Rapid molecular evolution of *Spiroplasma* symbionts of *Drosophila*

Michael Gerth[1*], Humberto Martinez-Montoya[2], Paulino Ramirez[3], Florent Masson[4] , Joanne S. Griffin[1], Rodolfo Aramayo[5], Stefanos Siozios[1], Bruno Lemaitre[4], Mariana Mateos[6] & Gregory D.D. Hurst[1]

[1] Institute of Infection, Veterinary and Ecological Sciences, University of Liverpool, Liverpool, UK

[2] Laboratory of Genetics and Comparative Genomics, Unidad Académica Multidisciplinaria Reynosa Aztlán, Universidad Autónoma de Tamaulipas, Reynosa, Mexico

[3] Department of Cell Systems and Anatomy, University of Texas Health San Antonio, San Antonio, Texas, USA

[4] Global Health Institute, School of Life Sciences, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland

[5] Department of Biology, Texas A&M University, College Station, Texas, USA

[6] Department of Ecology and Conservation Biology, Texas A&M University, College Station, Texas, USA

[*]Author for correspondence.

Current address: Department of Biological and Medical Sciences, Oxford Brookes University, Oxford, UK mgerth@brookes.ac.uk

Abstract

Spiroplasma are a group of Mollicutes whose members include plant pathogens, insect pathogens, and endosymbionts of animals. *Spiroplasma* phenotypes have been repeatedly observed to be spontaneously lost in *Drosophila* cultures, and several studies have documented a high genomic turnover in *Spiroplasma* symbionts and plant pathogens. These observations suggest that *Spiroplasma* evolves quickly in comparison to other insect symbionts. Here, we systematically assess evolutionary rates and patterns of *Spiroplasma poulsonii*, a natural symbiont of *Drosophila*. We analysed genomic evolution of *s*Hy within flies, and *s*Mel within *in vitro* culture over several years. We observed that *S. poulsonii* substitution rates are among the highest reported for any bacteria, and around two orders of magnitude higher compared with other inherited arthropod endosymbionts. The absence of mismatch repair loci *mutS* and *mutL* is conserved across *Spiroplasma* and likely contributes to elevated substitution rates. Further, the closely related strains *s*Mel and *s*Hy (>99.5% sequence identity in shared loci) show extensive structural genomic differences, which potentially indicates a higher degree of host adaptation in *s*Hy, a protective symbiont of *Drosophila hydei*. Finally, comparison across diverse *Spiroplasma* lineages confirms previous reports of dynamic evolution of toxins, and identifies loci similar to the male-killing toxin Spaid in several *Spiroplasma* lineages and other endosymbionts. Overall, our results highlight the peculiar nature of *Spiroplasma* genome evolution, which may explain unusual features of its evolutionary ecology.

41 **Introduction**

42 Many bacterial lineages have evolved to become associates of animal hosts (McFall-Ngai et
43 al., 2013). In arthropods, such associations are the rule, and maternally inherited,
44 endosymbiotic bacteria are especially common and diverse (Duron et al., 2008; Weinert et al.,
45 2015). Prominent examples include obligate nutritional symbionts of blood and sap feeders
46 (Douglas, 2009), reproductive manipulators (Drew et al., 2019), and protective symbionts
47 (Oliver et al., 2014). Collectively, inherited symbionts of arthropods are highly diverse with
48 respect to potential benefits and costs induced, and their degree of adaptation to hosts. For
49 example, even within a single lineage of inherited symbionts - *Wolbachia* - there are
50 nutritional mutualists (Hosokawa et al., 2010), protective symbionts (Teixeira et al., 2008),
51 and reproductive manipulators (Dyson et al., 2002). Importantly, these symbiont-conferred
52 traits may be modulated by environmental factors (Corbin et al., 2017) and host genetic
53 background (Hornett et al., 2008).

54 *Spiroplasma* are small, helical bacteria that lack cell-walls. Like other Mollicutes
55 (Mycoplasma-like organisms), they are exclusively found in host association and are often
56 pathogenic (Razin, 2006). Among the first discovered spiroplasmas were plant pathogens
57 (reviewed in Saglio and Whitcomb, 1979) and an inherited symbiont of *Drosophila* with the
58 ability to shift sex ratios towards females (Poulson and Sakaguchi, 1961). Subsequently,
59 *Spiroplasma* symbionts were found in a number of different *Drosophila* species (Haselkorn,
60 2010), as well as many other arthropods (Gasparich, 2002). More recently, molecular
61 evidence was brought forward for *Spiroplasma* symbionts in non-arthropod animals
62 (Duperron et al., 2013; He et al., 2018). In insects, two phenotypes induced by inherited
63 *Spiroplasma* have been studied in detail (reviewed in Anbutsu and Fukatsu, 2011; Ballinger

64 and Perlman, 2019). Firstly, *Spiroplasma* can enhance survival rates of its hosts under
65 parasite or parasitoid attack. In *Drosophila neotestacea*, *Spiroplasma* inhibits growth and
66 reproduction of the parasitic nematode *Howardula*, and thus reverses nematode-induced
67 sterility (Jaenike et al., 2010). Further, *Spiroplasma* symbionts may greatly enhance
68 *Drosophila* survival when attacked by parasitoids (Xie et al., 2010, 2014). Both protective
69 phenotypes are mediated by *Spiroplasma* encoded RIP toxins (ribosome-inactivating
70 proteins), which target the attackers' ribosomes (Ballinger and Perlman, 2017; Hamilton et
71 al., 2014, 2015). These processes may be complementary to protection through competition
72 between *Spiroplasma* and parasitoids for lipids in the host hemolymph (Paredes et al., 2016).
73 Secondly, *Spiroplasma* kills males early in development in several species of *Drosophila*,
74 planthoppers (Sanada-Morimura et al., 2013), ladybird beetles (Tinsley and Majerus, 2006),
75 lacewings (Hayashi et al., 2016), and butterflies (Jiggins et al., 2000). The male-killing
76 phenotype is also linked to a toxin: in *Drosophila melanogaster*, a *Spiroplasma* protein
77 containing OTU (Ovarian Tumour-like deubiquitinase) and ankyrin domains was
78 demonstrated to kill male embryos, and thus termed *Spiroplasma poulsonii* androcidin
79 ("Spaid", Harumoto and Lemaitre, 2018).

80 *Spiroplasma* induced phenotypes were commonly observed to be dynamic compared with
81 other symbiont traits, with repeated observation of phenotype change within laboratory
82 culture over relatively short time frames. For example, spontaneous loss of male killing was
83 found in *Spiroplasma* symbionts of *Drosophila nebulosa* (Yamada et al., 1982), and
84 *Drosophila willistoni* (Ebbert, 1991), and spontaneous emergence of non male-killing
85 *Spiroplasma* in *Drosophila melanogaster* has occurred at least twice in a single culture
86 (Harumoto and Lemaitre, 2018; Masson et al., 2020). In addition, *Spiroplasma* symbionts
87 artificially transferred from their native host *Drosophila hydei* into *D. melanogaster* initially

88 caused pathogenesis, but evolved to become benign over just a few host generations
89 (Nakayama et al., 2015). Genomic analysis has further documented dynamic *Spiroplasma*
90 evolution driven by viral proliferation (Carle et al., 2010; Ku et al., 2013; Ye et al., 1996) and
91 by extensive transfer of genetic material between plasmids, and from plasmids to
92 chromosomes (Joshi et al., 2005; Mouches et al., 1984). Notably, plasmids in *Spiroplasma*
93 commonly encode the systems that establish their phenotypic effect on their host (Berho et
94 al., 2006). Examples include Spaid (Harumoto and Lemaitre, 2018) and RIP genes (Ballinger
95 and Perlman, 2017), and rapid plasmid evolution may therefore contribute to the high
96 phenotypic evolvability of *Spiroplasma*. Furthermore, elevated rates of evolution were found
97 in various *Mycoplasma* species (Delaney et al., 2012; Woese et al., 1984), which are
98 pathogens closely related to *Spiroplasma* (Gupta et al., 2019).

99 Bacteria adapting to hosts often follow similar evolutionary trajectories, and both increased
100 mutational rates and proliferation of mobile genetic elements are commonly observed in
101 diverse symbiotic taxa (McCutcheon and Moran, 2011). Therefore, spontaneous loss of
102 *Spiroplasma* phenotypes and high genomic turnover could be explained by evolutionary
103 mechanisms common to all symbiotic taxa. Although independent findings discussed above
104 suggest that *Spiroplasma* symbionts may evolve quickly, this has never been quantified, and
105 it has remained unclear how substitution rates compare to those of other symbionts. However,
106 rates and patterns of evolutionary change may have important implications for *Spiroplasma*
107 evolutionary ecology, and for its potential use in biological control programmes (Clark and
108 Whitcomb, 1984; Schneider et al., 2019).

109 In this study, we systematically investigate rates and patterns of *Spiroplasma* evolution. To
110 this end, we employed two strains of *Spiroplasma poulsonii*: 1) *s*Hy is a natural associate of

111 *Drosophila hydei* (Ota et al., 1979) and confers protection against parasitic wasps (Xie et al.,
112 2010). We monitored its evolution over ~10 years in fly culture, and determined changes
113 through denovo reference genome sequencing and re-sequencing of multiple isolates. 2) *s*Mel
114 occurs naturally in *Drosophila melanogaster*, where it kills male offspring (Montenegro et al.,
115 2005) and protects from parasitoids (Xie et al., 2014). Using the previously established
116 complete genome (Paredes et al., 2015) and cell-free culture (Masson et al., 2018), we re-
117 sequenced isolates spanning ~2.5 years of evolution. This approach enabled tracing
118 *Spiroplasma* evolution on various levels: within-strain (<10 years divergence), between
119 strains, and through comparison with other *Spiroplasma* genomes, across a whole clade of
120 arthropod symbionts. It further allowed comparison between symbionts evolving in a host and
121 in axenic culture. We observed that rates of molecular evolution and chromosomal
122 rearrangements in *Spiroplasma poulsonii* are substantially faster than in other inherited
123 bacteria, and that this has resulted in markedly different genomic organisation in *s*Hy and
124 *s*Mel. Our work thus highlights *Spiroplasma poulsonii* as an amenable, highly evolvable, but
125 potentially unpredictable model insect symbiont.

126

127 **Materials and Methods**

128 *Spiroplasma* strains and genome sequencing

129 We documented the evolution of the *Spiroplasma poulsonii* strains *s*Hy and *s*Mel in fly hosts
130 and in cell free culture, respectively (Tab. 1). sHy is one of the *Spiroplasma* symbionts
131 naturally associated with *Drosophila hydei* (Mateos et al., 2006). Our evolution experiments
132 covered ca. ten years and three variants of *s*Hy, all of which all originated from a single
133 isolate (Fig. 1). Over the course of the study, *Drosophila hydei* hosts were maintained on

134 standard corn meal agar diet (1% agarose, 8.5% sugar, 6% maize meal, 2% autolysed yeast,
135 0.25% nipagin), at 25°C, and a 12h:12h light:dark cycle. To isolate *Spiroplasma s*Hy-Liv
136 DNA from *D. hydei*, we followed a protocol by Paredes et al. (2015): We collected ca. 300 *D.*
137 *hydei* virgins and aged them for 4 weeks. The flies were then anesthetized with CO_2 and
138 pricked into the anterior sternopleurum using a sterile needle. Hemolymph was extracted via
139 centrifugation (10,000g, 5 minutes, at 4°C) through a 0.45µm filter (Corning Costar Spin-X).
140 To avoid hemolymph coagulation, we only extracted batches of 20 flies at a time and
141 immediately transferred extracted hemolymph into ice cold phosphate buffered saline
142 solution (1X PBS buffer; 137 mmol NaCl, 2.7 mmol KCl, 10 mmol Na2HPO4, 1.8 mmol
143 KH2PO4). Extracts of all batches were combined, and the PBS solution with *Spiroplasma*
144 cells was centrifuged (12,000g, 5 minutes, at 4°C). Supernatant PBS was discarded and the
145 cell pellet was subjected to DNA extraction using a phenol-chloroform protocol and
146 subsequent ethanol precipitation. DNA was eluted in water, and sequenced at the Centre for
147 Genomics Research at University of Liverpool. Libraries were prepared using the Nextera XT
148 kit and sequenced as 2x250bp run on an Illumina MiSeq sequencer.

149 DNA extraction for the *Spiroplasma* strain *s*Hy-Tx followed the protocol described above,
150 with the following modifications: Immediately after the piercing, 35-40 flies were placed into
151 0.5 ml microcentrifuge tubes previously pierced in the bottom, which was placed within a 1.5
152 ml microcentrifuge tube containing 20µl PBS, and centrifuged at 4,500g for 10 seconds to
153 collect hemolymph. DNA was recovered using a chloroform-ethanol procedure
154 (Supplementary protocol 1) and diluted in AE buffer (Qiagen). A pair-ended library was
155 constructed by Eureka Genomics (Hercules, CA). DNA was fragmented, end repaired, A'
156 tagged, ligated to adaptors, size-selected, and enriched with 25 cycles of PCR during which
157 an index was incorporated to the sample. Sample preparation was performed according to

158 Illumina's Multiplexing Sample Preparation Guide and Eureka Genomics' proprietary
159 method. The resulting library was subjected to Illumina HiSeq 2500 sequencing at the Texas
160 AgriLife Genomics and Bioinformatics Services Facility (College Station, TX).

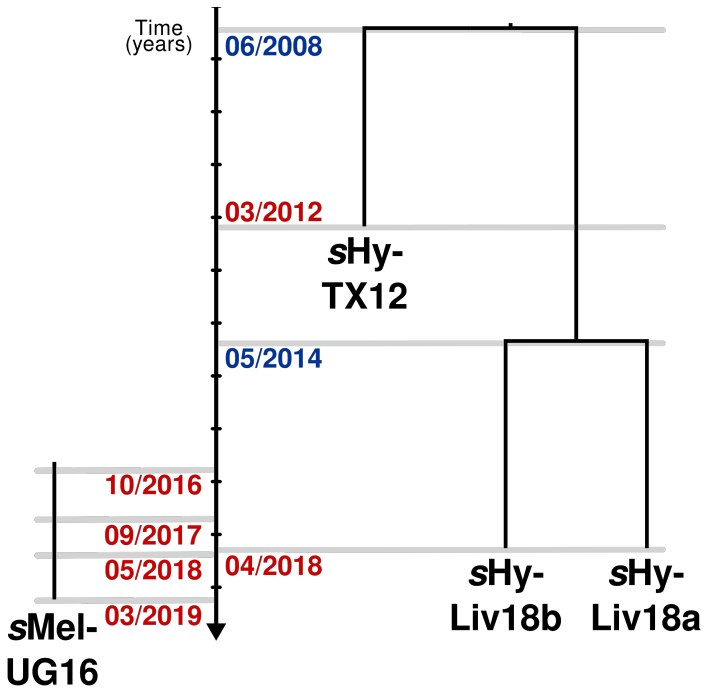
161 *s*Mel is a naturally occurring *Spiroplasma* symbiont of *D. melanogaster* best known for its
162 male-killing phenotype (Harumoto and Lemaitre, 2018; Montenegro et al., 2000). Cell free
163 culture of *s*Mel-Ug was recently accomplished (Masson et al., 2018), and we used a time
164 series of this culture to investigate *s*Mel molecular rates of evolution outside of host tissues.
165 The time series covered 29 months in total, and sequencing was performed at four time points
166 (Fig. 1). Illumina DNA sequencing was performed at MicrobesNG (Birmingham, UK) using
167 the "standard service". Culturing, passaging, and DNA extraction was performed as described
168 in Masson et al. (2018).

169 **Tab. 1:** *Spiroplasma poulsonii* strains investigated in this study. Note that we follow the
170 naming scheme suggested by Ballinger & Perlman (2017).

Name used in the current study		Natural host	Establishment in fly or cell-free culture	DNA extracted from, date	Also known as
***s*Hy**	*s*Hy-Tx12	*Drosophila hydei*	01/2004 (Mateos et al., 2006)	Hemolymph extract, 03/2012	Hap_1, TEN104-106
	*s*Hy-Liv18a	*Drosophila hydei*	06/2008 (copy of *s*Hy-Tx12)	Hemolymph extract, 04/2018	-
	*s*Hy-Liv18b	*Drosophila hydei*	03/2014 (copy of *s*Hy-Liv18A)	Hemolymph extract, 04/2018	-
***s*Mel**	*s*Mel-Ug16	*Drosophila melanogaster*	10/2016 (Masson et al., 2018)	Cell free culture, 4 time points 2016–18	MSRO-Uganda
	*s*Mel-Br18	*Drosophila melanogaster*	07/1997 (Montenegro et al., 2000)	Whole fly, 03/2018	MSRO-RED42

171

172



173

174 **Fig. 1:** Overview on *Spiroplasma poulsonii* strains investigated in this study. Three
175 *s*Hy strains were sampled from *Drosophila hydei* cultures over a total timespan of ten
176 years. Blue dates denote splitting of *D. hydei* cultures, red dates mark *Spiroplasma*
177 extraction for sequencing. For *s*Mel, four points of a time series of axenic culture were
178 sequenced (HL - start of culture, P14 - after 14 passages, P57 - after 57 passages,
179 P79 - after 79 passages). Here, red dates correspond to the date of isolation for
180 sequencing.

181

182 We further sequenced a single isolate of sMel-Br deriving from a strain established in 1997
183 (Tab. 1). For Illumina sequencing, 1-2 grams of whole body flies from *Spiroplasma*-positive
184 *D. melanogaster* were collected for CTAB-Phenol based DNA extraction. Illumina libraries
185 were prepared with Illumina Truseq DNA kit and sequenced on an Illumina NovaSeq-6000
186 S2 150 paired end flow cell by Texas AgriLife Genomics and Bioinformatics Services
187 Facility (College Station, TX). For Oxford Nanopore MinIon sequencing, heads were first
188 removed by immersing whole flies in liquid Nitrogen, and vigorous shaking on a metal sieve.
189 The headless specimens were immediately subjected to phenol-chloroform extraction.
190 Nanopore library preparation was performed using the Ligation Sequencing kit 1D (SQK-

191 LSK108), following precautionary measures to ensure long read lengths. The library was then
192 sequenced on MinION sequencing device (MIN-101B) with an SpotON Flow Cell Mark 1
193 (R9.5- FLO-MIN107.1). Base Calling was performed using Albacore version 2.0 (Oxford
194 Nanopore, Oxford, UK) and the reads were processed with Porechop version 0.24 (Wick et
195 al., 2017a) to remove adapter sequences. A draft assembly of *s*Mel-Br was created using
196 Illumina and Nanopore reads and Unicycler version 0.4.7 (Wick et al., 2017b). The assembly
197 was fragmented, but did contain 3 circular contigs with sequence similarity to other
198 *Spiroplasma* contigs.

199 <u>*s*Hy Reference genome sequencing and annotation</u>

200 In order to reconstruct a high quality reference genome for *s*Hy, we extracted DNA from 30
201 *Drosophila hydei* virgin females carrying the symbiont. To limit the impact of potentially
202 heterogeneous *Spiroplasma* populations within the flies, we used F3 individuals from a single
203 isofemale line for DNA extraction. The flies were anesthetized at -20°C, washed with 70%
204 ethanol, and briefly dried. The flies were then homogenized in G2 lysis buffer from the
205 Qiagen Genomic DNA Buffer Set using a 1ml dounce tissue grinder (DWK Life Science),
206 and DNA was extracted from the homogenate using a Qiagen Genomic-Tip 20/G following
207 the protocol modifications from Miller et al. (2018). A total of 4µg DNA was then used in the
208 "1D gDNA long reads without BluePippin protocol" library preparation protocol with the 1D
209 Ligation Sequencing Kit (SQK-LSK108, Oxford Nanopore, Oxford, UK). Sequencing was
210 performed on a MinION sequencer (Oxford Nanopore, Oxford, UK) using a single FLO-
211 MIN106 R9 flow cell for 48 hours, and the raw Nanopore signals were base-called using
212 Albacore version 2.3.1.

213 All reads passing the quality checks were used to create an assembly with Unicycler version
214 0.4.7 (Wick et al., 2017b). The resulting assembly contained 6 circular contigs (~15kb –
215 ~1.6Mbp) with high similarity to the previously published *s*Mel genome as determined by
216 BLAST+ version 2.9 searches (Camacho et al., 2009) in Bandage (Wick et al., 2015). We
217 also created a hybrid assembly with Unicycler using all long reads and the *s*Hy-Liv18a
218 Illumina reads. The hybrid assembly was more fragmented, but did contain 3 contigs which
219 spanned almost the entire 1.6Mbp circular contig of the long-read only assembly. Because
220 hybrid assemblies are typically more accurate than assemblies based on long reads only (De
221 Maio et al., 2019), we corrected the long read assembly using the hybrid contigs. This was
222 achieved by calling differences between the assemblies with the 'dnadiff' function
223 implemented in MUMmer version 3.18 (Kurtz et al., 2004), and correcting the assembly
224 accordingly. Finally, Pilon version 1.23 (Walker et al., 2014) was used to polish the corrected
225 assembly, again using the *s*Hy-Liv18a Illumina reads. After nine rounds of polishing with
226 Pilon, no further improvements could be observed and the assembly was considered final.

227 Rates of molecular evolution in *s*Hy and *s*Mel

228 We determined changes in *s*Hy and *s*Mel chromosomes over time by employing the Snippy
229 pipeline version 4.1.0 (Seemann, 2015). We used our newly assembled *s*Hy genome and the
230 previously sequenced *s*Mel genome (Masson et al., 2018) as references, annotated using the
231 Prokka pipeline version 1.13 with standard parameters and the *Spiroplasma* genetic code
232 (translation table 4, Seemann, 2014). *s*Hy and *s*Mel variants determined by Snippy were
233 filtered to only include positions which were covered by all *s*Hy or all *s*Mel Illumina libraries
234 with at least 5x coverage, respectively. Further, to compare rates of molecular evolution
235 between *Spiroplasma* and other microbial species, we counted the number of changes in third

236 codon positions of coding sequences (CDS). For both strains, we only considered CDS that
237 were entirely covered by sequencing depth of at least 5x in all Illumina libraries. Rates of
238 molecular evolution were calculated by: number of observed changes at 3rd codon positions /
239 number of considered 3rd codon positions / time in years over which the changes were
240 observed. For *s*Hy, using *s*Hy-Liv18a as reference, we calculated two rate estimates, one
241 based on comparison with *s*Hy-Liv18b, and one by comparing *s*Hy-Tx12 to the reference
242 (Fig. 1). For *s*Mel, the hemolymph extract from which the culture was established was used
243 as the reference ('HL', Fig. 1), and three rate estimates were calculated based on the changes
244 observed in the three sampled time points of the culture (Fig. 1).

245 Comparative genomics

246 We first compared the newly assembled *s*Hy genome with the most closely related, fully
247 sequenced *Spiroplasma* genome, *s*Mel. Both genomes were annotated using the Kyoto
248 Encyclopedia of Genes and Genomes (KEGG) and BlastKOALA (Kanehisa et al., 2016,
249 2006). Orthologous protein sequences determined using OrthoFinder version 2.2.3 (Emms
250 and Kelly, 2015). Synteny between the genomes was assessed through aligning the genomes
251 with Minimap2 version 2.15 (Li, 2018). Genome degradation was investigated by
252 documenting the number of unique KEGG numbers for *s*Hy and *s*Mel, all of which were
253 verified by BLAST+ searches. For each strain, we considered genes to be likely
254 pseudogenized or truncated if the longest CDS within an orthogroup spanned at most 60% of
255 the CDS length from the other strain.

256 Secondly, we compared *s*Hy with 17 *Spiroplasma* genomes from the Citri-Chrysopicola-
257 Mirum clade (Table S1, Supplementary material for full list of names and accession
258 numbers). All genomes were annotated using Prokka version 1.4.0, and orthology between

259 predicted CDS was established using OrthoFinder. We extracted single-copy orthologs
260 present in all strains, and aligned each locus separately with the L-INS-i method implemented
261 in Mafft version 7.450 (Katoh and Standley, 2013). Recombination was evaluated with
262 pairwise homoplasy index and window sizes of 100, 50, and 20 amino acids for each locus
263 (Bruen et al., 2006), and any locus showing evidence for recombination was excluded. We
264 used IQ-TREE version 1.6.10 (Nguyen et al., 2015) to reconstruct a core genome phylogeny
265 from the remaining 96 loci (covering 26,019 amino acid positions). Each locus was treated as
266 a separate partition with a distinct evolutionary rate (Chernomor et al., 2016), and optimal
267 models and number of partitions was estimated with IQ-TREE (Kalyaanamoorthy et al.,
268 2017; Lanfear et al., 2012). Branch support was assessed using 1,000 replicates of ultrafast
269 bootstraps (Hoang et al., 2018), and approximate likelihood ratio test (Guindon et al., 2010)
270 in IQ-TREE.

271 For all genomes, insertion sequence elements were annotated with Prokka. Prophage regions
272 were predicted with PhiSpy version 3.7.8, which uses typical prophage features ((e.g., gene
273 length, strand directionality, AT/GC skews, insertion points) for its predictions (Akhter et al.,
274 2012). We also used the Phaster web server, which uses sequence similarities and other
275 criteria to predict prophages (Arndt et al., 2016). We screened for the presence of toxin genes
276 implied in protective phenotypes (Ballinger and Perlman, 2017; Hamilton et al., 2015) and
277 male killing (Harumoto and Lemaitre, 2018). To this end, we downloaded UniProt sequence
278 alignments from the Pfam database (Finn et al., 2016) for the protein families RIP (PF00161,
279 ribosome inactivating protein) and OTU (PF02328, OTU-like cysteine protease),
280 respectively. We used these alignments as databases for searches with HMMER version 3.2.1
281 (Eddy, 2011), and protein sequences from all genomes as queries. Domain architecture of all
282 matching proteins was then determined using PfamScan with an e-value of 0.001 (Madeira et

283 al., 2019), SignalP 5.0 (Almagro Armenteros et al., 2019), and TMHMM 2.0 (Krogh et al.,
284 2001). RIP domains showed a high degree of divergence, and were therefore aligned to the
285 reference RIP HMM profile using hmmalign from the HMMER software package. From this
286 alignment, positions present in fewer than three sequences were removed, and a phylogeny
287 reconstructed with IQ-TREE. OTU domains were aligned using MAFFT, and a phylogeny
288 reconstructed with IQ-TREE.

289 Thirdly, we determined plasmid synteny across *Spiroplasma poulsonii* strains. In addition to
290 the plasmids of *s*Mel and *s*Hy, we also included a plasmid of *s*Neo and all circular contigs
291 from our *s*Mel-Br assembly. Plasmids were annotated using Prokka version 1.13 with a
292 protein database composed of all plasmid proteins available on NCBI GenBank. Plasmid
293 alignment and visualisation was performed with AliTV version 1.06 (Ankenbrand et al.,
294 2017).

295 For an extended set of 31 *Spiroplasma* genomes that also included strains from the Apis clade
296 (Table S1, supplementary material), we performed gene tree-species tree reconciliations to
297 investigate the evolutionary history of prophage loci. To this end, we annotated all genomes
298 and determined orthologous groups of loci as described above. Orthogroups were blasted
299 against a database containing all available *Spiroplasma* virus proteins on NCBI Protein (N =
300 192). Genes were classified as 'prophage related' if hits were at least 60% identical over 50%
301 of the length of any viral protein. Reconciliation was performed with the prophage loci using
302 GeneRax version 1.2.2, using maximum likelihood gene trees determined with IQ-TREE as
303 starting trees and the LG+G model for all loci. CRISPR/Cas systems and CRISPR arrays
304 were predicted using the tools CRISPRCasTyper version 1.2.1 (Russel et al., 2020) and
305 CRISPRIdentify (Mitrofanov et al., 2020), respectively.

306 Data visualisation

307 Figures were prepared in R version 3.6.2 (R Core Team, 2016) using the packages 'ape'
308 (Paradis and Schliep, 2019), 'cowplot' (Wilke, 2019), 'ggalluvial' (Brunson, 2019), 'ggplot2'
309 (Wickham, 2009), 'ggtree' (Yu et al., 2017), and 'ggridges' (Wilke, 2020). Phylogenetic trees
310 and domain architecture of toxin loci were visualised with EvolView version 3.0
311 (Subramanian et al., 2019).

312

313 **Results**

314 Rates and patterns of evolutionary change in *Spiroplasma poulsonii*

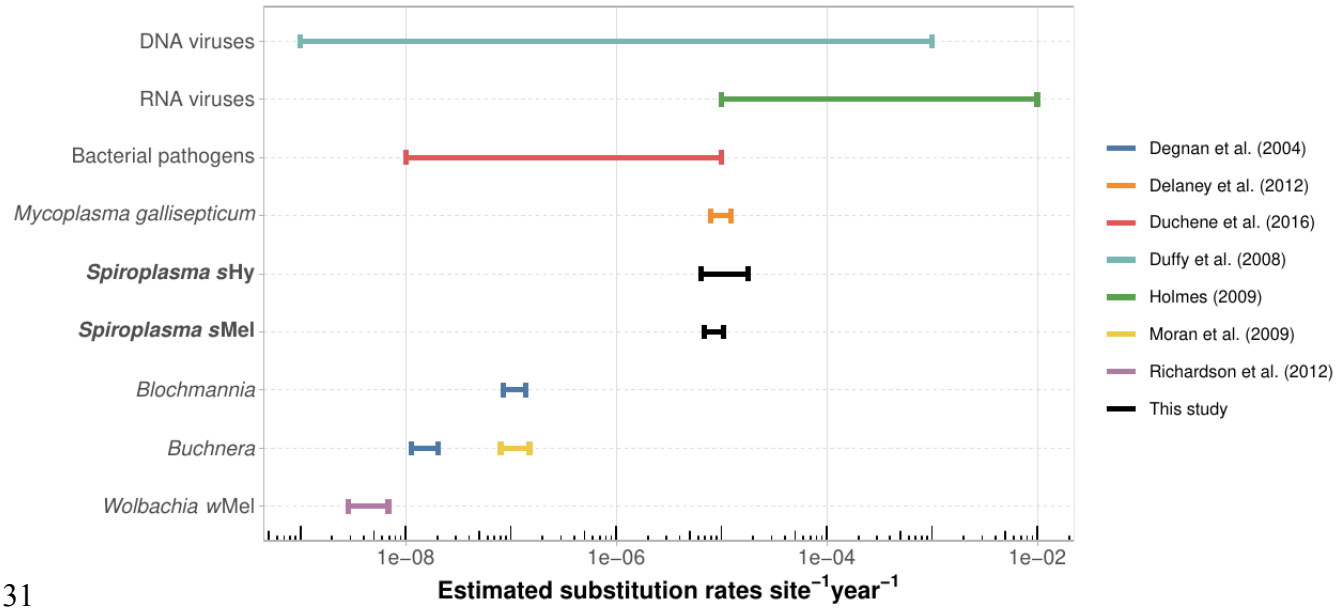
315 The *Spiroplasma poulsonii s*Hy reference genome comprises a single circular chromosome
316 (1,625,797bp) and five circular contigs (23,069bp – 15,710bp) with sequence similarities to
317 plasmids from other *Spiroplasma*. The chromosome contains 1,584 predicted coding
318 sequences (CDS), a single ribosomal RNA cluster, and 30 tRNAs. Overall, genome content
319 and metabolic capacities are highly similar to the previously sequenced, very closely related
320 *s*Mel genome (Paredes et al., 2015), and a detailed comparison follows further below.

321 To estimate rates of molecular evolution in *Spiroplasma poulsonii*, we measured
322 chromosome-wide changes in coding sequences of *Spiroplasma* from fly hosts (*s*Hy) and
323 axenic culture (*s*Mel) over time. Our estimates for *s*Hy and *s*Mel are overlapping and range
324 from 6.4×10^{-6} – 1.8×10^{-5} changes per position per year. This rate exceeds the rate reported for
325 other symbiotic bacteria such as *Wolbachia* and *Buchnera* by at least two orders of magnitude
326 (Fig. 2). Our estimate overlaps with rates calculated from some fast evolving human
327 pathogens (e.g., *Enterococcus faecium* and *Acinetobacter baumannii*), and with evolutionary

328 rates observed in the poultry pathogen *Mycoplasma gallisepticum* (which is closely related to

329 *Spiroplasma*). Indeed, *Spiroplasma* substitution rates fall at around the lower estimates for

330 RNA viruses (Fig. 2).



331

332 **Fig 2.** Comparison of estimated evolutionary rates across various microbes. Estimates
333 obtained in this study are highlighted in bold. All bacterial rates are from chromosomal
334 sequences only. DNA viruses - estimates as summarized in Duffy et al. (2008), including
335 ssDNA viruses and dsDNA viruses. RNA viruses - range of RNA virus substitution rates
336 from Holmes (2009). Bacterial pathogens - range of evolutionary rates estimated from
337 genome wide data of 16 bacterial pathogens (*Acinetobacter baumannii*, *Bordetella*
338 *pertussis*, *Enterococcus faecium*, *Klebsiella pneumoniae*, *Mycobacterium leprae*,
339 *Mycobacterium tuberculosis*, *Neisseria meningitidis*, *Pseudomonas aeruginosa*,
340 *Salmonella enterica*, *Shigella dysenteriae*, *Shigella sonnei*, *Staphylococcus aureus*,
341 *Streptococcus pneumoniae*, *Streptococcus pyogenes*, *Vibrio cholerae*, *Yersinia pestis*) as
342 determined by Duchêne et al. (2016). *Mycoplasma gallisepticum* - genome wide rate
343 estimated from multiple isolates over a 12 year period (Delaney et al., 2012).
344 *Blochmannia* and *Buchnera* (blue line) - based on 16S rDNA, *gidA*, and *groEL*
345 sequences, taken from Degnan et al. (2004). *Buchnera* (yellow line) - genome-wide rate
346 estimated from 7 isolates (Moran et al., 2009). *Wolbachia wMel* - genome-wide data
347 extracted from 179 *Drosophila melanogaster* SRA libraries (Richardson et al., 2012).

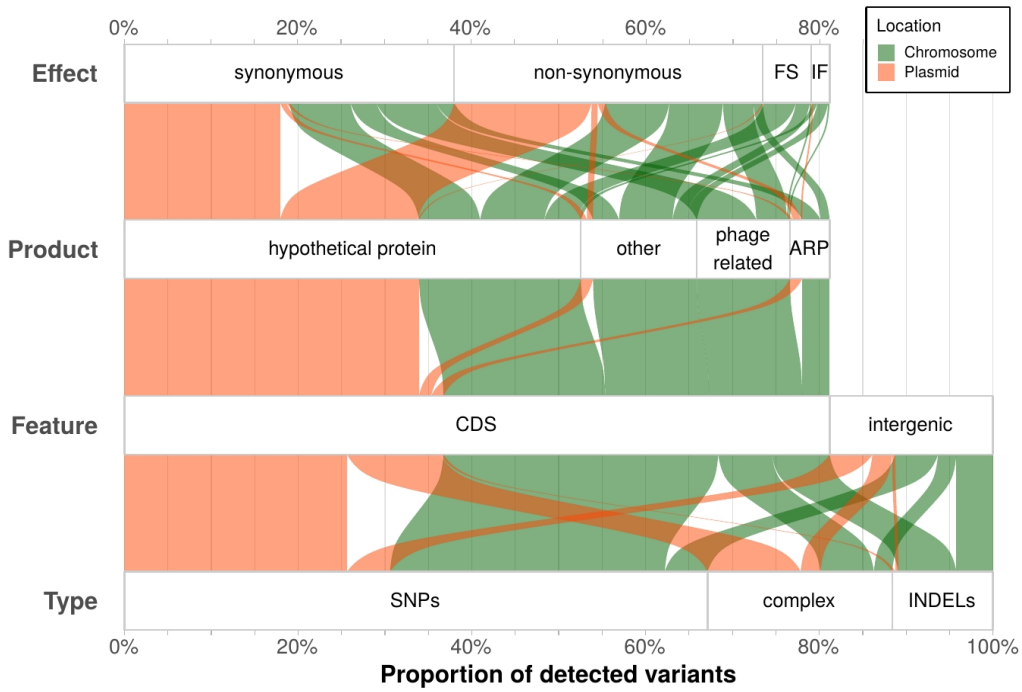
348 Over ~10 years of evolution, we observed a similar absolute number of variants on the

349 chromosome (~1.6 Mbp) and plasmids (~0.1 Mbp) of *s*Hy, i.e., relatively more changes on

350 the plasmids (Fig. 3, Table S3, supplementary material). Most variants in CDS affected

351 hypothetical proteins, but were enriched in prophage-associated loci and adhesin related
352 proteins (ARP, Fig. 3). Overall, *s*Hy variants in CDS were about equally often found to be
353 synonymous as non-synonymous (Fig. 3) and changes were biased towards GC > AT
354 substitutions (Fig. S1, Supplementary material). The changes in *s*Mel over ~2.5 years in
355 culture affected only 15 different CDS in total, of which four were ARPs, and three
356 lipoproteins (Table S2, supplementary material). Thus, the rates and patterns of evolutionary
357 change are similar between the axenically cultured *s*Mel and the host associated *s*Hy.

358



359

360 **Fig. 3:** Overview of all observed variants in *s*Hy (100% = 570 variants). All annotations
361 (Type, Feature, Product, Effect) are taken from results of the snippy pipeline.
362 Abbreviations: ARP - Adhesion related protein, CDS - predicted coding sequence,
363 complex – variants that span multiple nucleotides and/or SNPs, FS - frame shift, IF - in
364 frame, INDELs - insertions and deletions SNPs - single nucleotide polymorphisms.

365

366 Comparing the genomes of *s*Hy and *s*Mel revealed a notable contrast between the high degree
367 of nucleotide sequence identity on the one hand, and striking structural and gene content

368 differences on the other hand. In 604 single copy orthologues shared between the genomes,
369 average nucleotide identity was >99.5%. However, compared with *s*Mel, *s*Hy is reduced in
370 gene content (1,584 vs 2,388 genes), chromosome size (1.61 vs 1.88 Mbp), and coding
371 density (65% vs 82%). Further, *s*Hy contains fewer predicted insertion sequences (9 vs 89 in
372 *s*Mel) and intact prophages as predicted with Phaster (*s*Mel: 16 such regions spanning 426
373 Kb; *s*Hy: 2 regions, 14 Kb), but not PhiSpy (Fig. S2, Supplementary material). Both *s*Mel and
374 *s*Hy have a number of missing or truncated (i.e., potentially pseudogenized) genes when
375 compared with each other, but the level of genomic deterioration is higher in *s*Hy, and covers
376 a range of different genes (Fig. 4a, Table S4, supplementary material). Notable
377 pseudogenised or absent loci in *s*Hy include parts of the phosphotransferase system, in
378 particular the loci required for the uptake of N-Acetylmuramic Acid and N-acetylglucosamine
379 (see Table S3, Supplementary material for a full list). Further, while recA is truncated in
380 *s*Mel, the copy in *s*Hy appears complete and functional. As suggested by Paredes et al.
381 (2015), the loss of recA function in *s*Mel is therefore likely very recent.

382 Strikingly, there is evidence for a history of extensive chromosomal rearrangements since the
383 last common ancestor of *s*Hy and *s*Mel, and genome-wide synteny between the strains is low
384 (Fig. 4b). On average, syntenic blocks between the strains contained fewer predicted CDS for
385 *s*Hy (on average -0.3 CDS/kb, Fig. 4c), in line with a higher degree of genomic deterioration
386 in *s*Hy compared with *s*Mel. A comparison of plasmids across different *Spiroplasma*
387 *poulsonii* strains (*s*Hy, *s*Mel-Ug, *s*Mel-Br, and *s*Neo) revealed similar gene content across the
388 plasmids of different strains, but large differences in arrangement and number (Fig. S3,
389 Supplementary material).
390

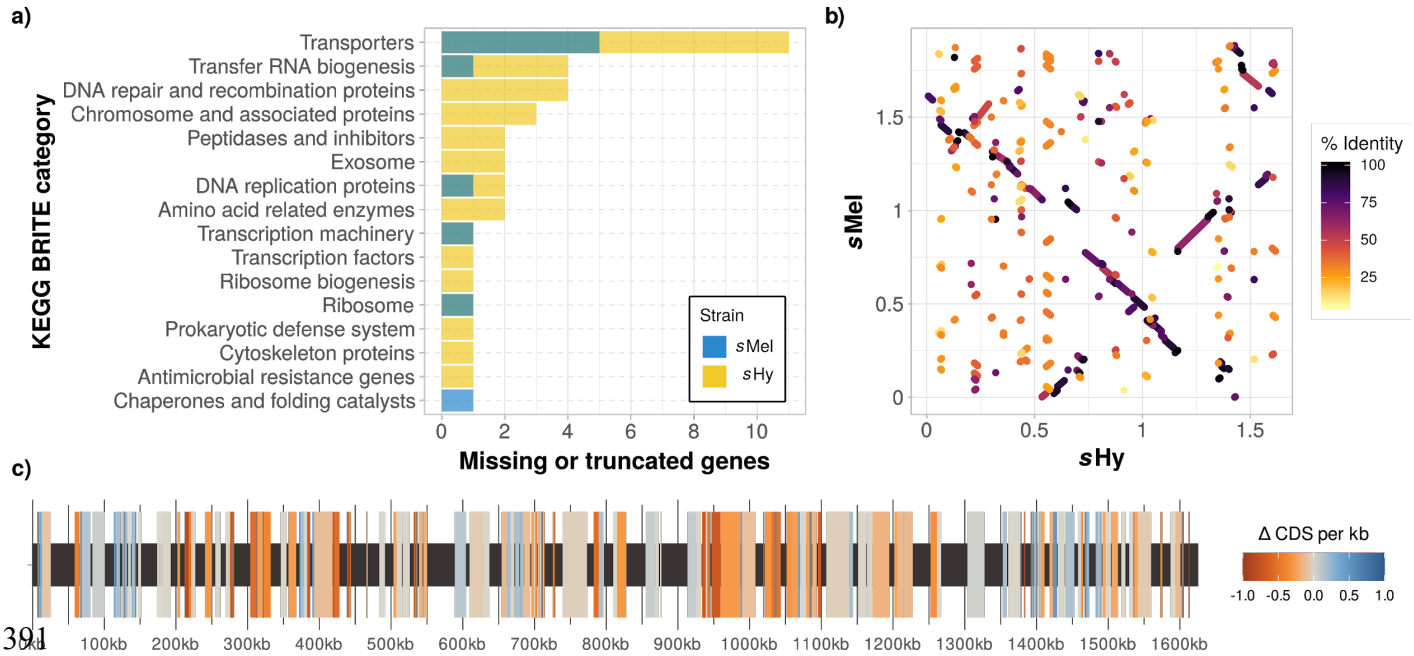


Fig 4: Comparison of genomic features between *s*Hy and *s*Mel. **a)** Number of genes with KEGG annotation that are present in one strain, but absent or truncated in the other. See Table S4, supplementary material for a complete list. **b)** Synteny as determined with minimap2. Syntenic between blocks are coloured by sequence similarity. **c)** Genomic map of *Spiroplasma poulsonii* *s*Hy with blocks syntenic to *s*Mel highlighted. The average number of predicted CDS per kb in these blocks is displayed as difference between *s*Hy and *s*Mel: positive and negative values indicate fewer CDS in *s*Hy or *s*Mel, respectively. Syntenic blocks were determined with a BLAST+ search using the *s*Mel chromosome sequence as query against a *s*Hy chromosome database, keeping a single best match for any *s*Hy region, and discarding hits below 1,000bp and under 95% nucleotide sequence similarity. Out of 635 syntenic blocks, 324 (spanning 852,730bp) contained fewer CDS in *s*Hy, and 164 (345,546bp) contained more CDS in *s*Hy, with 147 regions (334,084bp) showing no difference in CDS number.

Genome and toxin evolution in the genus *Spiroplasma*

Comparing *s*Hy with other sequenced strains of the *Spiroplasma* clades Citri, Poulsonii, Chrysopicola, and Mirum showed dynamic genome evolution within this genus of symbionts (Fig. 5). All investigated spiroplasmas have reduced genomes (~1.1–1.9Mbp), and the Poulsonii strains *s*Hy, *s*Mel and *s*Neo are among the strains with the largest main chromosomes (Fig. 5). Plasmid numbers range from 0–5, and *s*Hy has the highest number of plasmids of any of the investigated strains, although plasmid number is unclear for some of

414 the genomes in draft status, and *S. citri* may have seven plasmids (Saillard et al., 2008).
415 Prophage regions of varying sizes were predicted in all of the genomes (despite the lack of
416 clear homologues to viral sequences in some of these, Ku et al., 2013). Reconciliation of
417 prophage gene trees with the *Spiroplasma* species tree revealed that prophage proliferation
418 has likely happened relatively recently, and repeatedly in the Citri and Poulsonii clades (Fig.
419 S4, Supplementary material). Prophage loci are entirely absent, or very low in number, for
420 *Spiroplasma* strains that harbour CRISPR/Cas systems, or remnants thereof (Fig. S4,
421 Supplementary material). Further, *Spiroplasma* genome size correlated with the number of
422 insertion sequences (Fig. 5, Fig. S5, Supplementary material). The distribution of CDS
423 sequence lengths varies across the investigated genomes (Fig. 5), which may be explained by
424 differences in proportion of prophage regions, level of pseudogenization, and assembly
425 quality. Overall, our comparison indicates that *s*Hy may be more degraded than its closest
426 relatives *s*Neo and *s*Mel, with a smaller chromosome size, fewer predicted CDS and mobile
427 genetic elements, and the lowest coding density of these three (Fig. 5).

428 Spiroplasmas' most striking phenotypes in insects have been mechanistically linked to toxin
429 genes. For example, ribosome inactivating proteins (RIP) may protect *Spiroplasma* hosts by
430 cleaving ribosomal RNA of parasites and parasitoids (Hamilton et al., 2015). Five RIP loci
431 are present in *s*Mel (RIP1–5, of which 3–5 are almost identical copies). As expected, the
432 protective *s*Hy also encodes RIPs, but only has a single orthologue for RIP1. However, it
433 contains an additional RIP gene in three copies that appears to be absent in *s*Mel, and one of
434 these copies also contains ankyrin repeats (Fig. S6, Supplementary material).

435

436

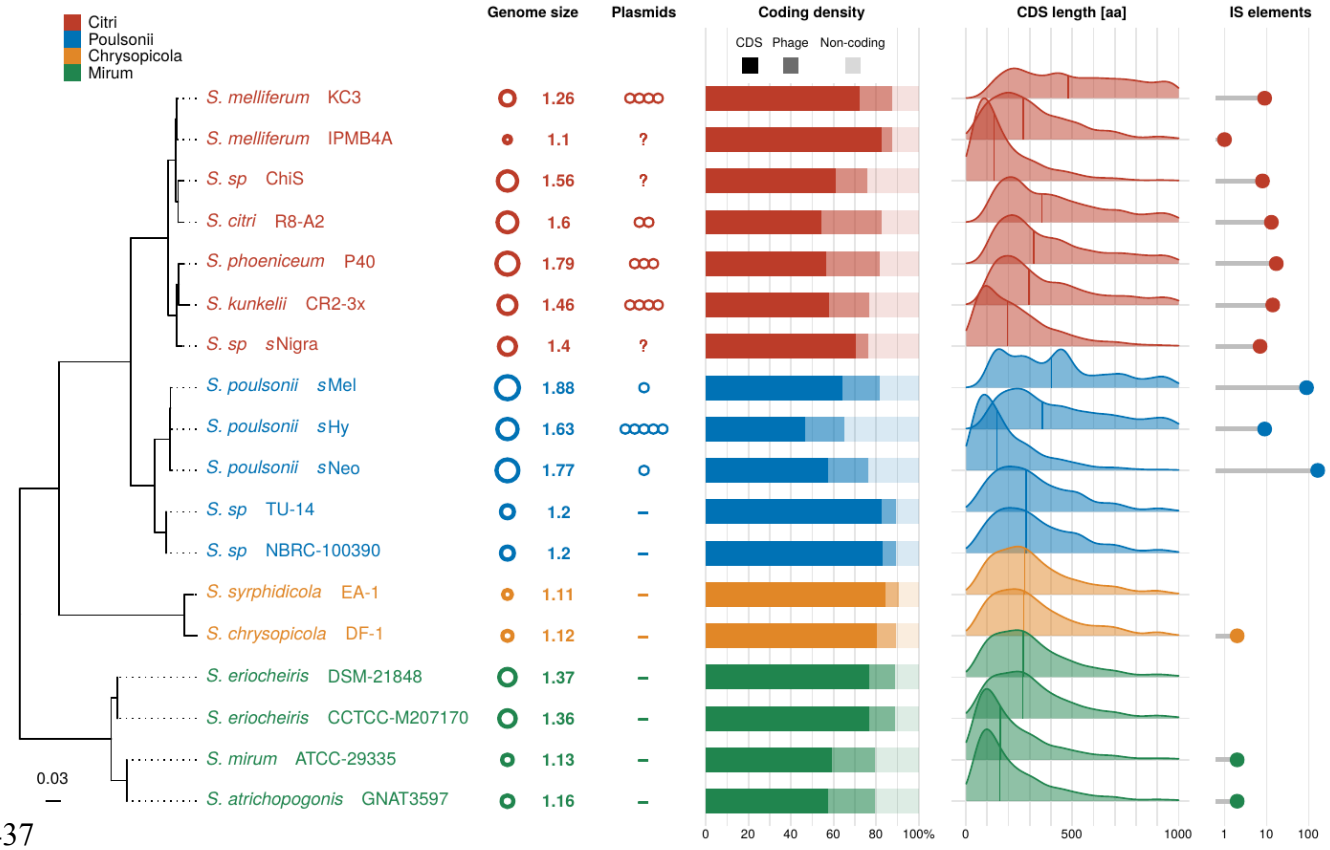
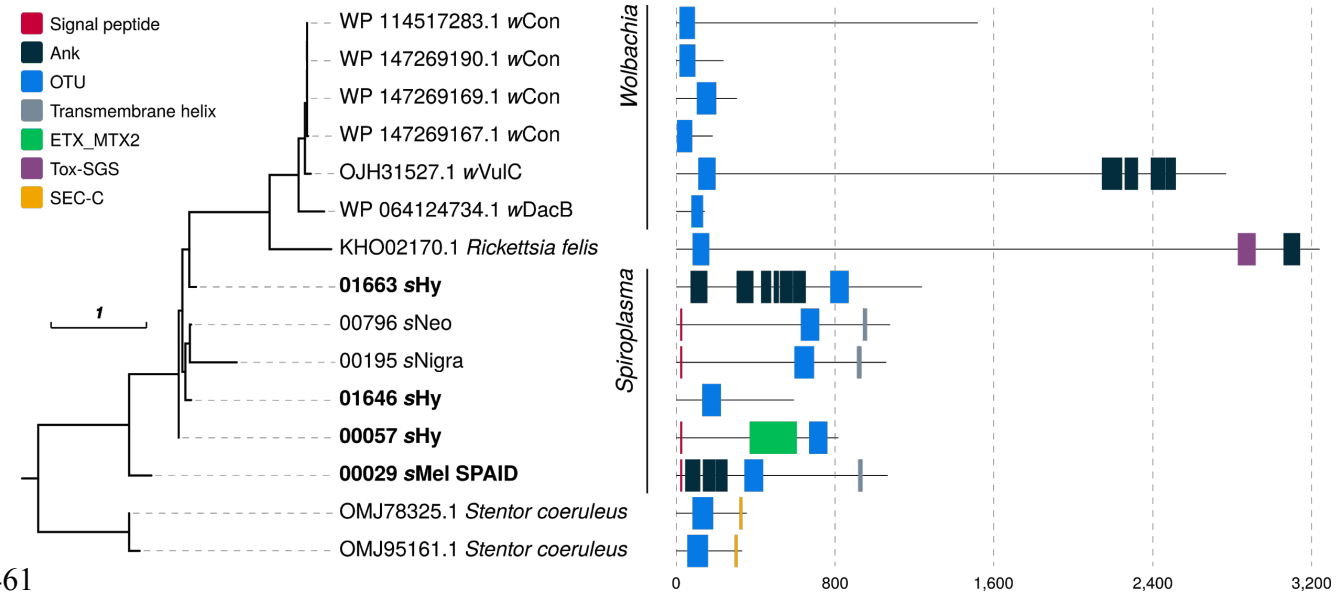


Fig 5: Genome properties of *Spiroplasma* strains from Citri, Poulsonii, Chrysopicola, and Mirum clades. Genome size and plasmid number was taken from the literature (Table S1, supplementary material). Coding density, CDS length and number of IS elements was determined with Prokka, and proportion of prophage was estimated with PhiSpy. *Spiroplasma* phylogeny is based on partitioned maximum likelihood analysis of 96 concatenated single copy genes present in all of the genomes (26,019 amino acid positions). All nodes were maximally supported by UFB and aLRT.

Further, the male killing phenotype of *s*Mel was recently established to be caused by *Spiroplasma* androcidin (Spaid), and both ankyrin repeats and a deubiquitinase domain (OTU) of this gene are necessary to induce male killing (Harumoto and Lemaitre, 2018). HMMER searches using OTU domain profiles revealed a number of *Spiroplasma* loci similar to Spaid (Fig. 6) which however lack its characteristic domain composition. For example, *s*Hy encodes three loci similar to Spaid: one lacks a signal peptide, one has no ankyrin repeats, and another encodes an epsilon-toxin like domain in addition to the OTU domain (Fig. 6). Other bacterial loci with notable similarities to the *Spiroplasma* OTU could only be

454 detected in the symbiotic taxa *Rickettsia* and *Wolbachia*. In the phylogenetic reconstruction of
455 OTU domains rooted with the eukaryotic sequences (from the ciliate *Stentor coeruleus*),
456 *Rickettsia* and *Wolbachia* loci are nested within the *Spiroplasma* sequences. This topology
457 suggests lateral gene transfer of OTU domain containing proteins from *Spiroplasma* to the
458 other intracellular taxa, although differences in genetic code between *Spiroplasma* and other
459 bacteria likely limit the probability for transfer of functional protein coding genes.

460



461

462 **Fig. 6:** Spaid-like proteins in *Spiroplasma* and other symbionts. Maximum likelihood tree
463 was reconstructed from an alignment of OTU domains (147 amino acid positions), and
464 domain predictions are based on PfamScan, SignalP, and TMHMM. Abbreviations
465 (corresponding to PFAM domains): Ank - ankyrin repeats, OTU - Ovarian tumor like
466 deubiquitinase, ETX_MTX2 - Clostridium epsilon toxin ETX/Bacillus mosquitocidal toxin
467 MTX2, Tox-SGS - Salivary gland secreted protein domain toxin, SEC-C - SEC-C nucleic
468 acid binding domain.

469

470

471 **Discussion**

472 *Spiroplasma*, an exceptionally fast evolving symbiont

473 The substitution rates we observed in *Spiroplasma* symbionts are among the highest reported
474 for any bacteria (Fig. 1). In a study comparing genome-wide evolutionary rates in bacterial
475 human pathogens (Duchêne et al., 2016), most taxa showed considerably slower rates, and
476 our estimate of *Spiroplasma* evolutionary rates overlaps with substitution rates of RNA
477 viruses (Holmes, 2009). A number of reasons have been proposed for elevated substitution
478 rates in bacterial symbionts and pathogens with small genome sizes: Firstly, host-associated
479 microbes may acquire essential metabolic intermediates from their hosts and thus need not
480 synthesise them (Moran, 2003). Obsolete metabolic genes subsequently accumulate now
481 neutral substitutions, resulting in pseudogenization. Secondly, reduced population sizes of
482 symbionts, together with bottlenecks at transmission events limit purifying selection of
483 deleterious mutations (Moran, 1996) and mobile genetic elements (Lynch, 2006) – both in
484 turn create further genomic deterioration. Thirdly, genomes of host-associated microbes are
485 AT rich (Rocha and Danchin, 2002), which makes introduction of errors through replication
486 slippage more likely (Moran et al., 2009). Fourthly, many symbionts lack DNA repair genes
487 (Moran et al., 2008), which may result in increased substitution rates.

488 All of these points very likely apply for *Spiroplasma poulsonii*: its metabolic capacities are
489 reduced compared with free living microbes, it has a high proportion of pseudogenes and
490 prophages, very high AT content, and lacks several DNA repair genes. However, many
491 symbiotic bacteria show similar trends of genome evolution (Burger and Lang, 2003), and it
492 is therefore somewhat puzzling that *Spiroplasma* stands out with this exceptionally high
493 substitution rate. Loss of DNA repair genes has been shown to co-occur with elevated

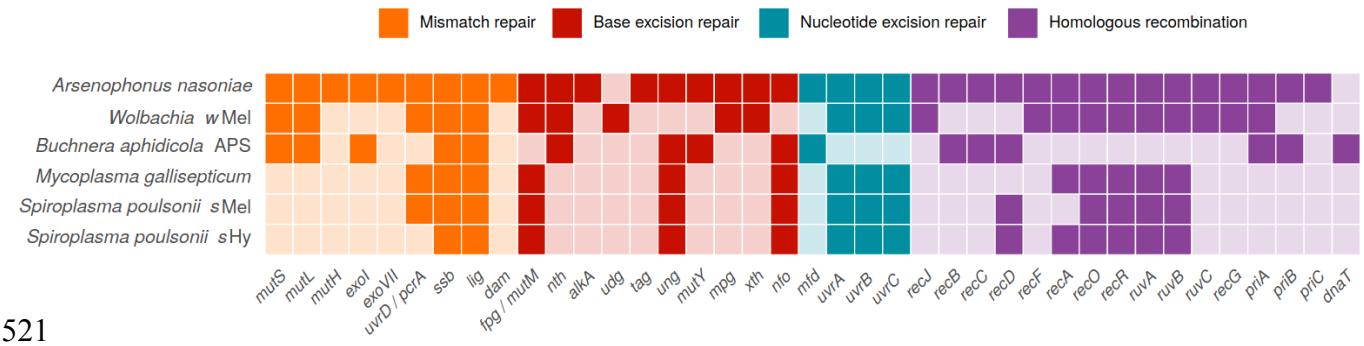
494 mutation rates in intracellular bacteria (Itoh et al., 2002; Koonin et al., 1996; Massey, 2008),
495 and may explain our observations. The loci encoding the mismatch repair proteins *mutS* and
496 *mutL* are universally lacking in *Spiroplasma*, but are present in the slower evolving symbionts
497 *Wolbachia w*Mel, and *Buchnera aphidicola* (Fig. 7, Richardson et al., 2012; Shigenobu et al.,
498 2000; Wu et al., 2004). In *E. coli*, hypermutator strains that have lost such loci have a ~10–
499 300 fold increase of spontaneous mutation rates (LeClerc et al., 1996; Lee et al., 2012), which
500 is comparable to the ~two orders of magnitude difference between substitution rates of
501 *Buchnera* and *Spiroplasma* (Fig. 2). Like *Spiroplasma*, *Mycoplasma* universally lacks the
502 mismatch repair loci *mutS*, *mutL*, and *mutH* (Carvalho et al., 2005; Chen et al., 2012). In line
503 with this observation, elevated evolutionary rates were hypothesized for mycoplasmas
504 (Woese et al., 1984), and in our comparison, *M. gallisepticum* appears to be the only bacterial
505 taxon with substitution rates similar to *Spiroplasma* (Delaney et al., 2012). Absence of DNA
506 mismatch repair pathway may thus be ancestral to Entomoplasmatales (Spiroplasmatacea +
507 Entomoplasmataceae) and contribute to the dynamic genome evolution across this taxon (Lo
508 et al., 2016; Rocha and Blanchard, 2002). Alternatively, increased substitutional rates caused
509 by the loss of these loci could have arisen multiple times independently in
510 Entomoplasmatales.

511 Further to absence of DNA repair genes causing elevated mutation rates, a recent comparative
512 study demonstrated a strong negative correlation between mutation rate and genome size in
513 free living and endosymbiotic bacteria (Bourguignon et al., 2020). This correlation is
514 however not apparent in the genomes of endosymbionts we have investigated. For example,
515 the considerably slower evolving *Buchnera* genomes are much smaller than *Spiroplasma*, and
516 *Wolbachia* would be predicted to have much larger genomes if their size was mainly
517 determined by mutational rates. This suggests that substitution rates alone are a poor predictor

518 for the sizes of here investigated genomes. Likely, these genome sizes result from an

519 interplay of multiple factors such as population size, patterns of DNA repair gene absence,

520 and mutational rates (Kuo et al., 2009; Marais et al., 2020).



521

522 **Fig. 7:** Absence (light colors) and presence (dark colors) of genes involved in DNA repair
523 in *Buchnera aphidicola*, *Wolbachia* wMel, *Mycoplasma gallisepticum*, and *Spiroplasma*
524 *poulsonii* sHy and sMel. *Arsenophonus nasoniae* is included as an example of a symbiont
525 with a relatively large genome and complete DNA repair pathways. Genes involved in
526 more than one pathway are listed only once, and those missing in all of the taxa are not
527 displayed. Data is taken from the KEGG pathway database (Kanehisa et al., 2006), and
528 annotation of sMel and sHy genomes as described in Materials and Methods section.

529

530 Our rate estimate is potentially biased by at least two factors. Firstly, we have only

531 investigated laboratory populations of *Spiroplasma poulsonii*. Each vertical transmission

532 event creates symbiont population bottlenecks potentially increasing genetic drift and thus

533 substitution rates (McCutcheon and Moran, 2011). Because the number of generations in

534 natural populations of the *Spiroplasma* host *Drosophila hydei* is lower compared with

535 laboratory reared hosts, vertical transmission events are rarer under natural conditions, and

536 substitution rates therefore potentially lower. Further, laboratory strains could experience

537 relaxed selection compared with natural symbiont populations. This may lead to higher

538 substitution rate estimates from laboratory populations compared with natural populations.

539 Secondly, substitution rates often appear larger when estimated over brief time periods (Ho et

540 al., 2005). Duchene et al. (2016) found that substitution rates measured over 10 years can be
541 up to one order of magnitude larger than those measured over 100 years. In agreement with
542 such bias, we found more variants in our *s*Mel culture after 19 months than for the 29 months
543 isolate (Table S3, supplementary material). The back mutations that likely have happened
544 between 19 and 29 months of our *s*Mel culture would go unnoticed when comparing genomes
545 over larger time scales.

546 More generally, it is difficult to estimate divergence times between *Spiroplasma* strains. The
547 substitution rates estimated in *s*Hy and *s*Mel would suggest the two strains have diverged
548 1,120–2,260 years ago. However, this estimate is unreliable due to a number of factors that
549 we cannot control for. For example, the number of generations per year for *Drosophila* hosts
550 of *Spiroplasma* differ depending on species and location, and is expected to be lower in the
551 wild compared with laboratory strains. Further, *Spiroplasma* may move between species, for
552 example via ectoparasitic mites (Jaenike et al., 2007). A number of partially sympatric
553 *Drosophila* species carry similar *Spiroplasma* strains (*D. hydei*, *D. melanogaster*, *D.*
554 *willistoni*, *D. neotestacea*) (Haselkorn, 2010), and with our data it is impossible to determine
555 the number and direction of potential *Spiroplasma* transfers between these species.

556 Reductive genome evolution in *Spiroplasma*

557 According to a widely supported model of genome reduction in symbiotic bacteria
558 (McCutcheon and Moran, 2011), the first stages of host restriction involve accumulation of
559 pseudogenes and mobile genetic elements, chromosomal rearrangements, increased
560 substitution rates, and excess deletions. Advanced stages of host association are accompanied
561 by further genome shrinkage, and purging of mobile genetic elements, which overall result in
562 more stable chromosomes. Using these characteristics, different levels of genome reduction

563 are apparent in the investigated *Spiroplasma* genomes, across which genome sizes correlate
564 positively with number of mobile genetic elements (plasmids, prophages, insertion sequences,
565 Fig. 5, Fig. S5, supplementary material). On this spectrum, *S. syrphidicola* and *S.*
566 *chrysopicola* have the most reduced genomes, without homologues of prophage loci, and high
567 levels of synteny between the two strains (Ku et al., 2013). It was therefore argued that
568 phages have likely invaded *Spiroplasma* only after the split of the Syrphidicola and
569 Citri+Poulsonii clades (Ku et al., 2013). Our prophage gene tree-species tree reconciliations
570 are in line with this hypothesis, but also indicate that prophage proliferation has largely
571 happened independently in different *Spiroplasma* lineages (Fig. S4, supplementary material).
572 CRISPR/Cas systems have multiple origins in *Spiroplasma* (Ipoutcha et al., 2019) and only
573 occur in strains lacking prophages (Fig. S4, supplementary material). While the absence of
574 antiviral systems often coincides with prophage proliferation (e.g., in the Citri clade), several
575 strains with compact, streamlined genomes lack CRISPR/Cas and prophages (e.g., TU-14 &
576 NBRC-100390, Fig. S4, supplementary material). These strains also show other hallmarks of
577 reduced symbiont genomes (small size, high coding density, lack of plasmids and
578 transposons, Fig. 5), which is in line with the model of genome reduction discussed above
579 and suggests prophage regions were purged from these genomes. Alternatively, these strains
580 may never have been exposed to phages.

581 Using the model of genome reduction during restriction to the host environment introduced
582 above, *Spiroplasma poulsonii*'s genome characteristics suggest that such restriction has
583 happened relatively recently. Although very closely related (99.5% sequence identity) and
584 found in very similar hosts, *s*Mel and *s*Hy differ markedly in coding density, genome size,
585 and proportion of prophage regions. Interestingly, PHASTER predicted 426 Kb of intact
586 phage regions in *s*Mel, but only 14 Kb for *s*Hy (Fig. S2, supplementary material). Both phage

587 proliferation in *s*Mel, and prophage loss in *s*Hy could have contributed to this. However,
588 when using the similarity agnostic tool PhiSpy, the predicted prophage regions were similar
589 in size between *s*Hy and *s*Mel (Fig. S2). This observation is compatible with degradation and/
590 or pseudogenization of prophage regions in *s*Hy, which would lead to reduced sequence
591 similarity to viral loci (and thus reduced detectability by PHASTER), but not entirely blur
592 prophage characteristics employed by PhiSpy (e.g., gene length, strand directionality, AT/GC
593 skews, insertion points, Akhter et al., 2012). Since the split of the lineages, *s*Hy has not only
594 lost prophage regions and insertion sequences compared with *s*Mel, but also several genes
595 that are often lost in host restricted Bacteria, such as parts of the phosphotransferase system
596 for the uptake of carbohydrates, one tRNA locus, and *uvrD* which plays a role in mismatch
597 repair, and nucleotide excision repair (Table S2, supplementary material).

598 Using signatures of genomic degradation as a proxy, our findings collectively suggest that
599 *s*Hy is in a more advanced stage of host restriction than *s*Mel. This may indicate co-
600 adaptation with the host as a result of the fitness benefits associated with *s*Hy under parasitoid
601 pressure, and the absence of detectable costs for carrying *s*Hy in *Drosophila hydei* (Osaka et
602 al., 2013; Xie et al., 2010, 2014). However, the *Spiroplasma* symbiont of *Drosophila*
603 *neotestacea s*Neo is also protective, does not cause obvious fitness costs (Jaenike et al.,
604 2010), but has a less reduced genome (Fig.5, Ballinger and Perlman, 2017). Further, it is also
605 possible that genome reduction in *s*Hy was mainly driven by stochastic effects or even by
606 adaptation to laboratory conditions, as we have not investigated contemporary *s*Hy from wild
607 *D. hydei* populations.

608 Implications for *Spiroplasma* evolutionary ecology

609 *Spiroplasma* evolutionary ecology shows several parallels to that of the most widely
610 distributed arthropod symbiont *Wolbachia*: both symbionts are found in a range of different
611 hosts (Duron et al., 2008), have the ability to invade novel hosts (Conner et al., 2017;
612 Haselkorn et al., 2009), may confer protection (Teixeira et al., 2008; Xie et al., 2010) but also
613 kill males (Hurst and Jiggins, 2000), and can spread across host populations swiftly
614 (Cockburn et al., 2013; Turelli and Hoffmann, 1991). However, there are also pronounced
615 differences between the symbionts: *Spiroplasma* rarely reaches the high infection frequencies
616 often observed in *Wolbachia* (Hilgenboecker et al., 2008; Watts et al., 2009), and is arguably
617 found in a more diverse host range that encompasses arthropods and plants (Regassa and
618 Gasparich, 2006), molluscs (Duperron et al., 2013), echinoderms (He et al., 2018), and
619 cnidarians (Viver et al., 2017). Further, *Wolbachia* show greatly reduced spontaneous
620 mutation rates compared with *Spiroplasma*, likely caused by a more complete set of DNA
621 repair genes (Fig. 7).

622 In theory, fast evolutionary rates should enable *Spiroplasma* to adapt to novel hosts quickly
623 (i.e., to reduce pathogenicity, and to maximise vertical transmission efficiency), and
624 experimental studies have found high horizontal transmission efficiency of *Spiroplasma*
625 (Haselkorn et al., 2013; Haselkorn and Jaenike, 2015; Nakayama et al., 2015). Consistent
626 with this, we found that genes implicated in host-symbiont compatibility and virulence (e.g.,
627 adhesion-related proteins, have evolved especially fast in our evolution experiments. For
628 example, adhesion-related proteins are important in cell invasion in other *Spiroplasma*
629 species (Béven et al., 2012; Dubrana et al., 2016; Hou et al., 2017) and are enriched for
630 evolutionary changes in *s*Hy and *s*Mel (Fig. 2). In addition, we documented dynamic

evolution and turnover of toxin loci, which are important for host fitness and symbiont compatibility (Fig. 6, Fig. S6, supplementary material). This genomic flexibility may contribute to *Spiroplasma*'s broader host range when compared with *Wolbachia*. However, elevated evolutionary rates also make deleterious changes more likely, and, in the absence of strong selection, may result in faster loss of symbionts (Jaenike, 2012). This may explain the generally low prevalence of *Spiroplasma* symbionts (Haselkorn, 2010), which seems to increase only when carrying the symbiont is associated with a large fitness benefit (Jaenike et al., 2010). In contrast, virtually identical *Wolbachia* strains ("superspreaders") are found in many different host species at very high frequencies (Turelli et al., 2018) – demonstrating that stationary genomes may be evolutionary advantageous. In summary, the nature of *Spiroplasma* genomic evolution likely contributes to its peculiar evolutionary ecology.

From a practical perspective, *Spiroplasma poulsonii* has many features of a desirable model for symbiont-host interactions: fast rates of evolution make it more likely that adaptation and spontaneous changes in phenotypes can be determined over short time scales, as has been observed previously (Harumoto and Lemaitre, 2018; Masson et al., 2020; Nakayama et al., 2015). On the other hand, fast evolutionary changes make experiments less predictable, and because stochastic effects become more pronounced, links of genomic changes with phenotypes may be obscured. Further, the generalisability of experimental results may be limited for extremely fast evolving symbionts. Our findings therefore also underline the importance of regular validation of laboratory symbiont strains through re-sequencing.

653 **Author contributions**

654 Designed and conceived project: MG, GDDH, MM. Experimental passage of *Spiroplasma* in
655 vivo: MG, JG, GDDH, MM. Experimental passage of *Spiroplasma* in vitro: FM, BL.
656 Symbiont DNA purification and sequencing: MG, PR, SS, RA, HMM, FM. Assembly and
657 annotation: MG, HMM, PR, RA. Molecular evolution analyses: MG. Wrote first draft of
658 paper: MG, MM, GDDH. All authors contributed in revising the manuscript and have read
659 and approved the final manuscript.

660

661 **Acknowledgements**

662 MG has received funding from the European Union's Horizon 2020 Research and Innovation
663 Program under Marie Sklodowska-Curie grant agreement 703379 and from EMBO through a
664 long term fellowship (ALTF 48-2015) co-funded by the European Commission
665 (LTFCOFUND2013, GA-2013-609409). Sequencing was supported by an award from the
666 University of Liverpool Technology Directorate Voucher Scheme, and by a Texas A&M
667 Genomics Seed Grant.

668

669 **Data accessibility**

670 Raw reads for sHy-Tx can be found under NCBI BioProject Accession PRJNA274591
671 (Biosample SAMN08102256, SRA SRR6326389). Reads from sMel-Br are stored under
672 PRJNA507275 (SAMN10488582). All other reads are available under NCBI BioProject
673 PRJNA640980. Alignments and supplementary protocol are available from Zenodo under the
674 DOI 10.5281/zenodo.3903208.

675

676 **References**

677 Akhter S, Aziz RK, Edwards RA. 2012. PhiSpy: a novel algorithm for finding prophages in bacterial
678 genomes that combines similarity- and composition-based strategies. *Nucleic Acids Res* **40**:e126.
679 Almagro Armenteros JJ, Tsirigos KD, Sønderby CK, Petersen TN, Winther O, Brunak S, von Heijne
680 G, Nielsen H. 2019. SignalP 5.0 improves signal peptide predictions using deep neural networks.
681 *Nat Biotechnol* **37**:420–423.
682 Anbutsu H, Fukatsu T. 2011. *Spiroplasma* as a model insect endosymbiont. *Environ Microbiol Rep*
683 **3**:144–153.
684 Ankenbrand MJ, Hohlfeld S, Hackl T, Förster F. 2017. AliTV—interactive visualization of whole
685 genome comparisons. *PeerJ Computer Science* **3**:e116.
686 Arndt D, Grant JR, Marcu A, Sajed T, Pon A, Liang Y, Wishart DS. 2016. PHASTER: a better, faster
687 version of the PHAST phage search tool. *Nucleic Acids Res* **44**:W16–21.
688 Ballinger MJ, Perlman SJ. 2019. The defensive *Spiroplasma*. *Curr Opin Insect Sci* **32**:36–41.
689 Ballinger MJ, Perlman SJ. 2017. Generality of toxins in defensive symbiosis: Ribosome-inactivating
690 proteins and defense against parasitic wasps in *Drosophila*. *PLoS Pathog* **13**:e1006431.
691 Berho N, Duret S, Danet J-L, Renaudin J. 2006. Plasmid pSci6 from *Spiroplasma citri* GII-3 confers
692 insect transmissibility to the non-transmissible strain *S. citri* 44. *Microbiology* **152**:2703–2716.
693 Béven L, Duret S, Batailler B, Dubrana M-P, Saillard C, Renaudin J, Arricau-Bouvery N. 2012. The
694 repetitive domain of ScARP3d triggers entry of *Spiroplasma citri into cultured cells of the vector*
695 *Circulifer haematoceps. PLoS One* **7**:e48606.
696 Bourguignon T, Kinjo Y, Villa-Martín P, Coleman NV, Tang Q, Arab DA, Wang Z, Tokuda G,
697 Hongoh Y, Ohkuma M, Ho SYW, Pigolotti S, Lo N. 2020. Increased Mutation Rate Is Linked to
698 Genome Reduction in Prokaryotes. *Curr Biol* **30**:3848–3855.e4.
699 Bruen TC, Philippe H, Bryant D. 2006. A simple and robust statistical test for detecting the presence
700 of recombination. *Genetics* **172**:2665–2681.
701 Brunson JC. 2019. ggalluvial: Alluvial Plots in "ggplot2."
702 Burger G, Lang BF. 2003. Parallels in genome evolution in mitochondria and bacterial symbionts.
703 *IUBMB Life* **55**:205–212.
704 Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+:
705 architecture and applications. *BMC Bioinformatics* **10**:421.
706 Carle P, Saillard C, Carrère N, Carrère S, Duret S, Eveillard S, Gaurivaud P, Gourgues G, Gouzy J,
707 Salar P, Verdin E, Breton M, Blanchard A, Laigret F, Bové J-M, Renaudin J, Foissac X. 2010.
708 Partial chromosome sequence of *Spiroplasma citri* reveals extensive viral invasion and important
709 gene decay. *Appl Environ Microbiol* **76**:3420–3426.
710 Carvalho FM, Fonseca MM, Batistuzzo De Medeiros S, Scortecci KC, Blaha CAG, Agnez-Lima LF.
711 2005. DNA repair in reduced genome: the *Mycoplasma* model. *Gene* **360**:111–119.
712 Chen L-L, Chung W-C, Lin C-P, Kuo C-H. 2012. Comparative analysis of gene content evolution in
713 phytoplasmas and mycoplasmas. *PLoS One* **7**:e34407.
714 Chernomor O, von Haeseler A, Minh BQ. 2016. Terrace Aware Data Structure for Phylogenomic
715 Inference from Supermatrices. *Syst Biol* **65**:997–1008.
716 Clark TB, Whitcomb RF. 1984. Pathogenicity of mollicutes for insects: possible use in biological
717 control. *Ann Microbiol* **135A**:141–150.
718 Cockburn SN, Haselkorn TS, Hamilton PT, Landzberg E, Jaenike J, Perlman SJ. 2013. Dynamics of
719 the continent-wide spread of a *Drosophila* defensive symbiont. *Ecol Lett* **16**:609–616.
720 Conner WR, Blaxter ML, Anfora G, Ometto L, Rota-Stabelli O, Turelli M. 2017. Genome
721 comparisons indicate recent transfer of *w*Ri-like *Wolbachia* between sister species *Drosophila*
722 *suzukii* and *D. subpulchrella*. *Ecol Evol* **7**:9391–9404.
723 Corbin C, Heyworth ER, Ferrari J, Hurst GDD. 2017. Heritable symbionts in a world of varying

temperature. *Heredity* **118**:10–20.
Degnan PH, Lazarus AB, Brock CD, Wernegreen JJ. 2004. Host-symbiont stability and fast evolutionary rates in an ant-bacterium association: cospeciation of camponotus species and their endosymbionts, candidatus blochmannia. *Syst Biol* **53**:95–110.
Delaney NF, Balenger S, Bonneaud C, Marx CJ, Hill GE, Ferguson-Noel N, Tsai P, Rodrigo A, Edwards SV. 2012. Ultrafast Evolution and Loss of CRISPRs Following a Host Shift in a Novel Wildlife Pathogen, *Mycoplasma gallisepticum*. *PLoS Genetics* **8**:e1002511.
De Maio N, Shaw LP, Hubbard A, George S, Sanderson ND, Swann J, Wick R, AbuOun M, Stubberfield E, Hoosdally SJ, Crook DW, Peto TEA, Sheppard AE, Bailey MJ, Read DS, Anjum MF, Walker AS, Stoesser N, On Behalf Of The Rehab Consortium. 2019. Comparison of long-read sequencing technologies in the hybrid assembly of complex bacterial genomes. *Microb Genom* **5**:e000294.
Douglas AE. 2009. The microbial dimension in insect nutritional ecology. *Funct Ecol* **23**:38–47.
Drew GC, Frost CL, Hurst GDD. 2019. Reproductive Parasitism and Positive Fitness Effects of Heritable Microbes. *eLS* **2**:1–8.
Dubrana MP, Béven L, Arricau-Bouvery N, Duret S, Claverol S, Renaudin J, Saillard C. 2016. Differential expression of *Spiroplasma citri* surface protein genes in the plant and insect hosts. *BMC Microbiol* **16**. doi:10.1186/s12866-016-0666-y
Duchêne S, Holt KE, Weill F-X, Le Hello S, Hawkey J, Edwards DJ, Fourment M, Holmes EC. 2016. Genome-scale rates of evolutionary change in bacteria. *Microb Genom* **2**:e000094.
Duffy S, Shackelton LA, Holmes EC. 2008. Rates of evolutionary change in viruses: patterns and determinants. *Nat Rev Genet* **9**:267–276.
Duperron S, Pottier M-A, Léger N, Gaudron SM, Puillandre N, Le Prieur S, Sigwart JD, Ravaux J, Zbinden M. 2013. A tale of two chitons: is habitat specialisation linked to distinct associated bacterial communities? *FEMS Microbiol Ecol* **83**:552–567.
Duron O, Bouchon D, Boutin S, Bellamy L, Zhou L, Engelstädter J, Hurst GDD. 2008. The diversity of reproductive parasites among arthropods: *Wolbachia* do not walk alone. *BMC Biol* **6**:27.
Dyson EA, Kamath MK, Hurst G. 2002. *Wolbachia* infection associated with all-female broods in *Hypolimnas bolina* (Lepidoptera: Nymphalidae): evidence for horizontal transmission of a butterfly male killer. *Heredity* **88**:166–171.
Ebbert MA. 1991. The interaction phenotype in the *Drosophila willistoni-Spiroplasma* symbiosis. *Evolution* **45**:971–988.
Eddy SR. 2011. Accelerated Profile HMM Searches. *PLoS Comput Biol* **7**:e1002195.
Emms DM, Kelly S. 2015. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol* **16**:157.
Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, Potter SC, Punta M, Qureshi M, Sangrador-Vegas A, Salazar GA, Tate J, Bateman A. 2016. The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res* **44**:D279–85.
Gasparich GE. 2002. Spiroplasmas: evolution, adaptation and diversity. *Front Biosci* **7**:d619–d640.
Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* **59**:307–321.
Gupta RS, Son J, Oren A. 2019. A phylogenomic and molecular markers based taxonomic framework for members of the order Entomoplasmatales: proposal for an emended order Mycoplasmatales containing the family Spiroplasmataceae and emended family Mycoplasmataceae comprised of six genera. *Antonie Van Leeuwenhoek* **112**:561–588.
Hamilton PT, Leong JS, Koop BF, Perlman SJ. 2014. Transcriptional responses in a *Drosophila* defensive symbiosis. *Mol Ecol* **23**:1558–1570.
Hamilton PT, Peng F, Boulanger MJ, Perlman SJ. 2015. A ribosome-inactivating protein in a Drosophila defensive symbiont. *Proc Natl Acad Sci U S A* **112**:350–355.
Harumoto T, Lemaitre B. 2018. Male-killing toxin in a bacterial symbiont of *Drosophila*. *Nature* **557**:252–255.

776 Haselkorn TS. 2010. The *Spiroplasma* heritable bacterial endosymbiont of *Drosophila*. *Fly* **4**:80–87.
777 Haselkorn TS, Cockburn SN, Hamilton PT, Perlman SJ, Jaenike J. 2013. Infectious adaptation:
778 potential host range of a defensive endosymbiont in *Drosophila*. *Evolution* **67**:934–945.
779 Haselkorn TS, Jaenike J. 2015. Macroevolutionary persistence of heritable endosymbionts:
780 acquisition, retention, and expression of adaptive phenotypes in *Spiroplasma*. *Mol Ecol* **14**:3752–
781 3765.
782 Haselkorn TS, Markow TA, Moran NA. 2009. Multiple introductions of the *Spiroplasma* bacterial
783 endosymbiont into *Drosophila*. *Molecular Ecology* **18**:1294–1305.
784 Hayashi M, Watanabe M, Yukuhiro F, Nomura M, Kageyama D. 2016. A Nightmare for Males? A
785 Maternally Transmitted Male-Killing Bacterium and Strong Female Bias in a Green Lacewing
786 Population. *PLoS ONE* **11**:e0155794.
787 He L-S, Zhang P-W, Huang J-M, Zhu F-C, Danchin A, Wang Y. 2018. The Enigmatic Genome of an
788 Obligate Ancient *Spiroplasma* Symbiont in a Hadal Holothurian. *Appl Environ Microbiol*
789 **84**:e01965–17.
790 Hilgenboecker K, Hammerstein P, Schlattmann P, Telschow A, Werren JH. 2008. How many species
791 are infected with *Wolbachia*?- a statistical analysis of current data. *FEMS Microbiol Lett*
792 **281**:215–220.
793 Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. 2018. UFBoot2: Improving the
794 Ultrafast Bootstrap Approximation. *Mol Biol Evol* **35**:518–522.
795 Holmes EC. 2009. The Evolution and Emergence of RNA Viruses. Oxford University Press.
796 Hornett EA, Duplouy AMR, Davies N, Roderick GK, Wedell N, Hurst GDD, Charlat S. 2008. You
797 Can't Keep a Good Parasite Down: Evolution of a Male-Killer Suppressor Uncovers
798 Cytoplasmic Incompatibility. *Evolution* **62**:1258–1263.
799 Hosokawa T, Koga R, Kikuchi Y, Meng X-Y, Fukatsu T. 2010. *Wolbachia* as a bacteriocyte-
800 associated nutritional mutualist. *Proc Natl Acad Sci U S A* **107**:769–774.
801 Ho SYW, Phillips MJ, Cooper A, Drummond AJ. 2005. Time dependency of molecular rate estimates
802 and systematic overestimation of recent divergence times. *Mol Biol Evol* **22**:1561–1568.
803 Hou L, Liu Y, Gao Q, Xu X, Ning M, Bi J, Liu H, Liu M, Gu W, Wang W, Meng Q. 2017.
804 *Spiroplasma eriocheiris* Adhesin-Like Protein (ALP) Interacts with Epidermal Growth Factor
805 (EGF) Domain Proteins to Facilitate Infection. *Front Cell Infect Microbiol*. **7**:13.
806 Hurst GDD, Jiggins FM. 2000. Male-killing bacteria in insects: mechanisms, incidence, and
807 implications. *Emerg Infect Dis* **6**:329–336.
808 Ipoutcha T, Tsarmpopoulos I, Talenton V, Gaspin C, Moisan A, Walker CA, Brownlie J, Blanchard
809 A, Thebault P, Sirand-Pugnet P. 2019. Multiple Origins and Specific Evolution of CRISPR/Cas9
810 Systems in Minimal Bacteria (Mollicutes). *Front Microbiol* **10**:2701.
811 Itoh T, Martin W, Nei M. 2002. Acceleration of genomic evolution caused by enhanced mutation rate
812 in endocellular symbionts. *Proc Natl Acad Sci U S A* **99**:12944–12948.
813 Jaenike J. 2012. Population genetics of beneficial heritable symbionts. *Trends Ecol Evol* **27**:226–232.
814 Jaenike J, Polak M, Fiskin A, Helou M, Minhas M. 2007. Interspecific transmission of endosymbiotic
815 *Spiroplasma* by mites. *Biol Lett* **3**:23–25.
816 Jaenike J, Unckless R, Cockburn SN, Boelio LM, Perlman SJ. 2010. Adaptation via Symbiosis:
817 Recent Spread of a *Drosophila* Defensive Symbiont. *Science* **329**:212–215.
818 Jiggins FM, Hurst GDD, Jiggins CD, von der Schulenburg JHG, Majerus MEN. 2000. The butterfly
819 *Danaus chrysippus* is infected by a male-killing *Spiroplasma* bacterium. *Parasitology* **120**:439–
820 446.
821 Joshi BD, Berg M, Rogers J, Fletcher J, Melcher U. 2005. Sequence comparisons of plasmids pBJS-O
822 of *Spiroplasma citri* and pSKU146 of *S. kunkelii*: implications for plasmid evolution. *BMC*
823 *Genomics* **6**:175.
824 Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. 2017. ModelFinder: fast
825 model selection for accurate phylogenetic estimates. *Nat Methods* **14**:587–589.
826 Kanehisa M, Goto S, Hattori M, Aoki-Kinoshita KF, Itoh M, Kawashima S, Katayama T, Araki M,
827 Hirakawa M. 2006. From genomics to chemical genomics: new developments in KEGG. *Nucleic*

828 *Acids Res* **34**:D354–7.
829 Kanehisa M, Sato Y, Morishima K. 2016. BlastKOALA and GhostKOALA: KEGG Tools for
830 Functional Characterization of Genome and Metagenome Sequences. *J Mol Biol* **428**:726–731.
831 Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7:
832 improvements in performance and usability. *Mol Biol Evol* **30**:772–780.
833 Koonin EV, Mushegian AR, Rudd KE. 1996. Sequencing and analysis of bacterial genomes. *Current*
834 *Biology* **6**:404–416.
835 Krogh A, Larsson B, von Heijne G, Sonnhammer EL. 2001. Predicting transmembrane protein
836 topology with a hidden Markov model: application to complete genomes. *J Mol Biol* **305**:567–
837 580.
838 Ku C, Lo W-S, Chen L-L, Kuo C-H. 2013. Complete genomes of two dipteran-associated
839 spiroplasmas provided insights into the origin, dynamics, and impacts of viral invasion in
840 spiroplasma. *Genome Biol Evol* **5**:1151–1164.
841 Kuo C-H, Moran NA, Ochman H. 2009. The consequences of genetic drift for bacterial genome
842 complexity. *Genome Res* **19**:1450–1454.
843 Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, Salzberg SL. 2004. Versatile
844 and open software for comparing large genomes. *Genome Biol* **5**:R12.
845 Lanfear R, Calcott B, Ho SYW, Guindon S. 2012. Partitionfinder: combined selection of partitioning
846 schemes and substitution models for phylogenetic analyses. *Mol Biol Evol* **29**:1695–1701.
847 LeClerc JE, Li B, Payne WL, Cebula TA. 1996. High mutation frequencies among *Escherichia coli*
848 and *Salmonella* pathogens. *Science* **274**:1208–1211.
849 Lee H, Popodi E, Tang H, Foster PL. 2012. Rate and molecular spectrum of spontaneous mutations in
850 the bacterium *Escherichia coli* as determined by whole-genome sequencing. *Proc Natl Acad Sci*
851 *U S A* **109**:E2774–83.
852 Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**:3094–3100.
853 Lo W-S, Huang Y-Y, Kuo C-H. 2016. Winding paths to simplicity: genome evolution in facultative
854 insect symbionts. *FEMS Microbiol Rev* **40**:855–874.
855 Lynch M. 2006. Streamlining and simplification of microbial genome architecture. *Annu Rev*
856 *Microbiol* **60**:327–349.
857 Madeira F, Park YM, Lee J, Buso N, Gur T, Madhusoodanan N, Basutkar P, Tivey ARN, Potter SC,
858 Finn RD, Lopez R. 2019. The EMBL-EBI search and sequence analysis tools APIs in 2019.
859 *Nucleic Acids Res* **47**:W636–W641.
860 Marais GAB, Batut B, Daubin V. 2020. Genome Evolution: Mutation Is the Main Driver of Genome
861 Size in Prokaryotes. *Curr Biol* **30**:R1083–R1085.
862 Massey SE. 2008. The proteomic constraint and its role in molecular evolution. *Mol Biol Evol*
863 **25**:2557–2565.
864 Masson F, Calderon-Copete S, Schüpfer F, Vigneron A, Rommelaere S, Garcia-Arraez MG, Paredes
865 JC, Lemaitre B. 2020. Blind killing of both male and female *Drosophila* embryos by a natural
866 variant of the endosymbiotic bacterium *Spiroplasma poulsonii*. *Cell Microbiol* **22**:e13156.
867 Masson F, Copete SC, Schüpfer F, Garcia-Arraez G, Lemaitre B. 2018. In Vitro Culture of the Insect
868 Endosymbiont *Spiroplasma poulsonii* Highlights Bacterial Genes Involved in Host-Symbiont
869 Interaction. *mBio* **9**:e00024-18.
870 Mateos M, Castrezana SJ, Nankivell BJ, Estes AM, Markow T a., Moran NA. 2006. Heritable
871 endosymbionts of *Drosophila*. *Genetics* **174**:363–376.
872 McCutcheon JP, Moran NA. 2011. Extreme genome reduction in symbiotic bacteria. *Nat Rev*
873 *Microbiol* **10**:13–26.
874 McFall-Ngai M, Hadfield MG, Bosch TCG, Carey HV, Domazet-Lošo T, Douglas AE, Dubilier N,
875 Eberl G, Fukami T, Gilbert SF, Hentschel U, King N, Kjelleberg S, Knoll AH, Kremer N,
876 Mazmanian SK, Metcalf JL, Nealson K, Pierce NE, Rawls JF, Reid A, Ruby EG, Rumpho M,
877 Sanders JG, Tautz D, Wernegreen JJ. 2013. Animals in a bacterial world, a new imperative for
878 the life sciences. *Proc Natl Acad Sci U S A* **110**:3229–3236.
879 Miller DE, Staber C, Zeitlinger J, Hawley RS. 2018. Highly Contiguous Genome Assemblies of 15

880 Species Generated Using Nanopore Sequencing. *G3- Genes Genom Genet* **8**:3131–3141.
881 Mitrofanov A, Alkhnbashi OS, Shmakov SA, Makarova KS, Koonin EV, Backofen R. 2020.
882 CRISPRIdentify: Identification of CRISPR arrays using machine learning approach- *bioRxiv*
883 2020.11.05.369512.
884 Montenegro H, Solferini VN, Klaczko LB, Hurst GDD. 2005. Male-killing *Spiroplasma* naturally
885 infecting *Drosophila melanogaster*. *Insect Mol Biol* **14**:281–287.
886 Montenegro H, Souza WN, da Silva Leite D, Klaczko LB. 2000. Male-killing selfish cytoplasmic
887 element causes sex-ratio distortion in *Drosophila melanogaster*. *Heredity* **85 Pt 5**:465–470.
888 Moran NA. 2003. Tracing the evolution of gene loss in obligate bacterial symbionts. *Curr Opin*
889 *Microbiol* **6**:512–518.
890 Moran NA. 1996. Accelerated evolution and Muller's rachet in endosymbiotic bacteria. *Proc Natl*
891 *Acad Sci U S A* **93**:2873–2878.
892 Moran NA, McCutcheon JP, Nakabachi A. 2008. Genomics and evolution of heritable bacterial
893 symbionts. *Annu Rev Genet* **42**:165–190.
894 Moran NA, McLaughlin HJ, Sorek R. 2009. The dynamics and time scale of ongoing genomic erosion
895 in symbiotic bacteria. *Science* **323**:379–382.
896 Mouches C, Barroso G, Gadeau A, Bové JM. 1984. Characterization of two cryptic plasmids from
897 *Spiroplasma citri* and occurrence of their DNA sequences among various spiroplasmas. *Ann Inst*
898 *Pasteur Microbiol* **135**:17–24.
899 Nakayama S, Parratt SR, Hutchence KJ, Lewis Z, Price TAR, Hurst GDD. 2015. Can maternally
900 inherited endosymbionts adapt to a novel host? Direct costs of *Spiroplasma* infection, but not
901 vertical transmission efficiency, evolve rapidly after horizontal transfer into *D. melanogaster*.
902 *Heredity* **114**:539–543.
903 Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: A fast and effective stochastic
904 algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* **32**:268–274.
905 Oliver KM, Smith AH, Russell JA. 2014. Defensive symbiosis in the real world - advancing
906 ecological studies of heritable, protective bacteria in aphids and beyond. *Funct Ecol* **28**:341–355.
907 Osaka R, Ichizono T, Kageyama D, Nomura M, Watada M. 2013. Natural variation in population
908 densities and vertical transmission rates of a *Spiroplasma* endosymbiont in *Drosophila hydei*.
909 *Symbiosis* **60**:73–78.
910 Ota T, Kawabe M, Oishi K, Poulson DF. 1979. Non-male-killing spiroplasmas in *Drosophila hydei*. *J*
911 *Hered* **70**:211–213.
912 Paradis E, Schliep K. 2019. ape 5.0: an environment for modern phylogenetics and evolutionary
913 analyses in R. *Bioinformatics* **35**:526–528.
914 Paredes JC, Herren JK, Schüpfer F, Lemaitre B. 2016. The Role of Lipid Competition for
915 Endosymbiont-Mediated Protection against Parasitoid Wasps in *Drosophila*. *mBio* **7**:e01006–16.
916 Paredes JC, Herren JK, Schüpfer F, Marin R, Claverol S, Kuo C-H, Lemaitre B, Béven L. 2015.
917 Genome Sequence of the *Drosophila melanogaster* Male-Killing *Spiroplasma* Strain MSRO
918 Endosymbiont. *mBio* **6**:e02437–14.
919 Poulson DF, Sakaguchi B. 1961. Nature of "sex-ratio" agent in *Drosophila*. *Science* **133**:1489–1490.
920 Razin S. 2006. The Genus Mycoplasma and Related Genera (Class Mollicutes) In: Dworkin M.,
921 Falkow S., Rosenberg E., Schleifer KH., Stackebrandt E., editor. The Prokaryotes. New York:
922 Springer. pp. 836–904.
923 R Core Team. 2016. R: A language and environment for statistical computing. R Foundation for
924 Statistical Computing, Vienna, Austria.
925 Regassa LB, Gasparich GE. 2006. Spiroplasmas: evolutionary relationships and biodiversity. *Front*
926 *Biosci* **11**:2983–3002.
927 Richardson MF, Weinert LA, Welch JJ, Linheiro RS, Magwire MM, Jiggins FM, Bergman CM. 2012.
928 Population Genomics of the *Wolbachia* Endosymbiont in *Drosophila melanogaster*. *PLoS Genet*
929 **8**:e1003129.
930 Rocha EPC, Blanchard A. 2002. Genomic repeats, genome plasticity and the dynamics of
931 *Mycoplasma* evolution. *Nucleic Acids Res* **30**:2031–2042.

932 Rocha EPC, Danchin A. 2002. Base composition bias might result from competition for metabolic
933 resources. *Trends Gene* **18**:291–294.
934 Russel J, Pinilla-Redondo R, Mayo-Muñoz D, Shah SA, Sørensen SJ. 2020. CRISPRCasTyper: An
935 automated tool for the identification, annotation and classification of CRISPR-Cas loci. *bioRxiv*
936 2020.05.15.097824.
937 Saglio PHM, Whitcomb RF. 1979. Diversity of wall-less prokaryotes in plant vascular tissue, fungi,
938 and invertebrate animals In: Whitcomb RF, Tully JG, editors. The Mycoplasmas. New York:
939 Academic Press. pp. 1–36.
940 Saillard C, Carle P, Duret-Nurbel S, Henri R, Killiny N, Carrère S, Gouzy J, Bové J-M, Renaudin J,
941 Foissac X. 2008. The abundant extrachromosomal DNA content of the *Spiroplasma* citri GII3-
942 3X genome. *BMC Genomics* **9**:195.
943 Sanada-Morimura S, Matsumura M, Noda H. 2013. Male Killing Caused by a *Spiroplasma* Symbiont
944 in the Small Brown Planthopper, *Laodelphax striatellus*. *J Hered* **104**:821–829.
945 Schneider DI, Saarman N, Onyango MG, Hyseni C, Opiro R, Echodu R, O'Neill M, Bloch D,
946 Vigneron A, Johnson TJ, Dion K, Weiss BL, Opiyo E, Caccone A, Aksoy S. 2019. Spatio-
947 temporal distribution of *Spiroplasma* infections in the tsetse fly (*Glossina fuscipes fuscipes*) in
948 northern Uganda. *PLoS Neglect Trop D* **13**:e0007340.
949 Seemann T. 2015. snippy: fast bacterial variant calling from NGS reads.
950 Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**:2068–2069.
951 Shigenobu S, Watanabe H, Hattori M, Sakaki Y, Ishikawa H. 2000. Genome sequence of the
952 endocellular bacterial symbiont of aphids *Buchnera* sp. APS. *Nature* **407**:81–86.
953 Subramanian B, Gao S, Lercher MJ, Hu S, Chen W-H. 2019. Evolview v3: a webserver for
954 visualization, annotation, and management of phylogenetic trees. *Nucleic Acids Res* **47**:W270–
955 W275.
956 Teixeira L, Ferreira A, Ashburner M. 2008. The bacterial symbiont *Wolbachia* induces resistance to
957 RNA viral infections in *Drosophila melanogaster*. *PLoS Biol* **6**:e1000002.
958 Tinsley MC, Majerus MEN. 2006. A new male-killing parasitism: *Spiroplasma* bacteria infect the
959 ladybird beetle *Anisosticta novemdecimpunctata* (Coleoptera: Coccinellidae). *Parasitology*
960 **132**:757–765.
961 Turelli M, Cooper BS, Richardson KM, Ginsberg PS, Peckenpaugh B, Antelope CX, Kim KJ, May
962 MR, Abrieux A, Wilson DA, Bronski MJ, Moore BR, Gao J-J, Eisen MB, Chiu JC, Conner WR,
963 Hoffmann AA. 2018. Rapid Global Spread of *w*Ri-like *Wolbachia* across Multiple *Drosophila*.
964 *Curr Biol* **28**:963–971.e8.
965 Turelli M, Hoffmann AA. 1991. Rapid spread of an inherited incompatibility factor in California
966 *Drosophila*. *Nature* **353**:440–442.
967 Viver T, Orellana LH, Hatt JK, Urdiain M, Díaz S, Richter M, Antón J, Avian M, Amann R,
968 Konstantinidis KT, Rosselló-Móra R. 2017. The low diverse gastric microbiome of the jellyfish
969 *Cotylorhiza tuberculata* is dominated by four novel taxa. *Environ Microbiol* **19**:3039–3058.
970 Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J,
971 Young SK, Earl AM. 2014. Pilon: an integrated tool for comprehensive microbial variant
972 detection and genome assembly improvement. *PLoS One* **9**:e112963.
973 Watts T, Haselkorn TS, Moran NA, Markow TA. 2009. Variable Incidence of *Spiroplasma* Infections
974 in Natural Populations of *Drosophila* Species. *PLoS One* **4**:e5703.
975 Weinert LA, Araujo-Jnr EV, Ahmed MZ, Welch JJ. 2015. The incidence of bacterial endosymbionts
976 in terrestrial arthropods. *Proc R Soc B* **282**:20150249.
977 Wickham H. 2009. ggplot2: Elegant Graphics for Data Analysis. Springer Science & Business Media.
978 Wick RR, Judd LM, Gorrie CL, Holt KE. 2017a. Completing bacterial genome assemblies with
979 multiplex MinION sequencing. *Microb Genom* **3**:e000132.
980 Wick RR, Judd LM, Gorrie CL, Holt KE. 2017b. Unicycler: Resolving bacterial genome assemblies
981 from short and long sequencing reads. *PLoS Comput Biol* **13**:e1005595.
982 Wick RR, Schultz MB, Zobel J, Holt KE. 2015. Bandage: interactive visualization of de novo genome
983 assemblies. *Bioinformatics* **31**:3350–3352.

984 Wilke CO. 2020. ggridges: Ridgeline Plots in "ggplot2."
985 Wilke CO. 2019. cowplot: Streamlined Plot Theme and Plot Annotations for "ggplot2."
986 Woese CR, Stackebrandt E, Ludwig W. 1984. What are mycoplasmas: the relationship of tempo and
987 mode in bacterial evolution. *J Mol Evol* **21**:305–316.
988 Wu M, Sun LV, Vamathevan J, Riegler M, Deboy R, Brownlie JC, McGraw EA, Martin W, Esser C,
989 Ahmadinejad N, Wiegand C, Madupu R, Beanan MJ, Brinkac LM, Daugherty SC, Durkin AS,
990 Kolonay JF, Nelson WC, Mohamoud Y, Lee P, Berry K, Young MB, Utterback T, Weidman J,
991 Nierman WC, Paulsen IT, Nelson KE, Tettelin H, O'Neill SL, Eisen JA. 2004. Phylogenomics of
992 the reproductive parasite *Wolbachia pipientis w*Mel: a streamlined genome overrun by mobile
993 genetic elements. *PLoS Biol* **2**:e69.
994 Xie JL, Butler S, Sanchez G, Mateos M. 2014. Male killing *Spiroplasma* protects *Drosophila*
995 *melanogaster* against two parasitoid wasps. *Heredity* **112**:399–408.
996 Xie JL, Vilchez I, Mateos M. 2010. *Spiroplasma* Bacteria Enhance Survival of *Drosophila hydei*
997 Attacked by the Parasitic Wasp *Leptopilina heterotoma*. *PLoS One* **5**.
998 doi:10.1371/journal.pone.0012149
999 Xie JL, Winter C, Winter L, Mateos M. 2014. Rapid spread of the defensive endosymbiont
1000 *Spiroplasma* in *Drosophila hydei* under high parasitoid was pressure. *FEMS Microbiol Ecol*
1001 **91**:1–11.
1002 Yamada M-A, Nawa S, Watanabe TK. 1982. A mutant of SR organism (SRO) in *Drosophila* that does
1003 not kill the host males. *Jpn J Genet* **57**:301–305.
1004 Ye F, Melcher U, Rascoe JE, Fletcher J. 1996. Extensive chromosome aberrations in *Spiroplasma*
1005 *citri* strain BR3. *Biochem Genet* **34**:269–286.
1006 Yu G, Smith DK, Zhu H, Guan Y, Lam TT. 2017. ggtree : an r package for visualization and
1007 annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol*
1008 *Evol* **8**:28–36.

1009

1010 **Supplementary material**

1011

1012 **Figure S1.** Mutational bias in *s*Hy evolution. Rates were calculated by counting the

1013 corresponding numbers of snps and dividing by the total number of AT or GC positions in the

1014 *s*Hy genome.

1015 **Figure S2.** Predicted prophage regions in *s*Hy and *s*Mel by PHASTER (blue) and PhiSpy

1016 (yellow).

1017 **Figure S3.** Plasmid synteny in *Spiroplasma poulsonii* strains. Numbers on ticks correspond to

1018 nucleotide position on plasmids in kb. Sequence similarities (>=60% and spanning >= 2kb)

1019 between plasmids are indicated by ribbons. Abbreviations: ARP - adhesion related protein,

1020 OTU - ovarian tumor domain containing protein (in *s*Mel: Spaid), ParA - plasmid partitioning

1021 protein like, RIP - ribosome inactivating domain containing protein, Spiralin - spiralin like
1022 protein. Figure was created with AliTV.

1023 **Figure S4.** Prophage loci and CRISPR/Cas systems in *Spiroplasma*. **a)** Summary of prophage
1024 gene tree-species tree reconciliations. All events (gene speciations, duplications, transfers and
1025 losses) were inferred using GeneRax and are mapped onto the *Spiroplasma* phylogeny based
1026 on single copy orthologs present in all strains. **b)** CRISPR/Cas systems and arrays as
1027 predicted by CCtyper and CRISPRidentify, respectively. Maximum likelihood phylogeny is
1028 based on aligned Cas9 protein sequences, and was reconstructed using IQ-TREE. Note that
1029 CRISPR/Cas is absent in all Citri and Poulsonii strains. Several strains have lost or reduced
1030 parts of the CRISPR/Cas system, which therefore are likely not functional.

1031 **Figure S5.** Correlations between genome size of investigated *Spiroplasma* strains and **a)** The
1032 number of IS elements as predicted by prokka; **b)** proportion of the genome encoding for
1033 prophages as predicted by PhiSpy. Abbreviations: R - Pearson's correlation coefficients; p -
1034 p-values.

1035 **Figure S6.** RIP loci in different *Spiroplasma* genomes. Maximum likelihood tree was
1036 calculated with IQ-TREE based on an alignment of RIP domains (288 amino acid positions),
1037 created using the hmmalign function of the HMMER software, and manually trimmed to
1038 exclude positions present in < 3 sequences. Domain prediction is based on PfamScan,
1039 SignalP, and TMHMM. Clades identified by Ballinger & Perlmann (2017) are indicated, and
1040 sequences from *s*Hy highlighted in bold font. UFB Bootstrap values ≥90 are shown on nodes.
1041 Abbreviations: SP - signal peptide, TM - transmembrane helix, Ank - ankyrin repeats, RIP -
1042 ribosome inactivating domain containing protein

1043 **Table S1.** *Spiroplasma* genomes included in comparative genomic analyses, with accession

1044 numbers, genome properties, and phenotypes induced in host.

1045 **Table S2.** Sequencing and mapping statistics for all *s*Hy and *s*Mel libraries used to infer

1046 variants.

1047 **Table S3.** List of changes observed in *s*Hy and *s*Mel over the course of the evolution

1048 experiments.

1049 **Table S4.** List of missing and truncated genes in *s*Hy and *s*Mel with KEGG annotation.