Genome Sequence of the *Drosophila melanogaster* Male-Killing Spiroplasma Strain MSRO Endosymbiont

Juan C. Paredes, a Jeremy K. Herren, a,b Fanny Schüpfner, a Ray Marin, c Stéphane Claverol, d © Chih-Horng Kuo, e Bruno Lemaitre, a Laure Béven f

Global Health Institute, School of Life Science, École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland; Martin Lüscher Emerging Infectious Diseases (ML-EID) Laboratory, Molecular Biology and Bioinformatics Unit, International Centre of Insect Physiology and Ecology (icpe), Nairobi, Kenya; Center for Integrative Genomics, University of Lausanne, Lausanne, Switzerland; Pôle Protéomique, Plateforme Génomique Fonctionnelle de Bordeaux, University of Bordeaux, Bordeaux, France; Institute of Plant and Microbial Biology, Academia Sinica, Taipei, Taiwan; INRA, University of Bordeaux, Unité Mixte de Recherche 1332, Villenave d’Ornon, France

**ABSTRACT** Spiroplasmas are helical and motile members of a cell wall-less eubacterial group called *Mollicutes*. Although all spiroplasmas are associated with arthropods, they exhibit great diversity with respect to both their modes of transmission and their effects on their hosts; ranging from horizontally transmitted pathogens and commensals to endosymbionts that are transmitted transovarially (i.e., from mother to offspring). Here we provide the first genome sequence, along with proteomic validation, of an endosymbiotic inherited *Spiroplasma* bacterium, the *Spiroplasma poulsonii* MSRO strain harbored by *Drosophila melanogaster*. Comparison of the genome content of *S. poulsonii* with that of horizontally transmitted spiroplasmas indicates that *S. poulsonii* has lost many metabolic pathways and transporters, demonstrating a high level of interdependence with its insect host. Consistent with genome analysis, experimental studies showed that *S. poulsonii* metabolizes glucose but not trehalose. Notably, trehalose is more abundant than glucose in *Drosophila* hemolymph, and the inability to metabolize trehalose may prevent *S. poulsonii* from overproliferating. Our study identifies putative virulence genes, notably, those for a chitinase, the H$_2$O$_2$-producing glycerol-3-phosphate oxidase, and enzymes involved in the synthesis of the eukaryote-toxic lipid cardiolipin. *S. poulsonii* also expresses on the cell membrane one functional adhesion-related protein and two divergent spiralin proteins that have been implicated in insect cell invasion in other spiroplasmas. These lipoproteins may be involved in the colonization of the *Drosophila* germ line, ensuring *S. poulsonii* vertical transmission. The *S. poulsonii* genome is a valuable resource to explore the mechanisms of male killing and symbiont-mediated protection, two cardinal features of many facultative endosymbionts.

**IMPORTANCE** Most insect species, including important disease vectors and crop pests, harbor vertically transmitted endosymbiotic bacteria. These endosymbionts play key roles in their hosts’ fitness, including protecting them against natural enemies and manipulating their reproduction in ways that increase the frequency of symbiont infection. Little is known about the molecular mechanisms that underlie these processes. Here, we provide the first genome draft of a vertically transmitted male-killing *Spiroplasma* bacterium, the *S. poulsonii* MSRO strain harbored by *D. melanogaster*. Analysis of the *S. poulsonii* genome was complemented by proteomics and ex vivo metabolic experiments. Our results indicate that *S. poulsonii* has reduced metabolic capabilities and expresses divergent membrane lipoproteins and potential virulence factors that likely participate in *Spiroplasma*-host interactions. This work fills a gap in our knowledge of insect endosymbionts and provides tools with which to decipher the interaction between *Spiroplasma* bacteria and their well-characterized host *D. melanogaster*, which is emerging as a model of endosymbiosis.
invade hemolymph and organs, including the salivary glands, muscles, nerves, and connective tissue (5, 8, 9). The “infectious” category includes pathogens of shrimp and crabs that pose serious problems in the aquaculture industry, as well as pathogens of bees, such as Spiroplasma melliferum and S. apis (8–12). In addition, three Spiroplasma species (S. citri, S. kunkeli, and S. phoeniceum) are known plant pathogens that proliferate in the plant phloem and are vectored by phloem-feeding insects (3–5). Some, notably S. citri, can be grown in vitro and are amenable to genetic studies, unlike maternally transmitted strains, which are highly fastidious. To date, the genomes of 11 Spiroplasma species have been totally or partially sequenced, including two plant pathogens, S. citri and S. kunkeli (13, 14); two bee pathogens, S. melliferum and S. apis (10–12); commensals of deerflies and syrphid flies, S. chrysopelicola and S. syrphidicola, respectively (15); and four commensals of mosquitoes S. tawaenense, S. culicicola, S. diminutum, and S. sabaudienese (16, 17). Spiroplasma genomes are small, ranging from 1 to 2 Mb, and AT rich. To date, no vertically transmitted Spiroplasma species have been sequenced. This represents the last major lineage of vertically transmitted insect endosymbiont to have its genome sequenced. Sequencing of the S. poulsonii genome required overcoming challenges associated with the sequencing and alignment of genomes that are highly AT rich and overrun with viruses and other mobile elements. Here we report the draft genome sequence of the Drosophila melanogaster endosymbiont S. poulsonii MSRO (here referred to as S. poulsonii).

Together with Wolbachia bacteria, Spiroplasma bacteria are the only known natural inherited symbionts of Drosophila flies (18). At least 16 Drosophila species have been found to be infected with inherited Spiroplasma spp. (19). S. poulsonii MSRO was isolated from an infected D. melanogaster female collected in Uganda (20). S. poulsonii resides primarily in the hemolymph of D. melanogaster and has only been observed inside cells as it is being transmitted vertically in the oocyte cytoplasm. Vertical transmission occurs by subversion of the yolk transport and uptake pathway (21). Although horizontal transmission of S. poulsonii does occur over evolutionary time scales, it is rare (22). Within natural populations of D. melanogaster, S. poulsonii prevalence is variable but generally less than 5% (20, 23). In other species of Drosophila, related S. poulsonii strains are found over a broad prevalence range, in some cases as high as 80% (19, 24). S. poulsonii MSRO affects its D. melanogaster host in two striking ways. Like a number of other inherited endosymbionts, it manipulates host reproduction in order to increase the frequency of infected female hosts. It does this by killing males at the embryonic stage (23). Male killing has evolved independently in a number of inherited symbionts, including some strains of Wolbachia and Rickettsia (25, 26). A number of inherited Spiroplasma strains kill male flies, beetles, butterflies, and aphids (19, 27–29). The mechanistic basis of male killing induced by Spiroplasma bacteria, or by any inherited symbionts, remains largely enigmatic. This phenotype has captured the imagination of many biologists, as it is quite incredible that the symbiont is able to specifically recognize and kill one sex of its host at such an early stage. Notably, recent sequencing of the genome of butterfly male-killing Wolbachia sp. strain wBoll-b has revealed some fast-evolving sequences and horizontal gene transfer associated with viral sequences not present in other, closely related, genomes of non-male-killing Wolbachia strains (30).

Recently, there has also been established that several Spiroplasma strains, including a number of S. poulsonii strains, can protect their hosts against eukaryotic parasites. For example, the S. poulsonii strain that infects Drosophila neotestacea protects its host against nematode infection (24) and the strains infecting D. hydei (SPHY) and D. melanogaster (MSRO) protect their hosts against parasitoid wasps (31, 32). There is much interest in determining how symbionts confer protection, including understanding how symbionts target natural enemies without causing collateral damage in their hosts. In the case of S. poulsonii MSRO, there appear to be only minor effects on infected female D. melanogaster longevity and rates of egg laying. Intriguingly, old symbiont-infected females show an apparent lack of coordination and tremors (33).

In this study, we purified S. poulsonii strain MSRO from D. melanogaster hemolymph and sequenced its genome. We provide the first sequenced genome of an inherited Spiroplasma bacterium. By comparing the S. poulsonii genome content with the genome contents of closely related horizontally transmitted (i.e., nonendosymbiotic) Spiroplasma bacteria, our study provides insights into the molecular mechanisms involved in the endosymbiotic mode of life and provides an important resource for deciphering the mechanism of male killing and symbiont-mediated protection.

**RESULTS AND DISCUSSION**

Sequencing and assembly of the genome of S. poulsonii strain MSRO. Male-killing S. poulsonii strain MSRO was initially isolated from an infected D. melanogaster female collected in Uganda (20) and subsequently transferred to and maintained in an Oregon-R wild-type strain (34). Although a strain of S. poulsonii (DW-1) infecting Drosophila willistoni was once cultivated in cell-free medium (35), it was subsequently lost. Currently, there is no cell-free culture of endosymbiotic Spiroplasma bacteria. Instead, bacterial cells were obtained by manual collection of hemolymph from infected female flies with a microinjector, followed by filtration through a 0.45-μm-pore-size filter, which is expected to block host cells and other possible bacterial contaminants but not Spiroplasma bacteria (36). Three micrograms of pure DNA was extracted by the classical phenol-chloroform DNA extraction technique. The sequence generated by a PacBio RSII smart cell sequencing run produced an assembly of 12 contigs. To determine the total genome length and to position the 12 contigs, we performed an in silico alignment with optical mapping (OpGen) technology. This allowed us to obtain a schematic view of the genome and place contigs 1 to 10 along a single circular chromosome. The presence of numerous viral sequences was the main obstacle preventing a complete genome assembly. Contigs 1 to 10 covered ~93% (1.76 Mb) of the ~1.9-Mb genome of S. poulsonii chromosome length predicted by optical mapping (Fig. 1). The estimated size of the S. poulsonii chromosome is similar to the estimated genome size of S. citri strain gii3, 1.82 Mb (14), and larger than those of other spiroplasmas of the same clade, 1.38 and 1.55 Mb for S. melliferum IPMB4A (11) and S. kunkelii (37), respectively. Nevertheless, it is significantly smaller than the 2-Mb size estimated by pulsed-field gel electrophoresis for S. poulsonii strain DW-1 (38). Variation in the number of prophage sequences is thought to be partially responsible for the difference in chromosome size between these two S. poulsonii strains. It is notable that S. citri and related genomes contain numerous viral sequences and range from 1.6 to 1.9 Mb (39–41).

The draft S. poulsonii genome has a G+C content of 26.6%,
### Draft assembly of the S. poulsonii MSRO genome

**A** Locations of the 12 contigs on the S. poulsonii chromosome as determined by optical mapping with MapSolver software (OpGen). Contigs A and B could not be placed on the chromosome. (B) Schematic representation of the S. poulsonii chromosome. Contigs (in yellow) were placed in order on the basis of the in silico alignment by optical mapping technology. The estimated length of the chromosome is ~1.89 Mb. GC content in depicted in blue. (C) Genome assembly statistics. The table shown was modified from reference 15. The superscript letter a indicates that for S. chrysopicola, S. melliferum, and S. poulsonii, putative pseudogenes were annotated with the "pseudo" tag in gene feature as suggested by the NCBI GenBank guidelines and were not included in the total number of protein-coding genes. For S. citri, 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. The superscript letter b indicates that most of the plectrovirus-related regions were excluded from the S. melliferum IPMB4A assembly because of unresolvable polymorphism, resulting in a lower number of plectroviral genes. The genome of S. chrysopicola does not contain any identifiable plectroviral fragments; this lineage was likely to have diverged prior to the plectroviral invasion of the common ancestor of S. citri, S. melliferum, and S. poulsonii.

#### Table: Genome assembly statistics

<table>
<thead>
<tr>
<th></th>
<th>S. chrysopicola DF-1</th>
<th>S. citri GII3-3X</th>
<th>S. melliferum IPMB4A</th>
<th>S. poulsonii MSRO</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of chromosomal contigs</td>
<td>1</td>
<td>39</td>
<td>24</td>
<td>10</td>
</tr>
<tr>
<td>Combined size of chromosomal contigs (bp)</td>
<td>1,123,322</td>
<td>1,525,756</td>
<td>1,098,846</td>
<td>1,757,846</td>
</tr>
<tr>
<td>Estimated chromosomal size (bp)</td>
<td>-</td>
<td>1,820,000</td>
<td>1,380,000</td>
<td>1,890,000</td>
</tr>
<tr>
<td>Estimated coverage (%)</td>
<td>-</td>
<td>83.8</td>
<td>79.6</td>
<td>93</td>
</tr>
<tr>
<td>G+C content (%)</td>
<td>28.8</td>
<td>25.9</td>
<td>27.5</td>
<td>26.5</td>
</tr>
<tr>
<td>Coding density (%)</td>
<td>89</td>
<td>80.2</td>
<td>85.1</td>
<td>76.6</td>
</tr>
<tr>
<td>Protein-coding genes&lt;sup&gt;a&lt;/sup&gt;</td>
<td>1,009</td>
<td>1,905</td>
<td>932</td>
<td>2,129</td>
</tr>
<tr>
<td>Length distribution (Q1/Q2/Q3) (a.a.)</td>
<td>171/278/425</td>
<td>83/149/286</td>
<td>176/280/440</td>
<td>83/150/278</td>
</tr>
<tr>
<td>Plectrovirus proteins&lt;br&gt;&lt;sup&gt;b&lt;/sup&gt;</td>
<td>0</td>
<td>375</td>
<td>11</td>
<td>94</td>
</tr>
<tr>
<td>Hypothetical proteins&lt;br&gt;&lt;sup&gt;a&lt;/sup&gt;</td>
<td>394</td>
<td>519</td>
<td>337</td>
<td>859</td>
</tr>
<tr>
<td>Annotated pseudogenes&lt;br&gt;&lt;sup&gt;a&lt;/sup&gt;</td>
<td>6</td>
<td>401</td>
<td>12</td>
<td>82</td>
</tr>
<tr>
<td>rRNA operon</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>tRNA</td>
<td>33</td>
<td>32</td>
<td>32</td>
<td>31</td>
</tr>
<tr>
<td>Number of plasmids</td>
<td>0</td>
<td>7</td>
<td>0</td>
<td>0-2</td>
</tr>
</tbody>
</table>

**FIG 1** Draft assembly of the S. poulsonii MSRO genome. (A) Locations of the 12 contigs on the S. poulsonii chromosome as determined by optical mapping with MapSolver software (OpGen). Contigs A and B could not be placed on the chromosome. (B) Schematic representation of the S. poulsonii chromosome. Contigs (in yellow) were placed in order on the basis of the in silico alignment by optical mapping technology. The estimated length of the chromosome is ~1.89 Mb. GC content in depicted in blue. (C) Genome assembly statistics. The table shown was modified from reference 15. The superscript letter a indicates that for S. chrysopicola, S. melliferum, and S. poulsonii, putative pseudogenes were annotated with the "pseudo" tag in gene feature as suggested by the NCBI GenBank guidelines and were not included in the total number of protein-coding genes. For S. citri, 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. The superscript letter b indicates that most of the plectrovirus-related regions were excluded from the S. melliferum IPMB4A assembly because of unresolvable polymorphism, resulting in a lower number of plectroviral genes. The genome of S. chrysopicola does not contain any identifiable plectroviral fragments; this lineage was likely to have diverged prior to the plectroviral invasion of the common ancestor of S. citri, S. melliferum, and S. poulsonii. a.a., amino acids.
which is in agreement with a previous determination done with *S. poulsonii* DW-1 (38). The G+C content in *S. poulsonii* is similar to that observed in the other spiroplasmas belonging to the same clade (25.9 to 27.5%). G+C content variation in spiroplasmas from the Citri clade is usually attributed to the proportion of prophage sequences, which are AT biased. In general, insect bacterial symbiont genomes tend to be highly AT biased.

In addition to the 10 chromosomal contigs, two small contigs, A and B, could not be placed on the chromosome by optical mapping (Fig. 1A). These contigs correspond either to extrachromosomal DNA (plasmid) or to chromosomal sequences that are too small to be placed on the chromosome by the MapSolver software. *Spiroplasma* species that are members of the Citri clade, *S. melliferum* KC3 and *S. citri* GH3-3X (but not *S. melliferum* IPMB44A), contain plasmids that often carry potential virulence factors (10, 11, 14). The fact that contigs A and B contain several genes that are found in plasmids in other *Spiroplasma* species (42), like *Soj* and those for several ARPs (adhesion-related proteins), suggests that they could also be extrachromosomal DNA sequences. However, the facts that plasmids have not previously been found in *Spiroplasma* species compared with each other (15, 16). In the *S. poulsonii* genome, we found 94 predicted coding gene sequences corresponding to viral sequences from the *Plectro virus* SpV1 family (SpV1-R8A2, SpV1-C74, and SVTS2). Previous studies have revealed the presence of SpV3-like virus sequences in various *Spiroplasma* species infecting *Drosophila* flies. As the genome of an SpV3-like virus has not been sequenced, it could not be annotated in the *S. poulsonii* genome (43). The lengths of the viral sequences found in the *S. poulsonii* chromosome range from a few hundred base pairs to more than 3,000 bp. In addition, the *S. poulsonii* genome contains about 235 related genes encoding four different families of transposases, which are generally presumed to be involved in catalyzing integration and excision of DNA sequences. As previously reported for *S. melliferum* IPMB44A (11), a high level of gene decay and chromosomal rearrangements was observed in the *S. poulsonii* genome compared to genomes of *S. citri* and *S. melliferum* (see Fig. S1 in the supplemental material). This is likely a consequence of abundant viral sequences. Proliferation of viruses and other mobile elements is a common feature of the early evolution of insect symbiont genomes (44), and it should be noted that *Wolbachia*, the other endosymbiont of *Drosophila*, also harbors multiple virus and mobile-element sequences that are also thought to facilitate genome rearrangements (30, 45).

Horizontal transfer of genes can occur as a result of viral integration. It has been demonstrated that species-specific regions of the genomes of *S. citri* and *S. melliferum* were flanked by viral sequences, suggesting that they originated by horizontal transfer (15). For instance, several genes encoding transmembrane proteins in *S. melliferum* strain KC3 are flanked by viral sequences, forming a functional mobile genetic unit thought to be acquired from other bacteria (10). We did not uncover any functional gene units of this type in the genome of *S. poulsonii*. All of the sequences found in *S. poulsonii* were related to other spiroplasmas. This indicates that horizontal transfer between *S. poulsonii* and other bacterial endosymbionts associated with *Drosophila* (e.g., *Wolbachia*) is absent or rare.

**Gene annotations.** The genome of *S. poulsonii* encodes 31 tRNAs that are organized in five clusters and cover all 20 amino acids. It contains a single 16S–23S–50S ribosomal operon (Fig. 1C; see Table S1 in the supplemental material). The 16S rRNA gene sequence of *S. poulsonii* MSRO confirms its position in the *Poulsonii* clade and is identical to that of DW1 isolated from *D. willistoni*, NSRO isolated from *D. nebulosa*, and SPHY isolated from *D. hydei* (Fig. 2) (22, 46).

Among the 2,263 predicted genes found in the *S. poulsonii* draft genome, only 401 encode well-characterized proteins, 859 encode hypothetical proteins, 82 are pseudogenes, 94 are viral coding sequences, 3 encode rRNA, and 31 encode tRNAs. The length distribution of predicted amino acid sequences in *S. poulsonii* is lower than that in other spiroplasmas, indicating that the number of pseudogenes is probably underestimated. However, since these short hypothetical proteins have no functional domain/database hit, it is extremely difficult to determine their function (Fig. 1C). According to the clusters of homologous genes (COG) classification, ~81% of the *S. poulsonii* genes encode unknown proteins. Among the 401 well-characterized proteins identified in the genome, the three most highly represented classes of proteins were involved in translation/ribosome (class J) (~25%), replication/recombination/repair (class L) (~23%), and carbohydrate transport and metabolism (class G) (~8%), (Fig. 3). Of note, the large size of class L is due mainly to the presence of putative transposases that fall in this category (see Table S1 in the supplemental material).

The genome contains numerous genes involved in DNA replication, recombination, and translation, including *dhaA*, *dnaB*, *dnaC*, *dnaE*, *dnaG*, * dnaI*, * dnaJ*, * dnaK*, * dnaN*, *gyrA*, * gyrB*, * ruvB*, * ruvA*, * uvrA*, * uvrB*, * uvrC*, * uvrD*, * ligA*, * rhnB*, and * rhnC*. Additionally, it includes the complete Sec secretory system (*secA*, * secDF*, * secG*, * secY*, * ffh*, * ftsY*, and * vidC*). Most of the genes encoding mismatch repair proteins, such as *mutL*, *mutH*, *exol*, *exox*, and *exoJ*, are missing; only *mutM* is still present. The majority of the genes encoding proteins involved in homologous recombination are missing as well; only *recA*, * recE*, * recD*, * recO*, * recU*, * recR*, and * recT* were found; however, the *recE* and * recA* genes were truncated. As previously reported for *S. citri* and *S. melliferum* (11, 14), *S. poulsonii* has a premature stop codon in * recA*. It was hypothesized that a functional * recA* gene could facilitate homologous recombination and therefore genome rearrangements (47). The fact that * recA* is truncated seems to contradict this notion since extensive genomic rearrangements were found in *S. poulsonii* compared to other, closely related, spiroplasmas (see Fig. S1 in the supplemental material). However, there is some evidence that the pseudogenization of * recA* is recent (codon shifts have different locations on the * recA* gene in the different strains) and may be caused by independent events in each *Spiroplasma* lineage. Thus, the possibility cannot be excluded that the genome rearrangements we observed occurred before the loss of * recA* (48).

FIG 2 Phylogenetic tree of Spiroplasma species. This maximum-likelihood phylogenetic tree was inferred by using the 16S rRNA gene. GenBank accession numbers are listed in Materials and Methods. The values on internal branches are percentages of bootstrap support based on 1,000 replicates. Color codes represent the different clades, and the black arrow denotes the probable point of viral invasion of the lineage (15). Representative images of Spiroplasma hosts are shown to the right of each clade. Species previously sequenced are underlined. Solid underlining denotes an infectious pathogen, and dashed underlining denotes a mutualistic relationship with the insect host. Of note, all of the genomes sequenced are from horizontally transmitted S. poulsonii.

To gain further insights into the genomic association adapted with the endosymbiotic life, we compared the gene contents with those of the three horizontally transmitted spiroplasmas S. citri GII3-X, S. melliferum, and S. chrysopicola. We found that they all have 618 genes in common, among which more than a third correspond to putative virus coding sequences. In addition, S. poulsonii has 92, 21, and 18 genes in common specifically with S. citri, S. melliferum, and S. chrysopicola, respectively (Fig. 4). Besides genes coding for viral sequences, the most-represented classes common to all three spiroplasmas are translation/ribosome (J), replication/DNA recombination/repair (L), and nucleotide transport/metabolism (F). Among the genes absent from S. chrysopicola and S. syrphidicola but also present in the other closely related Spiroplasma species, we found genes involved in sugar metabolism (treA, fruR, fruF, depD, and kgdA) (see Table S3 in the supplemental material) (15). Thus, the evolutionary divergence of the Citri-Poulsonii and Chrysopicala clades was accompanied by divergence in genes involved in carbohydrate metabolism. Interestingly, among spiroplasmas belonging to the Citri clade, the COG class of which S. poulsonii lacks the most genes compared to S. citri and S. melliferum is also carbohydrate transport and metabolism (G) (see Table S3). This reduction in the number of genes involved in carbohydrate transport and metabolism could be related to the fact that S. poulsonii is an endosymbiotic Spiroplasma bacterium that is confined in its insect host while S. citri and S. melliferum can also be found outside their insect host. Similar to other members of the Citri-Chrysopicola-Mirum clade, S. poulsonii contains five copies of mreB (49), which encodes a functional analogue of actin in bacteria and has been shown to play an important role in the helical shape of the Spiroplasma cytoskeletal structure (50).

**Carbon and energy metabolism.** The S. poulsonii genome encodes the well-conserved F$_1$F$_0$ ATP synthase operon (atpB, atpE, atpF, atpH, atpA, atpD, and atpC), the entire glycolysis pathway, and four enzymes of the nonoxidative phase of the pentose cycle (rpe, 2 tkt, prs, and rpiB). Its genome also encodes the phosphonopyruvate phosphotransferase transport systems (PTSs) of glucose (ptsG), 2-O-a-mannosyl-d-glucose (mnpA), maltose, and ascorbate (ulaA) (Fig. 5A). It is notable that the genes encoding N-acetylglucosamine PTSs and FruA transporters were pseudogenes. We were not able to identify a trehalose transporter gene (treB), and the trehalase treA gene is truncated. This is an interesting finding because trehalose is the most abundant sugar of the Dro sophila hemolymph, followed by glucose and mannose (51).
The absence of trehalose transporters from *S. poulsonii* may be important in the regulation of symbiont densities, as it may limit endosymbiont overproliferation, which could be detrimental to its host. *treB* is found in *S. citri*, *S. apis*, and *S. melliferum* (11, 12, 14) but is absent from the *S. chrysopicola* and *S. syrphidicola* genomes (15). *S. apis* and *S. melliferum*, but not *S. chrysopicola* and *S. syrphidicola*, are pathogenic to their host (52). Another observation that is in line with this hypothesis is that *S. citri* is much more pathogenic than *S. poulsonii* when injected into *D. melanogaster* (34). The pseudogenization of two fructose transporter genes (*fruA*) is not surprising since fructose is found at very low levels in the fly hemolymph (51). We speculate that the evolutionary pressure to maintain these genes in *S. poulsonii* might be relaxed relative to spiroplasmas that infect plants (e.g., *S. citri*) because fructose is an important source of carbon in plant phloem (53).

A polymer found in fly hemolymph that could potentially be used as a source of nitrogen and carbon by *S. poulsonii* is chitin. The presence of a chitinase gene belonging to the GH18 superfamily suggests that *S. poulsonii* might be able to cleave different forms of chitin, chitosan, and peptidoglycan to produce N-acetylglucosamine (54). However, it is unlikely that chitin cleavage plays a role in metabolite acquisition, since a gene essential for N-acetylglucosamine metabolism, *nagA* (GlcNAC-6P deacetylase), is truncated because of a viral sequence insertion. Moreover, genes encoding the PTS involved in the import of N-acetylglucosamine are also absent from this draft genome (Fig. 5A; see Table S1 in the supplemental material).

*S. poulsonii* harbors a gene cluster involved in arginine metabolism (arcA, arcB, arc, and arcD) that is present in all sequenced spiroplasmas. However, in contrast to the *arcC* gene of other spiroplasmas, that of *S. poulsonii* is truncated. This suggests that arginine degradation via the arginine deiminase pathway is not required for *S. poulsonii* persistence in *Drosophila* (Fig. 5A; see Table S3). Production of ATP by arginine degradation is not essential in *S. citri* either (55). The lack of a functional arcC gene in the *S. poulsonii* genome would prevent the metabolism of carbamoyl-phosphate and thus the synthesis of NH₃ for ATP production. While a complete arginine deiminase pathway is not essential for *S. poulsonii*, we cannot rule out the possibility that carbamoyl-phosphate could be the substrate of another enzyme and that arginine could instead be implicated in energy production.

The main classes of lipids that are present in *Spiroplasma* bacteria are phosphatidylglycerols, cardiolipins, sterols, and sphingolipids (56). Cardiolipins are a class of lipids found exclusively in
eubacteria and in the mitochondria of eukaryotic cells and play a role in the curvature of the bacterial membrane (57). In contrast to other major lipidic components of Spiroplasma membranes, cardiolipins are not present at detectable levels in Drosophila hemolymph and therefore would need to be synthesized by S. poulsonii (58). A more detailed investigation of the lipid metabolic capacity of S. poulsonii revealed a lack of enzymes involved in the citric acid cycle and fatty acid elongation. S. poulsonii does, however, appear to be able to modify some phospholipids. The presence of a phospholipase gene (pldB) and genes involved in the synthesis of phosphatidylglycerol and diphosphatidylglycerol (cardiolipin) indicates that S. poulsonii has the capacity to generate phospholipids from fatty acids or diacylglyceride (DAG) precursors imported from the external environment. The cardiolipin synthetase enzyme gene cls is present in S. poulsonii as in all sequenced genomes of spiroplasmas of the Citri-Chrysopicola-Mirum clade. Genomic analysis confirms that S. poulsonii can synthetize cardiolipin from DAG extracted from the hemolymph, consistent with a previous study (33) (Fig. 5B). S. poulsonii also harbors the complete set of enzymes implicated in terpenoid C55 synthesis (dxs, dxr, ispD, ispE, ispF, ispG, and ispH), which are involved in the nonmevalonate pathway resulting in the synthesis of farnesyl pyrophosphate. This complete pathway is also found in all Spiroplasma genomes of the Apis and Citri-Chrysopicola-Mirum clades (with the exception of S. citri, where ispE could not be identified, [14]). As was found to be the case for other spiroplasmas, these genes are dispersed throughout the S. poulsonii chromosome.

We utilized an in vitro assay to validate our analysis of the metabolic capacities of S. poulsonii. A previous study has shown that the helicoidal shape and motility of spiroplasmas require the maintenance of an electrochemical transmembrane gradient, which consumes ATP (39). We therefore used S. poulsonii shape

FIG 5 S. poulsonii sugar metabolism and virulence genes. (A) Metabolic scheme representing enzymes and transporters encoded by the S. poulsonii genome that are involved in sugar metabolism. Enzymes or transporters that are absent from or nonfunctional in S. poulsonii are red. (B) Potential entomopathogenic factors and surface proteins encoded by the S. poulsonii genome. Three potential virulence factors (chitinase A, GlpO, and cardiolipin) may play a role in S. poulsonii-mediated protection against parasitoid wasps. Two spiralin-like proteins and one ARP might play important roles in endosymbiosis.
deformation as an assay to monitor the ability of *S. poulsonii* to maintain a helicoidal shape and hence its capacity to efficiently generate ATP from several metabolites, including glucose, fructose, trehalose, or arginine. *S. poulsonii* cells were extracted from *D. melanogaster* hemolymph and incubated in phosphate buffer alone or complemented with different sugars or arginine. Bacterial shape was then monitored by fluorescence microscopy. Representative images (A) and percentages of helical cells compared to the total cell counts after 7 days (B) are shown. *S. poulsonii* maintains its helical shape for 1 week in the presence of glucose but not fructose, trehalose, or arginine. Shown is the mean ± the standard error of the mean of data pooled from three independent experiments (*n* = 100; ***, *P* < 0.0001).

**Potential virulence factors.** Because *S. poulsonii* is known to protect flies against parasitoid attacks, we searched the *S. poulsonii* genome for factors that might play a role in the killing of parasitic wasps that develop inside fly larvae. The *S. poulsonii* genome contains one gene coding for a chitinase A that could play a role in *Spiroplasma-Drosophila* interactions (Fig. 5B). One potential target of this chitinase is the basal membranes surrounding the ovaries that *S. poulsonii* must cross in order to reach the germ line and be vertically transmitted (21). This chitinase could be a virulence factor but could also be beneficial to the host by playing a role in *Spiroplasma*-mediated protection against parasitoid wasps by targeting the chitin-based cuticle of this hymenoptera.

As previously reported for *S. melliferum* KC3, *S. taiwanense*, and *S. culicicola*, the *S. poulsonii* genome contains the gene that encodes glycerol-3-phosphate oxidase (*glpO*) (10, 16, 17). This gene is found in a cluster between *glpK* and *glpF*, which encode a kinase and a membrane channel involved in glycerol metabolism, respectively. These three genes (*glpF, glpO*, and *glpK*) are absent from *S. citri* and *S. melliferum* IPMB4A (11, 14). *GlpO* has been characterized as a virulence factor in *Mycoplasma mycoides* subsp. *mycoides*, inducing tissue inflammation (60). The two substrates of GlpO are glycerol-3-phosphate and oxygen, while the two products of the catalyzed reaction are glycerone phosphate and hydrogen peroxide (H$_2$O$_2$). H$_2$O$_2$ has been reported to be involved in the cellular immune response of insect against parasitoid wasp eggs (61). We therefore hypothesize that GlpO could be involved in protection against parasitoid wasps (Fig. 5B) and possibly in the pathological effects observed in old flies harboring *S. poulsonii* (33).

As described above, our genomic analysis confirms that *S. poulsonii* has the set of enzymes to generate cardiolipin from diacylglycerol in fly hemolymph (see Table S1 in the supplemental material; Fig. 5B). A number of studies have demonstrated links between cardiolipin and cell death (62), which raises the possibil-
Adhesins and surface membrane proteins. Like other members of the class Mollicutes, Spiroplasma bacteria lack a cell wall and their plasma membrane is directly exposed to the external environment. In Mollicutes, many membrane-associated components, including transmembrane proteins and lipoproteins (i.e., acylated proteins at their N-terminal cysteinyl residue), have been shown to be involved in pathogenesis or virulence or to interact with the host immune system (63). In order to identify S. poulsonii surface components that may be putatively involved in interactions with its Drosophila host, we performed a proteomic analysis of the membrane and soluble fraction of S. poulsonii isolated from Drosophila hemolymph. Protein expression profiles were established following TX-114 phase partition and SDS-PAGE separation and further characterized for protein identities by liquid chromatography-tandem mass spectrometry. This analysis led us to identify more than 300 proteins that were validated by our genome annotation. The protein content of the detergent phase was enriched in amphiphilic proteins and contained 24 proteins for which a prokaryotic membrane lipoprotein lipid attachment site (PROSITE PS51527) was predicted. The detected lipoproteins represented 45% of the predicted lipoproteome (53 putative lipoproteins encoded by genes present in the genome; see Tables S1 and S2 in the supplemental material). The large number of lipoproteins exposed in S. poulsonii membrane suggests that S. poulsonii contains a large repertoire of lipoproteins that are likely to shape its surface and to contribute to its capacity to be an endosymbiont. Spiralin is a lipoprotein known only from Spiroplasma bacteria. It is the most abundant lipoprotein found at the surface of S. citri and S. melliferum (64, 65) and has been implicated in the abilities of S. citri to adhere to insect organs and to invade cells (66). While spiralin is highly conserved among spiroplasmas from the Citri clade (about 60% identity among the S. citri, S. melliferum, and S. kunkeli spiralins), the spiralin A-encoding gene (referred to as Spiralin A) found in S. poulsonii (v1c12390) is only 40% identical to those for other, closely related, Spiroplasma spiralins (see Fig. S2A in the supplemental material). Interestingly, the genome of S. poulsonii encodes another spiralin-related protein (v1c09300) that we named spiralin B. This lipoprotein is only ~25% similar to the S. melliferum and S. citri spiralins and S. poulsonii spiralin A (see Fig. S2A). To our surprise, proteomic analysis indicates that spiralin B, not spiralin A, was the most abundant protein detected in the S. poulsonii membrane (relative abundance, >70%). These divergent lipoproteins (spiralins A and B) may play a role in S. poulsonii vertical transmission, including colonization of the female germ line.

We also found a homolog of an ARP that was previously implicated in the ability of S. citri to invade insect cells (67). Genes encoding ARPs have been found in all spiroplasmas in the Citri clade and are found on plasmids (10, 42). As stated earlier, the putative ARP gene in S. poulsonii (v1c22630) was found on contig B, which may be extrachromosomal. Proteomic analysis suggests that this protein is abundant in S. poulsonii membrane (see Table S2 in the supplemental material). The S. poulsonii protein is 60 to 66% similar to S. citri ARPs (ScARPs) (see Fig. S2B) and, as is known to be the case for all characterized ScARPs, contains a signal peptide that is predicted to be cleaved after alanine 23. In addition, the presence of a hydrophobic transmembrane alpha helix located close to the C terminus is predicted in the S. poulsonii protein. Seven repetitions of a 36-amino-acid domain in the N-terminal hydrophilic region correspond to an NCL-1–HT2A–Lin–41 repeat domain that is found in numerous pathogenic bacteria and is known to be involved in eukaryotic cell invasion. This structural motif is found in several ScARPs but also in proteins mediating the invasion of host brain epithelia by Mycobacterium tuberculosis (68). In light of this, it is probable that the S. poulsonii ARP may also play a role in tissue invasion.

Conclusions. We sequenced the first genome of a vertically transmitted Spiroplasma species, S. poulsonii. Using PacBio technology and optical mapping, we obtained a draft assembly of the S. poulsonii MSRO genome composed of 12 contigs and covering 93% of the total chromosome length. The high number of repeated viral sequences integrated in the chromosome was the main obstacle to obtaining the complete chromosomal sequence. The size of the genome, the presence of many viral sequences, and the evidence of rearrangement indicate that S. poulsonii has retained many genomic features already described for spiroplasmas of the Citri clade. The absence of any sequence deriving from other groups of bacteria indicates that the mechanisms of male killing and vertical transmission have evolved independently in this lineage. The presence of many virus sequences and eroding genes indicates that S. poulsonii, like other facultative endosymbionts (Wolbachia, Sodalis, Serratia) has specific genome features distinct from those of both free-living bacteria and the highly compact obligate endosymbiont genomes (30, 45, 69–71).

Analysis of metabolic pathways reveals striking differences between S. poulsonii and horizontally transmitted spiroplasmas. S. poulsonii has lost many metabolic pathways and transporters, indicating a high level of metabolic integration with its insect host, similar to many insect endosymbiont genomes. Consistent with genome-based predictions, we observed that S. poulsonii metabolizes glucose but not trehalose, fructose, or arginine. The absence of a trehalose transporter is striking given that this sugar is the most abundant one in Drosophila hemolymph. We speculate that this adaptation may prevent S. poulsonii from overproliferating and damaging its host. It was recently demonstrated that the amount of hemolymphatic DAG limits S. poulsonii titers in the hemolymph (33). It was speculated that this greater dependence on lipids and lesser dependence on sugar prevent endosymbiont overproliferation under conditions of host nutrient limitation, since lipid availability depends heavily on the fly’s nutritional state. Our results confirmed that S. poulsonii lacks most of the lipid metabolic enzymes and therefore must acquire lipids intact from its host. Notably, S. poulsonii has the genomic capacity to produce the bacterium-specific lipid cardiolipin. It also possesses the complete nonmevalonate terpenoid synthesis pathway. S. poulsonii also has an incomplete set of enzymes required for arginine metabolism. Although S. poulsonii has a low impact on the general fitness of its host, very high titers of Spiroplasma bacteria in old flies are associated with defects suggestive of neuronal dysfunction (33). This indicates that the S. poulsonii genome codes for factors that have the capacity to be harmful to insects.

S. poulsonii is also a protective endosymbiont in that it can mediate the protection of its host against infestation by parasitoid wasps (31, 32). One possible explanation for this protection is that “immune priming” by S. poulsonii induces a stronger basal immune response in the fly and hence more efficient elimination of wasp eggs/larvae. However, S. poulsonii does not activate a humoral or cellular immune response in the fly (34). The S. poulsonii
genomic reveals three putative candidate virulence genes as possible virulence factors for the wasp: the gene for the chitinase Chia, the gene for the H₂O₂-producing enzyme GlpO, and the gene for the toxic lipid cardiolipin. It is noteworthy that virulence factors have been described as playing a possible role in Spiroplasma-mediated protection against nematodes (72). Another possible hypothesis that cannot be excluded is that Spiroplasma-mediated protection is mediated by nutrient competition. Endosymbiont proliferation depletes the lipid content of the fly hemolymph (33), and both spiroplasmas and wasp larvae colonize this fly compartment. One could speculate that Spiroplasma growth reduces the concentration of nutrients that are essential for wasp larval development.

We have identified several plasma membrane proteins that mediate the interaction with the host, notably, two spiralin-related proteins quite divergent from other Spiroplasma spiralin and an ARP. These adhesion proteins might be implicated in the interactions with yolk protein and the yolkless receptor on the exterior of the oocyte membrane, which result in Spiroplasma endocytosis (21). The S. poulsonii membrane proteins involved in mediating these interactions are likely to be spiralins and ARPs; however, the details of these interactions have not yet been elucidated. The current genome could be used as a tool to facilitate the genome annotation of other endosymbiotic spiroplasmas. Further comparisons of non-male-killing and male-killing Spiroplasma genomes could help us to understand the molecular mechanism underlying this phenotype. In conclusion, the completion of the S. poulsonii genome fills a major gap in our knowledge of insect endosymbionts and provides an important resource to decipher the interaction between Spiroplasma bacteria and their well-characterized insect host, D. melanogaster.

MATERIALS AND METHODS
Genomic DNA preparation. Cell-free cultivation of spiroplasmas is not currently possible. To obtain the required amount of genomic DNA to generate a genomic library (~3 µg), we extracted ~50 µl of pure fly hemolymph with a Nanoject microinjector (Drummond Scientific) from 1-month-old infected flies. The hemolymph extracts were immediately mixed with 500 µl of pure fly hemolymph, 5 µl of the bacterial contaminant or Drosophila oocyte involves interactions with yolk protein and the yolkless receptor on the exterior of the oocyte membrane, which result in Spiroplasma endocytosis (21). The S. poulsonii membrane proteins involved in mediating these interactions are likely to be spiralins and ARPs; however, the details of these interactions have not yet been elucidated. The current genome could be used as a tool to facilitate the genome annotation of other endosymbiotic spiroplasmas. Further comparisons of non-male-killing and male-killing Spiroplasma genomes could help us to understand the molecular mechanism underlying this phenotype. In conclusion, the completion of the S. poulsonii genome fills a major gap in our knowledge of insect endosymbionts and provides an important resource to decipher the interaction between Spiroplasma bacteria and their well-characterized insect host, D. melanogaster.

MATERIALS AND METHODS
Genomic DNA preparation. Cell-free cultivation of spiroplasmas is not currently possible. To obtain the required amount of genomic DNA to generate a genomic library (~3 µg), we extracted ~50 µl of pure fly hemolymph with a Nanoject microinjector (Drummond Scientific) from 1-month-old infected flies. The hemolymph extracts were immediately mixed with 500 µl of phosphate-buffered saline (PBS) on ice to avoid melanization. Bacteria were pelleted, washed twice with 500 µl of PBS, and then filtered through a 0.45-µm filter to eliminate other potential bacterial contaminants or Drosophila hemolymph cells (36). Bacteria were subsequently pelleted. For optical mapping, the bacterial pellet was sent to OpGen. For sequencing, DNA was extracted by phenol-chloroform extraction. The DNA library was prepared with the Pacific Biosciences DNA Template prep kit 2.0 (3 to 10 kb) with 3 smrtcells (V3/H11001) 1 smrtcell (V3/H9262 P4-C2) 2 movies, HGAP-3 Assembly was performed with HGAP (hierarchical genome assembly process) version 3 from the PacBio SMRT analysis pipeline smrtanalysis-2.2.0, mean coverage of 186.39, length cutoff of 9,450, 47 molecules real-time (SMRT) MagBead kit. The loading concentration was 180 min with the P4-C2 chemistry kit with size selection with the single-molecule real-time (SMRT) MagBead kit. The loading concentration was 0.09 nM in an SMRT cell v3 with the PacBio RS II sequencing machine. Assembly was performed with HGAP (hierarchical genome assembly process) version 3 from the PacBio SMRT analysis pipeline smrtanalysis-2.1.1. The mean coverage is 186.39 with the following parameters: 1 smrtcell 1 movie + 1 smrtcell (V3/H9262) 2 movies, HGAP-3 smrtanalysis-2.2.0, mean coverage of 186.39, length cutoff of 9,450, 47 polished contigs, a maximum contig length of 504,367, an N50 contig length of 163,695, and a contig length sum of 2,354,399. In the PacBio reference assembly were detected and corrected by aligning 454 reads (GATC Biotech AG Titanium Sequencing kit XLR70, genomic library for GS FLX sequencing). At each position of the genome, we computed the proportions of mapped reads supporting the reference base and a potential alternative base. If the alternative base was supported by at least 70% of the reads, the reference was updated to the new more abundant alternative base. Only uniquely mapping reads ≥200 bp long were considered in the analysis, and a coverage of at least 10 reads at a given position was required to proceed with the correction. Alignments were obtained with Bowtie 2. Important metabolic truncated genes were confirmed by PCR and Sanger sequencing.

Optical mapping comparative with other genomes. Whole-genome mapping was performed by OpGen. Contigs obtained by HGAP alignment were placed into optical mapping by in silico alignment with OpGen MapSolver Software 3.2.0. Figure 1B was obtained with Geneious 7.0.6 software. To examine the extent of genome rearrangement in Spiroplasma and other, related, bacteria, we used MAUVE 2.3.1 (73), S. melliferum IPMB4A (GenBank accession no. AMGI01000000), and S. citri GII-3X (GenBank accession no. AM285301 to AM285339).

Spiroplasma phylogenetic relationships were reconstructed by the maximum-likelihood method with a Tamura-Nei model in Geneious. In addition to S. poulsonii MSRO, 16S rRNA sequences from the following Spiroplasma bacteria were included in the analysis: S. saudaviensis (Ar-1343) (NR_025710), S. floricola (23-6) (NR_025703), S. aitc (ATCC 33834) (GJ993267), S. mirum (NR_014955), S. chrysopilica (DF-1) (NR_025699), S. pyrphycida (EA-1) (NR_025711), S. penaei (SHIRMP) (NR_043177), S. insolitum (M55) (NR_025705.1, Spiroplasma strain NSRO from D. nebulosa (AB434486), a Spiroplasma species from D. hydei (TU705) (FJ657200), S. poulsonii MSRO (UGA) (FJ657180), S. phoeniceum (P40) (NR_043178), S. kunkeli (DQ319068), S. citri (GI-3X) (AM285316), S. melliferum (IPMB4A) (JQ347516), S. melliferum BC-3 (NR_025756), S. mojavensis (OPN040756) (FJ657216), a Spiroplasma species from D. aldrichi (TU1106) (FJ657236), and S. wheeleri (cat0806B) (FJ657228). The 16S rRNA of a Spiroplasma species from D. neoteuctaceae was kindly provided by Steve Perlman. Mycoplasma genitalium (G37) (NR_026155) was used as the outgroup. We resampled the tree by bootstrapping (1,000 replicates). Alignment of the spiralin sequences of S. melliferum IPMB4A (ELLA4999), S. citri (AAB06629), and S. kunkeli (AAB03468) was performed with the CONNET algorithm.

Genome analysis and annotation. The procedures used for genome analysis and annotation were based on those described in our previous studies (11, 15) (see Text S1 in the supplemental material for more information).

Electrochemical transmembrane gradient maintenance test. Five hundred nanoliters of pure S. poulsonii was extracted and diluted in 200 µl of sodium phosphate buffer (50 mM) alone or complemented with 50 mM glucose, fructose, trehalose, or arginine. Images were taken every day for 7 days. To observe S. poulsonii in fly hemolymph, 5 µl of the sample was mixed with 5 µl of PBS containing 0.02 mM SYTO9 (Invitrogen). Slides were then mounted and observed on a Zeiss Axio Imager Z1. Images were captured with an AxioCam MRm camera and AxioVision software.

Nucleotide sequence accession number. This Whole Genome Shotgun project has been deposited at DDBJ/EMBL/GenBank under accession no. JTLV00000000. The version described in this paper is version JTLV01000000 (http://www.ncbi.nlm.nih.gov/nuccore/JTLV00000000).

SUPPLEMENTAL MATERIAL
Supplemental material for this article may be found at http://mbio.asm.org/lookup/suppl doi:10.1128/mBio.02437-14/-DCSupplemental.

Text S1, DOCX file, 0.1 MB.
Figure S1, EPS file, 0.8 MB.
Figure S2, EPS file, 2.5 MB.
Table S1, XLSX file, 0.2 MB.
Table S2, XLSX file, 0.1 MB.
Table S3, XLSX file, 1.1 MB.

mbio.asm.org  on December 14, 2017 - Published by mbio.asm.org Downloaded from
ACKNOWLEDGMENTS

We thank Steve Perlman and Elodie Ramond for stimulating discussions. We thank Keith Hartshman, Maroun Bou Sleiman, and Petr Dostál for biostats help and the Lausanne Genomic Technologies Facility (UNIL Switzerland) for DNA sequencing.

The Lemaitre lab work was funded by the ERC advanced grant (339970), Chih-Horng Kuo was supported by research grants from Academia Sinica and the National Science Council of Taiwan (NSC 101-2621-B-001-004-MY3). Laure Béven was supported by grants from Structure Fédérative de Recherche Biologie Intégrative et Ecologie.

REFERENCES


