Comparison of Metabolic Capacities and Inference of Gene Content Evolution in Mosquito-Associated Spiroplasma diminutum and S. taiwanense

Wen-Sui Lo1,2,3, Chuan Ku1, Ling-Ling Chen1, Tean-Hsu Chang1, and Chih-Horng Kuo1,2,4,*

1Institute of Plant and Microbial Biology, Academia Sinica, Taipei, Taiwan
2Molecular and Biological Agricultural Sciences Program, Taiwan International Graduate Program, National Chung Hsing University and Academia Sinica, Taipei, Taiwan
3Graduate Institute of Biotechnology, National Chung Hsing University, Taichung, Taiwan
4Biotechnology Center, National Chung Hsing University, Taichung, Taiwan
*Corresponding author: E-mail: chk@gate.sinica.edu.tw.

Accepted: July 16, 2013

Data deposition: The genome sequences reported in this study have been deposited at DDBJ/EMBL/GenBank under the accessions CP005074–CP50076.

Abstract

Mosquitoes are hosts of several Spiroplasma species that belong to different serogroups. To investigate the genetic mechanisms that may be involved in the utilization of similar hosts in these phylogenetically distinct bacteria, we determined the complete genome sequences of Spiroplasma diminutum and S. taiwanense for comparative analysis. The genome alignment indicates that their chromosomal organization is highly conserved, which is in sharp contrast to the elevated genome instabilities observed in other Spiroplasma lineages. Examination of the substrate utilization strategies revealed that S. diminutum can use a wide range of carbohydrates, suggesting that it is well suited to living in the gut (and possibly the circulatory system) of its mosquito hosts. In comparison, S. taiwanense has lost several carbohydrate utilization genes and acquired additional sets of oligopeptide transporter genes through tandem duplications, suggesting that proteins from digested blood meal or lysed host cells may be an important nutrient source. Moreover, one glycerol-3-phosphate oxidase gene (glpO) was found in S. taiwanense but not S. diminutum. This gene is linked to the production of reactive oxygen species and has been shown to be a major virulence factor in Mycoplasma mycoides. This finding may explain the pathogenicity of S. taiwanense observed in previous artificial infection experiments, while no apparent effect was found for S. diminutum. To infer the gene content evolution at deeper divergence levels, we incorporated other Mollicutes genomes for comparative analyses. The results suggest that the losses of biosynthetic pathways are a recurrent theme in these host-associated bacteria.

Key words: Mollicutes, Spiroplasma diminutum, Spiroplasma taiwanense, genome, mosquito, virulence factor.

Introduction

The complete genome sequence of an organism provides biologists with the opportunity to examine the presence or absence of certain genes that may explain its phenotype. For this reason, comparative analysis of genomes between related organisms with phenotypic differences is a powerful tool to investigate the underlying genetic mechanisms. In this work, we chose two mosquito-associated bacteria in the genus Spiroplasma as the study system and utilized a comparative genomics approach to infer their metabolic differentiations and gene content evolution.

Taxonomically, the genus Spiroplasma is described as a group of helical, motile, and wall-less bacteria in the class Mollicutes (Whitcomb 1981; Gasparich et al. 2004; Regassa and Gasparich 2006; Gasparich 2010). Similar to other members of this class, such as the vertebrate-pathogenic Mycoplasma and the plant-pathogenic Candidatus Phytoplasma, all characterized Spiroplasma species are found to be associated with eukaryotic hosts. Most commonly, spiroplasmas are associated with insects, such as various flies and mosquitoes in the order Diptera or various beetles in the order Coleoptera (Hackett et al. 1992;
Gasparich et al. 2004). Although most of these insect-associated spiroplasmas are not known to have any apparent effect on their hosts (Gasparich 2010), a small number of Spiroplasma lineages have been found to be either beneficial or pathogenic. For example, several uncultivated spiroplasmas can provide protection against parasitic nematodes (Jaenike et al. 2010), parasitoid wasps (Xie et al. 2010, 2011), or fungal pathogens (Łukasik et al. 2013) in their Drosophila or aphid hosts. Alternatively, notable examples of harmful spiroplasmas include the honeybee-pathogenic Spiroplasma melliferum (Clark et al. 1985) and S. apis (Mouches et al. 1983), the male-killing spiroplasmas in Drosophila and other insects (Williamson et al. 1999; Hurst and Jiggins 2000; Anbutsu and Fukatsu 2003; Tabata et al. 2011), and the mosquito-pathogenic S. culicicola and S. taiwanense (Humphery-Smith et al. 1991a, 1991b; Vazeille-Falcoz et al. 1994; Phillips and Humphery-Smith 1995). Because of their insect pathogenicity and relatively high host specificity, these spiroplasmas may be developed into biocontrol agents for insect pests (Anbutsu and Fukatsu 2011).

For biological control of insect pests, much attention has been given to mosquitoes because of the public health concerns (Federici et al. 2003). To date four Spiroplasma species have been isolated from mosquitoes, including S. culicicola from the salt marsh mosquito Aedes sollicitans collected in New Jersey, USA (Hung et al. 1987), S. sabaudiense from a mixed pool of A. sticticus and A. vexans collected in the French Northern Alps (Abalain-Colloc et al. 1987), and two species from mosquitoes collected in Taiwan: S. taiwanense from Culex tritaeniorhynchus (Abalain-Colloc et al. 1988) and S. diminutum from C. annulus and C. tritaeniorhynchus (Williamson et al. 1996). Interestingly, artificial infection experiments revealed that the Spiroplasma species exhibit different levels of pathogenicity toward their mosquito hosts. While S. diminutum can replicate inside A. albopictus, the infection does not reduce the host lifespan (Vorms-Le Morvan et al. 1991). In contrast, infection of the yellow fever mosquito A. aegypti by S. taiwanense significantly reduces the survival of larvae (Humphery-Smith et al. 1991a) and the lifespan of adult females (Humphery-Smith et al. 1991b; Vazeille-Falcoz et al. 1994). A histopathological study that used Anopheles stephensi as the host has shown that S. taiwanense can replicate both extra- and intra-cellularly in the host hemolymph, hemocytes, thoracic flight muscles, neural system, and other tissues (Phillips and Humphery-Smith 1995). Moreover, the infected mosquitoes exhibit loss of flight ability and reduced mobility, which are linked to extensive cell lysis and polysaccharide depletions in the thoracic flight muscles. Finally, cytadsorption of S. taiwanense was associated with the swelling and subsequent lysis of A. albopictus C6/36 cells in vitro (Chastel and Humphery-Smith 1991).

To investigate the genetic mechanisms that may explain the differences in pathogenicity toward their mosquito hosts in previous artificial infection experiments, we determined the complete genome sequences of S. diminutum and S. taiwanense in this study for comparative analysis. In addition to providing candidate genes for future characterization of virulence factors, comparisons with other available genome sequences, such as the honeybee-pathogenic S. melliferum (Alexeev et al. 2012; Lo et al. 2013) and the vertebrate-pathogenic Mycoplasma species (Sasaki et al. 2002; Thiaucourt et al. 2011), can further improve our understanding of genome evolution in these host-associated bacteria.

Materials and Methods

Molecular Phylogenetic Inference

To infer the evolutionary relationship among the Spiroplasma lineages of interest, we used 16S rDNA and DNA-directed RNA polymerase subunit beta (rpoB) to construct a molecular phylogeny. The sequences were obtained from the NCBI nucleotide database (Benson et al. 2012) and the corresponding accession numbers are provided in supplementary table S1, Supplementary Material online. These two genes were aligned separately using MUSCLE v3.8 (Edgar 2004) with the default settings and concatenated into a single dataset with 6,585 aligned nucleotide sites. A maximum likelihood phylogeny was inferred using PhyML v3.0 (Guindon and Gascuel 2003) with the GTR + I + G model and six substitution rate categories. To estimate the levels of clade support, we generated 1,000 nonparametric bootstrap samples using the SEQBOOT program of PHYLIp v3.69 (Felsenstein 1989). For the species in the Apis clade, including the four mosquito-associated Spiroplasma species, we collected the information of host association from the literature and provided a summary in supplementary table S2, Supplementary Material online.

Strain Source and DNA Preparation

The two focal bacterial strains, S. diminutum CUAS-1 (ATCC 49235) and S. taiwanense CT-1 (ATCC 43302), were obtained from the American Type Culture Collection (ATCC). The freeze-dried culture samples were processed according to the protocol provided by ATCC. Briefly, the samples were rehydrated by adding 5 ml ATCC 988 medium, titrated by serial dilution, and incubated in 30 °C without shaking until the medium turned yellow. The minimum concentration that showed spiroplasma growth was then transferred into R2 medium (Moulder et al. 2002) for DNA extraction using the Wizard Genomic DNA Purification Kit (Promega, USA). For each DNA sample, we amplified the 16S rDNA using the primer pair 8F (5′-agagggtgagtcctgctcag-3′) (Turner et al. 1999) and 1492R (5′-ggtacctgtgtagcacttt-3′) (Ohman et al. 2010) for Sanger sequencing to confirm the sample identity and that no contamination has occurred.
Genome Sequencing and Assembly

To determine the genome sequences of *S. diminutum* and *S. taiwanense*, we used a commercial service provider (Yourgene Bioscience, Taipei, Taiwan) for whole-genome shotgun sequencing with the 101-bp reads produced on the Illumina HiSeq 2000 platform (Illumina, USA). The procedure for de novo assembly was based on that described previously (Chung et al. 2013; Lo et al. 2013). Briefly, raw reads were quality-trimmed and filtered based on usable length. The resulting high quality reads were used as the input for the assembler of choice to produce draft assemblies (more details below). Subsequently, the draft assemblies were improved using an iterative procedure until the chromosomes and plasmids were sequenced to completion. For each iteration, we mapped all raw reads to the existing scaffolds using BWA v0.6.2 (Li and Durbin 2009) and visualized the results with IGV v2.1.24 (Robinson et al. 2011). Paired reads that extended the existing contigs or supported the linkage between contigs were used to improve the assembly. The MPILEUP program in the SAMTOOLS v0.1.18 package (Li et al. 2009) was used to identify polymorphic sites. Primer walking and additional Sanger sequencing were used to fill the gaps and to verify the assembly.

For *S. taiwanense*, we utilized one paired-end library (insert size = 192 bp, 47,312,605 read-pairs, approximately 9.6 Gb of raw data). The initial de novo assembly was performed using VELVET v1.2.07 (Zerbino and Birney 2008) with the parameters k-mer, expected coverage, and coverage cutoff set to 89, 1200, and 100, respectively. For *S. diminutum*, we utilized one paired-end library (insert size = 178 bp, 44,436,475 pairs, approximately 9.0 Gb of raw data) and one mate-pair library (insert size = ~4.1 kb, 18,273,021 pairs, approximately 3.7 Gb of raw data). The initial de novo assembly was performed using ALLPATH-LG release 42781 (Gnerre et al. 2011) to take advantage of the availability of the mate-pair library. A subset of raw reads was randomly selected from each library to represent ~50× coverage for the initial draft assembly as suggested by the assembler documentation.

Annotation and Comparative Analysis

The procedures for genome annotation and comparative analysis were based on those described previously (Ku et al. 2013; Lo et al. 2013). The complete genome sequences were processed using RNAmmer (Lagesen et al. 2007), tRNAscan-SE (Lowe and Eddy 1997), and PRODIGAL (Hyatt et al. 2010) for gene predictions. The protein-coding genes were annotated based on the single-copy orthologous genes in the *S. melliferum* IPMB4A genome (Lo et al. 2013) identified by OrthoMCL (Li et al. 2003) with a BLASTP (Altschul et al. 1997; Camacho et al. 2009) e-value cutoff of 1 × 10^{-15}. The 259 homologous gene clusters that contain one single orthologous gene from each of the species compared were used to infer a species phylogeny. The concatenated alignment contains 104,376 aligned amino acid sites and was used for PhyML analysis with the LG substitution model (Le and Gascuel 2008). The clade supports were inferred by using 1,000 bootstrap samples. After obtaining the species phylogeny, the phylogenetic distribution pattern of homologous gene clusters was inferred based on the presence/absence of genes in each of the species compared.

To compare the chromosomal organization between different *Spiroplasma* species, we utilized MAUVE v2.3.1 (Darling et al. 2010) for genome alignment. To estimate the genome-wide nucleotide sequence divergence level, we identified the single-copy orthologs in each genome pair using OrthoMCL (Li et al. 2003) with a BLASTN (Altschul et al. 1997; Camacho et al. 2009) e-value cutoff of 1 × 10^{-15}. The corresponding sequences were aligned using MUSCLE v3.8 (Edgar 2004) with the default settings and concatenated into a single alignment for each pair. The DNADIST program of PHYLIP v3.69 (Felsenstein 1989) was used to calculate the sequence identity.

For the gene content comparison with honeybee-associated *S. melliferum*, we merged the two draft genomes available for this species (Alexeev et al. 2012; Lo et al. 2013) into a pan-genome to better represent its gene repertoire. For the comparison with other Mollicute lineages, we selected *Mycoplasma mycoides* subsp. capri LC str. 95010 (GenBank accession number NC_015431) (Thiaucourt et al. 2011) and *Mesorolium florae* L1 (NC_006055) to represent the Mycoides-Entomoplasmataceae clade, which is the sister group to the Apis clade that contain *S. diminutum* and *S. taiwanense* (Gasparich et al. 2004). Additionally, *M. pene-trans* HF-2 (NC_004432) (Sasaki et al. 2002) was used as the outgroup for this comparison because it has the highest number of protein-coding genes among the *Mycoplasma* species with complete genome sequences available. For these gene content comparisons, the homologous gene clusters were identified using OrthoMCL (Li et al. 2003) with a BLASTP (Altschul et al. 1997; Camacho et al. 2009) e-value cutoff of 1 × 10^{-15}. The 259 homologous gene clusters that contain one single orthologous gene from each of the species compared were used to infer a species phylogeny. The concatenated alignment contains 104,376 aligned amino acid sites and was used for PhyML analysis with the LG substitution model (Le and Gascuel 2008). The clade supports were inferred by using 1,000 bootstrap samples. After obtaining the species phylogeny, the phylogenetic distribution pattern of homologous gene clusters was inferred based on the presence/absence of genes in each of the species compared.
Results and Discussion

Molecular Phylogeny of Mosquito-Associated Spiroplasma Species

The maximum likelihood phylogeny inferred using the concatenated alignment of 16S rDNA and rpoB (fig. 1) is mostly congruent with a previous study that used only 16S rDNA and the maximum parsimony method (Gasparich et al. 2004). The major inconsistencies are the placements of S. corruscae, S. turonicum, S. litorale, and S. taiwanense. These species were thought to be sisters of the S. apis–S. montanense clade (Gasparich et al. 2004) but our results provided alternative placements with low levels of bootstrap support. Because molecular phylogenies inferred using a limited number of loci are often problematic, future improvements on the availability of molecular markers are required to resolve these uncertainties.

Despite these uncertainties within the Apis clade, it is clear that the four mosquito-associated Spiroplasma species are quite divergent. This observation is consistent with the results from serotyping, which placed S. culicicola, S. diminutum, S. sabelaunense, and S. taiwanense in groups X, XXV, XIII, and XXII, respectively (Gasparich et al. 2004). Taken together, these results suggest that the association with mosquito hosts may have evolved independently among these Spiroplasma species. The comparison between S. diminutum and S. taiwanense is of particular interest because these two species were both isolated from mosquitoes collected in Taiwan during 1980–1981 and appeared to overlap in their native host range. The three characterized strains of S. taiwanense (CT-1T, CT-2, and CT-3) were all isolated from C. annulus and C. tritaeniorhynchus (Abalain-Colloc et al. 1988). The two characterized strains of S. diminutum, CUAS-1T and CT-4, were isolated from C. annulus and C. tritaeniorhynchus, respectively (Williamson et al. 1996).

Genome Sequences of S. diminutum and S. taiwanense

The genomes of S. diminutum and S. taiwanense were sequenced to completion in this study (table 1 and fig. 2). Both genomes contain a circular chromosome that is ~1.0 Mb in size (S. diminutum: 945,296 bp; S. taiwanense: 1,075,140 bp). The S. taiwanense genome contains a circular plasmid that is 11,138 bp in size and encodes 11 protein-coding genes (1 SO-like protein and 10 hypothetical proteins); no plasmid was found in the S. diminutum genome. The chromosomal GC contents are consistent with previous estimates obtained using biochemical methods, with S. diminutum having a GC content of 25.5% (Williamson et al. 1996) and S. taiwanense having a GC content of 23.9% (Abalain-Colloc et al. 1988). Both genomes contain a single ribosomal RNA gene cluster, which corresponds to the highest peak observed in the GC content plot (fig. 2; ~711–716 kb in S. diminutum and ~859–864 kb in S. taiwanense). Both genomes encode 29 tRNA genes, which are fewer than those found in S. citri and S. melliferum (table 1).

The genome alignment between S. diminutum and S. taiwanense indicates that their chromosomes are largely syntenic except for a ~122 kb inversion that encompasses the putative replication terminus (fig. 3A). This conservation in chromosomal organization was surprising because these two species are relatively divergent, with an average genome-wide nucleotide sequence identity of 76.1% (calculated based on 652 single-copy orthologous genes shared between these two genomes, the concatenated alignment contains a total of 668,307 aligned nucleotide sites). For comparison, the closely related S. citri and S. melliferum in the Citri clade have an average genome-wide nucleotide sequence identity of 99.0% (based on 696 genes and 691,679 sites), yet exhibit extensive rearrangements (fig. 3B). This genome instability in the Citri clade may be explained by the presence of highly repetitive plectroviral fragments (table 1), which may have promoted their genome instability (Ye et al. 1996; Ku et al. 2013; Lo et al. 2013).

Despite the similarities described above, close inspections of the S. diminutum–S. taiwanense comparison reveal several intriguing differences. First, most of the genome-specific regions in these two species are located near the putative replication terminus (fig. 2), suggesting that these regions are hotspots for molecular evolution by accelerated sequence divergence or horizontal gene transfers. Intriguingly, this clustering of species-specific genes was not found in a comparison between S. chrysopicola and S. syrphidicola (Ku et al. 2013). It is unclear whether this difference was due to the fact that these two species pairs are sampled from different Spiroplasma clades or because the divergence levels are quite different (i.e., the average genome-wide nucleotide identity between S. chrysopicola and S. syrphidicola is ~92.2%, which is much higher than the S. diminutum–S. taiwanense comparison). Second, while no pseudogene was found in the S. diminutum genome, we identified 54 putative pseudogenes with premature stop codons and/or frameshift indels in S. taiwanense (table 1). These pseudogenes include those involved in carbohydrate uptake (treB, fruA, celB, nagB, and sgaB), carbohydrate metabolism (gplX, scrB, bgl, and lacG), and homologous recombination (ruvA and ruvB). Additionally, S. taiwanense contains many more long intergenic regions (~300 bp) than S. diminutum (87 vs. 32), which may harbor highly degraded pseudogenes that cannot be easily identified by sequence similarity searches. This increase in pseudogene numbers is similar to those found in the genomes of recent or facultative pathogens (Ochman and Davalos 2006). Furthermore, the observed genome degradations suggest that S. taiwanense may have a smaller effective population size than S. diminutum, which is consistent with the field isolation records that S. taiwanense has a narrower natural host range (Abalain-Colloc et al. 1988; Williamson et al. 1996). Consequently, the smaller effective
FIG. 1.—Molecular phylogeny of spiroplasmas. The maximum likelihood tree was inferred based on the concatenated alignment of the 16S ribosomal RNA gene and RNA polymerase subunit beta (rpoB). The numbers on the internal branches indicate the percentage of bootstrap support based on 1,000 replicates (only values > 70% are shown). The sequences from *Bacillus subtilis* are included as the outgroup. The two species with genome sequences reported in this study (i.e., *Spiroplasma diminutum* and *S. taiwanense*) are highlighted in bold. The hosts of the *Spiroplasma* species in the *Apis* clade are labeled, with the host genus name inside the parentheses.
Complete Genome Sequences of Spiroplasma diminutum and S. taiwanense

| Strain               | Spiroplasma diminutum | Spiroplasma taiwanense | Spiroplasma melliferum |
|----------------------|-----------------------|------------------------|------------------------|
| Genome Accession     | CP005074              | CP005076               | AMG01000001-A MG01000002 |
| Number of chromosomal contigs | 1                     | 1                      | 24                     |
| Estimation of the size of the chromosome (bp) | 954,296               | 1,075,140              | 1,098,846              |
| Estimated覆盖率 (%) | —                     | —                      | 79.6                   |
| Estimated GC content (%) | 25.5                  | 23.9                   | 27.5                   |
| Coding density (%)   | 92.7                  | 82.5                   | 85.1                   |
| Protein-coding genes | 858                   | 991                    | 932                    |
| Length distribution (Q1/Q2/Q3) (a.a.) | 177/283/443           | 137/247/397            | 176/280/440            |
| Plectrovirus proteins | 0                     | 1                      | 11                     |
| Hypothetical proteins | 210                   | 467                    | 337                    |
| Annotated pseudogenes | 05                    | 4                      | 1                      |
| rRNA operon          | 1                     | 1                      | 1                      |
| tRNA genes           | 29                    | 46                     | 32                     |
| Number of plasmids   | 0                     | 1                      | 4                      |

*For S. diminutum, S. taiwanense, and S. melliferum, putative pseudogenes were annotated with the “pseudo” tag in gene features as suggested by the NCBI GenBank guidelines and were not counted in the total number of protein-coding genes. For S. melliferum KC3 and S. citri GII3-3X, putative pseudogenes were annotated by adding the term “truncated” in the CDS product description field and were included in the total number of protein-coding genes.

Comparison of Substrate Utilization Strategies

To investigate the genetic mechanisms that may be involved in utilizing mosquito hosts and the possible explanations for differences in the pathogenicity inferred from previous artificial infection experiments (Chastel and Humphery-Smith 1991; Humphery-Smith et al. 1991a, 1991b; Vorms-Le Morvan et al. 1991; Vazeille-Falcoz et al. 1994; Phillips and Humphery-Smith 1995), we compared the substrate utilization strategies of S. diminutum and S. taiwanense based on their annotated transporters and metabolic enzymes (fig. 4). The results indicate that both species are capable of importing and utilizing glucose, fructose, and N-acetylgalactosamine (GlcNAc). However, the genes involved in the utilization of trehalose (treA and treB), cellobiose (celB), sucrose (scrB and scrK), and N-acetylmuramic acid (MurNAc; murP and murQ) are found in S. diminutum but not S. taiwanense. Among these substrates, cellobiose and MurNAc may be derived from algae and bacteria that are consumed by mosquito larvae, sucrose is the major carbohydrate in nectar and plant sap consumed by adult mosquitoes, and trehalose is the most abundant sugar in insect hemolymph (Becker et al. 1996; Blatt and Roces 2001). The flexible sugar usage capacity suggests that S. diminutum is well suited to the environment in mosquito gut and may be capable of living in the host circulatory system as well.

Most of these S. diminutum-specific genes appear to have been lost in the S. taiwanense genome through pseudogenization (see above). The loss of trehalose utilization genes (treA and treB) suggests that S. taiwanense may face limited carbohydrate supplies in host hemolymph, which is consistent with the observation that S. taiwanense cells often display postexponential morphologies in the hemolymph of infected Ano. stephensi (Phillips and Humphery-Smith 1995). Intriguingly, we found that the S. taiwanense genome encodes a copy of glycerol-3-phosphate oxidase (gpo), which can be used to produce hydrogen peroxide (H2O2) and reactive oxygen species (ROS). This gene has been shown to be a major virulence factor that causes host tissue inflammation and cell population size has resulted in elevated levels of genetic drift and increased accumulation of slightly deleterious mutations (Kuo et al. 2009; Kuo and Ochman 2009, 2010). Interestingly, a similar pattern of genome degradation is also observed in the pathogenic S. citri and S. melliferum (table 1), both of which have lost the recombinase A gene (recA) that is required for DNA repair by homologous recombination (Mañas et al. 1996; Carle et al. 2010; Alexeev et al. 2012; Lo et al. 2013). In contrast, these DNA repair-related genes (e.g., recA, ruvA, ruvB, etc.) are still intact in the S. diminutum genome, which may explain why this genome has the lowest incidence of pseudogenes and the highest coding density among the Spiroplasma genomes reported to date (Carle et al. 2010; Alexeev et al. 2012; Lo et al. 2013; Ku et al. 2013).
death in M. mycoides (Pilo et al. 2005, 2007) and may contribute to the tissue damage (Phillips and Humphery-Smith 1995) and higher mortality rates (Humphery-Smith et al. 1991a, 1991b; Vazeille-Falcoz et al. 1994) observed in S. taiwanense-infected mosquitoes. It will be interesting to examine the timing and tissue-specificity of glpO activation and to investigate the link to stress responses in future empirical studies.

In contrast to the deficiencies in carbohydrate utilization, S. taiwanense may be more efficient in oligopeptide uptake compared with S. diminutum. The gene cluster that encodes for oligopeptide ABC transporters appears to have experienced tandem duplications and exists in three copies on the S. taiwanense chromosome (~820–847 kb). In addition to the lysed host cells, the digested blood meal in the gut of female mosquitoes can provide abundant substrates for these transporters. Taken together, although S. diminutum and S. taiwanense are both associated with Culex mosquitoes in Southeast Asia (Abalain-Colloc et al. 1988; Williamson et al. 1996), their substrate utilization strategies for utilizing these closely related hosts appear to be quite different.

Gene Content Comparison with the Honeybee-Associated S. melliferum

Two previously published genome sequences of the honeybee-associated S. melliferum (Alexeev et al. 2012; Lo et al. 2013) provide an opportunity for comparative analysis of gene content between two major clades of Spiroplasma (fig. 1). A three-way comparison among S. melliferum–S. diminutum–S. taiwanense revealed that these species
shared a total of 472 homologous gene clusters (fig. 5 and supplementary table S3, Supplementary Material online). In addition to the essential genes conserved across all bacterial genomes such as those involved in DNA replication, transcription, translation, and other fundamental cell processes (Koonin 2003; Lapierre and Gogarten 2009; Chen et al. 2012), we found that these spiroplasmas all have the glycolysis pathway to convert phosphorylated sugars into pyruvate for energy generation, the nonmevalonate pathway (dxs, dxr, ispD, ispF, ispG, and ispH) to synthesize isopentenyl pyrophosphate (IPP) for terpenoid backbone, and oligopeptide ABC transporters (oppA, oppB, oppC, oppD, and oppF) to import amino acids for peptide synthesis. Furthermore, these genomes contain the genes required for nucleotide biosynthesis from nucleobases (adenine, guanine, uracil, and xanthine) and a nucleoside (thymidine). The presence of these genes is in agreement with the previous findings that spiroplasmas have more flexible metabolic capabilities compared with mycoplasmas and phytoplasmas (Carle et al. 2010; Chen et al. 2012; Lo et al. 2013), which may contribute to their lower degree of host dependence.

Other than the metabolic genes and transporters described above, these insect-associated spiroplasmas shared several genes related to oxidative stress resistance such as those involved in iron–sulfur (Fe–S) cluster synthesis (sufS, sufU, sufB, sufC, and sufD). The organization of this suf operon is conserved within Spiroplasma and other Gram-positive bacteria, while distinct from those found in Gram-negative bacteria (Riboldi et al. 2009). Additionally, these spiroplasmas all have the thiol peroxidase (tpx), which has been shown to be important in protecting Enterococcus faecalis cells inside mouse macrophages (La Carbona et al. 2007).
together, these genes may protect these insect-associated bacteria against the reactive oxygen intermediates generated by the host immune system (Cerenius et al. 2008).

In terms of species-specific gene clusters, *S. melliferum* has the highest number compared with *S. diminutum* and *S. taiwanense* (435, 134, and 281, respectively). While most of these species-specific genes are annotated as hypothetical proteins with unknown functions, some have more detailed annotation for inferring the functional significance. For example, *S. melliferum* has the entire gene set for arginine catabolism (*arcA*, *arcB*, and *arcC*), which is consistent with the biochemical assay results that this species can hydrolyze arginine (Clark et al. 1985) whereas *S. diminutum* and *S. taiwanense* cannot (Abalain-Colloc et al. 1988; Williamson et al. 1996). This ability for arginine hydrolysis can contribute to energy generation and provide organic nitrogen, which allows for more flexible metabolisms and may promote cell growth when other energy sources are limited (Pereyre et al. 2009). Moreover, *S. melliferum* has the gene set for uridine monophosphate (UMP) synthesis (*pyrB*, *pyrC*, *pyrD*, *pyrE*, and *pyrF*), which may reduce its dependence on the host for nucleotides. Finally, a large number of *S. melliferum*-specific genes are originated from plectroviral invasion of this genome and the associated horizontal gene transfer (Alexeev et al. 2012; Lo et al. 2013).

One important finding from this among-species comparison is related to the variable patterns of carbohydrate uptake and utilization. Extending the results from the *S. diminutum*–*S. taiwanense* comparison as discussed above, we found that the phosphotransferase system (PTS) transporters for importing glucose and fructose appear to be conserved among the spiroplasmas characterized to date. Although the PTS transporter for importing GlcNAc (*nagE*) is shared by these three species, it was not found in the draft genome assembly of the phytopathogenic *S. citri* (Carle et al. 2010; Lo et al. 2013). It is

---

**Fig. 4.**—Sugar uptake and utilization. Comparison of the phosphotransferase system (PTS) transporters and enzymes involved in sugar uptake and utilization between *S. diminutum* and *S. taiwanense*. Gene names are color-coded according to their patterns of presence/absence (gray: shared; blue: *S. diminutum*-specific; red: *S. taiwanense*-specific). DHAP, dihydroxyacetone phosphate; G3P, glycerol 3-phosphate; GlcNAc, *N*-acetylglucosamine; MurNAc, *N*-acetylmuramic acid; ROS, reactive oxygen species.

**Fig. 5.**—Comparative analysis of gene content among *Spiroplasma* species. The numbers of shared and species-specific homologous gene clusters from a three-species comparison are shown in the Venn diagram. Gene names in the metabolic map are color-coded based on their patterns of presence/absence among the three species compared. DMAPP, dimethylallyl pyrophosphate; GlcNAc, *N*-acetylglucosamine; IPP, isopentenyl pyrophosphate; MurNAc, *N*-acetylmuramic acid; PRPP, phosphoribosyl pyrophosphate; PTS, phosphotransferase system.
not clear whether the absence of this gene in S. citri was due to true loss or the incompleteness of its draft genome assembly. The pattern for sucrose uptake was unclear for the same reason as well because while the corresponding gene (scyA) was not found in either S. melliferum or S. citri, this gene may reside in the unassembled parts of these two genomes. Nonetheless, the availability of the complete genome sequence of S. taiwanense suggests that the ability to utilize trehalose, cellobiose, and MurNAC is dispensable.

Comparison with the Mycoides-Entomoplasmataceae Clade and Inference of Gene Content Evolution

The genus Spiroplasma is known to be a paraphyletic group with the Mycoides-Entomoplasmataceae clade (containing M. mycoides and other nonhelical species assigned to the genera Mesoplasma and Entomoplasm) as its descendants (Gasparich et al. 2004). Because the Apis clade (containing the S. diminutum and S. taiwanense reported in this study) is the sister group to the Mycoides-Entomoplasmataceae clade (fig. 1), the availability of these two new genome sequences provides an opportunity to infer the gene content evolution among these bacteria.

To investigate this question, we identified 259 single-copy genes shared among selected Mollicutes genomes for phylogenetic inference. The organismal phylogeny inferred from the concatenated alignment based on the maximum likelihood method received 100% bootstrap support on all internal nodes (fig. 1) and is consistent with our current understanding of Mollicutes evolution (Gasparich et al. 2004). Using this phylogeny as the framework, we inferred putative events of gene gains and losses based on the pattern of gene presence and absence in each of the genome compared (fig. 6) and is consistent with our current understanding of Mollicutes evolution (Gasparich et al. 2004).

Although it is reasonable to hypothesize that some of the putative gene gains may have contributed to important functions, such inference was difficult because most of the lineage-specific genes are annotated as hypothetical proteins without functional description. Rather, the main finding from this analysis is that losses of biosynthetic pathways appear to be a recurrent theme among these host-associated bacteria (Ochman and Davalos 2006; McCutcheon and Moran 2011). For example, the genes involved in arginine catabolism and UMP synthesis as described above appear to have been lost in the common ancestor of the Apis and Mycoides-Entomoplasmataceae clades. Moreover, the genes involved in the synthesis of IPP and Fe–S cluster appear to have been lost in the common ancestor of the Mycoides-Entomoplasmataceae clade.

Finally, we found that all Spiroplasma genomes characterized to date have at least five copies of mreB (Ku et al. 2013), which encodes the cell shape determining protein MreB and has been linked to the helical morphology of these bacteria (Kurner et al. 2005). However, this gene is present as a single copy gene in the Mes. florum genome and was not found in either of the Mycoplasma genomes. Because this gene was found in several Firmicutes genomes but not most of the Mollicutes genomes (Chen et al. 2012), it is possible that this gene was acquired by the common ancestor of spiroplasmas (possibly through horizontal gene transfer). Subsequently, gene family expansion by duplication occurred and allowed for subfunctionalization (and possibly neofunctionalization) of different copies, which contributed to the distinct helical shape of spiroplasma cells. The losses of these genes in the common ancestor of the Mycoides-Entomoplasmataceae clade are likely to be responsible for the reversion back to nonhelical shape of these descendants of spiroplasmas.

Conclusions

In summary, this study provides the first set of complete genome sequences for two Spiroplasma species in the Apis clade, which is the most diverse group within this genus. The conservation in chromosome organization suggests that these sequences may be used as the references for future genomic studies in related species. Through comparative analysis at different phylogenetic depths, we identified several genetic

![Fig. 6.—Phylogenetic distribution pattern of homologous gene clusters. The organismal phylogeny is inferred from the concatenated protein alignment of 259 single-copy genes shared by all species. All internal nodes received 100% bootstrap support based on 1,000 replicates and maximum likelihood inference. The numbers in parentheses below species names indicate the number of homologous gene clusters found in each species. The numbers above a branch and preceded by a “+” sign indicate the number of homologous gene clusters that are uniquely present in all daughter lineages; the numbers below a branch and preceded by a “–” sign indicate the number of homologous gene clusters that are uniquely absent. For example, 127 gene clusters are shared by S. diminutum and S. taiwanense and do not contain a homolog from all four other species compared; similarly, three gene clusters are missing in these two Spiroplasma species but are present in all four other species.](image)
mechanisms that may explain the results of previous phenotypic characterizations (metabolism, pathogenicity, etc.). For future work, genomic characterizations and functional studies that include other mosquito-associated spiroplasmas can further improve our understanding of the diverse genetic mechanisms of utilizing similar hosts among these phylogenetically distinct bacteria. Additionally, more comprehensive evaluations of the pathogenicity of each Spiroplasma species in different mosquitoes, particularly the native hosts, are required to investigate bacterium–host interactions. At a deeper divergence level, genomic characterization of the basal Ixodidae clade is required to shed light on the genome evolution in the genus Spiroplasma and its nonhelical descendants.

**Supplementary Material**

Supplementary tables S1–S4 are available at Genome Biology and Evolution online (http://www.gbe.oxfordjournals.org/).

**Acknowledgments**

The authors thank the DNA Analysis Core Laboratory (Institute of Plant and Microbial Biology, Academia Sinica) for providing Sanger sequencing service. This work was supported by the research grants from the Institute of Plant and Microbial Biology at Academia Sinica and the National Science Council of Taiwan [NSC 101-2621-B-001-004-MY3] to C.H.K.

**Literature Cited**

Abalain-Colloc ML, et al. 1987. Spiroplasma sabaudiense sp. nov. from mosquitoes collected in France. Int J Syst Microbiol. 37:260–265.

Abalain-Colloc ML, et al. 1988. Spiroplasma taiwanense sp. nov. from Culex tritaeniorhynchus mosquitoes collected in Taiwan. Int J Syst Microbiol. 38:103–107.

Alexeev D, et al. 2012. Application of Spiroplasma melliferum proteogenomic profiling for the discovery of virulence factors and pathogenicity mechanisms in host-associated spiroplasmas. J Proteome Res. 11:224–236.

Altschul SF, et al. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25:3389–3402.

Anbutsu H, Fukatsu T. 2003. Population dynamics of male-killing and nonmale-killing spiroplasmas in Drosophila melanogaster. Appl Environ Microbiol. 69:1428–1434.

Anbutsu H, Fukatsu T. 2011. Spiroplasma as a model insect endosymbiont. Environ Microbiol Rep. 3:144–153.

Becker A, Schlöder P, Steele JE, Wegener G. 1996. The regulation of trehalose metabolism in insects. Experientia 52:433–439.

Benson DA, et al. 2012. GenBank. Nucleic Acids Res. 40:D48–D53.

Blatt J, Roecs F. 2001. Haemolymph sugar levels in foraging honeybees (Apis mellifera carnica): dependence on metabolic rate and in vivo measurement of maximal rates of trehalose synthesis. J Exp Biol. 204:2709–2716.

Camacho C, et al. 2009. BLAST+: architecture and applications. BMC Bioinformatics. 10:421.

Carle P, et al. 2010. Partial chromosome sequence of Spiroplasma citri reveals extensive viral invasion and important gene decay. Appl Environ Microbiol. 76:3420–3426.

Cerénius L, Lee BL, Söderhäll K. 2008. The proPO-system: pros and cons for its role in invertebrate immunity. Trends Immunol. 29:263–271.

Chastel C, Hymephy-Smith I. 1991. Mosquito spiroplasmas. Adv Dis Vector Res. 7:149–206.

Chen L-L, Chung W-C, Lin C-P, Kuo C-H. 2012. Comparative analysis of gene content evolution in phytoplastas and mycoplasmas. PLoS One 7:e34407.

Chung W-C, Chen L-L, Lo W-S, Lin C-P, Kuo C-H. 2013. Comparative analysis of the peanut witches’-broom phytoplasma genome reveals horizontal transfer of potential mobile units and effectors. PLoS One 8:e62770.

Clark TB, et al. 1985. Spiroplasma melliferum, a new species from the honeybee (Apis mellifera). Int J Syst Bacteriol. 35:296–308.

Darling AE, Mau B, Perna NT. 2010. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. PLoS One 5:e11147.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32:1792–1797.

Federici BA, Park H-W, Bideshi DK, Wirth MC, Johnson JJ. 2003. Recombinant bacteria for mosquito control. J Exp Biol. 206:3877–3885.

Felsenstein J. 1989. PHYLP2: Phylogeny Inference Package (Version 3.2). Cladistics 5:164–166.

Gasparich GE. 2010. Spiroplasmas and phytoplastas: microbes associated with plant hosts. Biologicals 38:193–203.

Gasparich GE, et al. 2004. The genus Spiroplasma and its non-helical descendants: phylogenetic classification, correlation with phenotype and roots of the Mycoplasma clade. Int J Syst Evol Microbiol. 54:893–918.

Gnerre S, et al. 2011. High-quality draft assemblies of mammalian genomes from massively parallel sequence data. Proc Natl Acad Sci U S A 108:1513–1518.

Guindon S, Gascuel O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol. 52:696–704.

Hackett KJ, et al. 1992. Lampryridae (Coleoptera): A plethora of mollicute associations. Microb Ecol. 23:181–193.

Humphrey-Smith I, Gruet O, Chastel C. 1991a. Pathogenicity of Spiroplasma taiwanense for larval Aedes aegypti mosquitoes. Med Vet Entomol. 5:229–232.

Humphrey-Smith I, Gruet O, Le Goff F, Chastel C. 1991b. Spiroplasma (Mollicutes: Spiroplasmataceae) pathogenic for Aedes aegypti and Anopheles stephensi (Diptera: Culicidae). J Med Entomol. 28:219–222.

Hung SHY, Chen TA, Whitcomb RF, Tully JG, Chen YX. 1987. Spiroplasma culicicola sp. nov. from the salt marsh mosquito Aedes sollicitans. Int J Syst Bacteriol. 37:365–370.

Hurst GD, Jiggins FM. 2000. Male-killing bacteria in insects: mechanisms, incidence, and implications. Emerg Infect Dis. 6:329–336.

Hyatt D, et al. 2010. Prodigal: Prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics 11:119.

Jaenike J, Uncleess R, Cockburn SN, Boelio LM, Perlman SJ. 2010. Adaptation via symbiosis: recent spread of a Drosophila defensive symbiont. Science 329:212–215.

Kaneki H, Muto S. 2000. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 28:27–30.

Kaneki H, Muto S, Furumichi M, Tanabe M, Hirakawa M. 2010. KEGG for representation and analysis of molecular networks involving diseases and drugs. Nucleic Acids Res. 38:D355–D360.

Koonin EV. 2003. Comparative genomics, minimal gene-sets and the last universal common ancestor. Nat Rev Microbiol. 1:127–136.

Krzywinski M, et al. 2009. Circos: an information aesthetic for comparative genomics. Genome Res. 19:1639–1645.
Complete Genome Sequences of *Spiroplasma diminutum* and *S. taiwanense*

Phillips RN, Humphrey-Smith I. 1995. The histopathology of experimentally induced infections of *Spiroplasma taiwanense* (class: Mollicutes) in *Anopheles stephensi* mosquitoes. J Invertebr Pathol. 66:185–195.

Pilo P, Frey J, Vilei EM. 2007. Molecular mechanisms of pathogenicity of *Mycoplasma mycoides* subspp. *mycoides* SC. Vet J. 174:513–521.

Pilo P, et al. 2005. A metabolic enzyme as a primary virulence factor of *Mycoplasma mycoides* subspp. *mycoides* Small Colony. J Bacteriol. 187: 6824–6831.

Regassa LB, Gasparich GE. 2006. Spiroplasmas: evolutionary relationships and biodiversity. Front Biosci. 11:2983–3002.

Ribolli GP, Verli H, Frazzon J. 2009. Structural studies of the *Enterococcus faecalis* SufU [Fe-S] cluster protein. BMC Biochem. 10.3.

Robinson JT, et al. 2011. Integrative genomics viewer. Nat Biotechnol. 29: 24–26.

Sasaki Y, et al. 2002. The complete genomic sequence of *Mycoplasma penetrans*, an intracellular bacterial pathogen in humans. Nucleic Acid Res. 30:5293–5300.

Tabata J, et al. 2011. Male killing and incomplete inheritance of a novel *Spiroplasma* in the moth *Ostrinia zagulaeae*. Microb Ec. 61:254–263.

Tatusov RL, Koonin EV, Lipman DJ. 1997. A genomic perspective on protein families. Science 278:631–637.

Tatusov R, et al. 2003. The COG database: an updated version includes eukaryotes. BMC Bioinformatics 4:41.

Thiaucourt F, et al. 2011. *Mycoplasma mycoides*, from 'mycoides Small Colony' to 'capri'. A microevolutionary perspective. BMC Genomics 12:114.

Turner S, Pryer KM, Miao VPW, Palmer JD. 1999. Investigating deep phylogenetic relationships among cyanoarchaea and plastids by small subunit rRNA sequence analysis. J Eukaryot Microbiol. 46:327–338.

Vazelle-Falcoz M, Percher-Merien A-M, Rodhain F. 1994. Experimental infection of *Aedes aegypti* mosquitoes, sucking mice, and rats with four mosquito spiroplasmas. J Invertebr Pathol. 63:37–42.

Vorms-Le Morvan J, Vazelle-Falcoz M-C, Rodhain F, Chastel C. 1991. Infection expérimentale de moustiques *Aedes albopictus* par une souche de spiroplasmes isolée de *Culex annulatus* a Taiwan. Bull Soc Pathol Exot. 84:15–24.

Whitcomb RF. 1981. The biology of spiroplasmas. Ann Rev Entomol. 26: 397–425.

Williamson DL, et al. 1996. *Spiroplasma diminutum* sp. nov., from *Culex annulatus* mosquitoes collected in Taiwan. Int J Syst Bacteriol. 46: 229–233.

Williamson DL, et al. 1999. *Spiroplasma poulsonii* sp. nov., a new species associated with male-lethality in *Drosophila willistoni*, a neotropical species of fruit fly. Int J Syst Bacteriol. 49:611–618.

Xie J, Tiner B, Vilchez I, Mateos M. 2011. Effect of the *Drosophila* endosymbiont *Spiroplasma* on parasitoid wasp development and on the reproductive fitness of wasp-attacked fly survivors. Evol Ecol. 25: 1065–1079.

Xie J, Vilchez I, Mateos M. 2010. *Spiroplasma* bacteria enhance survival of *Drosophila hydei* attacked by the parasitic wasp *Leptopilina heterotoma*. PLoS One 5:e12149.

Ye F, Melcher UJ, Rascoe JE, Fletcher J. 1996. Extensive chromosome aberrations in *Spiroplasma citri* Strain BR3. Biochem Genet. 34: 269–286.

Zerbino DR, Birney E. 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. 18:821–829.

Associate editor: Nancy Moran