

# A genome sequence survey of the mollicute corn stunt spiroplasma *Spiroplasma kunkelii*

Xiaodong Bai, Saskia A. Hogenhout \*

Department of Entomology, Ohio Agricultural Research and Development Center (OARDC), Ohio State University, Wooster, OH 44691, USA

Received 19 November 2001; received in revised form 21 January 2002; accepted 7 February 2002

First published online 12 March 2002

## Abstract

The mollicute corn stunt spiroplasma (*Spiroplasma kunkelii*) is a leafhopper-transmitted pathogen of maize. Sequencing of the ~1.6-Mb genome of *S. kunkelii* was initiated to aid understanding the genetic basis of spiroplasma interactions with their plant and leafhopper hosts. In total, 144 712 nucleotides of non-redundant, high-quality *S. kunkelii* genome sequence were obtained. Sequence tags were searched against the Mycoplasmataceae and *Bacillus/Clostridium* databases. Results showed that, in addition to spiroplasma phage SpV1 DNA insertions, spiroplasma genomes harbor more purine and amino acid biosynthesis, transcription regulation, cell envelope and DNA transport/binding genes than Mycoplasmataceae genomes. This investigation demonstrates that survey sequencing is an efficient procedure for gene discovery and genome characterization. The results of the *S. kunkelii* sequencing project are available at the Spiroplasma WebPage at <http://www.oardc.ohio-state.edu/spiroplasma/genome.htm>. © 2002 Federation of European Microbiological Societies. Published by Elsevier Science B.V. All rights reserved.

**Keywords:** Genome sequence; Leafhopper; Mollicute; Gene loss; Evolution; Gram-positive bacterium; Plant pathogen; *Spiroplasma kunkelii*

## 1. Introduction

The mollicute *Spiroplasma kunkelii* is a member of the family Spiroplasmataceae within the order Mycoplasmales. Spiroplasmas are primarily associated with insects and plants in epiphytic, symbiotic or pathogenic interactions. Three spiroplasma species evolved as plant pathogens: the citrus stubborn spiroplasma *Spiroplasma citri*, the corn stunt spiroplasma (CSS) *S. kunkelii*, and the periwinkle yellowing spiroplasma *Spiroplasma phoeniceum*. Plant-pathogenic spiroplasmas are restricted to the sieve tubes of their plant hosts and are transmitted from plant to plant by phloem-feeding leafhoppers in a persistent propagative manner [1–3]. CSS is one of the most important threats to maize. Typical symptoms of CSS infection include chlorosis, stunted plants with reduced internode length and proliferation of ears that do not mature [1].

Mollicutes are thought to have diverged from a Gram-positive *Clostridium*-like ancestor and differ phenotypically from other bacteria in their minute size (0.3–0.5 µm) and lack of cell wall [4,5]. Genomes of Mollicutes are smaller in size than those of most other prokaryotes as a result of degenerative or reductive evolution. However, gene loss in spiroplasmas was not as extensive as in other members of the class Mollicutes [4,6]. Interestingly, the spiroplasma morphology differs from that of other Mollicutes. All members within the genus *Spiroplasma* have pleomorphic shapes varying from spherical or slightly ovoid, 100–250 nm, to helical fragments that are about 120 nm in diameter and 2–4 µm long during active growth and up to 15 µm in later stages of growth, whereas members of other mollicute genera typically have a spheroidal to ovoid shape and commonly do not have a helical elongated stage in their life cycle.

The genomes of *Mycoplasma genitalium*, *Mycoplasma pneumoniae*, *Ureaplasma urealyticum* and *Mycoplasma pulmonis* of the family Mycoplasmataceae within the order Mycoplasmales have been sequenced to completion [7–10], and close to seven full genome sequences from the *Bacillus/Clostridium* group, the most closely related walled bacteria to the class Mollicutes, are available as well. The sample-sequencing project of the *S. kunkelii* genome and

\* Corresponding author. Tel.: +1 (330) 263 3730;  
Fax: +1 (330) 263 3686.

E-mail address: hogenhout.1@osu.edu (S.A. Hogenhout).

subsequent comparison of *S. kunkelii* sequence data with Mycoplasmataceae species and *Bacillus/Clostridium* sequence databases as described herein revealed interesting differences in gene content between *S. kunkelii* and members of the Mycoplasmataceae.

## 2. Materials and methods

### 2.1. Selection of the *S. kunkelii* strain

The *S. kunkelii* strain CSS-M was selected for genome sequencing. The strain was originally isolated from infected corn plants in Tlaltizapan, Mexico in 1992 [11]. It has been maintained at the Ohio Agricultural Research and Development Center (OARDC) by serial transfers with the CSS leafhopper vector (*Dalbulus maidis*) as described in Ebbert and Nault [12]. CSS-M was isolated from infected corn stems, propagated in liquid LD8A3 medium, and plated onto LD8A3 agar plates [13]. A culture derived from a single colony was used for genomic DNA isolation. The *S. kunkelii* clone was transmitted to maize seedlings (*Zea mays* L. 'Early Sunglow') by *D. maidis* [1], indicating the clone kept its characteristics of leafhopper transmission and pathogenesis of plants.

### 2.2. Construction of *S. kunkelii* genomic DNA libraries

Genomic DNA was isolated from *S. kunkelii* using the Qiagen (Valencia, CA, USA) Whole Genomic DNA isolation kit following the manufacturer's procedures. The isolated genomic DNA was used for library construction. The DNA was digested to completion with *Eco*RI or *Hind*III, ligated into appropriately digested, phosphatase-treated pUC18, and transformed into electrocompetent XLblue *Escherichia coli* (Stratagene, La Jolla, CA, USA) cells. For construction of a random sheared DNA library, DNA was fragmented with the Hydroshear™ (GeneMachines) into pieces with a distribution centered on 1.5 kb. Blunt-ended DNA was then ligated into pPCR-Script Amp Sk(+) plasmid (Stratagene), and plasmid DNA was introduced into chemically competent XL10-Gold Kan *E. coli* following the manufacturer's procedures (Stratagene). Insert-carrying plasmids were identified in transformants by detecting white colonies after growth on X-Gal/IPTG [14].

### 2.3. Sequencing and sequence analysis

Colonies were grown overnight at 37°C in single wells of 96-well microtiter plates containing 150 µl LB freeze (4 mM MgSO<sub>4</sub>, 360 mM K<sub>2</sub>HPO<sub>4</sub>, 132 mM KH<sub>2</sub>PO<sub>4</sub>, 17 mM Na-citrate, 68 mM (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 4.4% glycerol in LB, pH 7.0) and 100 µg ml<sup>-1</sup> ampicillin and transferred to LB agar plates containing 100 µg ml<sup>-1</sup> ampicillin after 18 h

using a 96-well plate replicator. The inoculated agar plates were sent to MWG-Biotech (High Point, NC, USA) for one-pass sequencing of the inserts using the M13 forward and reverse primers for pUC18, and T7 and T3 primers for pPCR-Script Amp Sk(+) plasmids on an ABI377 automatic sequencer. Trace files were analyzed with the PHRED and CROSS\_MATCH algorithms of MacPhred/Phrap [15,16] to translate the ABI377 chromatogram data of the sequence files into accurate quality information for each base call and detection of plasmid sequences, respectively. Plasmid sequences were removed from each sequence tag and high quality sequence data (>20 phred score) were collected into a database and searched against the non-redundant (nr) database at National Center for Biotechnology Information (NCBI) using nucleotide–nucleotide BLAST (blastn) or the translating BLAST (blastx) algorithms [17]. To screen for redundant sequence tags, the *S. kunkelii* sequence database was also searched against itself with the blastn algorithm. Nucleotide sequences with significant similarities ( $E$ -value  $\leq 10^{-5}$ ) to sequences in the NCBI database were collected, translated into proteins and searched against the full non-redundant protein database and non-redundant databases of Mycoplasmataceae and *Bacillus/Clostridium* at NCBI with the protein–protein BLAST (blastp) algorithm. All sequence analyses were performed on local Linux workstations.

## 3. Results and discussion

### 3.1. Library construction, sequencing and sequence analysis

To confirm the identity of the isolated DNA, the spiralin gene was amplified using primers described by Foissac et al. [18]. The nucleotide sequence of the spiralin gene amplification product was identical to that reported earlier [18], thus confirming the identity of the CSS-M clone of *S. kunkelii* (data not shown). Insert sizes of clones from the shotgun library ranged from 0.5 to 4 kb and clones from the *Eco*RI or *Hind*III libraries contained fragment sizes ranging from 150 bp to 10 kb.

In total, 94 inserts from the *Eco*RI and *Hind*III libraries, and 188 inserts from the sheared DNA library were sequenced from flanking primer sites after which the sequences were collected into a database. Low quality and cloning vector sequences were removed from the database. The database was then searched against itself with the blastn algorithm to analyze redundancy. Mollicute genome sequences show the presence of highly repeated regions and spiroplasma genomes have many copies of spiroplasma phage SpV1 DNA [19]. Therefore, redundant sequence tags were not assembled into contigs but within each set of redundant clones one sequence tag from the forward and reverse direction with best phred quality scores were kept in the database whereas others were re-

Table 1  
Sequence tags with significant similarity ( $E$ -value  $\leq 10^{-5}$ ) to spiroplasma virus SpV1 and *S. citri* putative virulence proteins

Identity	Sequence tag ID	Acc. No. of best hit	$E$ -value
<i>Spiroplasma virus SpVI ORFs</i>			
ORF1, capsid protein	MSAC_D10.y	9626113	2e-19
	MSAD_C09.y	1143020	4e-94
	MSAD_C12.x	9626113	3e-62
	MSAD_B02.x	1143020	e-117
ORF3, transposase gene	MEAA_B07.y	1143021	2e-39
	MHAA_H01.y	9626114	3e-14
	MSAC_A12.y	1143018	1e-14
	MSAD_C03.y	1143013	1e-46
ORF2	MSAC_C02.y	9626110	2e-26
	MSAC_D02.y	9626110	2e-26
	MSAD_H02.y	P15893	2e-42
ORF4	MSAD_H02.x	U28974	3e-21
ORF5	MSAD_A08.x	9626111	3e-12
ORF7	MHAA_C12.y	P15898	3e-10
ORF14	MHAA_G02.y	1143012	1e-09
	MSAD_E11.x	U28972	8e-08
ORFa	MHAA_H07.x	1143008	3e-29
<i>S. citri putative virulence proteins</i>			
P123	MSAD_A01.x	T28663	7e-61
P58	MHAA_A07.x	7482012	4e-22
P54	MHAA_A07.y	7482011	1e-64
P18	MHAA_F03.y	7482010	2e-24

Deduced protein sequences were searched against the non-redundant database at NCBI. Identity and sequence tag identity (ID) are indicated and for each sequence tag the accession number and  $E$ -value of entry with highest similarity are listed.

moved. This resulted in a database of 144 712 nucleotides (396 sequence tags) of non-redundant high-quality ( $> 20$  phred score) *S. kunkelii* genome sequences representing 9% of the *S. kunkelii* genome, based on an estimated genome size of 1600 kb [6]. All sequences were deposited in the random single pass read genome survey sequence database (dbGSS) of GenBank (accession Nos. BH234783 to BH235178).

The 396 sequence tags were searched against the complete NCBI nr database with the blastn and blastx algorithms. The overall percentage of sequence tags with significant similarity ( $E$ -value  $\leq 10^{-5}$ ) to open reading frames (ORFs) in the NCBI nr database was  $\sim 40\%$  (150/396), which is in agreement with previous findings that biological functions can be assigned to  $\sim 50\%$  of the ORFs in completed genome sequencing projects [20].

### 3.2. DNA phage sequences

Unlike the Mycoplasmatataceae, spiroplasma genomes harbor many spiroplasma phage SpV1 DNA insertions [7–10,19,21,22]. In this survey  $\sim 5\%$  (17/396) of the *S. kunkelii* sequence tag database had significant similarity to spiroplasma virus SpV1 DNA (Table 1). The percentage of phage sequences in the *S. kunkelii* sequence tag database is comparable to the 7% DNA phage sequences found in the genome of the Gram-negative leafhopper-transmitted vascular plant pathogen, *Xylella fastidiosa* [20].

### 3.3. Spiroplasma-specific sequences

In total, 133 sequence tags had significant similarities to prokaryotic and/or eukaryotic protein sequences in the NCBI nr database. Included were four sequences unique to spiroplasmas with similarity to putative *S. citri* virulence genes encoding P123, P58, P54, or P18 (Table 1) [23,24]. These genes are part of a 9.5-kb *S. citri* genome segment that is deleted from a non-transmissible line of *S. citri*.

### 3.4. Comparative genome analysis

As a preliminary assessment of to what extent the *S. kunkelii* genome content differs from those of Mycoplasmatataceae species, sequence tags were translated into proteins to ensure that the sequences were part of ORFs and, subsequently, the protein sequences were searched against the Mycoplasmatataceae, *Bacillus/Clostridium* and complete nr protein databases of GenBank (Table 2). The Mycoplasmatataceae database was selected because it contains the full genome sequences of three mycoplasma and one ureaplasma species [7–10], whereas the *Bacillus/Clostridium* database was selected because it contains many completed genome sequences and *Bacillus/Clostridium* species are thought to be closest walled relatives to Mollicutes [4,25]. Interesting gene content differences among *S. kunkelii*, and Mycoplasmatataceae and *Bacillus/Clostridium* species are discussed below.

Table 2  
Sequence tags with significant similarities ( $E$ -value  $\leq 10^{-5}$ ) to GenBank nr protein sequences

Sequence tag ID	Identity	Sequence length (aa)	Best entry Mycoplasmataceae, accession No. ( $E$ -value)	Best entry <i>Bacillus/Clostridium</i> , accession No. ( $E$ -value)	Best entry GenBank organism, accession No. ( $E$ -value)
<i>Amino acid biosynthesis</i>					
MEAA_A03.y	Thymidylate kinase	96	14089465 (5e-08)	2632295 (3e-12)	<i>B. subtilis</i> , 16077096 (2e-10)
MSAC_C05.y	Folylpolyglutamate synthase/dihydrofolate synthetase ( <i>folC</i> )	127	—	4930039 (3e-09)	<i>Str. pneumoniae</i> , 15900133 (4e-09)
MSAD_B10.x	Methionine aminopeptidase (MAP) (peptidase M)	179	14089981 (3e-37)	11131429 (3e-41)	<i>B. halodurans</i> , 15612719 (1e-35)
PE_10.x	Serine hydroxymethyl transferase (glyA)	239	1673936 (2e-78)	12723496 (5e-84)	<i>Str. pneumoniae</i> , 15902972 (2e-64)
<i>Cell envelope</i>					
MHAA_E07.x	Cell shape determining protein (MreB-like protein)	172	—	10176363 (2e-33)	<i>B. halodurans</i> , 15616301 (4e-07)
PH_05.x	GcpE protein	102	—	1730252 (3e-26)	<i>B. subtilis</i> , 16079569 (1e-27)
<i>Fatty acid and phospholipid metabolism</i>					
MEAA_D03.x	Probable <i>N</i> -acetylglucosamine-6-phosphate deacetylase	73	14089782 (1e-06)	A69664 (4e-08)	<i>St. aureus</i> , 15893481 (9e-06)
MHAA_C12.x	1-Acyl- <i>sn</i> -glycerol-3-phosphate acyltransferase	112	13508038 (2e-10)	10174252 (0.008)	<i>M. pulmonis</i> , 15828585 (2e-6)
MHAA_H07.x	orfa	65	14089575 (1e-09)	2633961 (8e-12)	<i>S. citri</i> , 1143008 (3e-29)
MSAD_G09.y	1-Acyl- <i>sn</i> -glycerol-3-phosphate acyltransferase	117	13508038 (5e-22)	2633289 (3e-05)	<i>M. pneumoniae</i> , 13508038 (6e-20)
<i>Cellular processes</i>					
MEAA_C12.y	DnaK protein (hsp 70)	189	8920287 (2e-51)	P45554 (5e-58)	<i>E. rhusiopathiae</i> , 1169374 (4e-58)
MHAA_B03.x	GTP-binding membrane protein (LepA)	80	6899301 (2e-17)	12724067 (1e-18)	<i>La. lactis</i> , 15673090 (2e-05)
<i>Energy metabolism</i>					
MEAA_B07.x	Dihydroliipoamide dehydrogenase	43	1674136 (1e-10)	12722900 (3e-07)	<i>Pseudomonas putida</i> , 1706442 (7e-16)
MEAA_C10.x	Phosphomannomutase (PMM)	133	1352196 (3e-19)	C69835 (8e-29)	<i>B. halodurans</i> , 15613669 (8e-19)
MHAA_E06.y	Glycerol-3-phosphate dehydrogenase	167	14089537 (2e-15)	1146220 (8e-27)	<i>St. aureus</i> , 15924464 (2e-24)
MHAA_A06.y	ATP synthase $\beta$ chain	213	14089679 (1e-94)	10176378 (1e-86)	<i>M. pulmonis</i> , 15828737 (1e-76)
MHAA_D08.y	Fructose-biphosphate aldolase	107	12044873 (2e-23)	10944298 (1e-21)	<i>Cl. acetobutylicum</i> , 15894114 (2e-25)
MHAA_F07.y	Transketolase	63	14089925 (6e-08)	7328298 (6e-06)	<i>M. pulmonis</i> , 14089925 (7e-08)
MHAA_G06.x	Pyruvate kinase	82	14089653 (1e-07)	ns	<i>S. citri</i> , 2384686 (3e-17)
MSAC_G02.y	ATP synthase B chain precursor	88	2146068 (1e-11)	12061042 (0.001)	<i>M. pneumoniae</i> , 13508341 (1e-04)
MSAD_B06.x	Phosphopyruvate hydratase/enolase	55	14089932 (2e-17)	8670811 (8e-16)	<i>A. aeolicus</i> , 6015091 (1e-14)
<i>Purines, pyrimidines, nucleosides, and nucleotides</i>					
MEAA_B12.x	Adenylosuccinate lyase	188	—	WZB5SDS (2e-37)	<i>Lactobacillus sakei</i> , 15217116 (1e-32)
MEAA_C08.x	Adenylosuccinate synthetase	123	—	10176653 (1e-30)	<i>S. citri</i> , 1709937 (1e-49)

Deduced protein sequences were blastp searched against the non-redundant database and the Mycoplasmataceae and *Bacillus/Clostridium* protein databases at NCBI. Sequence tag identity (ID) and deduced amino acid (aa) length are indicated and for each sequence tag accession numbers and  $E$ -values of entries with highest similarities are listed. The organism of entry with highest similarity is listed for the non-redundant (nr) database search results. *A.*, *Aquifex*; *B.*, *Bacillus*; *C.*, *Chlamydia*; *Ca.*, *Campylobacter*; *Chl.*, *Chlorobium*; *Cl.*, *Clostridium*; *E.*, *Erysipelothrix*; *En.*, *Enterococcus*; *G.*, *Geobacillus*; *L.*, *Listeria*; *La.*, *Lactococcus*; *M.*, *Mycoplasma*; *My.*, *Mycobacterium*; *S.*, *Spiroplasma*; *St.*, *Staphylococcus*; *Str.*, *Streptococcus*; *T.*, *Thermotoga*; *U.*, *Urealyticum*; *V.*, *Vibrio*; *X.*, *Xylella*; *Y.*, *Yersinia*. —, no significant hit and sequence is absent; ns,  $E$ -value  $> 10^{-5}$  but sequence is present in genomes of one or more members of the Mycoplasmataceae or *Bacillus/Clostridium* group.

Table 2 (continued)

Sequence tag ID	Identity	Sequence length (aa)	Best entry Mycoplasmataceae, accession No. ( <i>E</i> -value)	Best entry <i>Bacillus/Clostridium</i> , accession No. ( <i>E</i> -value)	Best entry GenBank organism, accession No. ( <i>E</i> -value)
MEAA_D08.y	Deoxyguanosine kinase	131	–	586859 (1e-08)	<i>M. mycoides</i> , 16040925 (7e-36)
MHAA_B06.x	Thymidine kinase	106	14089558 (4e-16)	2636243 (4e-26)	<i>B. subtilis</i> , 16080759 (9e-21)
MHAA_C05.y	Cytidine deaminase	113	D53312 (2e-24)	10173982 (3e-19)	<i>M. pirum</i> , 1345713 (3e-22)
MHAA_H10.y	Deoxyguanosine kinase	144	–	4033719 (3e-11)	<i>M. mycoides</i> , 16040925 (2e-32)
MSAC_D04.y	Adenine phosphoribosyltransferase	149	14089736 (2e-29)	12723536 (9e-33)	<i>T. maritima</i> , 15644136 (2e-33)
MSAC_F08.x	GMP synthetase (glutamine amidotransferase)	89	–	3483135 (2e-50)	<i>X. fastidiosa</i> , 15839020 (6e-56)
<i>Regulatory functions</i>					
MHAA_A11.y	RNA polymerase $\sigma$ factor (rpoD)	114	12045103 (2e-06)	O66381 (1e-07)	<i>Cl. acetobutylicum</i> , 15894582 (3e-06)
MSAC_A08.x	Transcriptional regulator involved in nitrogen regulation (NifR3 family)	121	–	10172709 (1e-29)	<i>B. halodurans</i> , 15612660 (3e-28)
MSAD_H07.y	Predicted transcription regulator SinR	80	–	10174744 (3e-06)	<i>Cl. acetobutylicum</i> , 15894128 (3e-19)
<i>Replication</i>					
MEAA_B05.y	DNA-directed DNA polymerase I	193	–	A32949 (3e-35)	<i>Str. pyogenes</i> , 15674390 (2e-32)
MEAA_F12.x	DNA gyrase subunit B	104	14089786 (4e-13)	2558946 (4e-16)	<i>M. capricolum</i> , 17008093 (2e-19)
MHAA_G05.y	Chain A, helicase product complex	171	14090113 (7e-11)	2781090 (4e-17)	<i>G. stearothermophilus</i> , 9257172 (1e-18)
MHAA_H04.x	DNA-directed RNA polymerase $\beta$ subunit	96	600226 (5e-21)	12724825 (3e-19)	<i>S. citri</i> , 1350848 (3e-46)
MSAC_A09.x	ParA family protein	141	12045330 (2e-06)	9968459 (4e-09)	<i>S. citri</i> , 10432498 (7e-16)
MSAC_B02.y	Glucose-inhibited division protein A	78	14089666 (2e-22)	P25812 (7e-18)	<i>M. pulmonis</i> , 14089666 (3e-20)
MSAC_C06.y	Cell division protein FtsH	174	14090194 (2e-43)	S66099 (4e-42)	<i>M. pulmonis</i> , 15829250 (9e-44)
MSAD_B12.y	DNA primase	96	13508092 (1e-15)	664755 (6e-25)	<i>L. innocua</i> , 16800560 (4e-24)
MSAD_G04.y	ATP-dependent helicase PcrA	112	14090183 (4e-15)	P56255 (6e-16)	<i>My. tuberculosis</i> , 15840373, (2e-12)
<i>Transcription</i>					
MHAA_E11.x	Transcription antitermination factor (nusG)	112	14089595 (1e-06)	O08386 (2e-19)	<i>L. monocytogenes</i> , 16802292 (2e-07)
MHAA_H05.y	Polynucleotide phosphorylase (PNPase)	206	–	1184680 (3e-82)	<i>B. subtilis</i> , 16078732 (4e-71)
MSAC_H02.y	DNA-directed RNA polymerase $\alpha$ chain	114	6601578 (5e-22)	12725120 (1e-18)	<i>M. capricolum</i> , 629301 (1e-28)
MSAD_B09.x	Transcription antitermination protein NusG	68	ns	12725158 (5e-09)	<i>Str. coelicolor</i> , 1709420 (2e-06)
<i>Translation</i>					
MEAA_B02.x	ATP-dependent protease (lon-protease)	122	1674198 (8e-14)	B42375 (1e-15)	<i>V. cholerae</i> , 15641922 (1e-13)
MEAA_B09.x	50S ribosomal protein L21	38	14089744 (8e-05)	12724034 (3e-08)	<i>Str. pyogenes</i> , 15674860 (4e-08)
MEAA_B09.y	Valine-tRNA ligase	121	1351181 (3e-20)	10175660 (9e-28)	<i>St. aureus</i> , 15927242 (2e-21)
MEAA_C04.y	Cysteinyl tRNA synthetase	76	1351147 (5e-12)	12724882 (6e-12)	<i>Cl. stricklandii</i> , 6899996 (4e-10)
MEAA_C06.x	50S ribosomal protein L3	40	14090004 (4e-08)	P42920 (2e-05)	<i>M. capricolum</i> , 132957 (2e-06)
MEAA_D03.y	50S ribosomal protein L2	109	14090000 (7e-36)	P04257 (3e-36)	<i>M. capricolum</i> , 71083 (4e-20)
MEAA_D09.x	Translation initiation factor 2 (infB)	108	2497279 (5e-40)	10175033 (3e-35)	<i>M. genitalium</i> , 12044994 (7e-49)
MEAA_G12.y	Prolyl-tRNA synthetase	217	14089596 (7e-63)	13633967 (5e-64)	<i>B. burgdorferi</i> , 15594747 (1e-50)
MHAA_A08.y	Translation elongation factor G (EF-G)	180	14089842 (5e-78)	10172743 (3e-92)	<i>B. halodurans</i> , 15612694 (2e-80)
MHAA_A11.x	Hypothetical protein similar to O-sialoglycoprotein endopeptidase	68	14089531 (2e-18)	1945110 (1e-19)	<i>St. aureus</i> , 15927624 (8e-15)
MHAA_B08.y	50S ribosomal protein L4	38	2766504 (9e-36)	S24364 (3e-52)	<i>M. capricolum</i> , 132981 (2e-45)

Table 2 (continued)

Sequence tag ID	Identity	Sequence length (aa)	Best entry Mycoplasmataceae, accession No. ( <i>E</i> -value)	Best entry <i>Bacillus/Clostridium</i> , accession No. ( <i>E</i> -value)	Best entry GenBank organism, accession No. ( <i>E</i> -value)
MHAA_C07.x	Asparaginyl-tRNA synthetase	119	14090186 (1e-28)	12724857 (4e-14)	<i>Cl. acetobutylicum</i> , 15896505 (1e-32)
MHAA_D07.y	Seryl-tRNA synthetase	90	1361847 (8e-26)	12724729 (9e-28)	<i>A. aeolicus</i> , 15605830 (7e-24)
MHAA_C09.x	50S ribosomal protein L5	88	3844757 (2e-29)	4512416 (3e-33)	<i>B. halodurans</i> , 15612709 (5e-27)
MHAA_C09.y	30S ribosomal protein S3	59	3914904 (1e-04)	ns	<i>S. citri</i> , O31161 (7e-26)
MHAA_C11.x	30S ribosomal protein S8	85	14089988 (4e-17)	P56209 (7e-17)	<i>M. capricolum</i> , 134021 (2e-23)
MHAA_C11.y	50S ribosomal protein L17 (fragment)	119	14089975 (8e-27)	P07843 (3e-31)	<i>M. capricolum</i> , 7674204 (1e-38)
MHAA_D12.y	50S ribosomal protein L19	66	14089881 (1e-14)	10175098 (6e-20)	<i>B. halodurans</i> , 15615041 (1e-18)
MHAA_E03.y	Isoleucyl-tRNA synthetase	101	14090082 (3e-12)	437916 (2e-23)	<i>St. aureus</i> , 1174521 (1e-24)
MHAA_E05.y	Threonyl-tRNA synthetase	132	13508292 (1e-38)	143766 (8e-42)	<i>U. urealyticum</i> , 13358098 (5e-45)
MSAC_A11.x	Glutamyl-tRNA synthetase	131	13508417 (4e-16)	289282 (4e-20)	<i>B. subtilis</i> , 16077160 (2e-15)
MSAC_B10.x	Phenylalanyl-tRNA synthetase $\beta$ chain	175	ns	40054 (3e-06)	<i>C. pneumoniae</i> , BAA98801.1 (4e-10)
MSAC_C04.y	Isoleucyl-tRNA synthetase	180	14090082 (2e-32)	10175165 (3e-41)	<i>La. sakei</i> , 15487790 (5e-27)
MSAC_H02.y	DNA-directed DNA polymerase ( $\alpha$ chain)	107	6601579 (2e-25)	10172773 (2e-30)	<i>M. capricolum</i> , 629301 (1e-28)
MSAD_B06.y	Tryptophanyl-tRNA synthetase	145	14090160 (7e-26)	10175491 (2e-34)	<i>B. halodurans</i> , 15615433 (1e-28)
MSAD_B12.x	Glycyl-tRNA synthetase	201	6899491 (1e-63)	4584090 (1e-30)	<i>St. aureus</i> , 15924555 (1e-55)
MSAD_E10.x	Heat shock protein GroEL	83	12045254 (4e-23)	12723267 (5e-35)	<i>En. faecalis</i> , 15625350 (1e-33)
MSAD_H08.x	Ribosomal large subunit pseudouridine synthase B	95	14089751 (7e-04)	410137 (1e-15)	<i>B. subtilis</i> , 466190 (1e-31)
PE_05.x	Histidyl-tRNA synthetase	220	12044885 (7e-29)	3915057 (4e-45)	<i>L. innocua</i> , 16800623 (2e-07)
PE_14.y	Peptide chain release factor 1 (RF-I)	144	1350577 (5e-45)	S55437 (2e-44)	<i>M. capricolum</i> , 2500137 (6e-55)
PE_21.y	50S ribosomal protein L2	195	14090000 (5e-66)	P04257 (5e-72)	<i>M. capricolum</i> , 71083 (7e-77)
PH_04.x	Glycyl-tRNA synthetase	68	14089865 (1e-17)	4584090 (2e-24)	<i>B. cereus</i> , 4584090 (2e-17)
PS_02.y	30S ribosomal protein S17	85	14089993 (4e-22)	P23828 (7e-31)	<i>S. citri</i> , 3122807 (3e-39)
<i>Transport and binding proteins</i>					
MEAA_F10.x	Phosphotransferase EII (PTS system)	138	14089430 (8e-14)	2633144 (6e-15)	<i>M. capricolum</i> , 530422 (9e-15)
MEAA_G05.y	Phosphate ABC transporter, permease protein	177	1361743 (4e-25)	4530449 (1e-30)	<i>V. cholerae</i> , 15600843 (1e-23)
MEAA_H04.x	ABC transporter	168	2146659 (1e-51)	12723139 (8e-46)	<i>U. urealyticum</i> , 13358103 (2e-47)
MSAC_A07.y	Methylgalactosidase permease ATP-binding protein	229	4914644 (3e-47)	12724309 (6e-47)	<i>U. urealyticum</i> , 13357571 (2e-36)
MSAC_A08.y	ABC transporter, ATP-binding protein	70	12044917 (3e-15)	12724060 (1e-12)	<i>Cl. acetobutylicum</i> , 15894109 (2e-13)
MSAC_C02.x	Highly similar to phosphotransferase system (PTS) fructose-specific enzyme IIABC component	87	1045736 (1e-05)	2633811 (4e-09)	<i>L. innocua</i> , 16801491 (5e-07)
MSAC_D03.y	ABC transporter, ATP-binding protein	115	14089609 (2e-11)	10173618 (3e-21)	<i>T. maritima</i> , 15643786 (5e-19)
MSAC_D06.x	Highly similar to Mg(2+) transport ATPase	140	14089568 (2e-15)	12724231 (8e-31)	<i>L. monocytogenes</i> , 16804726 (5e-20)
MSAC_F10.y	Similar to ABC transporter (ATP-binding protein)	131	14090034 (7e-36)	D70009 (5e-39)	<i>B. subtilis</i> , 16080207 (2e-38)
MSAD_A04.y	Similar to ABC transporter ATP-binding protein – oligopeptide transport	108	14089827 (9e-17)	S11153 (5e-31)	<i>Str. pneumoniae</i> , 15903745 (3e-30)
MSAD_C02.y	Phosphotransferase system, glucose-specific IIABC component	111	14089430 (4e-32)	66867 (2e-21)	<i>M. pulmonis</i> , 14089430 (4e-32)
MSAD_D03.x	Oligopeptide permease (ATP-binding protein)	50	13507956 (8e-07)	1420862 (2e-10)	<i>Str. pyogenes</i> , 15674468 (6e-09)
MSAD_D12.y	Transfer complex protein trsK protein (traK)	75	–	6470167 (1e-09)	<i>B. anthracis</i> , 6470167 (1e-07)
MSAD_E06.y	Cation-transporting P-ATPase	125	14089568 (3e-05)	12724231 (4e-21)	<i>La. lactis</i> , 15673239 (3e-16)

Table 2 (continued)

Sequence tag ID	Identity	Sequence length (aa)	Best entry Mycoplasmataceae, accession No. ( <i>E</i> -value)	Best entry <i>Bacillus/Clostridium</i> , accession No. ( <i>E</i> -value)	Best entry GenBank organism, accession No. ( <i>E</i> -value)
MSAD_F05.x	Phosphate ABC transporter, permease protein	96	13508349 (7e-14)	4530449 (2e-20)	<i>Str. pneumoniae</i> , 15901902 (8e-19)
<i>Other categories</i>					
MEAA_A02.x	Amidase	62	2146059 (1e-06)	ns	<i>M. capricolum</i> , 530426 (6e-10)
MEAA_A03.x	Conserved hypothetical	165	14090195 (1e-08)	467456 (6e-04)	<i>U. urealyticum</i> , 13357633 (2e-10)
MEAA_A06.y	SpoE family protein/cell division protein	141	–	S09411 (2e-31)	<i>Str. pneumoniae</i> , 15900761 (2e-33)
MEAA_B03.y	Probable GTP-binding protein	56	13508214 (5e-12)	1146219 (9e-17)	<i>B. subtilis</i> , 1730915 (6e-15)
MEAA_D11.x	Nitrogen fixation protein NifU	69	–	10176042 (6e-13)	<i>B. halodurans</i> , 15615981 (7e-11)
MEAA_D12.y	Predicted SAM-dependent methyltransferase	145	1045939 (7e-10)	12724027 (1e-16)	<i>Cl. acetobutylicum</i> , 12724027 (1e-16)
MEAA_E07.x	RNA-binding Sun protein	100	–	2633946 (4e-08)	<i>B. subtilis</i> , 16078637 (4e-04)
MEAA_E08.y	Conserved hypothetical protein	98	–	10173873 (1e-21)	<i>St. aureus</i> , 10173873 (1e-21)
MEAA_E09.x	Hypothetical protein	112	–	–	<i>Chl. tepidum</i> , 10039641 (2e-24)
MEAA_E12.x	199 aa long conserved hypothetical protein	125	S73881 (3e-05)	P54501 (1e-08)	<i>B. subtilis</i> , P54501 (1e-08)
MEAA_F10.y	Similar to putative phosphoprotein phosphatase	56	7109691 (5e-06)	10175125 (1e-06)	<i>L. monocytogenes</i> , 10175125 (1e-06)
MHAA_A09.x	Probable thiol peroxidase	77	14090123 (6e-09)	P72500 (9e-12)	<i>Str. pneumoniae</i> , 15901486 (1e-09)
MHAA_A10.y	Hypothetical protein	152	1674179 (7e-10)	2634923 (7e-13)	<i>A. aeolicus</i> , 7451802 (2e-13)
MHAA_B12.x	Conserved GTP-binding protein	124	14089767 (2e-09)	12724592 (2e-08)	<i>L. monocytogenes</i> , 14089767 (2e-09)
MHAA_B12.y	tRNA $\delta$ (2) isopentenylpyrophosphate transferase	75	–	13701103 (3e-11)	<i>St. aureus</i> , 15924294 (2e-11)
MHAA_C06.y	P115-like ( <i>Mycoplasma hyorhinae</i> ) ABC transporter ATP-binding protein	168	14090129 (8e-52)	10175107 (2e-46)	<i>M. pulmonis</i> , 14090129 (8e-52)
MHAA_D09.x	Probable thiol peroxidase	56	–	P31307 (5e-07)	<i>Cl. acetobutylicum</i> , 15896549 (5e-07)
MHAA_E04.x	Acyl carrier protein phosphodiesterase (ACP phosphodiesterase)	118	14089726 (2e-39)	2619052 (6e-09)	<i>M. pulmonis</i> , 14089726 (2e-39)
MHAA_F06.y	Partitioning or sporulation protein (ParA) (soj protein)	109	ns	9968459 (3e-12)	<i>L. monocytogenes</i> , 9968459, (3e-12)
MHAA_F12.x	Conserved hypothetical protein	87	14089574 (3e-09)	12723043 (2e-10)	<i>Str. pyogenes</i> , 12723043 (2e-10)
MHAA_G03.y	Probable type I restriction enzyme restriction chain	134	14090092 (7e-06)	13700111 (2e-24)	<i>St. aureus</i> , 15923185 (8e-27)
MHAA_G08.y	Exodeoxyribonuclease V ( $\alpha$ subunit)	84	14090197 (2e-07)	2635193 (4e-12)	<i>C. pneumoniae</i> , 15835659 (9e-13)
MHAA_H02.x	Conserved hypothetical protein	99	–	2635763 (6e-29)	<i>St. aureus</i> , 15923836 (2e-26)
MSAC_A06.x	Conserved hypothetical protein	195	–	12724713 (3e-34)	<i>Y. pestis</i> , 16121243 (2e-29)
MSAC_B11.y	Hypothetical protein	140	3845056 (6e-16)	12724031 (2e-19)	<i>La. lactis</i> , 12724031 (2e-19)
MSAC_C09.y	Conserved hypothetical protein	174	13508006 (5e-11)	13027335 (5e-21)	<i>St. aureus</i> , 13027335, (5e-21)
MSAC_E03.x	Conserved hypothetical protein	43	–	7429432 (3e-05)	<i>Synechocystis</i> sp. PCC 680 7444728, (4e-06)
MSAC_H01.x	Conserved hypothetical protein	128	150165 (5e-09)	10175107 (3e-10)	<i>Str. pneumoniae</i> , 10175107 (3e-10)
MSAD_E03.x	Nitroreductase	151	–	7432647 (2e-05)	<i>Ca. jejuni</i> , 15792391 (2e-08)
MSAD_F02.y	Hypothetical protein	133	P75273 (2e-06)	5420109 (3e-12)	<i>Str. thermophilus</i> , 5420109 (3e-12)
MSAD_F01.x	Hypothetical 35.3 kDa protein, SLR1819	91	–	P37497 (1e-05)	<i>Synechocystis</i> sp. PCC 6803, P73709, (2e-07)
MSAD_G02.x	Conserved hypothetical protein	99	14090099 (3e-17)	7328260 (8e-15)	<i>M. pulmonis</i> , 7328260 (8e-15)
PH_01.y	BH2415 – unknown conserved protein	82	–	10175035 (4e-19)	<i>B. halodurans</i> , 10175035 (4e-19)
PH_05.y	Hypothetical protein in fibril gene 3' region	315	12045292 (6e-14)	–	<i>S. citri</i> , P27712 (e-101)

### 3.5. Amino acid, purine, pyrimidine, nucleoside and nucleotide metabolism

Mycoplasmataceae species lack most genes involved in de novo biosynthesis of pyrimidines, purines and amino acids [7–10]. However, in contrast to mycoplasmas and *U. urealyticum*, the *S. kunkelii* genome seems to harbor the nucleotide and/or amino acid biosynthesis genes encoding adenylosuccinate lyase, adenylosuccinate synthase, GMP synthase, deoxyguanosine kinase, and foylpolylglutamate synthase/dihydrofolate synthetase (*folC*) (Table 2). Adenylosuccinate lyase is a tetrameric enzyme involved in de novo synthesis of inosine monophosphate (IMP) and adenosine monophosphate [26], adenylosuccinate synthase catalyzes the first step in de novo biosynthesis of AMP [27], and guanine monophosphate (GMP) synthase catalyzes the last step from IMP into GMP [26]. Deoxyadenosine/deoxyguanosine kinase and deoxyadenosine/deoxycytidine kinase are required, together with thymidine kinase, for deoxynucleotide synthesis in *Lactobacillus acidophilus* [28]. Interestingly, the deoxyguanosine kinase gene is present in the mollicute *Mycoplasma mycoides*. Within the order Mycoplasmatales, *M. mycoides* belongs to the Entomoplasmataceae, a family more closely related to the Spiroplasmataceae than the Mycoplasmataceae [4]. The *folC* gene product is essential for production of glycine, methionine, purine and thymidine [29]. These data suggest that *S. kunkelii* can synthesize more amino acids and nucleotides de novo than Mycoplasmataceae species do, which is in agreement with experimental evidence that spiroplasma culturing media are less complex than those of the culturable mycoplasmas [13,25].

### 3.6. Cell envelope

The sequence data indicate that *S. kunkelii* harbors at least two cell envelope biosynthesis genes that are absent from members of the Mycoplasmataceae. The *gcpE* gene is involved in the acetylation of peptidoglycans and isoprenoid biosynthesis and is broadly distributed in eubacteria and plants [30,31]. MreB is a cytoskeletal protein and forms a filamentous helical structure close to the cell surface of eubacteria, and has an actin-like role in bacterial cell morphogenesis [32]. The clear morphological differences between spiroplasmas and Mycoplasmataceae and our finding that the *mreB* gene is absent from Mycoplasmataceae genomes but present in *S. kunkelii* suggest that MreB may have a critical role in the unique helical cell structure of spiroplasmas.

### 3.7. Regulatory functions

Our sequence data show that the *S. kunkelii* regulatory mechanisms are more complex than those of the Mycoplasmataceae. Three genes were identified encoding the regulatory proteins NifR3, SinR and PNPase that were absent

in the three sequenced mycoplasmas and *U. urealyticum* but present in Firmicutes. NifR3 is important for the regulation of the dormant and vegetative cell stages of the ciliate *Sterkiella histriomuscorum* [33]. The function of NifR3 in bacteria is not known. SinR is involved in the transition of a vegetative stage to sporulation in *Bacillus subtilis* in response to nutrient depletion [34]. Spiroplasmas do not make spores, but are extremely pleomorphic. It is tempting to speculate that NifR3 and SinR may be involved in *S. kunkelii* cell shape regulation as a response to nutrient availability. A third regulatory protein, polynucleotide phosphorylase (PNPase) is responsible for mRNA decay, translation activation and transcript stabilization in *B. subtilis* [35,36]. The loss of PNPase is lethal for *E. coli*, but affects only competence development in *B. subtilis* [37,38] and may affect competence of *S. kunkelii* as well. The discovery of these regulatory factors in *S. kunkelii* is surprising as, thus far, members of the Mycoplasmataceae are known to lack major regulators of gene expression [7,8,39,40].

### 3.8. Replication

One surprising finding was that the DNA polymerase I protein of *S. kunkelii* did not match the DNA polymerases of Mycoplasmataceae, whereas it had significant similarity to the DNA polymerase I proteins of *Streptococcus* species (*E*-values:  $2e^{-35}$  and  $2e^{-32}$ , sequence tag MEAA\_B05.y, Table 2). Closer analysis revealed that the 193 amino acid sequence tag of *S. kunkelii* was similar to the C-terminal polymerase domain of DNA polymerase I. In contrast, putative DNA polymerases I of *M. genitalium* (GenBank accession No. I64228), *M. pneumoniae* (S73784), *U. urealyticum* (C82895) and *M. pulmonis* (CAC13893) are ~300 amino acids in size and consist of the N-terminal 5'-3' exonuclease part (proofreading) part but lack the C-terminal 3'-5' exonuclease and polymerase domains (Klenow fragment) of the enzyme [10]. This finding suggests that, unlike mycoplasmas and *U. urealyticum*, the *S. kunkelii* *polA* gene may encode the full-length DNA polymerase I protein including the proofreading and Klenow domains similarly to that of *Streptococcus pneumoniae* [41].

### 3.9. Transport and binding proteins

In contrast to Mycoplasmataceae, the *S. kunkelii* genome harbors at least one copy of a *traK* homologue. *S. kunkelii* *traK* has the highest similarity to *traK* of the *B. anthracis* virulence plasmid pX02.09 (Table 2) [42]. This conserved protein family binds DNA and couples plasmid to membrane proteins for transport to the mating cell and/or are pathogenicity factors involved in transport of virulence factors to the extracellular environment of bacteria [43,44]. The function of *S. kunkelii* TraK protein remains to be investigated.

Two *S. kunkelii* sequence tags (MSAC\_C02.x and MSAD\_C02.y, Table 2) harbor sequences similar to fructose permease of the phosphoenolpyruvate:fructose phosphotransferase system (fructose PTS). Mutagenesis of the operon encoding fructose PTS proteins in another leafhopper-transmitted plant-pathogenic spiroplasma, *S. citri*, significantly decreases plant pathogenicity [45]. The most likely explanation is that utilization of fructose in the plant sieve tubes by *S. citri* may interfere with the normal physiology of the plant causing chlorosis, stunting and wilting [45]. This may be true for *S. kunkelii* in sieve tubes of corn plants as well. Homologues of fructose PTS proteins were also identified in Mycoplasmataceae and other Firmicutes (Table 2).

### 3.10. Genes in other categories

The *S. kunkelii* genome harbors at least one copy of a *spoIIIE* homologue that is not found in the Mycoplasmataceae genome sequenced so far (sequence tag MEAA\_A06.y, Table 2). In *B. subtilis*, the *spoIIIE* gene product is involved in the coordination of chromosome segregation and clearing DNA from the site of division during septum formation [46] and, therefore, is likely to be involved in *S. kunkelii* cell division.

A *nifU*-like gene of 228 nucleotides in length was identified in this sequence project and harbors solely the C-terminal conserved domain containing two conserved cysteines, whereas functional iron–sulfur cluster-binding NifU proteins contain additional middle domains with four conserved cysteines (sequence tag MEAA\_D11.x, Table 2) [47–49]. Several smaller NifU-like genes are also found in the nitrogen fixing *Rhodobacter* and *Azotobacter* species and single gene mutagenesis studies show that they are not essential for survival or nitrogen fixation of bacteria [50]. The functions of these shorter NifU-like genes are not known.

A sequence similar to the oxygen-insensitive NAD(P)H nitroreductase was found in the *S. kunkelii* database (sequence tag MSAD\_E03.x, Table 2). This enzyme catalyzes the reduction of a variety of nitroaromatic compounds to highly toxic metabolites [51]. Although absent from the mycoplasmas and *U. urealyticum* genomes, it is found in the small (~650 kb) genome of the insect vectored apple proliferation phytoplasma (gi405516) [52]. It is noteworthy that phytoplasmas are the only other group of Mollicutes that infect plants causing characteristic chlorosis and stunting symptoms.

Two sequence tags have identity to the 20 kDa PsaD thiol peroxidase proteins of *Streptococcus* species [53,54]. Tag MHAA\_A09.x contains the N-terminal part of this protein, whereas MHAA\_D09.x harbors the C-terminal end. In *Str. pneumoniae*, the *psaD* gene is located downstream from the *psa* locus with the *psaA*, *psaB* and *psaC* genes encoding an ABC-type Mn permease complex [54]. Mutagenesis of each of four *psa* genes resulted in penicillin tolerance, defective adhesion and reduced transformation efficiency of *Str. pneumoniae* [54]. The *psaA* gene encodes an adhesin-like surface protein, and *psaA* and *psaD* related genes were identified in *Streptococcus sanguis*, *Streptococcus parasanguis* and *Streptococcus gordonii* [53].

Several sequence tags have identity to conserved hypothetical proteins that are lacking from the mycoplasmas and *U. urealyticum* genomes sequenced thus far (Other categories, Table 2). We found only one sequence tag with identity to Mollicute sequences but not those of the *Bacillus/Clostridium* group (sequence tag PH\_05.y, Table 2). The deduced protein sequence of this tag is a homologue of a hypothetical protein encoded by a gene in the downstream region of the fibril gene region of *S. citri* [55]. The fibril protein is important for the helical cell shape and motility of spiroplasmas [56,57] and the gene encoding it is lacking from the genomes of the oval-shaped mycoplasmas and *U. urealyticum* [7–10]. Because the hypothet-

Table 3  
*S. kunkelii* sequence tags with similarity to rRNA genes

Sequence tag ID	Identity	Accession No., organism	E-value
MEAA_E09.x	16S rDNA	46914, <i>S. citri</i>	0.0
		175961, <i>S. poulsonii</i>	0.0
		175965, <i>S. citri</i>	0.0
		175964, <i>S. apis</i>	0.0
		175967, <i>S. mirum</i>	0.0
		175969, <i>S. monobiae</i>	0.0
		175962, <i>S. taiwanense</i>	e-180
		175970, <i>S. diabroticae</i>	e-179
		175963, <i>S. gladiatoris</i>	e-171
		175473, <i>Entomoplasma melaleuca</i>	e-166
		MHAA_F02.y	16S rDNA, 16S/23S spacer region, 23S rDNA
4456860, <i>Spiroplasma</i> sp.	e-151		
2707198, <i>S. citri</i>	e-125		
5821442, <i>M. putrefaciens</i>	2e-19		

*S. kunkelii* sequence tags were searched against the full GenBank nr nucleotide database with the blastn algorithm. Identities, accession numbers and organisms, and E-values of the first 10 and four entries of respectively MEAA\_E09.x and MHAA\_F02.y, the only sequence tags with similarities to rRNA genes, are listed.

ical protein gene is localized near the fibril protein gene [55] and is unique to Mollicutes (Table 2), this hypothetical protein may be an important constituent of the mollicute cytoskeleton.

### 3.11. Ribosomal RNA genes

Clones MEAA\_E09 and MHAA\_F02 contained part of the 16S and 23S ribosomal RNA (rRNA) genes and the 16S–23S internal spacer with closest similarity to rRNA gene regions from *S. citri*, as is expected from the *S. kunkelii* phylogenetic position [4] (Table 3). *S. kunkelii* rRNA genes have not been sequenced previously.

### 3.12. Conclusions

In summary, our data show that, in addition to the large number of spiroplasma phage DNA insertions, *S. kunkelii* also harbors more amino acid and nucleotide biosynthesis, transcription regulation, cell envelope and DNA transport/binding genes than the genomes of the Mycoplasmataceae species do. Our data also demonstrate that genome comparisons among Mollicutes are extremely informative because of their small genome sizes, broad host range, differences in morphology, and well-defined biology. In addition to the already completed genome sequences of four Mycoplasmataceae species, several genome sequence projects of Mollicutes in other families are ongoing including those of *M. mycoides* and *Mycoplasma capricolum* in the family Entomoplasmataceae (<http://www.ncbi.nlm.nih.gov/PMGifs/Genomes/bact.html>), and *S. kunkelii* (<http://www.genome.ou.edu/spiro.html>) and *S. citri* (<http://www.cwu.edu/~verheys/s.citri/>). Genome comparison of species within a family, among families within the class Mollicutes and between Mollicutes and Firmicutes should prove extremely valuable.

### Acknowledgements

The authors thank Dr. Margareth Redinbaugh for carefully reading the manuscript and Dr. Robert Davis for help with establishing *S. kunkelii* in vitro cultures at the OARDC. This research was funded by the OARDC research enhancement and competitive grants program.

### References

- [1] Nault, L.R. (1980) Maize bushy stunt and corn stunt: a comparison of disease symptoms, pathogen host ranges and vectors. *Phytopathology* 70, 709–712.
- [2] Purcell, A.H. (1982) Insect vector relationship with procaryotic plant pathogens. *Annu. Rev. Phytopathol.* 20, 397–417.
- [3] Markham, P.G. (1983) Spiroplasmas in leafhoppers: a review. *Yale J. Biol. Med.* 56, 745–751.

- [4] Bové, J.M. and Garnier, M. (1997) In: *Developments in Plant Pathology, Pathogen and Microbial Contamination Management in Micropropagation*, Vol. 12 (Cassels, A.C., Ed.), pp. 45–60. Kluwer Academic Publishers, Dordrecht.
- [5] Razin, S., Yogev, D. and Naot, Y. (1998) Molecular biology and pathogenicity of mycoplasmas. *Microbiol. Mol. Biol. Rev.* 62, 1094–1156.
- [6] Bové, J.M. (1997) Spiroplasmas: infectious agents of plants, arthropods and vertebrates. *Wien. Klin. Wochenschr.* 109, 604–612.
- [7] Fraser, C.M., Gocayne, J.D., White, O., Adams, M.D., Clayton, R.A., Fleischmann, R.D., Bult, C.J., Kerlavage, A.R., Sutton, G. and Kelley, J.M. et al. (1995) The minimal gene complement of *Mycoplasma genitalium*. *Science* 270, 397–403.
- [8] Himmelreich, R., Hilbert, H., Plagens, H., Pirkel, E., Li, B.C. and Herrmann, R. (1996) Complete sequence analysis of the genome of the bacterium *Mycoplasma pneumoniae*. *Nucleic Acids Res.* 24, 4420–4449.
- [9] Glass, J.I., Lefkowitz, E.J., Glass, J.S., Heiner, C.R., Chen, E.Y. and Cassell, G.H. (2000) The complete sequence of the mucosal pathogen *Ureaplasma urealyticum*. *Nature* 407, 757–762.
- [10] Chambaud, I., Heilig, R., Ferris, S., Barbe, V., Samson, D., Galisson, F., Moszer, I., Dybvig, K., Wroblewski, H., Viari, A., Rocha, E.P. and Blanchard, A. (2001) The complete genome sequence of the murine respiratory pathogen *Mycoplasma pulmonis*. *Nucleic Acids Res.* 29, 2145–2153.
- [11] Ebbert, M.A. and Nault, L.R. (2001) Survival in *Dalbulus* leafhopper vectors improves after exposure to maize stunting pathogens. *Entomol. Exp. Appl.* 100, 311–324.
- [12] Ebbert, M.A. and Nault, L.R. (1994) Improved overwintering ability in *Dalbulus maidis* (Homoptera: Cicadellidae) vectors infected with *Spiroplasma kunkelii* (Mycoplasmatales: Spiroplasmataceae). *Environ. Entomol.* 23, 634–644.
- [13] Lee, I.-M. and Davis, R.E. (1989) Serum-free media for cultivation of spiroplasmas. *Can. J. Microbiol.* 35, 1092–1099.
- [14] Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) *Molecular Cloning: a Laboratory Manual*, 2nd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- [15] Ewing, B., Hillier, L., Wendl, M.C. and Green, P. (1998) Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res.* 8, 175–185.
- [16] Ewing, B. and Green, P. (1998) Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res.* 8, 186–194.
- [17] Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990) Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410.
- [18] Foissac, X., Bové, J.M. and Saillard, C. (1997) Sequence analysis of *Spiroplasma phoeniceum* and *Spiroplasma kunkelii* spiralin genes and comparison with other spiralin genes. *Curr. Microbiol.* 35, 240–243.
- [19] Bébéar, C.-M., Aullo, P., Bové, J.M. and Renaudin, J. (1996) *Spiroplasma citri* virus SpV1: characterization of viral sequences present in the spiroplasmal host chromosome. *Curr. Microbiol.* 32, 134–140.
- [20] Simpson, A.J., Reinach, F.C., Arruda, P., Abreu, F.A., Acencio, M., Alvarenga, R., Alves, L.M., Araya, J.E., Baia, G.S., Baptista, C.S., Barros, M.H., Bonaccorsi, E.D., Bordin, S., Bové, J.M., Briones, M.R., Bueno, M.R., Camargo, A.A., Camargo, L.E., Carraro, D.M., Carrer, H., Colauto, N.B., Colombo, C., Costa, F.F., Costa, M.C., Costa-Neto, C.M., Coutinho, L.L., Cristofani, M., Dias-Neto, E., Docena, C., El-Dorry, H., Facincani, A.P., Ferreira, A.J., Ferreira, V.C., Ferro, J.A., Fraga, J.S., Franca, S.C., Franco, M.C., Frohme, M., Furlan, L.R., Garnier, M., Goldman, G.H., Goldman, M.H., Gomes, S.L., Gruber, A., Ho, P.L., Hoheisel, J.D., Junqueira, M.L., Kemper, E.L., Kitajima, J.P., Krieger, J.E., Kuramae, E.E., Laigret, F., Lambais, M.R., Leite, L.C., Lemos, E.G., Lemos, M.V., Lopes, S.A., Lopes, C.R., Machado, J.A., Machado, M.A., Madeira, A.M., Madeira, H.M., Marino, C.L., Marques, M.V., Martins, E.A., Martins, E.M., Matsukuma, A.Y., Menck, C.F., Miracca, E.C., Miyaki, C.Y., Monteiro-Vitorello, C.B., Moon, D.H., Nagai, M.A., Nascimento, A.L., Netto, L.E., Nhani Jr., A., Nobrega,

- F.G., Nunes, L.R., Oliveira, M.A., de Oliveira, M.C., de Oliveira, R.C., Palmieri, D.A., Paris, A., Peixoto, B.R., Pereira, G.A., Pereira Jr., H.A., Pesquero, J.B., Quaggio, R.B., Roberto, P.G., Rodrigues, V., de, M.R.A.J., de Rosa Jr., V.E., de Sa, R.G., Santelli, R.V., Sawasaki, H.E., da Silva, A.C., da Silva, A.M., da Silva, F.R., da Silva Jr., W.A., da Silveira, J.F., Silvestri, M.L., Siqueira, W.J., de Souza, A.A., de Souza, A.P., Terenzi, M.F., Truffi, D., Tsai, S.M., Tshako, M.H., Vallada, H., Van Sluys, M.A., Verjovski-Almeida, S., Vettore, A.L., Zago, M.A., Zatz, M., Meidanis, J. and Setubal, J.C. (2000) The genome sequence of the plant pathogen *Xylella fastidiosa*. *Nature* 406, 151–157.
- [21] Ye, F., Laigret, F. and Bove, J.M. (1994) A physical and genomic map of the prokaryote *Spiroplasma melliferum* and its comparison with the *Spiroplasma citri* map. *C. R. Acad. Sci.* 317, 392–398.
- [22] Ye, F., Laigret, F., Carle, P. and Bove, J.M. (1995) Chromosomal heterogeneity among various strains of *Spiroplasma citri*. *Int. J. Syst. Bacteriol.* 45, 729–734.
- [23] Ye, F., Melcher, U., Rascoe, J.E. and Fletcher, J. (1996) Extensive chromosome aberrations in *Spiroplasma citri* strain BR3. *Biochem. Genet.* 34, 269–286.
- [24] Fletcher, J., Wayadande, A., Melcher, U. and Fengchun, Y. (1998) The phytopathogenic mollicute-insect vector interface: a closer look. *Phytopathology* 88, 1351–1358.
- [25] Razin, S. (1994) DNA probes and PCR in diagnosis of mycoplasma infections. *Mol. Cell. Probes* 8, 497–511.
- [26] Mantsala, P. and Zalkin, H. (1992) Cloning and sequence of *Bacillus subtilis purA* and *guaA*, involved in the conversion of IMP to AMP and GMP. *J. Bacteriol.* 174, 1883–1890.
- [27] Honzatko, R.B. and Fromm, H.J. (1999) Structure-function studies of adenylosuccinate synthase from *Escherichia coli*. *Arch. Biochem. Biophys.* 370, 1–8.
- [28] Ma, G.T., Hong, Y.S. and Ives, D.H. (1995) Cloning and expression of the heterodimeric deoxyguanosine kinase/deoxyadenosine kinase of *Lactobacillus acidophilus* R-26. *J. Biol. Chem.* 270, 6595–6601.
- [29] Singer, S., Ferone, R., Walton, L. and Elwell, L. (1985) Isolation of a dihydrofolate reductase-deficient mutant of *Escherichia coli*. *J. Bacteriol.* 164, 470–472.
- [30] Rather, P.N., Solinsky, K.A., Paradise, M.R. and Parojcic, M.M. (1997) *arcC*, an essential gene involved in density-dependent regulation of the 2'-N-acetyltransferase in *Providencia stuartii*. *J. Bacteriol.* 179, 2267–2273.
- [31] Campos, N., Rodriguez-Concepcion, M., Seemann, M., Rohmer, M. and Boronot, A. (2001) Identification of *gcpE* as a novel gene of the 2-C-methyl-D-erythritol 4-phosphate pathway for isoprenoid biosynthesis in *Escherichia coli*. *FEBS Lett.* 488, 170–173.
- [32] Jones, L.J., Carballido-Lopez, R. and Errington, J. (2001) Control of cell shape in bacteria: helical, actin-like filaments in *Bacillus subtilis*. *Cell* 104, 913–922.
- [33] Tourancheau, A.B., Morin, L., Yang, T. and Perasso, R. (1999) Messenger RNA in dormant cells of *Sterkiella histriomuscorum* (Oxytrichiade): identification of putative regulatory gene transcripts. *Protist* 150, 137–147.
- [34] Gaur, N.K., Oppenheim, J. and Smith, I. (1991) The *Bacillus subtilis sin* gene, a regulator of alternate developmental processes, codes for a DNA-binding protein. *J. Bacteriol.* 173, 678–686.
- [35] Wang, W. and Bechhofer, D.H. (1996) Properties of a *Bacillus subtilis* polynucleotide phosphorylase deletion strain. *J. Bacteriol.* 178, 2375–2382.
- [36] Oussenko, I.A. and Bechhofer, D.H. (2000) The *yvaJ* gene of *Bacillus subtilis* encodes a 3'-to-5' exonuclease and is not essential in a strain lacking polynucleotide phosphorylase. *J. Bacteriol.* 182, 2639–2642.
- [37] Donovan, W.P. and Kushner, S.R. (1986) Polynucleotide phosphorylase and ribonuclease II are required for cell viability and mRNA turnover in *Escherichia coli* K-12. *Proc. Natl. Acad. Sci. USA* 83, 120–124.
- [38] Luttinger, A., Hahn, J. and Dubnau, D. (1996) Polynucleotide phosphorylase is necessary for competence development in *Bacillus subtilis*. *Mol. Microbiol.* 19, 343–356.
- [39] Himmelreich, R., Plagens, H., Hilbert, H., Reiner, B. and Herrmann, R. (1997) Comparative analysis of the genomes of the bacteria *Mycoplasma pneumoniae* and *Mycoplasma genitalium*. *Nucleic Acids Res.* 25, 701–712.
- [40] Weiner III, J., Herrmann, R. and Browning, G.F. (2000) Transcription in *Mycoplasma pneumoniae*. *Nucleic Acids Res.* 28, 4488–4496.
- [41] Lopez, P., Martinez, S., Diaz, A., Espinosa, M. and Lacks, S.A. (1989) Characterization of the *polA* gene of *Streptococcus pneumoniae* and comparison of the DNA polymerase I it encodes to homologous enzymes from *Escherichia coli* and phage T7. *J. Biol. Chem.* 264, 4255–4263.
- [42] Okinaka, R., Cloud, K., Hampton, O., Hoffmaster, A., Hill, K., Keim, P., Koehler, T., Lamke, G., Kumano, S., Manter, D., Martinez, Y., Ricke, D., Svensson, R. and Jackson, P. (1999) Sequence, assembly and analysis of pX01 and pX02. *J. Appl. Microbiol.* 87, 261–262.
- [43] Errington, J., Bath, J. and Wu, L.J. (2001) DNA transport in bacteria. *Nat. Rev. Mol. Cell. Biol.* 2, 538–545.
- [44] Christie, P.J. and Vogel, J.P. (2000) Bacterial type IV secretion: conjugation systems adapted to deliver effector molecules to host cells. *Trends Microbiol.* 8, 354–360.
- [45] Gaurivaud, P., Danet, J.L., Laigret, F., Garnier, M. and Bove, J.M. (2000) Fructose utilization and phytopathogenicity of *Spiroplasma citri*. *Mol. Plant-Microbe Interact.* 13, 1145–1155.
- [46] Bath, J., Wu, L.J., Errington, J. and Wang, J.C. (2000) Role of *Bacillus subtilis* SpoIIIE in DNA transport across the mother cell-prespore division septum. *Science* 290, 995–997.
- [47] Ouzounis, C., Bork, P. and Sander, C. (1994) The modular structure of NifU proteins. *Trends Biochem. Sci.* 19, 199–200.
- [48] Nishio, K. and Nakai, M. (2000) Transfer of iron-sulfur cluster from NifU to a ferredoxin. *J. Biol. Chem.* 275, 22615–22618.
- [49] Agar, J.N., Yuvaniyama, P., Jack, R.F., Cash, V.L., Smith, A.D., Dean, D.R. and Johnson, M.K. (2000) Modular organization and identification of a mononuclear iron-binding site within the NifU protein. *J. Biol. Inorg. Chem.* 5, 167–177.
- [50] Masepohl, B., Angermuller, S., Hennecke, S., Hubner, P., Moreno-Vivian, C. and Klipp, W. (1993) Nucleotide sequence and genetic analysis of the *Rhodobacter capsulatus* ORF6-nifUI SVW gene region: possible role of NifW in homocitrate processing. *Mol. Gen. Genet.* 238, 369–382.
- [51] Bryant, C. and DeLuca, M. (1991) Purification and characterization of an oxygen-insensitive NAD(P)H nitroreductase from *Enterobacter cloacae*. *J. Biol. Chem.* 266, 4119–4125.
- [52] Jarausch, W., Saillard, C., Heliott, B., Garnier, M. and Dosba, F. (2000) Genetic variability of apple proliferation phytoplasmas as determined by PCR-RFLP and sequencing of a non-ribosomal fragment. *Mol. Cell. Probes* 14, 17–24.
- [53] Kolenbrander, P.E., Andersen, R.N. and Ganeshkumar, N. (1994) Nucleotide sequence of the *Streptococcus gordonii* PK488 coaggregation adhesin gene, *scaA*, and ATP-binding cassette. *Infect. Immun.* 62, 4469–4480.
- [54] Novak, R., Braun, J.S., Charpentier, E. and Tuomanen, E. (1998) Penicillin tolerance genes of *Streptococcus pneumoniae*: the ABC-type manganese permease complex Psa. *Mol. Microbiol.* 29, 1285–1296.
- [55] Williamson, D.L., Renaudin, J. and Bové, J.M. (1991) Nucleotide sequence of the *Spiroplasma citri* fibril protein gene. *J. Bacteriol.* 173, 4353–4362.
- [56] Trachtenberg, S. (1998) Mollicutes – wall-less bacteria with internal cytoskeletons. *J. Struct. Biol.* 124, 244–256.
- [57] Trachtenberg, S. and Gilad, R. (2001) A bacterial linear motor: cellular and molecular organization of the contractile cytoskeleton of the helical bacterium *Spiroplasma melliferum* BC3. *Mol. Microbiol.* 41, 827–848.