This is a post-review version of the following article:

Evolution and diversity of inherited Spiroplasma in Myrmica ants

Matthew J. Ballinger, Logan D. Moore and Steve J. Perlman

February 2018

The final publication is available at:

https://doi.org/10.1128/AEM.02299-17

Citation for this paper:

Accepted Manuscript:

Evolution and diversity of inherited Spiroplasma in Myrmica ants

Matthew J. Ballinger\textsuperscript{a,1}, Logan D. Moore\textsuperscript{a}, and Steve J. Perlman\textsuperscript{a}.

\textsuperscript{a}Department of Biology, University of Victoria, Victoria, BC, Canada V8W 2Y2

\textsuperscript{1}To whom correspondence should be addressed: mattball@uvic.ca
Abstract (204 words)

Microbial partners play an important role in the biology and ecology of animals. In insects, maternally-transmitted symbionts are especially common and can have host effects ranging from reproductive manipulation to nutrient provisioning and defense against natural enemies. In this study, we report a genus-wide association of Myrmica ants with the inherited bacterial symbiont, Spiroplasma. We screen Myrmica ants collected from the wild, including the invasive European fire ant, Myrmica rubra, and find an extraordinarily high prevalence of this symbiont – 8 of 9 species, 42 of 43 colonies, and 250 of 276 individual workers were harboring Spiroplasma – only one host species was uninfected. In our screens, each host species carried a distinct Spiroplasma strain, and none were infected with more than one strain. All symbionts belong to the citri clade, allied most closely with pathogenic Spiroplasma of corn crops and honeybees, and there is strong evidence of host-symbiont persistence across evolutionary timescales. Genome sequencing of two Spiroplasma symbionts revealed candidate genes that may play a part in the symbiosis, a nutrient transporter absent from other Spiroplasma, and a ribosome-inactivating protein previously implicated in parasite defense. These results together suggest long-term, likely mutualistic relationships atypical of Spiroplasma-insect associations with potential significance for broad ecological interactions of Myrmica.
Animal-associated microbial symbionts can dramatically affect the biology of their hosts. Identification and characterization of these intimate partnerships remains an essential component of describing and predicting species interactions, especially for invasive host species. Ants perform crucial ecological functions as ecosystem engineers, scavengers, and predators, and ants in the genus *Myrmica* can be aggressive resource competitors and reach high densities in their native and invaded habitats. In this study, a novel symbiosis is identified between *Myrmica* ants and the facultative bacterial symbiont, *Spiroplasma*. Broad host distribution, high frequencies of infection, and host-symbiont codivergence over evolutionary timescales, an uncommon feature of *Spiroplasma* associations, suggest an important, likely mutualistic interaction. Genome sequencing identified highly divergent gene candidates that may contribute to *Spiroplasma*’s role as a possible defensive or nutritional partner in *Myrmica*. 
Introduction

It is now well established that most insects harbor maternally inherited bacterial endosymbionts that play critical roles in the ecology and evolution of their hosts (1).

Insect lineages that feed exclusively on nutrient-poor diets, such as plant sap or animal blood, typically host obligate nutritional endosymbionts that provide essential vitamins and amino acids. These obligate endosymbionts are often housed in specialized symbiont organs and show patterns of strict and ancient co-diversification with their hosts.

More common still are facultative inherited symbionts, of which the best known is *Wolbachia* (2). While these symbionts are transmitted almost exclusively through females over ecological timescales, they very rarely cospeciate with their hosts, and instead, repeatedly colonize new host lineages over evolutionary timescales via horizontal transmission. Although not essential for host survival or reproduction, many facultative inherited symbionts increase host fitness under certain conditions, for example by protecting their hosts against natural enemies or environmental stresses (3–8). Others manipulate their host's reproduction in order to increase the frequency of symbiont-infected females (9–13). Five bacterial lineages are particularly widespread as facultative symbionts of insects. In addition to *Wolbachia*, these are *Arsenophonus*, *Cardinium*, *Rickettsia*, and *Spiroplasma* (14, 15). Initial surveys have found that these bacteria infect ~5-30% of insect species – this represents millions of infected species. However, most insect lineages have been poorly sampled, and even when infections have been reported, it is often not understood how these facultative symbionts affect host fitness or persist in host populations.
Spiroplasma is an incredibly diverse genus of bacteria that infect arthropods, with a wide range of fitness effects and transmission strategies (16). Many Spiroplasma are pathogenic, including pathogens of bees, crayfish, and plants (17–20). Even more prevalent are horizontally transmitted gut commensals that have been isolated from a wide range of insects, including beetles and flies (21, 22). Finally, maternal transmission has evolved independently in a number of Spiroplasma lineages. Vertically transmitted Spiroplasma can be found both inside and outside cells, often at high densities in insect hemolymph, as well as in ovarian tissues (23). While the effects of most vertically transmitted Spiroplasma are not known, a number of strains manipulate their host's reproduction by killing male embryos; male-killing Spiroplasma strains have been documented in butterflies, planthoppers, beetles, flies, and lacewings (10, 11, 24–26). Some Spiroplasma protect their hosts against natural enemies, with strains that infect aphids providing protection against pathogenic fungi (27), and strains that infect Drosophila flies protecting against parasitic wasps and nematodes (4, 8). Recent studies have implicated a diverse arsenal of toxins called ribosome-inactivating proteins (RIPs) in Drosophila defense (28, 29). Identification of a Spiroplasma-encoded RIP transcript in the publically-available transcriptome of the European invasive fire ant, Myrmica rubra, motivated closer examination of the relationship between Spiroplasma and this ant genus in the present study.

Although Spiroplasma infects a wide range of arthropods, few studies have examined a specific group of hosts in detail. The best studied inherited Spiroplasma are those that infect Drosophila. At least 18 species have been found to harbor Spiroplasma (8, 30, 31), with infection frequencies ranging from less than 5% to greater than 85%
Drosophila flies have been independently colonized by five different lineages of inherited Spiroplasma, from the citri, poulsonii, ixodetis, and tenebrosa clades (30). In this study, ant species in the genus Myrmica were surveyed for Spiroplasma. This genus also appears to be a hotspot for Spiroplasma infection, with all species except one infected at high frequency. Unlike Drosophila, however, Spiroplasma infecting Myrmica are all members of the citri clade, and there is a strong phylogenetic signal suggesting persistent host-symbiont associations across evolutionary timescales.

Results

Spiroplasma symbionts are widespread in Myrmica

Nine species of Myrmica, broadly distributed across the genus (Fig 1), were screened for Spiroplasma by PCR amplification of the ftsZ gene; all but one were positive (Table 1). Spiroplasma genes were also detected in all three publicly available Myrmica transcriptomes (M. rubra, M. ruginodis, and M. sulcinodis; NCBI BioProject PRJDB4088). Among screened species, the prevalence of infection was high, with 250 out of 276 individuals, and 42 out of 43 colonies, testing positive. COI failed to amplify or yield quality sequence from nine of the 284 DNA extractions and these samples were excluded from prevalence calculations. In one case, ftsZ amplified from a Myrmica sp. sample in which COI failed and this sample was conservatively excluded from subsequent analysis. For the two best sampled species, M. rubra and M. scabrinodis (two mtDNA haplotypes), infection frequencies were 86% and 96%, respectively, and all colonies were infected. Infection frequencies were similarly high in juvenile stages, with 9 of 10 larvae and 8 of 9 pupae from one M. scabrinodis colony testing positive. Many of the species in our data set are represented by a single colony or individual, yet in most
cases Spiroplasma was consistently detected despite limited sampling; however, high
prevalence for these species should not be assumed until it can be demonstrated
through a similarly thorough sampling effort.

To determine where Spiroplasma infection is localized, DNA extractions from the
head, thorax, gaster, and legs of adult Myrmica vandeli and M. scabrinodis were
screened; all tissue types were positive, indicating Spiroplasma is present in the
hemolymph and is not restricted to the gut.

 Spiroplasma-host specificity and evolutionary relationships

Phylogenetic analysis of symbiont ftsZ sequences places all of the Myrmica
Spiroplasma strains in the citri clade (Fig 2A), although they are not monophyletic. No
Spiroplasma strain was shared between species. Two distinct ftsZ sequences were
recovered from the published transcriptome of M. ruginodis, suggesting a coinfection,
although this was not examined in greater detail, as M. ruginodis samples were not
screened. Myrmica Spiroplasma form three clades that appear to correspond with host
species groups, one with members of the scabrinodis group, one with members of the
fracticornis group and allies, and a third with M. rubra and M. alaskensis. Unlike the
Spiroplasma of Myrmica, those of other ants are more broadly distributed throughout
the genus Spiroplasma (Fig S1). ParaFit was used to perform a global test of host-
symbiont codivergence among all ten distinct host lineages that were screened plus M.
ruginodis. The null hypothesis of independent host and symbiont evolution, was not
rejected at a significance threshold of .05 (Fig 2B; p = .07). Exclusion of M. ruginodis
and its dual Spiroplasma strains resulted in rejection of the null hypothesis, though at a
marginally significant p = .02.
The most thorough sampling was from Lac Remoray, France, where 22, 2, and 3 colonies of *M. scabrinodis*, *M. vandeli*, and *Formica picea* were collected within meters of each other. *Myrmica scabrinodis* contained two distinct mitochondrial haplotypes (96.2% similar at COI), and each of these haplotypes harbored its own *Spiroplasma ftsZ* haplotype (99.2% similar at *ftsZ*). Sixteen colonies had one mitochondrial haplotype, six had the other, and no colony had both. Sanger sequencing of a fragment of the long wavelength rhodopsin gene, as well as Illumina sequencing, confirmed that these two mitochondrial haplotypes are one species (i.e. there were no differences in nuclear genes). The two *M. vandeli* colonies harbored a distinct *Spiroplasma* strain that was 99.0-99.2% similar at *ftsZ* to the *Spiroplasma* in *M. scabrinodis*. *Spiroplasma* was absent from the three *Formica picea* colonies (n=26 individuals), further highlighting the absence of lateral transfer of *Spiroplasma* symbionts among microsympatric hosts.

Lastly, two of the *M. scabrinodis* colonies were initially keyed as *M. martini*, a species that was only recently described based on complex morphometrics (34), with no clear morphological features distinguishing it from *M. scabrinodis* (and with the authors' discriminant function misclassifying 10% of individuals). Molecular data from both colonies – mitochondrial and nuclear loci, as well as the symbiont locus – are identical to those of the other 14 *M. scabrinodis* haplotype B colonies in our study, suggesting *M. martini* is not a valid species.

**Genome content of * Spiroplasma* symbionts of *Myrmica***

*Spiroplasma* genomes were sequenced from two host species, *M. vandeli* and *M. scabrinodis*. Eighty-seven and eighty-six million reads, respectively, were generated from the pooled DNA of five ants per species. Preliminary metagenomes were
assembled from low-GC reads (<31%) and consisted of almost exclusively ant, *Wolbachia*, and *Spiroplasma* contigs by blastp (Table 2). These preliminary *Spiroplasma* contigs were used to improve mapping and assembly of *Spiroplasma* reads in each final assembly (see methods). From the final assemblies, 481 contigs encoding 1,019 proteins and 402 contigs encoding 995 proteins were assigned to the *Spiroplasma* symbionts of *M. scabrinodis* and *M. vandeli*, respectively. 98.4% of *M. scabrinodis* proteins were also identified in *M. vandeli*, and 96.8% in the reciprocal comparison, suggesting that the majority of protein coding genes are represented in our *Spiroplasma* assemblies. As expected, the vast majority of these putative genes also yielded blastp hits to the genomes of *S. citri*, *S. kunkelii*, *S. melliferum* and *S. poulsonii* (Table 2). Genome read coverage for *M. scabrinodis* and *M. vandeli* respectively, was 10.4 and 17.2 (median coverage), and 9.1 and 13.6 (mode coverage) (Fig 3A and B). A majority fraction of the top blastp hits for each *Spiroplasma* assembly was to taxa belonging to the citri clade; 70.8% of 1,019 in *M. scabrinodis* and 68.9% of 995 genes in *M. vandeli* (Fig 3C and D). The species receiving the largest fraction of top hits was *Spiroplasma melliferum*, a honey bee pathogen closely allied with the plant pathogens *S. citri* and *S. kunkelii*. Genome sequencing facilitated a more thorough comparison of nucleotide identity between strains than the *ftsZ* locus alone – across 30 kb of syntenous coding and intergenic sequence the two share 95% identity.

Genes that are unique to these strains relative to other *Spiroplasma* taxa may hint toward the biological role of *Spiroplasma* in *Myrmica*. Hypothetical ORFs located on the same contig as a *Spiroplasma* gene were translated and queried by blastp and HMMER against the nr protein database and reference proteomes. Using a
conservative minimum of 600 nucleotides for ORF prediction, nine and eight candidates from *M. scabrinodis* and *M. vandeli*, respectively, were identified. All but one returned no significant similarity or domain conservation to known proteins. The exception encodes a nutrient transporter gene that is absent from all other sequenced *Spiroplasma* genomes: the substrate component of an energy-coupling factor (ECF) membrane transporter. No disruption in read coverage was evident between the ECF transporter and the *Spiroplasma* genes flanking it, it uses the *Mycoplasma/Spiroplasma* genetic code, and is also present in the transcriptome of *M. sulcinodis*, suggesting it is not an artifact of contaminating sequence reads in the assembly. PCR screens confirmed its presence in *M. vandeli* and both *M. scabrinodis* strains, but did not yield amplicons from *Myrmica* specimens from outside of the scabrinodis species group. Phylogenetic analysis alongside the most similar blastp matches and ECF transporter gene families of other *Spiroplasma* taxa placed the putative novel transporter on a long branch, distantly related to characterized families (Fig 4).

**Ribosome-inactivating proteins in *Myrmica Spiroplasma* symbionts**

Ribosome-inactivating protein (RIP) coding regions were identified in each of the *Spiroplasma* genomes and in the transcriptome of *M. rubra*. The RIP of *M. vandeli* is not predicted to encode a functional protein due to reading frame disruptions, while the RIPS of *M. scabrinodis* and *M. rubra* appear to encode intact ORFs with conserved active site residues, though the latter is only partially represented (~60% of the gene). Phylogenies of RIPS are not congruent with hosts (Fig 5), as was found in other *Spiroplasma* (35). The RIP of *M. scabrinodis* assembled into a 14 kb contig with an order of magnitude greater read coverage than that of *M. vandeli*, suggesting copy
number variation between the two (Fig 3A and B). Other genes encoded on this contig include those with strong amino acid similarity to proteins involved in type IV secretion systems used for conjugative DNA transfer and encoded on plasmids of S. citri and S. kunke Iii (36, 37), including soj, mob, and traE.

Presence of RIPs was confirmed by PCR for both M. vandeli and M. scabrinodis, while reactions targeting the M. rubra RIP failed to amplify from our Toronto, Vancouver, and Victoria samples. Detection of RIPs from individual colonies of scabrinodis group hosts varied by Spiroplasma strain. The RIP pseudogene was detected in workers from each of the two colonies of M. vandeli, while the intact RIP was detected in all of the colonies of M. scabrinodis bearing Spiroplasma haplotype B but none of the haplotype A colonies.

Discussion

In this study, Myrmica ants are shown to be a hotspot for Spiroplasma infection. Eight of nine species screened in the current study, as well as all three species with publicly available transcriptomes, harbor Spiroplasma. In addition, infection prevalence within species is high, with 42 of 43, and 250 of 276 colonies and individuals respectively, infected. No strains were shared between multiple ant species, while one species, M. ruginodis, hosted two different strains. Other broad surveys of symbionts in ants, using universal 16S ribosomal RNA primers, have also reported Spiroplasma infections from multiple species groups, including citri, ixodetis and platyhelix (Fig S1), in 27 of 95 species (38) and in 24 of 464 species (39). In this latter study, one half of infections were in the genus Polyrachis. Although these screens do not typically distinguish between inherited and horizontally transmitted Spiroplasma, or provide
information about prevalence within ant species, they suggest that *Spiroplasma* infections may not be uncommon in ants.

The perfect association between mitochondrial haplotype and *Spiroplasma* variant, which was found in our detailed screening of microsympatric colonies of *M. scabrinodis* and *M. vandeli*, is strong evidence for symbiont vertical transmission. Two mitochondrial haplotypes were found in *M. scabrinodis* (96% similar at COI), and these are perfectly associated with *Spiroplasma* haplotypes, as indicated by variation in the *ftsZ* gene (99% similar). Individuals (and colonies) with different mitochondrial haplotypes showed no differences in their nuclear genes, and it is not known how or why this mitochondrial polymorphism persists. This pattern of genetic variation is in contrast with *M. rubra*, where nuclear but *not* mitochondrial differences are associated with a queen reproductive polymorphism (40, 41). The close relative *M. vandeli* harbors a closely related *Spiroplasma* strain (95% similar to *M. scabrinodis* symbionts, 99% at *ftsZ* alone). It is interesting that these ants do not share or appear to exchange *Spiroplasma*, despite being found so close to each other, and is in contrast with other studies finding the exchange of *Wolbachia* between socially parasitic ants and those within the host colonies (42–44). Interestingly, one of these studies found that unlike *Wolbachia*, *Spiroplasma* strains were not exchanged between host and social parasite (44). Finally, the high infection prevalence in *Myrmica* larvae and pupae, and widespread tissue distribution, also suggest vertical transmission.

All *Spiroplasma* in the present study were from the citri clade. A 16S rRNA screen also found a strain from the citri clade infecting *Myrmica incompleta* (39). The *Myrmica* symbionts are not monophyletic; plant pathogens *S. kunkellii* and *S. citri*, bee
pathogenic *S. melliferum*, and symbionts of *Drosophila wheeleri*, *D. aldrichi*, and *D. mojavensis* are all nested within the group of *Myrmica Spiroplasma*. Although there is not strict cospeciation between *Myrmica* and their symbionts, there is strong phylogenetic signal, with three lineages of *Spiroplasma* each closely associated with a lineage of *Myrmica*, although much more detailed sampling and screening of *Myrmica* is required to determine how many independent acquisitions of *Spiroplasma* have occurred. This will be challenging, as *Myrmica* is a very diverse clade of ants whose taxonomy and evolutionary relationships are unresolved and sometimes controversial, with many cryptic species (40, 45–47). We know of no cases of cospeciation between *Spiroplasma* with their hosts, and in general cospeciation between facultative symbionts and their hosts is very rare (see ref (48) for a *Wolbachia* example). Most facultative symbionts are lost before their hosts speciate, and infect new hosts via horizontal transmission (30, 49, 50).

The patterns of association between *Spiroplasma* and *Myrmica* differ greatly from *Drosophila*, the best studied insect lineage that is commonly infected by inherited *Spiroplasma*. There, at least five lineages from four clades have colonized *Drosophila*, and hosts are often infected at low frequencies, largely depending on maternal transmission efficiency, as well as the fitness and phenotypic effects of the symbiont. For example, male-killing strains are found at low frequencies in host populations (51), whereas a strain that protects against a very common virulent nematode parasite occurs at high frequency (33).

Of course, the obvious next step is to determine what effects *Spiroplasma* might have on their *Myrmica* hosts. It is unlikely that these microbes are essential, as some
individuals (and one species) were uninfected. Obligate inherited symbionts of ants include *Blochmannia* that recycle nitrogen for their carpenter ant hosts (52); *Myrmica* ants, however, are primarily predaceous and not thought to feed on nutrient-limited diets (53). Also, no *Spiroplasma* are known to be obligate symbionts.

Perhaps *Spiroplasma* manipulates *Myrmica* reproduction, for example by killing males. It is challenging though to demonstrate sex ratio distortion in ants and other social Hymenoptera, as this would involve isolating symbiont-free queens, rearing colonies to produce reproductives, and comparing their sex ratios with those of infected colonies. As far as we are aware, only one study has shown a convincing link between symbionts and sex ratio distortion in ants (54). In that study, artificial selection on sex ratio in colonies of the pharaoh ant *Monomorium pharaonis* resulted in rapid changes in the frequency of *Wolbachia*.

Another possibility is that *Spiroplasma* persist in *Myrmica* by providing protection against natural enemies. *Myrmica* are commonly infected with parasites (55). In fact, new species of *Myrmica* have been erroneously described due to parasitic nematodes, because infected ants often look different from uninfected ones, with distended abdomens (56, 57). To explore the potential for protection, we sequenced *Spiroplasma* genomes and surveyed for ribosome-inactivating proteins (RIPs), toxins that are widespread and diverse in *Spiroplasma* and that have been implicated in defense against parasitic nematodes and wasps (28, 29). These toxins appear to evolve rapidly and exhibit elevated rates of gains and losses in *Spiroplasma*, making initial detection by PCR difficult, even with degenerate primers. However, our genome surveys uncovered two RIPs. One of these, in the *M. vandeli* symbiont, is a pseudogene, while
the other, in the *M. scabrinodis* symbiont, appears to lie on a plasmid that is present in
the strain associated with haplotype B, but not haplotype A. That the RIP toxins are
pseudogenized or found on plasmids suggests that their host associations are dynamic,
perhaps evolving in concert with changing pressures from natural enemies. In support
of this, we also uncovered a *Spiroplasma* RIP from the transcriptome of *M. rubra*
collected from its native range in Europe, but we could not detect it in N. American
colonies, suggesting that it also occurs on a plasmid and was lost when *M. rubra*
invaded N. America, perhaps due to enemy release. Of course, much work remains in
order to determine whether these RIPS might be protective and against what.

Genomes can also provide useful clues for understanding symbiont biology (1).
We searched the *Myrmica-Spiroplasma* metagenome for novel *Spiroplasma* genes, i.e.
genes that do not occur in any sequenced *Spiroplasma* genomes, of which 21 are
currently available, from four clades, including *S. melliferum*, *S. citri* and *S. kunkellii*
from the citri clade. All but one of the new genes were of unknown function, and there
were no new metabolic pathways uncovered, further suggesting that the symbiont does
not fill an obligate nutritional or metabolic role for its host. We identified a divergent ECF
S (substrate) component gene responsible for conferring substrate specificity to a
transport complex. In shared energy-coupling transport systems, an ATPase binding
cassette and transmembrane protein, the so-called AT module, is a universal
component of each transporter, with substrate specificity being conferred by S
component genes (58). Substrates include vitamins and transition metals, therefore, it is
possible that *Spiroplasma* symbionts of *Myrmica* are supplementing or siphoning
nutrients. Phylogenetic analysis can help to identify candidate protein functions among
members of specialized protein classes; however, this gene could not be conclusively attributed to a characterized substrate family, thus its role and importance in this symbiosis remain open questions.

Finding such prevalent inherited Spiroplasma in Myrmica opens up many interesting questions. The next step is to perform experiments comparing symbiont-infected and uninfected ants, following antibiotic treatment of lab colonies. A promising model would be the European fire ant, Myrmica rubra, one of the most invasive ant species globally. Interestingly, a little cited study from twenty years ago treated M. rubra lab colonies with antibiotics, and found an effect on ant growth and queen production (59).

Materials and Methods

Sample collection, DNA extraction, and Spiroplasma screening

Individuals from twenty-nine ant colonies from two sites that were 35 m apart (site 1: 46.7594 °N, 6.2527 °E, and site 2: 46.7595 °N, 6.2573 °E) in Réserve Naturelle du Lac de Remoray, France, were collected in August 2016. This area was already known to harbor a number of different Myrmica species. Species determinations were made by Mesut Koken, a local ant expert. The number of ants sampled from each colony ranged from 6-20; for one M. scabrinodis colony, seven larvae and ten pupae were also collected. Colony samples were stored separately in 95% ethanol. In addition, Myrmica samples were received from colleagues as whole ants in ethanol or as DNA extractions (see Acknowledgements and Table S1).

In preparation for DNA extraction, ants were removed from ethanol and air dried for 10 minutes. DNA was extracted from individual ants using the PrepMan Ultra
(Applied Biosystems) sample preparation reagent. To rule out the possibility that
Spiroplasma was an ectosymbiont harbored on the ant cuticle, all sampled individuals
from one M. scabrinodis colony were surface sterilized by submerging ant specimens in
2.5% bleach for two minutes, then submerging in 70% ethanol for four minutes, and
then rinsing in distilled water twice for three minutes. To test for systemic infection, the
head, thorax, gaster and legs of eight ants from one M. scabrinodis and one M. vandeli
colony were carefully dissected, and DNA was extracted and screened separately.
DNA extractions were screened for Spiroplasma by PCR using primers targeting
a 780 bp fragment of the single copy cell division protein gene ftsZ (F2: 5’
TGAACAGTCGCGTCAATAAA and R2: 5’ CCACCAGTAACATTAATAGCATCA
(30)). Some initial screens were also performed using primers targeting 300 bp of
Spiroplasma 16S rRNA (Spi16S-F: 5’ CCTGAGTAGTATGCTCGCAAGAG and Spi16S-
R: 5’ CCCACCTCTCCTCTAGCTTAC). Primers targeting the Myrmica host gene,
mitochondrially-encoded cytochrome oxidase subunit I (COI) were used as a positive
control for DNA quality and to sequence a region of Myrmica DNA for molecular
identification and analysis. Primer sequences are MyrmCOI F: 5’
TCGTTAGAATTAGATCTTGT and R: 5’ ATGAGAAATTAATCCAAATCCAG for
species in the scabrinodis group taxa, uMyrmicaCOI F: 5’
TAATTAATAATGAYCAAATTTATAATAC and R: 5’
GTRGRATTGCATTATATAGTTGC for all other Myrmica taxa, as well as
LCO1490: 5’ TAAACTCTGAGTGACCACAAAAATCA and HCO2198: 5’
GGTCAACAAATCATAAAGATATTTGG (60) for Formica. Primers targeting 500 bp of a
variable portion of the nuclear encoded long wavelength rhodopsin gene (LR143F: 5’
CACTGGTATCARTCGCACC32SAT LR672R and LR672R: 5'
CCRCAMGCWGTCATGTRCCTTC (47) were used to confirm that divergent M.
scabrinodis mitochondrial haplotypes corresponded to the same species. RIP primers
were scabRIP F: 5' GAGGAACATCAATTGAGTTCTTCT and R: 5'
AATCTTCATCTTTGACTTACCAC and vanRIP F: 5'
TCCTGGTTAGATAACTTTCTGTC and R: 5' ATTATTGAGTTGAGGTATCGC.
ECF transporter primers were Sp-ECF F: 5' CTAGCAGCTGTAATGTTAGCATTAAC
and R: 5' CTAATTCCACAGGCATAAAATAGTAG. Thermal cycling programs for ftsZ,
MyrmCOI, Spi16S, HCO/LCO, LR, RIPS, and ECFs were 35 cycles of 95° C for :30, 54°
C for :30 72° C for 1:15 and for uMyrmicaCOI 35 cycles of 95° C for :30, 49° C for :30
72° C for 1:15. PCR products were assessed by DNA gel electrophoresis on a 1%
agarose gel stained with ethidium bromide and visualized under UV light. One or more
ftsZ and COI amplicons per Myrmica species was sequenced with the Sanger method
by Sequetech (California, USA) using forward and reverse primers. At least one ftsZ
amplicon for each host taxon in our sample set was sequenced. Intrahost symbiont
diversity was examined by sequencing ftsZ and COI from all individuals from one M.
vandeli and two M. scabrinodis colonies (n=10, 9, 8 individuals, respectively).

Sequence processing and phylogenetic analysis
Primer regions were trimmed from the Sanger sequences by hand, yielding a
final product of 648 bases for ftsZ and 690 bases for COI. ftsZ sequences were used to
query nucleotide sequences deposited in the NCBI transcriptome shotgun assembly
(TSA) database. For Myrmica taxa with TSA hits to Spiroplasma ftsZ sequences, COI
sequences were also recovered from the transcriptome by blastn. Nucleotide and amino
acid sequences were aligned with MAFFT 7.309 (61). For each nucleotide alignment, the best substitution model for phylogenetic analysis was determined in jModelTest 2.1.7 (62). For ftsZ, the best model was HKY+I+G and for COI it was TPM2uf+G. RIP and ECF transporter phylogenies were built from amino acid alignments and used the LG substitution model. Maximum likelihood phylograms were constructed using PhyML 2.2.0 (63) implemented in Geneious R10. SH-like approximate likelihood ratio test scores were calculated for each branch. Alignments and phylogenies used to test for evidence of cophylogeny were generated as described above. Tests of codivergence were carried out with ParaFit (64) implemented in the R package ape (65). Briefly, patristic distance matrices were calculated for the host COI and symbiont ftsZ loci using the same models of nucleotide substitution as above. Included in these alignments were all the hosts that were positive for Spiroplasma by our PCR screens, ten distinct host mitochondrial lineages and their symbionts, plus M. ruginodis host and Spiroplasma sequences from the public transcriptome. Distance matrices were permuted randomly to create a distribution of data against which to test the null hypothesis of independent host and symbiont evolutionary histories. Matrices were permuted 999 times. Because infection frequency in M. ruginodis could not be determined nor could read contamination in the sequence read archive be ruled out, the codivergence analysis was carried out once with and once without this host. The tanglegram depicted in Fig 2 was visualized with TreeMap 3 (66).

**Spiroplasma genome DNA extractions, sequencing, and assembly**

To prepare samples for Illumina sequencing, genomic DNA was extracted from pools of five ants for each of two species in the *scabrinodis* clade, *Myrmica scabrinodis*
and *M. vandeli*, using the phenol-chloroform method. Short insert shotgun libraries were prepared and 125 bp paired-end reads were sequenced by Genome Québec (Montréal, Québec, Canada) on a HiSeq 2500 v4 system.

Preliminary metagenomes were assembled for each of the two *Myrmica* host species, *M. vandeli* and *M. scabrinodis* after filtering out reads with GC content greater than 31% and their pairs. The low GC data set was mapped to the mitochondrial genome sequence of *Myrmica scabrinodis* (NCBI Reference Sequence NC_026133) and removed. Reads were trimmed, filtered, and mapped with BBMap 37.36 (by Brian Bushnell, sourceforge.net/projects/bbmap/). Filtered reads were assembled de novo using SPAdes 3.10.1 (67). Open reading frames with a minimum length of 300 nucleotides were predicted and translated with Geneious R10 and compared against the nr protein database available on NCBI using blastp. All contigs encoding *Spiroplasma* genes, i.e. sequences for which the top blastp hit was to *Spiroplasma*, were retained as a preliminary genome assembly. *Spiroplasma* genes with higher GC content, though rare, were absent from these assemblies. To produce a more complete assembly for each symbiont strain, a second iteration of mapping and assembly was carried out. Using the original read sets, reads were quality trimmed, and GC filtered at 45%. All remaining reads that mapped to contigs in the preliminary assemblies, as well as to the sequenced genome of *Spiroplasma citri* and its plasmids were again assembled de novo using SPAdes. The blastp search was repeated to identify *Myrmica* and bacterial contigs that had not been removed through the mapping procedure. A collection of genes with top hits to either non-*Spiroplasma* or unclassified Entomoplasmatales taxa, 56 in *M. vandeli* and 53 in *M. scabrinodis*, were interpreted as
Spiroplasma genes nonetheless if they were encoded on Spiroplasma contigs and showed strong blastp hits to Spiroplasma taxa as well.

To identify genes unique to the Myrmica Spiroplasma, proteins with top blastp matches to Spiroplasma were annotated back onto the assembly contigs and compared against all predicted ORFs on each contig, i.e. most ORFs had two annotations, one as a predicted protein coding gene and one as a Spiroplasma blastp match. ORFs with only the former annotation were investigated individually with blastp, blastn, and HMMER to identify putative functions. RIPs were identified by tblastn to the final assemblies and to the three Myrmica transcriptomes and sequence read archives available as part of NCBI BioProject PRJDB4088.

Acknowledgments

We thank Megan Frederickson, Rob Higgins, Danielle Hoefele, Caroline Cameron, Corrie Moreau, and Jake Russell for donating Myrmica specimens, and Jocelyn Claude and Mesut Koken for assisting in collecting and identifying specimens. We also thank Ryan Gawryluk and Cuong Le for technical advice. This work was funded by a Sinergia grant from the Swiss National Science Foundation awarded to SJP and a Jamie Cassels Undergraduate Research Award from the University of Victoria to LDM.

Data Availability

Genomic DNA sequence reads and PCR amplicon sequences generated during this study (68) have been submitted to GenBank under BioProject PRJNA419549 and accession numbers MG558353-MG558456.

References


9. Hurst GDD, Jiggins FM, Hinrich Graf von der Schulenburg J, Bertrand D, West


Russell JA, Funaro CF, Giraldo YM, Goldman-Huertas B, Suh D, Kronauer DJC, Moreau CS, Pierce NE. 2012. A Veritable Menagerie of Heritable Bacteria from Ants, Butterflies, and Beyond: Broad Molecular Surveys and a Systematic


55. Witek M, Barbero F, Markó B. 2014. Myrmica ants host highly diverse parasitic
communities: from social parasites to microbes. Insectes Soc.


lentiviruses can account for phylogenetic similarity with the primate phylogeny.

67. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM,
Nikolenko SI, Pham S, Prjibelski AD, Pyshkin A V., Sirotkin A V., Vyahhi N, Tesler
G, Alekseyev MA, Pevzner PA. 2012. SPAdes: A New Genome Assembly
477.

68. Ballinger MJ, Moore LD, Perlman SJ. 2017. Metagenomic sequence reads of
Myrmica scabrinodis and Myrmica vandeli.
Table 1. *Myrmica* species or species groups screened for *Spiroplasma*

<table>
<thead>
<tr>
<th><em>Myrmica</em> species</th>
<th>Screened</th>
<th>Positive</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>individuals/colonies</td>
<td>individuals/colonies</td>
<td>%</td>
</tr>
<tr>
<td><em>alaskensis</em></td>
<td>16/1</td>
<td>16/1</td>
<td>100.0</td>
</tr>
<tr>
<td><em>fracticornis</em></td>
<td>8/1</td>
<td>0/0</td>
<td>0.0</td>
</tr>
<tr>
<td><em>rubra</em></td>
<td>68/12</td>
<td>58/12</td>
<td>85.2</td>
</tr>
<tr>
<td><em>scabrinodis</em> mtDNA type A</td>
<td>37/6</td>
<td>34/6</td>
<td>91.9</td>
</tr>
<tr>
<td><em>scabrinodis</em> mtDNA type B</td>
<td>126/16</td>
<td>123/16</td>
<td>97.6</td>
</tr>
<tr>
<td><em>sp.</em> 1 (group A; CSM0598 W1)</td>
<td>1/1</td>
<td>1/1</td>
<td>100.0</td>
</tr>
<tr>
<td><em>sp.</em> 2 (group A; CSM0202, CSM0296)</td>
<td>4/2</td>
<td>4/2</td>
<td>100.0</td>
</tr>
<tr>
<td><em>sp.</em> 3 (group B; CSM1794c)</td>
<td>1/1</td>
<td>1/1</td>
<td>100.0</td>
</tr>
<tr>
<td><em>sp.</em> 4 (group B; CSM1798)</td>
<td>2/1</td>
<td>1/1</td>
<td>50.0</td>
</tr>
<tr>
<td><em>vandeli</em></td>
<td>13/2</td>
<td>12/2</td>
<td>92.3</td>
</tr>
</tbody>
</table>

Table 2. Genome assembly statistics of novel *Spiroplasma* genomes

<table>
<thead>
<tr>
<th></th>
<th><em>Myrmica scabrinodis</em></th>
<th><em>Myrmica vandeli</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>Preliminary metagenomes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Contigs (&gt;300 nt)</td>
<td>113,714</td>
<td>107,962</td>
</tr>
<tr>
<td>N50</td>
<td>593</td>
<td>605</td>
</tr>
<tr>
<td>Wolbachia genes</td>
<td>338</td>
<td>574</td>
</tr>
<tr>
<td>Final <em>Spiroplasma</em> genomes</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Spiroplasma</em> contigs (&gt;300 nt)</td>
<td>481</td>
<td>402</td>
</tr>
<tr>
<td><em>Spiroplasma</em> N50</td>
<td>3477</td>
<td>4471</td>
</tr>
<tr>
<td><em>Spiroplasma</em> nucleotides</td>
<td>1,150,673</td>
<td>1,206,483</td>
</tr>
<tr>
<td><em>Spiroplasma</em> ORFs (&gt; 300 nt)</td>
<td>1,019</td>
<td>995</td>
</tr>
<tr>
<td>with matches to:</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>S. citri</em> genome</td>
<td>957 (93.9%)</td>
<td>916 (92.1%)</td>
</tr>
<tr>
<td><em>S. kunkelli</em> genome</td>
<td>928 (91.1%)</td>
<td>891 (89.5%)</td>
</tr>
<tr>
<td><em>S. melliferum</em> genome</td>
<td>895 (87.8%)</td>
<td>881 (88.5%)</td>
</tr>
<tr>
<td><em>S. poulsonii</em> genome</td>
<td>936 (91.8%)</td>
<td>907 (91.2%)</td>
</tr>
</tbody>
</table>

Figure 1. Diverse *Myrmica* ants harbor *Spiroplasma*

Maximum-likelihood phylogram of mitochondrially-encoded cytochrome oxidase subunit I (COI) nucleotide sequences of ants in the genus *Myrmica*. Gold type indicates taxa.
screened for *Spiroplasma*. Species groups are designated by alternately shaded boxes and labeled at the right side of the phylogram. Bold type marks a species group containing one or more screened taxa. Branches are labeled with SH-like approximate likelihood ratio test scores greater than 0.75.

**Figure 2.** *Spiroplasma* in *Myrmica* belong to the citri clade

(A) Maximum-likelihood phylogram of cell division protein *ftsZ* nucleotide sequences from *Spiroplasma* bacteria. Gold type indicates a taxon amplified from *Myrmica* ants in this study. Branches are labeled with SH-like approximate likelihood ratio test scores of 0.65 or higher. (B) Tanglegram depicting the pattern of cophylogeny between host and symbiont gene trees. Horizontal black lines between trees connect host taxa to symbiont taxa. Codivergence is not statistically significant as tested by ParaFit (p = .07).

**Figure 3.** Genome coverage and *Spiroplasma* gene assignment of *Myrmica* symbionts

Read coverage distribution graphed in a 0-50 range and 0-300 range for (A) *M. scabrinodis* and (B) *M. vandeli* *Spiroplasma* symbionts. Read coverage of the RIP-encoding contig for each symbiont is indicated with an arrow. Pie charts summarize gene assignments within each symbiont’s genome. The majority fraction of *Spiroplasma* genes in (C) *M. scabrinodis* and (D) *M. vandeli* *Spiroplasma* symbionts match best by blastp to the honeybee pathogen, *Spiroplasma melliferum*, and overall, most genes match to members of the citri clade: *Spiroplasma citri*, *S. kunkelii*, *S. melliferum*, and *S. poulsonii*. In the legend, right, each taxon label is accompanied by the number of top blastp matches from *M. scabrinodis* and *M. vandeli* symbionts, respectively.
Figure 4. A divergent ECF transporter in the genomes of *Myrmica Spiroplasma* symbionts

Maximum-likelihood phylogram of energy-coupling factor (ECF) transporter substrate component amino acid sequences. Tips are labeled with taxonomic identifiers and clades with protein family information, if available. Gold type indicates the novel ECF transporter of *Myrmica*-associated *Spiroplasma* symbionts, and red type indicates other ECF transporters identified in the genomes of these symbionts. Branches are labeled with SH-like approximate likelihood ratio test scores of 0.75 or higher.

Figure 5. Diversity of ribosome-inactivating proteins in *Myrmica Spiroplasma*

Maximum-likelihood phylogram of *Spiroplasma*-encoded ribosome inactivating protein (RIP) amino acid sequences. Tips are labeled with *Spiroplasma* species names or references to the host species harboring a *Spiroplasma* symbiont. Gold type indicates a RIP sequence identified from a *Myrmica*-associated *Spiroplasma* symbiont.
A

Mycoplasma mycoides
  Spiroplasma litorale
  Spiroplasma turonicum
  Spiroplasma cantharicola
  Spiroplasma diminutum
  Spiroplasma culicicola
  Spiroplasma apis
  Spiroplasma helicoides
  Spiroplasma sabauniense

1

Spiroplasma chrysoplicola
  Spiroplasma syphidicola
  Spiroplasma of Glossina fuscipes
  Spiroplasma of Glossina tachinoides
  Spiroplasma of Drosophila simulans
  Spiroplasma of Drosophila melanogaster
  Spiroplasma of Drosophila neotestaceae
  Spiroplasma of Drosophila hydei
  Spiroplasma eriocheiris
  Spiroplasma atrichopogonis

1

Spiroplasma of Myrmica sp. CSM1798
  Spiroplasma of Myrmica sp. CSM1794
  Spiroplasma of Myrmica sp. CSM0598
  Spiroplasma of Myrmica sp. CSM0298
  Spiroplasma of M. alaskensis
  Spiroplasma of M. rubra
  Spiroplasma of M. ruginodis TSA
  Spiroplasma of Drosophila aldrichi
  Spiroplasma of Drosophila wheeleri
  Spiroplasma of Drosophila mojavensis
  Spiroplasma kunkelii
  Spiroplasma citri
  Spiroplasma melliferum
  Spiroplasma of M. vandelii
  Spiroplasma of M. scabrinodis A
  Spiroplasma of M. scabrinodis B

0.2

B

Myrmica COI

Spiroplasma ftsZ

CSI0202
CSI0296
CSI0598 W1
CSI1794c
CSI1798
vandelii
scabrinodis A
alaskensis
ruginodis
rubra

CS1798c
CSI0202
CSI0296
CSI0598 W1
CSI1794c
CSI1798
vandelii
scabrinodis A
alaskensis
ruginodis
rubra
Spiroplasma of Myrmica scabrinodis

A

Count

0 10 20 30 40 50

Coverage

RIP

B

Count

0 10 20 30 40

Coverage

RIP

C

Top blastp hit
- *S. melliferum* (386, 350)
- *S. citri* (151, 145)
- *Spiroplasma* sp. (106, 98)
- *S. kunkelii* (101, 106)
- *S. poulsonii* (83, 85)
- mirum clade (56, 65)
- chrysricola clade (40, 46)
- apis clade (39, 38)
- *Entomoplasmales* (34, 34)
- *Mycoplasma* (9, 11, 9)
- *Entomoplasma* (5, 7)
- *Mesoplasma* (5, 4)
- *Spiroplasma* phage (3, 6)
- other citri clade (1, 0)

1019 proteins

D

995 proteins
Fig S1. Diversity of Spiroplasma in Myrmica and other ants
A maximum-likelihood phylogram of Spiroplasma 16S ribosomal DNA constructed from an alignment of approximately 400 nucleotide positions. Spiroplasma from Myrmica are shown in gold, those detected from other ant genera are shown in brown, and those of Drosophila are shown in dark green. Branches are labeled with SH-like approximate likelihood ratio test scores of 0.75 or greater.