685531 224534492	275 LIPELDFDIEGPNGFLVHLERRLL EKESLE	319 EIPHAVILIAEGAGQ
B Borrelia	Borrelia garinii Borrelia spielmanii Borrelia burgdorferi	219685531 224534492 218249692	275 LIPELDFDIEGPNGFLVHLERRLL EKESLE	319 EIPHAVILIAEGAGQ
B Borrelia (12/12) -	Borrelia garinii Borrelia spielmanii Borrelia burgdorferi Borrelia valaisiana Borrelia efectii	219685531 224534492 218249692 224531593	275 LIPELDFDIEGPNGFLVHLERRLL EKESLE	319 EIPHAVILIAEGAGQ
B Borrelia (12/12) –	Borrelia garinii Borrelia spielmanii Borrelia burgdorferi Borrelia valaisiana Borrelia afzelii Borrelia sp. 3/1	219685531 224534492 218249692 224531593 111115557 225551773	275 LIPELDFDIEGPNGFLVHLERRLL EKESLE	319 EIPHAVILIAEGAGQ
B Borrelia (12/12) –	Borrelia garinii Borrelia spielmanii Borrelia burgdorferi Borrelia atzelii Borrelia atzelii Borrelia sp. SV1 Barrelia crociduree	219685531 224534492 218249692 224531593 111115557 225551773 386859963	275 LIPELDFDIEGPNGFLVHLERRLL EKESLE	319 EIPHAVILIAEGAGQ
B Borrelia (12/12) -	Borrelia garinii Borrelia spielmanii Borrelia burgdorferi Borrelia afzelii Borrelia afzelii Borrelia crozidurae Tregonema succinifaciena	219685531 224534492 218249692 224531593 111115557 225551773 386859963 328949033	275 LIPELDFDIEGPNGFLVHLERRLL EKESLE	319 EIPHAVILIAEGAGQ
B Borrelia (12/12) –	Borrelia garinii Borrelia spielmanii Borrelia valaisiana Borrelia valaisiana Borrelia afzelii Borrelia sp. SV1 Borrelia crocidurae Treponema succinifaciens Treponema phagedenis	219685531 224534492 218249692 224531593 111115557 225551773 388659963 328949033 320536657	275 LIPELDFDIEGPNGFLVHLERRLL EKESLE	319 EIPHAVILIAEGAGQ
B Borrelia (12/12) —	Borrelia garinii Borrelia spielmanii Borrelia burgdorfori Borrelia valaisiana Borrelia afzelii Borrelia sp. SV1 Borrelia crocidurae Treponema phagedenis Treponema brenaborense	219685531 224534492 218249692 224531593 111115557 225551773 386859963 328949033 320536657 332297140	275 LIPELDFDIEGPNGFLVHLERRLL EKESLE	319 EIPHAVILIAEGAGQ 
B Borrelia (12/12) —	Borrelia garinii Borrelia spielmanii Borrelia burgdorferi Borrelia valaisiana Borrelia afzelii Borrelia sp. SV1 Borrelia crocidurae Treponema succinifaciens Treponema phagedenis Treponema azotonutricium	219685531 224534492 218249692 224531593 111115557 225551773 386859963 328949033 320536657 332297140 333993411	275 LIPELDFDIEGPNGFLVHLERRLL EKESLE 	319 EIPHAVILIAEGAGQ 
B Borrelia (12/12) –	Borrelia garinii Borrelia spielmanii Borrelia valaisiana Borrelia ditelii Borrelia atzelii Borrelia atzelii Borrelia crocidurae Treponema buccinifaciens Treponema bragedenis Treponema azotonutricium Treponema azotonutricium	219685531 224534492 218249692 224531593 111115557 225551773 386859963 328949033 320536657 332297140 333993411 339498823	275 LIPELDFDIEGPNGFLVHLERRLL EKESLE	319 EIPHAVILIAEGAGO 
B Borrelia (12/12) -	Borrelia garinii Borrelia spielmanii Borrelia valaisiana Borrelia valaisiana Borrelia afzelii Borrelia sp. SV1 Borrelia crocidurae Treponema piagedenis Treponema brennaborense Treponema azotonutricium Treponema caldaria Spirochaeta africana	219685531 224534492 218249692 224531593 111115557 326551773 386859963 32694903 320536657 332297140 333993411 333993411	275 LIPELDFDIEGPNGFLVHLERRLL EKESLE	319 EIPHAVILIAEGAGQ 
B Borrelia (12/12) -	Borrelia garinii Borrelia spielmanii Borrelia burgdorferi Borrelia atzelii Borrelia atzelii Borrelia op. 3V1 Borrelia crocidurae Treponema phagedenis Treponema phagedenis Treponema zotonutricium Treponema caldaria Spirochaeta africana Spharechaeta pleemorpha	219685531 224534492 218246692 224531593 111115557 225551773 386859953 326949033 326949033 326949033 32297140 333999411 339498823 373484532 359352066	275 LIPELDFDIEGPNGFLVHLERRLL 	319 EIPHAVILIAEGAGQ 
B Borrelia (12/12) -	Borrelia garinii Borrelia spielmanii Borrelia burgdorferi Borrelia drzelii Borrelia drzelii Borrelia crocidurae Treponema succinifaciens Treponema brennaborense Treponema azotonutricium Treponema azotonutricium Spiaerochaeta africana Sphaerochaeta pleomorpha Brachyspira murdochii	219688531 224534492 218249692 224531593 3111115557 225551773 326859963 320536657 332297140 333993411 339498823 373484532 35952086 296127381	275 LIPELDFDIEGPNGFLVHLERRLL 	319 EIPHAVILIAEGAGQ 
B Borrelia (12/12) -	Borrelia garinii Borrelia spielmanii Borrelia valaisiana Borrelia distiana Borrelia distelii Borrelia sp. SV1 Borrelia crocidurae Treponema buccinifaciens Treponema buccinifaciens Treponema brennaborense Treponema cotonutricium Treponema caldaria Spirochaeta africana Spirochaeta africana Sphaerochaeta pleomorpha Brachyspira murdochii Brachyspira intermedia	219685531 224534492 218249692 224531593 111115557 225551773 3868559653 328949033 320536657 332297140 339993411 339498823 373484532 359352066 296127381 384208173	275 LIPELDFDIEGPNGFLVHLERRLL EKESLE 	319 EIPHAVILIAEGAGQ 
B Borrelia (12/12) -	Borrelia garinii Borrelia spielmanii Borrelia burgdorfori Borrelia dzelii Borrelia atzelii Borrelia op. SV1 Borrelia cocidurae Treponema phagedenis Treponema zatonutricium Treponema caldaria Spirochaeta africana Sphaerochaeta pleomorpha Brachyspira intermedia Turneriella parva	219685531 224534492 218246692 224531593 3111115557 225551773 386859963 32694903 32694903 32694903 3299740 333993411 33999411 339994823 773444532 359352086 296127381 36420817 36420817 392401968	275 LIPELDFDIEGPNGFLVHLERRLL 	319 EIPHAVILIAEGAGQ 
B Borrelia (12/12) -	Borrelia garinii Borrelia spielmanii Borrelia burgdorferi Borrelia valaisiana Borrelia atzelii Borrelia atzelii Borrelia sp. SV1 Borrelia crocidurae Treponema succinifaciens Treponema brennaborense Treponema aztonutricium Treponema aztonutricium Spiaerochaeta pleomorpha Brachyspira murdochii Brachyspira intermedia Turneriella parva Leptonema illini	219685531 224534492 218249692 224531593 3111115557 225551773 32685993 320536657 332297140 333993411 339498823 373484552 359352086 296127361 384208173 392401968 48859579	275 LIPELDFDIEGPNGFLVHLERRLL EKESLE 	319 EIPHAVILIAEGAGQ 
B Borrelia (12/12) -	Borrelia garinii Borrelia spielmanii Borrelia valaisiana Borrelia dizelii Borrelia dizelii Borrelia crocidurae Treponema auccinifaciens Treponema buccinifaciens Treponema catdaria Spinochaeta africana Sphaerochaeta pleomorpha Brachyspira murdochii Brachyspira intermedia Turmeriella parva Leptonema illini Leptospira licerasiae	219688531 224534492 218249692 224531593 111115557 225551773 326865963 320536657 332297140 339993411 339498823 373484532 359352086 296127381 384208173 392401968 488855579 359888124	275 LIPELDFDIEGPNGFLVHLERRLL EKESLE	319 EIPHAVILIAEGAGQ 
B Borrelia (12/12) -	Borrelia garinii Borrelia spielmanii Borrelia burgdorferi Borrelia dzelii Borrelia atzelii Borrelia op. SV1 Borrelia cocidurae Treponema succinifaciens Treponema brenaborense Treponema caldaria Spirochaeta africana Sphaerochaeta africana Sphaerochaeta africana Brachyspira intermedia Turneriella parva Leptonema illini Leptospira licerasiae Leptospira noguchii	219685531 224534492 224531593 111115557 225551773 386859963 32694903 32694903 3229740 333993411 33999411 33999411 33999413 359352086 296127381 84208173 392401968 48859579 359968124 359726595	275 LIPELDFDIEGPNGFLVHLERRLL EKESLE	319 EIPHAVILIAEGAGQ 
B Borrelia (12/12) -	Borrelia garinii Borrelia spielmanii Borrelia burgdorferi Borrelia valaisiana Borrelia atzelii Borrelia otzelii Borrelia otzelii Borrelia crocidurae Treponema phagedenis Treponema phagedenis Treponema caldaria Spiarconaeta africana Sphaerochaeta africana Sphaerochaeta africana Sphaerochaeta africana Sphaerochaeta africana Sphaerochaeta pleemorpha Brachyspira murdochii Brachyspira intermedia Turneriella parva Leptonema illini Leptospira spantareai	219685531 224534492 218246692 224531593 111115557 225551773 326859953 32297140 333993411 339498823 373484532 255952086 296127381 384208173 392401968 48855579 359688124 48855579	275 LIPELDFDIEGPNGFLVHLERRLL EKESLE	319 EIPHAVILIAEGAGQ 
B Borrelia (12/12) -	Borrelia garinii Borrelia spielmanii Borrelia valdisiana Borrelia valdisiana Borrelia drželii Borrelia croidurae Treponema succinifaciens Treponema brennaborense Treponema azotonutricium Treponema azotonutricium Spiaerochaeta pleomorpha Brachyspira intermedia Turneriella parva Leptospira licerasiae Leptospira licerasiae Leptospira santarosai Desulfarculus baarsii	219688531 224534492 218249692 224531593 3265551773 326855963 320536657 332297140 333995411 339498823 373484532 35952066 296127381 384208173 392401968 488855579 359688124 359723595	275 LIPELDFDIEGPNGFLVHLERRLL EKESLE	319 EIPHAVILIAEGAGQ 
B Borrelia (12/12) -	Borrelia garinii Borrelia spielmanii Borrelia burgdorferi Borrelia dzelii Borrelia azelii Borrelia op. SV1 Borrelia op. SV1 Borrelia cocidurae Treponema succinifaciens Treponema brenaborense Treponema caldaria Spirochaeta africana Sphaerochaeta africana Sphaerochaeta africana Brachyspira intermedia Turneriella parva Leptonema illini Leptospira licerasiae Leptospira santarosai Desulfarculus baaresii Law, intracellularia	219685531 224534492 224531593 111115557 225551773 386859963 32694903 32694903 3229740 333993411 33999411 33999411 33999413 359352086 296127361 29592086 296127361 3592086124 359723595 359603664 295923595 359603664 295923595	275 LIPELDFDIEGPMGFLVHLERRLL EKESLE	319 EIPHAVILIAEGAGQ 
B Borrelia (12/12) -	Borrelia garinii Borrelia spielmanii Borrelia valaisiana Borrelia valaisiana Borrelia atzelii Borrelia otzelii Borrelia otzelii Borrelia crocidurae Treponema succinifaciens Treponema piagedenis Treponema zatonutricium Treponema caldaria Spirochaeta africana Sphaerochaeta pieemorpha Brachyspira murdochii Brachyspira murdochii Brachyspira murdochii Brachyspira intermedia Turneriella parva Leptonema illini Leptospira santarocai Desulfarculus baarsii Law, intracelularis Sorangium cellulosum Desulfarculus pareis	219685531 224534492 218246692 224531593 111115557 225551773 386859953 32297140 333999411 339496823 373444532 296127381 384208173 392401968 488659579 359688124 35972595 55968364 302342145 2598686497 162450529	275 LIPELDFDIEGPNGFLVHLERRLL EKESLE	319 EIPHAVILIAEGAGQ 
B Borrelia (12/12) -	Borrelia garinii Borrelia spielmanii Borrelia valaisiana Borrelia valaisiana Borrelia atzelii Borrelia atzelii Borrelia orooidurae Treponema brennaborense Treponema brennaborense Treponema aztonutricium Treponema aztonutricium Treponema aztonutricium Sphaerochaeta pleomorpha Brachyspira murdochii Brachyspira murdochii Brachyspira intermedia Turneriella parva Leptonema illini Leptospira liceraine Leptospira neguchi Leptospira santarosai Desulfarculus baarsii Law. intracellularia Sorangium cellulosum Desulfoyibrio vulgaris Gender. obscurve	219685531 224534492 218249692 224531593 326551773 326655963 32297140 333992411 339498823 373484552 296127381 384208173 392401958 48855579 359688124 35972559 359688124 35972559 3596886497 162450529 218885867	275 LIPELDFDIEGPNGFLVHLERRLL EKESLE	319 EIPHAVILIAEGAGQ 
B Borrelia (12/12) -	Borrelia garinii Borrelia spielmanii Borrelia burgdorfori Borrelia valaisiana Borrelia azelii Borrelia op. SVI Borrelia op. SVI Borrelia orocidurae Treponema brenaborense Treponema brenaborense Treponema caldaria Spirochaeta africana Sphaerochaeta africana Sphaerochaeta africana Sphaerochaeta africana Sphaerochaeta africana Brachyspira intermedia Turneriella parva Leptonema illini Leptospira licerasiae Leptospira nigentii Leptospira natoresii Leptospira santarosai Desulfarculus baarsii Law. intracellularis Sorangium cellulosum Desulfovibrio vulgaris Geoder. obscurus	219685531 224534492 224531593 3111115557 225551773 386859963 32294903 322934903 32293403 32293403 3229740 333993411 33993411 33993411 33993411 33993411 33993411 33993413 359352086 296127361 359352086 296127361 359352086 29612736 35968364 48855579 35968364 359725595 35968364 359725595 35968364 9496497 162450529 218888667 284891246	275 LIPELDFDIEGPMGFLVHLERRLL EKESLE	319 EIPHAVILIAEGAGQ 
B Borrelia (12/12) - Other Spirochaetes - (0/36) Other Bacteria (0/>250) -	Borrelia garinii Borrelia spielmanii Borrelia spielmanii Borrelia valaisiana Borrelia atzelii Borrelia op. SVI Borrelia op. SVI Borrelia cocidurae Treponema phagedenis Treponema caldaria Spirochaeta africana Spharechaeta africana Spharechaeta africana Spharechaeta africana Spharechaeta africana Spirochaeta africana Spharechaeta africana Spharechaeta africana Brachyspira murdochii Brachyspira murdochii Brachyspira intermedia Turneriella parva Leptonema illini Leptospira santarosai Desulfarculus baarsii Law. intracellularis Sorangium cellulosum Desulfovibrio vulgaris Geoder. obscurus Aeromicrobium marinum Amveol. methanolica	219685531 224534492 218246692 224531593 325551773 386859953 32297140 33399940823 373484532 359352086 296127381 384208173 392401988 488855579 359688124 35972505 35968124 302342145 94986497 162450529 21886567 284991246 311742057	275 LTPELDFDIEGPMGFLVHLERRLL EKESLE	319 EIPHAVILIAEGAGQ 

FIGURE 7 | (A) Partial sequence alignment of the protein

6-phosphofructokinase (pyrophosphate) containing a 1 amino acid insert in a conserved region that is specifically present in the species from the genera *Treponema, Spirochaeta, and Sphaerochaeta*, but not found in any other

sequenced bacteria. **(B)** Partial sequence alignment of phosphofructokinase containing a 6 amino acid insert that is specific for the genera *Borrelia*. Sequence information for other CSIs showing similar specificities is provided in **Table 5** and in Supplemental Figures 14–30.

Table 5	<b>Conserved Signature</b>	Indels that are spec	ific for groups withi	n the family Spirochaetaceae.
---------	----------------------------	----------------------	-----------------------	-------------------------------

Protein name	Gene name	GI Number	Figure Number	Specificity	Indel size	Indel position
6-phosphofructokinase (pyrophosphate)	pfp	15639102	Figure 7A	Treponema, Spirochaeta and Sphaerochaeta	1 aa ins	148–184
Bifunctional Hpr kinase/phosphatase	hprK	3322886	Supplemental Figure 14	Treponema, Spirochaeta and Sphaerochaeta	1 aa ins	183–221
30S ribosomal protein S13	rpsM	302337499	Supplemental Figure 15	Treponema, Spirochaeta and Sphaerochaeta	1 aa del	1–39
Phosphofructokinase	pfk	219685531	Figure 7B	Borrelia	6 aa ins	275–319
50S ribosomal protein L4	rplD	224534698	Supplemental Figure 16	Borrelia	1 aa ins	103–136
tRNA pseudouridine 55 synthase	truB	203284699	Supplemental Figure 17	Borrelia	2 aa ins	143–178
Translation elongation factor Tu	tuf	203284386	Supplemental Figure 18	Borrelia	1 aa del	330–369
Histidyl-tRNA synthetase	hisS	187918014	Supplemental Figure 19	Borrelia	1 aa del	273–301
Seryl-tRNA synthetase	serS	187918098	Supplemental Figure 20	Borrelia	1 aa del	231–264
Spoiiij-associtated protein	jag	219684344	Supplemental Figure 21	Borrelia	3 aa ins	114–154
Nicotinate phosphoribosyltransferase	pncB	187918492	Supplemental Figure 22	Borrelia	1 aa del	134–159
Ribose 5-phosphate isomerase	rpiA	119953435	Supplemental Figure 23	Borrelia	1 aa ins	86–110
Ribonuclease Z	rnz	195941574	Supplemental Figure 24	Borrelia	2 aa ins	64–94
Hypothetical protein BGAFAR04_0762	-	386859948	Supplemental Figure 25	Borrelia	1 aa ins	206–236
Signal recognition particle, subunit FFH/SRP54	-	119953471	Supplemental Figure 26	Borrelia	1 aa ins	374–412
Hypothetical protein BSV1_0075	-	15594416	Supplemental Figure 27	Borrelia	1 aa del	52–97
Aspartyl/glutamyl-tRNA amidotransferase subunit A	gatA	119953137	Supplemental Figure 28	Borrelia	1 aa ins	364–402
Ribosomal RNA methyltransferase	rlmE	203284234	Supplemental Figure 29	Borrelia	1 aa ins	15–48
LysM domain/M23/M37 peptidase domain protein	-	224534310	Supplemental Figure 30	Borrelia	1 aa ins	320–365

in protein-protein interactions (Akiva et al., 2008; Singh and Gupta, 2009; Gupta, 2010). Thus, the CSI identified in FlgC likely plays an important role in the cellular functions of the flagellar basal-body.

The phylum Spirochaetes contains 4 main lineages (viz. Spirochaetaceae, Brachyspiraceae, Leptospiraceae, and Brevinemataceae). These lineages have historically been distinguished from each other by their biochemical characteristics and their 16S rRNA gene sequences (Harwood and Canale-Parola, 1984; Paster et al., 1991; Paster, 2011a). In this study we have also identified 22 CSIs in a diverse range of proteins that are specific to each of the main sequenced lineages of the phylum Spirochaetes (viz. Spirochaetaceae, Brachyspiraceae, and Leptospiraceae), which serve to distinguish these lineages from themselves and all other bacteria. Seven of these identified CSIs were specific for the family Spirochaetaceae, 6 CSIs were identified that were specific for the family Brachyspiraceae, and 5 CSIs were identified that were specific to the family Leptospiraceae. Each of these lineages also branch distinctly and are separated by long branches in both 16S rRNA based and concatenated protein based phylogenetic trees (Figures 1, 2). This molecular and phylogenetic evidence supports the current division of these lineages. However, the large number of CSIs discovered for each of these groups and their genetic distances suggests that these lineages may represent higher taxonomic divisions (viz. orders or classes) than currently recognized. It is noteworthy that two of the CSIs that are specific for the Brachyspiraceae family and one that is specific for the Leptospiraceae are again found in flagella-related proteins (viz. FlgK, FlgB, FlgG) indicating that there might be interesting

differences in the structures and/or functions of flagella within the Spirochaete families.

The family Spirochaetaceae, which contains the genera Borrelia, Clevelandina, Cristispira, Diplocalyx, Hollandina, Pillotina, Sphaerochaeta, Spirochaeta, and Treponema, is the most diverse of the lineages within the phylum Spirochaetes (Paster, 2011b; Euzéby, 2013). The interrelationships between the genera within this family are not reliably resolved by 16S rRNA sequence analysis (Paster, 2011b) (Figure 2). In this study we have identified 19 CSIs which serve to delineate at least certain relationships within the family Spirochaetaceae. Three of the CSIs identified are specifically found in members of the genera Sphaerochaeta, Spirochaeta, and Treponema and 16 additional CSIs were identified that are specifically found in members of the genus Borrelia. These CSIs suggest that the genera Sphaerochaeta, Spirochaeta, and Treponema shared a common ancestor distinct from the members of the genus Borrelia. In our concatenated protein phylogenetic tree, the genera Sphaerochaeta, Spirochaeta and Treponema formed a well-supported monophyletic clade, which was separated from the members of the genus Borrelia by a long branch, supporting the relationship delineated by these CSIs. Both of these two clades also exhibit considerable phylogenetic diversity. The clade consisting of genera Sphaerochaeta, Spirochaeta, and Treponema contains a number of distinct smaller subclades while the members of the genus Borrelia form two highly distinct clades in the phylogenetic trees. However, further work to identify molecular markers will be required to determine the significance of the branching of these subclades. The genus Cristispira has not had its genome sequenced, but it branches



with the members of the genus *Borrelia* reliably in 16S rRNA based phylogenetic trees suggesting that some, if not all, of the *Borrelia* specific CSIs identified in this study may also be found in *Cristispira* (Paster, 2011b) (**Figure 2**). The remaining members of the family *Spirochaetaceae* (viz. *Clevelandina*, *Diplocalyx*, *Hollandina*, and *Pillotina*) have been identified in the hindguts of termite and cockroaches but have yet to be isolated and grown in

pure or mixed culture. The current placement of the identified members of *Clevelandina*, *Diplocalyx*, *Hollandina*, and *Pillotina* in distinct genera within the family *Spirochaetaceae* is ambiguous and based largely on analyses of morphological characteristics (Bermudes et al., 1988). No genome or 16S rRNA sequences are currently available from these genera for phylogenetic analysis. However, the observations presented in this report suggest that the family *Spirochaetaceae* contains at least two distinct monophyletic groups: one consisting of the genera *Sphaerochaeta*, *Spirochaeta*, and *Treponema* and another consisting of the genera *Borrelia* and *Cristispira*.

## TAXONOMIC IMPLICATIONS

The results presented here show that the main lineages of the phylum Spirochaetes are evolutionarily distinct. The families Spirochaetaceae, Brachyspiraceae, and Leptospiraceae are distinguished from each other and all other bacteria by large numbers of identified CSIs in widely distributed proteins. Additionally, these three families branch distinctly in both 16S rRNA based and concatenated protein based phylogenetic trees. The results presented here also show that the family Spirochaetaceae consists of two distinct monophyletic groups. The distinctiveness of these groups is supported by both molecular evidence, in the form of the large numbers of discovered CSIs, and phylogenetic analyses. Additionally, both of these distinct groups exhibit a large amount of phylogenetic diversity which is currently not reflected in their taxonomy. The current taxonomic organization of the phylum Spirochaetes places all of the main lineages (viz. Spirochaetaceae, Brachyspiraceae, Leptospiraceae, and Brevinemataceae) into a single order. However, to adequately recognize both distinctiveness of the main lineages within the phylum Spirochaetes and the distinctiveness and diversity of the two main groups within the family Spirochaetaceae, the main lineages of the phylum Spirochaetes would have to have their taxonomic rank increased. To recognize the distinctiveness of both the main lineages within the phylum Spirochaetes and the two main groups within the family Spirochaetaceae we are proposing a taxonomic rearrangement of the phylum as follows: We propose that the family Leptospiraceae be transferred to the novel order Leptospiriales ord. nov. within the class Spirochaetia, the family Brachyspiraceae be transferred to the novel order Brachyspiriales ord. nov. within the class Spirochaetia, the family Brevinemataceae be transferred to the novel order Brevinematales ord. nov. within the class Spirochaetia, and that the genera Borrelia and Cristispira be transferred to the novel family Borreliaceae fam. nov. within the order Spirochaetales (Figure 8). The emended descriptions of the order Spirochaetales and the family Spirochaetaceae, as well as a description of the new taxonomic groups Leptospiriales ord. nov., Brachyspiriales ord. nov., Brevinematales ord. nov., and Borreliaceae fam. nov. are provided below.

# EMENDED DESCRIPTION OF THE ORDER Spirochaetales (BUCHANAN, 1917)

The order contains two families, *Spirochaetaceae* and *Borreliaceae*, of which *Spirochaetaceae* is the type family. Organisms are helical or coccoid,  $0.1-75 \,\mu$ m in diameter and  $3.5-250 \,\mu$ m in length. Cells do not have hooked ends. Cells may possess flagella. Periplasmic flagella overlap in the central region of the cell. The diamino acid component of the peptidoglycan is L-ornithine. Anaerobic, facultatively anaerobic, or microaerophilic. Organisms are Chemo-organotrophic and utilize carbohydrates or amino acids as carbon and energy sources. Both free living and host associated members. The G + C content of the DNA is 27–66 (mol%). The type genus is *Spirochaeta* (Ehrenberg, 1835).

Organisms from this order are distinguished from all other Bacteria by the conserved signature indels (CSIs) described in this report in the following proteins: Alanyl-tRNA synthetase, Phosphoribosylpyrophosphate synthetase, SecY preprotein translocase, peptide chain release factor 2, DNA mismatch repair protein MutS, and DNA mismatch repair protein MutL.

## EMENDED DESCRIPTION OF THE FAMILY *Spirochaetaceae* (SWELLENGREBEL 1907 EMEND. ABT ET AL., 2012)

The family contains seven genera, *Clevelandina*, *Diplocalyx*, *Hollandina*, *Pillotina*, *Sphaerochaeta*, *Spirochaeta*, and *Treponema* of which *Spirochaeta* is the type genus. Organisms are helical or coccoid,  $0.1-75 \,\mu\text{m}$  in diameter and  $5-250 \,\mu\text{m}$  in length. Cells do not have hooked ends. Cells may possess flagella. Periplasmic flagella overlap in the central region of the cell. Cells can be anaerobic or facultatively anaerobic. The diamino acid component of the peptidoglycan is L-ornithine. Organisms are chemoorganotrophic and utilize carbohydrates or amino acids as carbon and energy sources. Both free living and host associated members. The G + C content of the DNA is 36–66 (mol%).

Organisms from this family are distinguished from all other bacteria by the CSIs described in this report in the following proteins: 6-phosphofructokinase (pyrophosphate), bifunctional Hpr kinase/phosphatase, and 30S ribosomal protein S13.

## DESCRIPTION OF Borreliaceae fam. nov.

*Borreliaceae* (Bor.re'li.a'ce.ae. N.L. fem. n. *Borrelia* type genus of the family; -aceae ending to denote a family; M.L. fem. pl. n. *Borreliaceae* the *Borrelia* family).

The family contains two genera, *Borrelia* and *Cristispira* of which *Borrelia* is the type genus. Organisms are helical,  $0.2-3 \,\mu$ m in diameter and  $3-180 \,\mu$ m in length. Cells do not have hooked ends. Periplasmic flagella overlap in the central region of the cell. Cells are motile, host-associated, and microaerophilic. The diamino acid component of the peptidoglycan is L-ornithine. Organisms are chemo-organotrophic and utilize carbohydrates or amino acids as carbon and energy sources. The G + C content of the DNA is 27–32 (mol%).

Organisms from this family are distinguished from all other Bacteria by the CSIs described in this report in the following proteins: Phosphofructokinase, 50S ribosomal protein L4, tRNA pseudouridine 55 synthase, Translation elongation factor-Tu, Histidyl-tRNA synthetase, Seryl-tRNA synthetase, Spoiiij-associtated protein, Nicotinate phosphoribosyltransferase, Ribose 5-phosphate isomerase, Ribonuclease Z, Hypothetical protein BGAFAR04\_0762, Signal recognition particle subunit FFH/SRP54, Hypothetical protein BSV1\_0075, Aspartyl/glutamyl-tRNA amidotransferase subunit A, Ribosomal RNA methyltransferase, and a LysM domain/M23/M37 peptidase domain protein.

## DESCRIPTION OF Brachyspiriales ord. nov.

*Brachyspiriales* (Bra.chy.spi.ra'les. N.L. fem. n. Brachyspira type genus of the order; suff. *-ales* ending to denote an order; N.L. fem. pl. n. *Brachyspiriales* the order of *Brachyspira*).

The order contains the type family *Brachyspiraceae*. Organisms are helical,  $0.2-0.4 \,\mu$ m in diameter and  $2-11 \,\mu$ m in length. Cell

ends may be blunt or pointed and do not have hooked ends. Periplasmic flagella overlap in the central region of the cell. Cells are motile, host-associated, and obligately anaerobic and aero-tolerant. The diamino acid component of the peptidoglycan is L-ornithine. Organisms are Chemo-organotrophic and utilize monosaccharides, disaccharides, the trisaccharide trehalose, and amino sugars as carbon and energy sources. The G + C content of the DNA is 24–28(mol%). The type genus is *Brachyspira* (Hovind-Hougen et al., 1982).

Organisms from this order are distinguished from all other bacteria by the CSIs described in this report in the following proteins: Flagellar hook-associated protein FlgK, DNA polymerase I, Valyl-tRNA synthetase, ATP-dependent protease La, and Glutamyl-tRNA amidotransferase subunit B. The description of the family *Brachyspiraceae* is the same as that of the order *Brachyspiriales*.

## DESCRIPTION OF Brevinematales ord. nov.

*Brevinematales* (Bre.vi.ne.ma.ta'les. N.L. fem. n. *Brevinema -atos* type genus of the order; suff. *-ales* ending to denote an order; N.L. fem. pl. n. *Brevinematales* the order of *Brevinema*).

The description of the order is the same as the description of the type family, *Brevinemataceae*.

## DESCRIPTION OF *Leptospiriales* ord. nov.

*Leptospiriales* (Lep.to.spi.ra'les. N.L. fem. n. *Leptospira* type genus of the order; suff. *-ales* ending to denote an order; N.L. fem. pl. n. *Leptospiriales* the order of *Leptospira*).

## REFERENCES

- Abt, B., Göker, M., Scheuner, C., Han, C., Lu, M., Misra, M., et al. (2013). Genome sequence of the thermophilic fresh-water bacterium Spirochaeta caldaria type strain (H1 <sup>T</sup>), reclassification of Spirochaeta caldaria and Spirochaeta stenostrepta in the genus Treponema as Treponema caldaria comb. nov, Treponema stenostrepta comb. nov, and Treponema zuelzerae comb. nov., and emendation of the genus Treponema. Stand. Genomic Sci. 8. doi: 10.4056/sigs.3096473. [Epub ahead of print].
- Abt, B., Han, C., Scheuner, C., Lu, M., Lapidus, A., Nolan, M., et al. (2012). Complete genome sequence of the termite hindgut bacterium Spirochaeta coccoides type strain (SPN1<sup>T</sup>), reclassification in the genus Sphaerochaeta as Sphaerochaeta coccoides comb. nov. and emendations of the family Spirochaetaceae and the genus Sphaerochaeta. Stand. Genomic Sci. 6, 194. doi: 10.4056/sigs.2796069
- Adeolu, M., and Gupta, R. S. (2013). Phylogenomics and molecular signatures for the order *Neisseriales*: proposal for division of the order *Neisseriales* into the emended family *Neisseriaceae*

and *Chromobacteriaceae* fam. nov. *Antonie Van Leeuwenhoek* 104, 1–24. doi: 10.1007/s10482-013-9920-6

- Adler, B., and de la Peña Moctezuma, A. (2010). *Leptospira* and leptospirosis. *Vet. Microbiol.* 140, 287–296. doi: 10.1016/j.vetmic.2009.03.012
- Akiva, E., Itzhaki, Z., and Margalit, H. (2008). Built-in loops allow versatility in domain? Domain interactions: lessons from selfinteracting domains. *Proc. Natl. Acad. Sci. U.S.A.* 105, 13292. doi: 10.1073/pnas.0801207105
- Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., et al. (1997). Gapped, BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25, 3389–3402. doi: 10.1093/nar/25.17.3389
- Anthony, N. E., Blackwell, J., Ahrens, W., Lovell, R., and Scobey, M. W. (2013). Intestinal spirochetosis: an Enigmatic disease. *Dig. Dis. Sci.* 58, 202–208. doi: 10.1007/s10620-012-2305-2
- Baldauf, S. L., and Palmer, J. D. (1993). Animals and fungi are each other's closest relatives: congruent evidence from multiple proteins. *Proc. Natl. Acad. Sci. U.S.A.* 90, 11558–11562. doi: 10.1073/pnas.90. 24.11558

The order contains the type family *Leptospiraceae*. Organisms are helical, 0.1–0.3  $\mu$ m in diameter and 2–11  $\mu$ m in length. Cell have hooked ends. Periplasmic flagella do not overlap in the central region of the cell. Cells are motile. The diamino acid component of the peptidoglycan is  $\alpha$ , $\epsilon$ -diaminopimelic acid. Obligately aerobic or microaerophilic. Organisms are Chemo-organotrophic and long-chain fatty acids or long-chain fatty alcohols as carbon and energy sources. Both free living and host associated members. The G + C content of the DNA is 33–55 (mol%). The type genus is *Leptospira* (Noguchi, 1917).

Organisms from this order are distinguished from all other Bacteria by the CSIs described in this report in the following proteins: 50S Ribosomal protein L14, 30S Ribosomal protein S2, Alanyl-tRNA synthetase, Flagellar basal-body rod protein FlgG, and Flagellar filament core protein FlaB. The description of the family *Leptospiraceae* is the same as that of the order *Leptospiriales*.

# **ACKNOWLEDGMENTS**

This work was supported by a MRI-ORF Water Round research grant.

# SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://www.frontiersin.org/Evolutionary\_and\_Genomic\_Mic robiology/10.3389/fmicb.2013.00217/abstract

- Bellgard, M. I., Wanchanthuek, P., La, T., Ryan, K., Moolhuijzen, P., Albertyn, Z., 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. doi: 10.1371/journal.pone.0004641
- Bermudes, D., Chase, D., and Margulis, L. (1988). Morphology as a basis for taxonomy of large Spirochetes symbiotic in wood-eating cockroaches and termites: *Pillotina* gen. nov., nom. rev; *Pillotina* calotermitidis sp. nov., nom. rev; *Diplocalyx* gen. nov., nom. rev; *Diplocalyx* calotermitidis sp. nov., nom. rev; Hollandina gen. nov., nom. rev; Hollandina pterotermitidis sp. nov., nom. rev; and Clevelandina reticulitermitidis gen. nov., sp. nov. Int. J. Syst. Bacteriol. 38, 291–302. doi: 10.1099/00207713-38-3-291
- Bhandari, V., Ahmod, N. Z., Shah, H. N., and Gupta, R. S. (2013). Molecular signatures for the *Bacillus* species: demarcation of the *Bacillus* subtilis and *Bacillus cereus* clades in molecular terms and proposal to limit the placement of new species into the genus *Bacillus*. Int. J. Syst. Evol. Microbiol. 63, 2712–2726. doi: 10.1099/ijs.0.048488-0

- Buchanan, R. E. (1917). Studies in the nomenclature and classification of bacteria. II. The primary subdivisions of the Schizomycetes. *J. Bacteriol.* 2, 155–164.
- Bulach, D. M., Zuerner, R. L., Wilson, P., Seemann, T., McGrath, A., Cullen, P. A., et al. (2006). Genome reduction in *Leptospira* borgpetersenii reflects limited transmission potential. Proc. Natl. Acad. Sci. U.S.A. 103, 14560–14565. doi: 10.1073/pnas.0603979103
- Casjens, S. R., Fraser-Liggett, C. M., Mongodin, E. F., Qiu, W. G., Dunn, J. J., Luft, B. J., et al. (2011). Whole genome sequence of an unusual *Borrelia burgdorferi* sensu lato isolate. *J. Bacteriol.* 193, 1489–1490. doi: 10.1128/JB.01521-10
- Castresana, J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Biol. Evol.* 17, 540–552. doi: 10.1093/oxfordjournals.molbev.a026334
- Cavalier-Smith, T. (2002). The neomuran origin of archaebacteria, the negibacterial root of the universal tree and bacterial megaclassification. *Int. J. Syst. Evol. Microbiol.* 52, 7–76.
- Chaconas, G. (2005). Hairpin telomeres and genome plasticity in

*Borrelia*: all mixed up in the end. *Mol. Microbiol.* 58, 625–635. doi: 10.1111/j.1365-2958.2005.04872.x

- Chaconas, G., and Kobryn, K. (2010). Structure, function, and evolution of linear replicons in *Borrelia. Annu. Rev. Microbiol.* 64, 185–202. doi: 10.1146/annurev.micro.112408. 134037
- Chou, L. F., Chen, Y. T., Lu, C. W., Ko, Y. C., Tang, C. Y., Pan, M. J., et al. (2012). Sequence of *Leptospira santarosai* serovar Shermani genome and prediction of virulence-associated genes. *Gene.* 511, 364–370. doi: 10.1016/j.gene. 2012.09.074
- Ciccarelli, F. D., Doerks, T., Von Mering, C., Creevey, C. J., Snel, B., and Bork, P. (2006). Toward automatic reconstruction of a highly resolved tree of life. *Science* 311, 1283–1287. doi: 10.1126/science.1123061
- Cutler, S. J. (2010). Relapsing fever: a forgotten disease revealed. J. Appl. Microbiol. 108, 1115–1122. doi: 10.1111/j.1365-2672.2009.04598.x
- Dai, Q., Restrepo, B. I., Porcella, S. F., Raffel, S. J., Schwan, T. G., and Barbour, A. G. (2006). Antigenic variation by *Borrelia hermsii* occurs through recombination between extragenic repetitive elements on linear plasmids. *Mol. Microbiol.* 60, 1329–1343. doi: 10.1111/j.1365-2958.2006.05177.x
- Dworkin, M. S., Schwan, T. G., and Anderson, D. E. Jr., and Borchardt, S. M. (2008). Tick-borne relapsing fever. *Infect. Dis. Clin. North Am.* 22, 449. doi: 10.1016/j.idc.2008. 03.006
- Ehrenberg, C. G. (1835). "Dritter Beitrag zur Erkenntniss grosser Organisation in der Richtung des kleinsten Raumes," in Abhandlungen der Preussischen Akademie der Wissenschaften (Berlin) aus den Jahre 1833–1835, 143–336.
- Elbir, H., Gimenez, G, Robert, C., Bergström, S., Cutler, S., Raoult, D., et al. (2012). Complete genome sequence of *Borrelia crocidurae*. *J. Bacteriol.* 194, 3723–3724. doi: 10.1128/JB.00118-12
- Ellen, R. P., and Galimanas, V. B. (2005). Spirochetes at the forefront of periodontal infections. *periodontology 2000* 38, 13–32. doi: 10.1111/j.1600-0757.2005.00108.x
- Euzéby, J. P. (2013). List of Prokaryotic names with Standing in Nomenclature. Available online at: www.bacterio.net
- Fraser, C. M., Casjens, S., Huang, W. M., Sutton, G. G., Clayton, R., Lathigra, R., et al. (1997).

Genomic sequence of a Lyme disease spirochaete, *Borrelia burgdorferi. Nature* 390, 580–586. doi: 10.1038/37551

- Gao, B., and Gupta, R. S. (2012a). Microbial systematics in the post-genomics era. *Antonie Van Leeuwenhoek* 101, 45–54. doi: 10.1007/s10482-011-9663-1
- Gao, B., and Gupta, R. S. (2012b). Phylogenetic framework and molecular signatures for the main clades of the phylum Actinobacteria. *Microbiol. Mol. Biol. Rev.* 76, 66–112. doi: 10.1128/MMBR.05011-11
- Gerbase, A. C., Rowley, J. T., Heymann, D. H. L., Berkley, S. F. B., and Piot, P. (1998). Global prevalence and incidence estimates of selected curable STDs. Sex. Transm. Infect. 74, S12.
- Glöckner, G., Lehmann, R., Romualdi, A., Pradella, S., Schulte-Spechtel, U., Schilhabel, M., et al. (2004). Comparative analysis of the *Borrelia garinii* genome. *Nucleic Acids Res.* 32, 6038–6046. doi: 10.1093/nar/gkh953
- Gogarten, J. P., Doolittle, W. F., and Lawrence, J. G. (2002). Prokaryotic evolution in light of gene transfer. *Mol. Biol. Evol.* 19, 2226–2238. doi: 10.1093/oxfordjournals.molbev.a004046
- Goodner, B., Hinkle, G., Gattung, S., Miller, N., Blanchard, M., Qurollo, B., et al. (2001). Genome sequence of the plant pathogen and biotechnology agent Agrobacterium tumefaciens C58. Science 294, 2323–2328. doi: 10.1126/science.1066803
- Griffiths, E., and Gupta, R. S. (2004). Signature sequences in diverse proteins provide evidence for the late divergence of the order *Aquificales*. *Int. Microbiol.* 7, 41–52.
- Gupta, R. S. (1998). Protein phylogenies and signature sequences: a reappraisal of evolutionary relationships among archaebacteria, eubacteria, and eukaryotes. *Microbiol. Mol. Biol. Rev.* 62, 1435.
- Gupta, R. S. (2009). Protein signatures (molecular synapomorphies) that are distinctive characteristics of the major cyanobacterial clades. *Int. J. Syst. Evol. Microbiol.* 59, 2510.
- Gupta, R. S. (2010). Molecular signatures for the main phyla of photosynthetic bacteria and their subgroups. *Photosyn. Res.* 104, 357–372. doi: 10.1007/s11120-010-9553-9
- Gupta, R. S., Chander, P., and George, S. (2012). Phylogenetic framework and molecular signatures for the class *Chloroflexi* and its different clades; proposal for division of the class *Chloroflexi* class. nov. into

the suborder *Chloroflexineae* subord. nov., consisting of the emended family *Oscillochloridaceae* and the family *Chloroflexaceae* fam. nov., and the suborder *Roseiflexineae* subord. nov., containing the family *Roseiflexaceae* fam. nov. *Antonie Van Leeuwenhoek* 103, 99–119. doi: 10.1007/s10482-012-9790-3

- Gupta, R. S., Chen, W. J., Adeolu, M., and Chai, Y. (2013). Molecular signatures for the class Coriobacterija and its different clades; Proposal for division of the class Coriobacteriia into the emended order Coriobacteriales, containing the emended family Coriobacteriaceae and Atopobiaceae fam. nov., and Eggerthellales ord. nov., containing the family Eggerthellaceae fam. nov. Int. I. Syst. Evol. Microbiol. doi: 10.1099/ijs.0. 048371-0. [Epub ahead of print].
- Gupta, R. S., and Griffiths, E. (2002). Critical issues in bacterial phylogeny. *Theor. Popul. Biol.* 61, 423–434. doi: 10.1006/tpbi.2002.1589
- Håfström, T., Jansson, D. S., and Segerman, B. (2011). Complete genome sequence of *Brachyspira intermedia* reveals unique genomic features in *Brachyspira* species and phage-mediated horizontal gene transfer. *BMC Genomics* 12:395. doi: 10.1186/1471-2164-12-395
- Han, C., Gronow, S., Teshima, H., Lapidus, A., Nolan, M., Lucas, S., et al. (2011). Complete genome sequence of *Treponema succinifaciens* type strain (6091<sup>T</sup>). *Stand. Genomic Sci.* 4, 361. doi: 10.4056/sigs.1984594
- Harris, J. K., Kelley, S. T., Spiegelman, G. B., and Pace, N. R. (2003). The genetic core of the universal ancestor. *Genome Res.* 13, 407–412. doi: 10.1101/gr.652803
- Harwood, C. S., and Canale-Parola, E. (1984). Ecology of spirochetes. Annu. Rev. Microbiol. 38, 161–192. doi: 10.1146/annurev.mi.38.100184. 001113
- Homma, M., DeRosier, D. J., and Macnab, R. M. (1990). Flagellar hook and hook-associated proteins of *Salmonella typhimurium* and their relationship to other axial components of the flagellum. *J. Mol. Biol.* 213, 819–832. doi: 10.1016/S0022-2836(05)80266-9
- Hovind-Hougen, K., Birch-Andersen, A., Hendrik-Nielsen, R., Orholm, M., Pedersen, J. O., Teglbaerg, P. S., et al. (1982). Intestinal spirochetosis: morphological characterization and cultivation of the spirochete Brachyspira aalborgi gen. nov., sp. nov. J. Clin. Microbiol. 6, 1127–1136.

- Jeanmougin, F., Thompson, J. D., Gouy, M., Higgins, D. G., and Gibson, T. J. (1998). Multiple sequence alignment with Clustal, X. *Trends Biochem. Sci.* 23, 403. doi: 10.1016/S0968-0004(98)01285-7
- Jones, D. T., Taylor, W. R., and Thornton, J. M. (1992). The rapid generation of mutation data matrices from protein sequences. *Comput. Appl. Biosci.* 8, 275–282.
- Kobryn, K. (2007). "The linear hairpin replicons of Borrelia burgdorferi," in *Microbial Linear Plasmids*, eds F. Meinhardt, and R. Klassen (Heidelberg: Springer), 117–140.
- Lescot, M., Audic, S., Robert, C., Nguyen, T. T., Blanc, G., Cutler, S. J., et al. (2008). The genome of *Borrelia recurrentis*, the agent of deadly louse-borne relapsing fever, is a degraded subset of tick-borne *Borrelia duttonii*. *PLoS Genet.* 4:e1000185. doi: 10.1371/journal.pgen.1000185
- Li, C., Wolgemuth, C. W., Marko, M., Morgan, D. G., and Charon, N. W. (2008). Genetic analysis of spirochete flagellin proteins and their involvement in motility, filament assembly, and flagellar morphology. *J. Bacteriol.* 190, 5607–5615. doi: 10.1128/JB.00319-08
- Lin, C., den Bakker, H. C., Suzuki, H., Lefébure, T., Ponnala, L., Sun, Q., et al. (2013). Complete genome sequence of the porcine strain *Brachyspira pilosicoli* P43/6/78<sup>T</sup>. *Genome Announc.* 1. doi: 10.1128/ genomeA.00215-12. [Epub ahead of print].
- Ludwig, W., and Klenk, H. P. (2001). "Overview: a phylogenetic backbone and taxonomic framework for prokaryotic systematics," in *Bergey's Manual of Systematic Bacteriology*, ed G. M. Garrity (New York, NY: Springer), 49–65.
- Macnab, R. M. (2003). How bacteria assemble flagella. Annu. Rev. Microbiol. 57, 77–100. doi: 10.1146/ annurev.micro.57.030502.090832
- Mavromatis, K., Yasawong, M., Chertkov, O., Lapidus, A., Lucas, S., Nolan, M., et al. (2010). Complete genome sequence of *Spirochaeta smaragdinae* type strain (SEBR 4228<sup>T</sup>). *Stand. Genomic Sci.* 3, 136. doi: 10.4056/sigs.1143106
- Nau, R., Christen, H. J., and Eiffert, H. (2009). Lyme disease: ?current state of knowledge. *Dtsch. Arztebl. Int.* 106, 72. doi: 10.3238/arztebl.2009.0072
- NCBI. (2013). NCBI Genome Database. Available online at: http://www. ncbi.nlm.nih.gov/genome/
- Noguchi, H. (1917). Spirochaeta icterohaemorrhagiae in American wild

rats and its relation to the Japanese and European strains. *J. Exp. Med.* 25, 755–763.

- Olsen, I., Paster, B. J., and Dewhirst, F. E. (2000). Taxonomy of spirochetes. *Anaerobe* 6, 39–57. doi: 10.1006/anae.1999.0319
- Paster, B. J. (2011a). "Phylum XV. Spirochaetes Garrity and Holt (2001)," in *Bergey's Manual of Systematic Bacteriology*, eds D. J. Brenner, N. R. Krieg, G. M. Garrity, and J. T. Staley (New York, NY: Springer), 471.
- Paster, B. J. (2011b). "Family I. Spirochaetaceae Swellengrebel 1907, 581AL," in *Bergey's Manual* of Systematic Bacteriology, eds D. J. Brenner, N. R. Krieg, G. M. Garrity, and J. T. Staley (New York, NY: Springer), 473–531.
- Paster, B. J., and Dewhirst, F. E. (2000). Phylogenetic foundation of spirochetes. J. Mol. Microbiol. Biotechnol. 2, 341–344.
- Paster, B. J., Dewhirst, F. E., Weisburg, W. G., Tordoff, L. A., Fraser, G. J., Hespell, R. B., et al. (1991). Phylogenetic analysis of the spirochetes. J. Bacteriol. 173, 6101–6109.
- Pati, A., Sikorski, J., Gronow, S., Munk, C., Lapidus, A., Copeland, A., et al. (2010). Complete genome sequence of *Brachyspira murdochii* type strain (56–150<sup>T</sup>). Stand. Genomic Sci. 2, 260.
- Picardeau, M., Bulach, D. M., Bouchier, C., Zuerner, R. L., Zidane, N., Wilson, P. J., et al. (2008). Genome sequence of the saprophyte *Leptospira biflexa* provides insights into the evolution of *Leptospira* and the pathogenesis of leptospirosis. *PloS ONE* 3:e1607. doi: 10.1371/ journal.pone.0001607
- Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., et al. (2013). The, SILVA ribosomal, RNA gene database

project: improved data processing and web-based tools. *Nucleic Acids Res.* 41, D590–D596. doi: 10.1093/nar/gks1219

- Ren, S. X., Fu, G., Jiang, X. G., Zeng, R., Miao, Y. G., Xu, H., et al. (2003). Unique physiological and pathogenic features of *Leptospira interrogans* revealed by whole-genome sequencing. *Nature* 422, 888–893. doi: 10.1038/ nature01597
- Rokas, A., and Holland, P. W. H. (2000). Rare genomic changes as a tool for phylogenetics. *Trends Ecol. Evol.* 15, 454–459. doi: 10.1016/S0169-5347(00)01967-4
- Rokas, A., Williams, B. L., King, N., and Carroll, S. B. (2003). Genome-scale approaches to resolving incongruence in molecular phylogenies. *Nature* 425, 798–804. doi: 10.1038/nature02053
- Schutzer, S. E., Fraser-Liggett, C. M., Qiu, W. G., Kraiczy, P., Mongodin, E. F., Dunn, J. J., et al. (2012). Whole-genome sequences of *Borrelia bissettii*, *Borrelia valaisiana*, and *Borrelia spielmanii*. J. Bacteriol. 194, 545–546. doi: 10.1128/JB.06263-11
- Seshadri, R., Myers, G. S. A., Tettelin, H., Eisen, J. A., Heidelberg, J. F., Dodson, R. J., et al. (2004). Comparison of the genome of the oral pathogen *Treponema denticola* with other spirochete genomes. *Proc. Natl. Acad. Sci.* U.S.A. 101, 5646–5651. doi: 10.1073/pnas.0307639101
- Singh, B., and Gupta, R. S. (2009). Conserved inserts in the Hsp60 (GroEL) and Hsp70 (DnaK) proteins are essential for cellular growth. *Mol. Genet. Genomics* 281, 361–373. doi: 10.1007/s00438-008-0417-3
- Smajs, D., Norris, S. J., and Weinstock, G. M. (2012). Genetic diversity

in Treponema pallidum: implications for pathogenesis, evolution and molecular diagnostics of syphilis and yaws. *Infect. Genet. Evol.* 12, 191–202. doi: 10.1016/j.meegid.2011.12.001

- Smajs, D., Zobaniková, M., Strouhal, M., Cjková, D., Dugan-Rocha, S., Pospísilová, P., et al. (2011). Complete genome sequence of *Treponema paraluiscuniculi*, strain Cuniculi A: the loss of infectivity to humans is associated with genome decay. *PLoS ONE* 6:e20415. doi: 10.1371/journal.pone.0020415
- Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., and Kumar, S. (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* 28, 2731–2739. doi: 10.1093/molbev/ msr121
- Tavaré, S. (1986). "Some probabilistic and statistical problems in the analysis of, DNA sequences," in *Lectures* on Mathematics in the Life Sciences, ed R. M. Miura (Providence, RI: American Mathematical Society), 57–86.
- Visser, M. B., and Ellen, R. P. (2011). New insights into the emerging role of oral spirochaetes in periodontal disease. *Clin. Microbiol. Infect.* 17, 502–512. doi: 10.1111/j.1469-0691.2011.03460.x
- Whelan, S., and Goldman, N. (2001). A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Mol. Biol. Evol.* 18, 691–699. doi: 10.1093/oxfordjournals.molbev. a003851
- Wu, D., Hugenholtz, P., Mavromatis,
   K., Pukall, R., Dalin, E., Ivanova, N.
   N., et al. (2009). A phylogeny-driven genomic encyclopaedia of

Bacteria and Archaea. Nature 462, 1056–1060. doi: 10.1038/ nature08656

Zhong, J., and Barbour, A. G. (2004). Cross species hybridization of a *Borrelia burgdorferi* DNA array reveals infection and culture associated genes of the unsequenced genome of the relapsing fever agent *Borrelia hermsii. Mol. Microbiol.* 51, 729–748. doi: 10.1046/j.1365-2958.2003.03849.x

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 17 June 2013; paper pending published: 04 July 2013; accepted: 11 July 2013; published online: 30 July 2013.

Citation: Gupta RS, Mahmood S and Adeolu M (2013) A phylogenomic and molecular signature based approach for characterization of the phylum Spirochaetes and its major clades: proposal for a taxonomic revision of the phylum. Front. Microbiol. **4**:217. doi: 10.3389/fmicb.2013.00217

This article was submitted to Frontiers in Evolutionary and Genomic Microbiology, a specialty of Frontiers in Microbiology.

Copyright © 2013 Gupta, Mahmood and Adeolu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.