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. In arthropods, *Spiroplasma* are found across a broad host range, but typically with lower incidence than other bacteria with similar ecology, such as *Wolbachia* or *Rickettsia*. *Spiroplasma* symbionts of *Drosophila* are best known as male-killers and protective symbionts, and both phenotypes are mediated by *Spiroplasma*-encoded toxins. *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. 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 markedly increased compared with other symbionts. 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 may be explained by 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.

Introduction

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

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

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

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

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

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

In this study, we systematically investigate rates and patterns of *Spiroplasma* evolution. To this end, we employed two strains of *Spiroplasma poulsonii*: 1) *s*Hy is a natural associate of *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 (J. Xie et al., 2014). Using the previously established
116 complete genome (Paredes et al., 2015) and cell-free culture (Masson et al., 2018), we
117 re-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 **Materials and Methods**

127 *Spiroplasma* strains and genome sequencing

128 We documented the evolution of the *Spiroplasma poulsonii* strains *s*Hy and *s*Mel in fly hosts
129 and in cell free culture, respectively (Tab. 1). sHy is one of the *Spiroplasma* symbionts
130 naturally associated with *Drosophila hydei (Mateos et al., 2006)*. Our evolution experiments
131 covered ca. ten years and three variants of *s*Hy, all of which all originated from a single
132 isolate (Fig. 1). Over the course of the study, *Drosophila hydei* hosts were maintained on
133 standard corn meal agar diet (1% agarose, 8.5% sugar, 6% maize meal, 2% autolysed yeast,

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

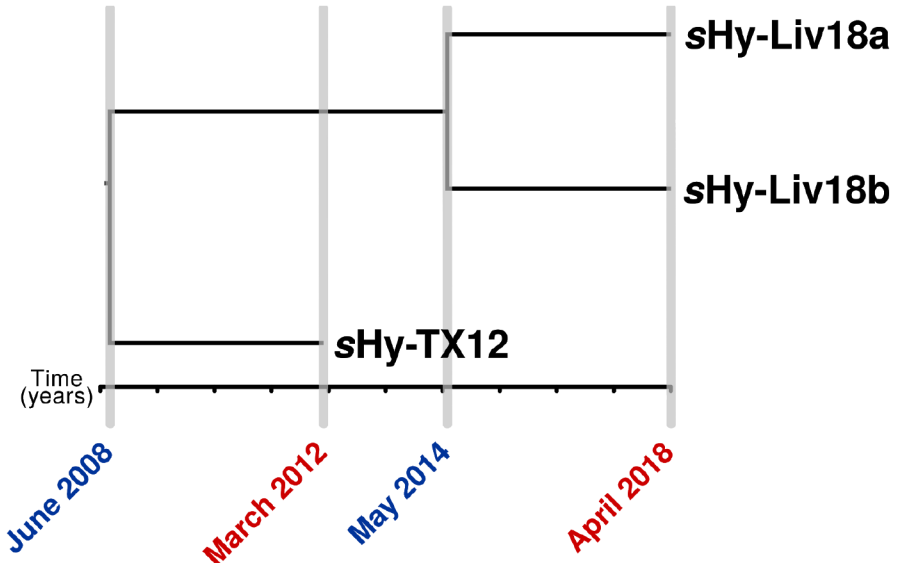
148 DNA extraction for the *Spiroplasma* strain *s*Hy-Tx followed the protocol described above,
149 with the following modifications: Immediately after the piercing, 35-40 flies were placed into
150 0.5 ml microcentrifuge tubes previously pierced in the bottom, which was placed within a 1.5
151 ml microcentrifuge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

158 method. The resulting library was subjected to Illumina HiSeq 2500 sequencing at the Texas

159 AgriLife Genomics and Bioinformatics Services Facility (College Station, TX).

160 **Tab. 1:** *Spiroplasma poulsonii* strains investigated in this study. Note that we follow the
161 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
sHy	sHy-Tx12	*Drosophila hydei*	01/2004 (Mateos et al., 2006)	Hemolymph extract, 03/2012	Hap_1, TEN104-106
	sHy-Liv18a	*Drosophila hydei*	06/2008 (copy of sHy-Tx12)	Hemolymph extract, 04/2018	-
	sHy-Liv18b	*Drosophila hydei*	03/2014 (copy of sHy-Liv18A)	Hemolymph extract, 04/2018	-
sMel	sMel-Ug16	*Drosophila melanogaster*	10/2016 (Masson et al., 2018)	Cell free culture, 4 time points 2016–18 (Tab. 2)	MSRO-Uganda
	sMel-Br18	*Drosophila melanogaster*	07/1997 (Montenegro et al., 2000)	Whole fly, 03/2018	MSRO-RED42



162 **Fig. 1:** Overview on *Spiroplasma poulsonii* sHy symbionts sampled from *Drosophila*
163 *hydei* cultures. The evolution of three strains was investigated over a total timespan of
164 ten years. Blue dates denote splitting of *D. hydei* cultures, red dates mark
165 *Spiroplasma* extraction for sequencing.

166 *s*Mel is a naturally occurring *Spiroplasma* symbiont of *D. melanogaster* best known for its
167 male-killing phenotype (Harumoto and Lemaitre, 2018; Montenegro et al., 2000). Cell free
168 culture of *s*Mel-Ug was recently accomplished (Masson et al., 2018), and we used a time
169 series of this culture to investigate *s*Mel molecular rates of evolution outside of host tissues.
170 The time series covered 29 months in total, and sequencing was performed at four time points
171 (Table 2). Illumina DNA sequencing was performed at MicrobesNG (Birmingham, UK) using
172 the "standard service". Culturing, passaging, and DNA extraction was performed as described
173 in Masson et al. (2018).

174 **Tab 2:** Timeline for *Spiroplasma poulsonii* *s*Mel culturing and sequencing.

ID	Date of isolation for sequencing	Age (months)
HL	11/10/2016	0 (start of culture)
P14	09/02/2017	4
P57	23/05/2018	19
P79	25/03/2019	29

175 We further sequenced a single isolate of sMel-Br deriving from a strain established in 1997
176 (Tab. 1). For Illumina sequencing, 1-2 grams of whole body flies from *Spiroplasma*-positive
177 *D. melanogaster* were collected for CTAB-Phenol based DNA extraction. Illumina libraries
178 were prepared with Illumina Truseq DNA kit and sequenced on an Illumina NovaSeq-6000
179 S2 150 paired end flow cell by Texas AgriLife Genomics and Bioinformatics Services
180 Facility (College Station, TX). For Oxford Nanopore MinIon sequencing, heads were first
181 removed by immersing whole flies in liquid Nitrogen, and vigorous shaking on a metal sieve.
182 The headless specimens were immediately subjected to phenol-chloroform extraction.
183 Nanopore library preparation was performed using the Ligation Sequencing kit 1D
184 (SQK-LSK108), following precautionary measures to ensure long read lengths. The library
185 was then sequenced on MinION sequencing device (MIN-101B) with an SpotON Flow Cell

186 Mark 1 (R9.5- FLO-MIN107.1). Base Calling was performed using Albacore version 2.0
187 (Oxford Nanopore, Oxford, UK) and the reads were processed with Porevhop version 0.24
188 (Wick et al., 2017a) to remove adapter sequences. A draft assembly of sMel-Br was created
189 using Illumina and Nanopore reads and Unicycler version 0.4.7 (Wick et al., 2017b). The
190 assembly was fragmented, but did contain 3 circular contigs with sequence similarity to other
191 *Spiroplasma* contigs.

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

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

206 All reads passing the quality checks were used to create an assembly with Unicycler version
207 0.4.7 (Wick et al., 2017b). The resulting assembly contained 6 circular chromosomes (~15 kb
208 – ~1.6 Mbp) with high similarity to the previously published *s*Mel genome as determined by

209 BLAST+ version 2.9 searches (Camacho et al., 2009) in Bandage (Wick et al., 2015). We also
210 created a hybrid assembly with Unicycler using all long reads and the *s*Hy-Liv18a Illumina
211 reads. The hybrid assembly was more fragmented, but did contain 3 contigs which spanned
212 almost the entire 1.6 Mbp circular contig of the long-read only assembly. Because hybrid
213 assemblies are typically more accurate than assemblies based on long reads only (De Maio et
214 al., 2019), we corrected the long read assembly using the hybrid contigs. This was achieved
215 by calling differences between the assemblies with the 'dnadiff' function implemented in
216 MUMmer version 3.18 (Kurtz et al., 2004), and correcting the assembly accordingly. Finally,
217 Pilon version 1.23 (Walker et al., 2014) was used to polish the corrected assembly, again
218 using the *s*Hy-Liv18a Illumina reads. After nine rounds of polishing with Pilon, no further
219 improvements could be observed and the assembly was considered final.

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

221 We determined changes in *s*Hy and *s*Mel chromosomes over time by employing the Snippy
222 pipeline version 4.1.0 (Seemann, 2015). We used our newly assembled *s*Hy genome and the
223 previously sequenced *s*Mel genome (Masson et al., 2018) as references, annotated using the
224 Prokka pipeline version 1.13 with standard parameters and the *Spiroplasma* genetic code
225 (translation table 4, Seemann, 2014). *s*Hy and *s*Mel variants determined by Snippy were
226 filtered to only include positions which were covered by all *s*Hy or all *s*Mel Illumina libraries
227 with at least 5x coverage, respectively. Further, to compare rates of molecular evolution
228 between *Spiroplasma* and other microbial species, we counted the number of changes in third
229 codon positions of coding sequences (CDS). For both strains, we only considered CDS that
230 were entirely covered by sequencing depth of at least 5x in all Illumina libraries. Rates of
231 molecular evolution were calculated by: number of observed changes at 3rd codon positions /

232 number of considered 3rd codon positions / time in years over which the changes were
233 observed. For *s*Hy, using *s*Hy-Liv18a as reference, we calculated two rate estimates, one
234 based on comparison with *s*Hy-Liv18b, and one by comparing *s*Hy-Tx12 to the reference
235 (Fig. 1). For *s*Mel, the hemolymph extract from which the culture was established was used
236 as the reference ('HL', Tab. 2), and three rate estimates were calculated based on the changes
237 observed in the three sampled time points of the culture (Tab. 2).

238 Comparative genomics

239 We first compared the newly assembled *s*Hy genome with the most closely related, fully
240 sequenced *Spiroplasma* genome, *s*Mel. Both genomes were annotated using the Kyoto
241 Encyclopedia of Genes and Genomes (KEGG) and BlastKOALA (Kanehisa et al., 2016,
242 2006) and orthologous protein sequences determined using OrthoFinder version 2.2.3 (Emms
243 and Kelly, 2015). Synteny between the genomes was assessed through aligning the genomes
244 with Minimap2 version 2.15 (Li, 2018). Genome degradation was investigated by
245 documenting the number of unique KEGG numbers for *s*Hy and *s*Mel, all of which were
246 verified by BLAST+ searches. For each strain, we considered genes to be likely
247 pseudogenized or truncated if the longest CDS within an orthogroup spanned at most 60% of
248 the CDS length from the other strain.

249 Secondly, we compared *s*Hy with 17 *Spiroplasma* genomes from the
250 Citri-Chrysopicola-Mirum clade (Table S1, Supplementary material for full list of names and
251 accession numbers). All genomes were annotated using Prokka version 1.4.0, and orthology
252 between predicted CDS was established using OrthoFinder. We extracted single-copy
253 orthologs present in all strains, and aligned each locus separately with the L-INS-i method
254 implemented in Mafft version 7.450 (Katoh and Standley, 2013). Recombination was

255 evaluated with pairwise homoplasy index and window sizes of 100, 50, and 20 amino acids
256 for each locus (Bruen et al., 2006), and any locus showing evidence for recombination was
257 excluded. We used IQ-TREE version 1.6.10 (Nguyen et al., 2015) to reconstruct a core
258 genome phylogeny from the remaining 96 loci (covering 26,019 amino acid positions). Each
259 locus was treated as a separate partition with a distinct evolutionary rate (Chernomor et al.,
260 2016), and optimal models and number of partitions was estimated with IQ-TREE
261 (Kalyaanamoorthy et al., 2017; Lanfear et al., 2012). Branch support was assessed using
262 1,000 replicates of ultrafast bootstraps (Hoang et al., 2018), and approximate likelihood ratio
263 test (Guindon et al., 2010) in IQ-TREE.

264 For all genomes, insertion sequence elements were annotated with Prokka. Prophage regions
265 were predicted with PhiSpy version 3.7.8, which uses typical prophage features (e.g., gene
266 length, strand directionality, AT/GC skews, insertion points, Akhter et al., 2012) for its
267 predictions (Akhter et al., 2012). We also used the Phaster web server, which uses sequence
268 similarities and other criteria to predict prophages (Arndt et al., 2016). We screened for the
269 presence of toxin genes implied in protective phenotypes (Ballinger and Perlman, 2017;
270 Hamilton et al., 2015) and male killing (Harumoto and Lemaitre, 2018). To this end, we
271 downloaded UniProt sequence alignments from the Pfam database (Finn et al., 2016) for the
272 protein families RIP (PF00161, ribosome inactivating protein) and OTU (PF02328, OTU-like
273 cysteine protease), respectively. We used these alignments as databases for searches with
274 HMMER version 3.2.1 (Eddy, 2011), and protein sequences from all genomes as queries.
275 Domain architecture of all matching proteins was then determined using PfamScan with an
276 e-value of 0.001 (Madeira et al., 2019), SignalP 5.0 (Almagro Armenteros et al., 2019), and
277 TMHMM 2.0 (Krogh et al., 2001). RIP domains showed a high degree of divergence, and
278 were therefore aligned to the reference RIP HMM profile using hmmalign from the HMMER

279 software package. From this alignment, positions present in fewer than three sequences were
280 removed, and a phylogeny reconstructed with IQ-TREE. OTU domains were aligned using
281 MAFFT, and a phylogeny reconstructed with IQ-TREE.

282 Thirdly, we determined plasmid synteny across *Spiroplasma poulsonii* strains. In addition to
283 the plasmids of *s*Mel and *s*Hy, we also included a plasmid of *s*Neo and all circular contigs
284 from our *s*Mel-Br assembly. Plasmids were annotated using Prokka version 1.13 with a
285 protein database composed of all plasmid proteins available on NCBI GenBank. Plasmid
286 alignment and visualisation was performed with AliTV version 1.06 (Ankenbrand et al.,
287 2017).

288 Data visualisation

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

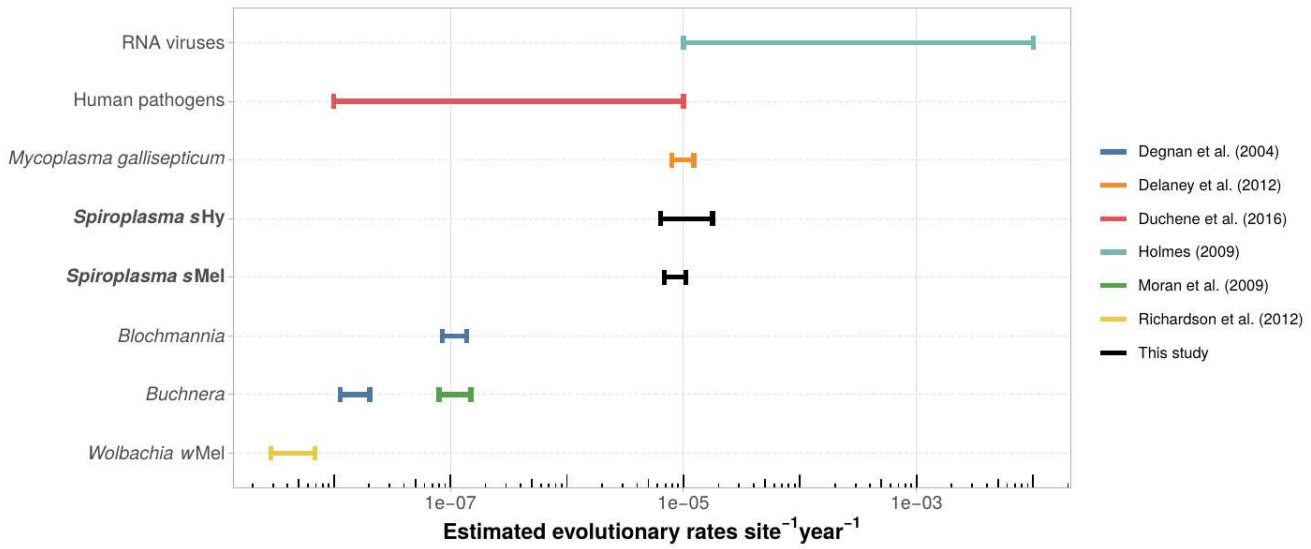
294 **Results**

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

296 The *Spiroplasma poulsonii s*Hy reference genome comprises a single circular chromosome
297 (1,625,797bp) and five circular contigs (23,069bp – 15,710bp) with sequence similarities to
298 plasmids from other *Spiroplasma*. The chromosome contains 1,584 predicted coding
299 sequences (CDS), a single ribosomal RNA cluster, and 30 tRNAs. Overall, genome content

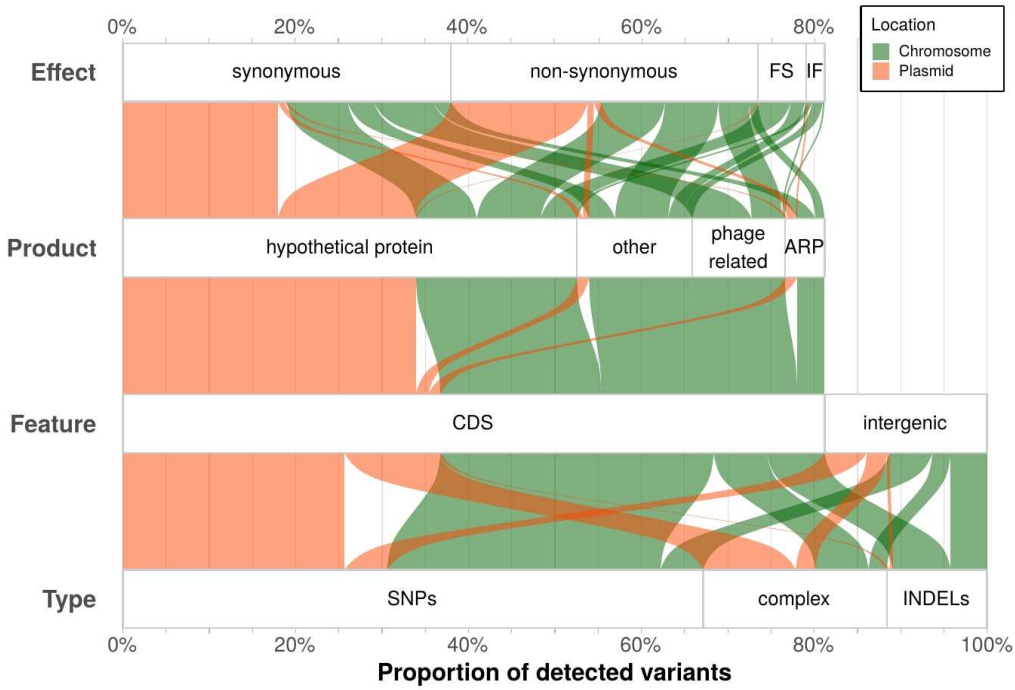
300 and metabolic capacities are highly similar to the previously sequenced, very closely related
301 *s*Mel genome (Paredes et al., 2015), and a detailed comparison follows further below.

302 To estimate rates of molecular evolution in *Spiroplasma poulsonii*, we measured
303 chromosome-wide changes in coding sequences of *Spiroplasma* from fly hosts (*s*Hy) and
304 axenic culture (*s*Mel) over time. Our estimates for *s*Hy and *s*Mel are overlapping and range
305 from $6.4 \times 10^{-6} - 1.8 \times 10^{-5}$ changes per position per year. This rate exceeds the rate reported for
306 other symbiotic bacteria such as *Wolbachia* and *Buchnera* by at least two orders of magnitude
307 (Fig. 2). Our estimate overlaps with rates calculated from some of fastest evolving human
308 pathogens (e.g., *Enterococcus faecium* and *Acinetobacter baumannii*), and with evolutionary
309 rates observed in the poultry pathogen *Mycoplasma gallisepticum* (which is closely related to
310 *Spiroplasma*). Indeed, *Spiroplasma* substitution rates fall at around the lower estimates for
311 viruses (Fig. 2).



312 **Fig 2.** Comparison of estimated evolutionary rates across various microbes. For details
313 on how evolutionary rates were calculated, please refer to the original studies and the
314 Materials and Methods section.

315 Over ~10 years of evolution, we observed a similar absolute number of variants on the
316 chromosome (~1.6 Mbp) and plasmids (~0.1 Mbp) of *s*Hy, i.e., relatively more changes on
317 the plasmids (Fig. 3, Table S2, supplementary material). Most variants in CDS affected
318 hypothetical proteins, but were enriched in prophage-associated loci and adhesin related
319 proteins (ARP, Fig. 3). Overall, *s*Hy variants in CDS were about equally often found to be
320 synonymous as non-synonymous (Fig. 3) and changes were biased towards GC > AT
321 substitutions (Fig. S1, Supplementary material). The changes in *s*Mel over ~2.5 years in
322 culture affected only 15 different CDS in total, of which four were ARPs, and three
323 lipoproteins (Table S2, supplementary material).



324 **Fig. 3:** Overview of all observed variants in *s*Hy (100% = 570 variants). All annotations
325 (Type, Feature, Product, Effect) are taken from results of the snippy pipeline.

326 Comparing the genomes of *s*Hy and *s*Mel revealed a notable contrast between the high degree
327 of nucleotide sequence identity on the one hand, and striking structural and gene content
328 differences on the other hand. In 604 single copy orthologues shared between the genomes,

329 average nucleotide identity was >99.5%. However, compared with *s*Mel, *s*Hy is reduced in
330 gene content (1,584 vs 2,388 genes), chromosome size (1.61 vs 1.88 Mbp), and coding
331 density (65% vs 82%). Further, *s*Hy contains fewer predicted insertion sequences (9 vs 89 in
332 *s*Mel) and intact prophages as predicted with Phaster (*s*Mel: 16 such regions spanning 426
333 Kb; *s*Hy: 2 regions, 14 Kb), but not PhiSpy (Fig. S2, Supplementary material). Both *s*Mel and
334 *s*Hy have a number of missing or truncated (i.e., potentially pseudogenized) genes when
335 compared with each other, but the level of genomic deterioration is higher in *s*Hy, and covers
336 a range of different genes (Fig. 4a, Table S3, supplementary material). Notable pseudogenised
337 or absent loci in *s*Hy include parts of the phosphotransferase system, in particular the loci
338 required for the uptake of N-Acetylmuramic Acid and N-acetylglucosamine (see Table S3,
339 Supplementary material for a full list).

340 Strikingly, there is evidence for a history of extensive chromosomal rearrangements since the
341 last common ancestor of *s*Hy and *s*Mel. The levels of synteny between the strains is much
342 lower than expected at the observed sequence similarity (Fig. 4b). Similarly, a comparison of
343 plasmids across different *Spiroplasma poulsonii* strains (*s*Hy, *s*Mel-Ug, *s*Mel-Br, and *s*Neo)
344 revealed similar gene content across the plasmids of different strains, but large differences in
345 arrangement and number (Fig. S3, Supplementary material).

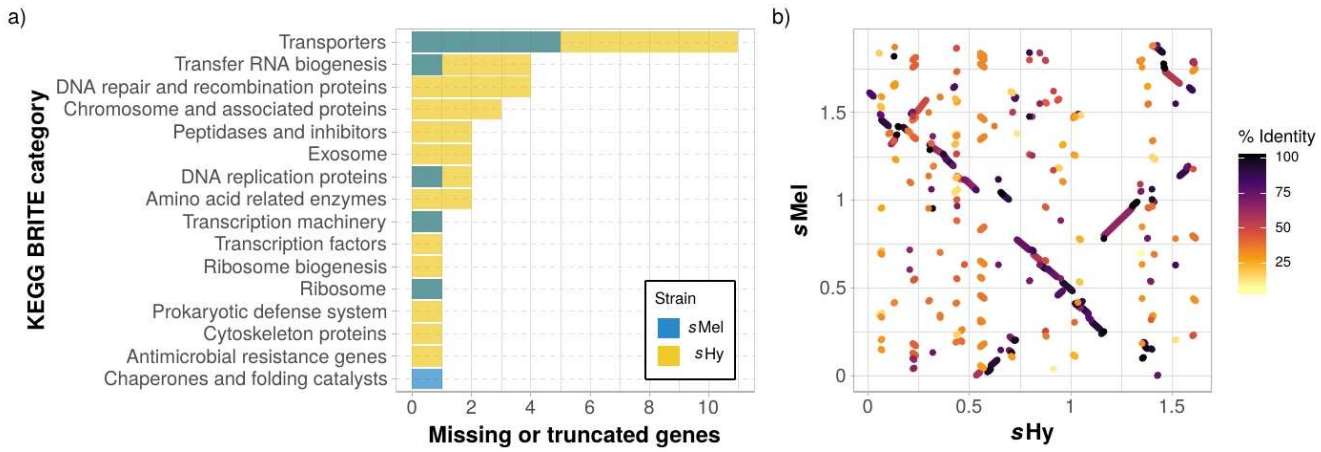


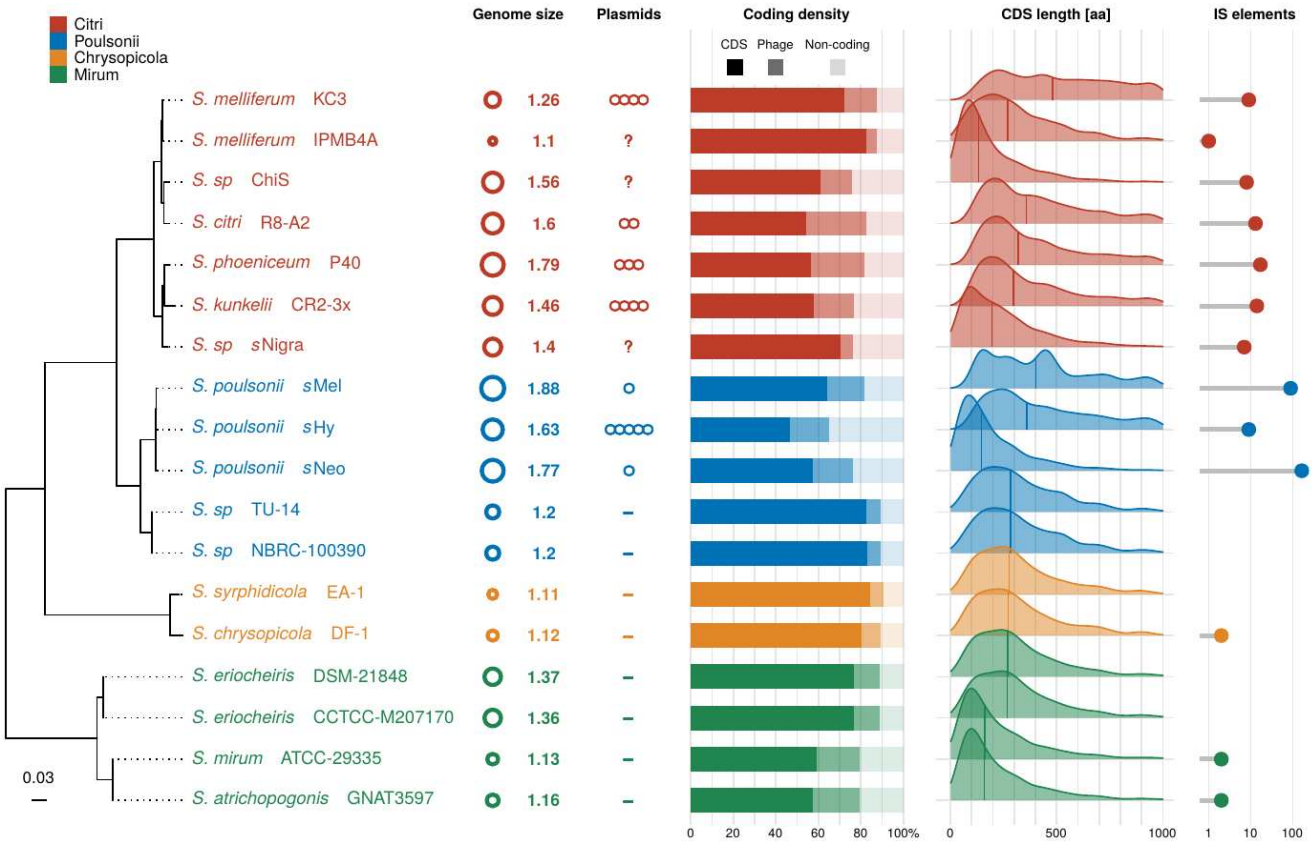
Fig 4: Comparison of *s*Hy and *s*Mel. **a)** Number of genes with KEGG annotation that are present in one strain, but absent or truncated in the other. See Table S3, supplementary material for a complete list. **b)** Synteny as determined with minimap2.

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 the genomes in draft status, and *S. citri* may have seven plasmids (Saillard et al., 2008). Prophage regions of varying sizes were predicted in all of the genomes (despite the lack of clear homologues to viral sequences in some of these, Ku et al., 2013), and strains with especially high numbers of prophage associated loci show lower median CDS length when compared to strains with very small predicted prophage regions. Further, in the bigger *Spiroplasma* genomes, the number of insertion sequences was typically higher (Fig. 5). The comparison indicates that *s*Hy may be more degraded than its closest relatives *s*Neo and

363 *s*Mel, with a smaller chromosome size, fewer predicted CDS and mobile genetic elements,

364 and the lowest coding density of these three (Fig. 5).



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

372 Spiroplasmas' most striking phenotypes in insects have been mechanistically linked to toxin

373 genes. For example, ribosome inactivating proteins (RIP) may protect *Spiroplasma* hosts by

374 cleaving ribosomal RNA of parasites and parasitoids (Hamilton et al., 2015). Five RIP loci

375 are present in *s*Mel (RIP1–5, of which 3–5 are almost identical copies). As expected, the

376 protective *s*Hy also encodes RIPs, but only has a single orthologue for RIP1. However, it

377 contains an additional RIP gene in three copies that appears to be absent in *s*Mel, and one of
378 these copies also contains ankyrin repeats (Fig. S4, Supplementary material).

379 Further, the male killing phenotype of *s*Mel was recently established to be caused by
380 *Spiroplasma* androcidin (Spaid), and both ankyrin repeats and a deubiquitinase domain
381 (OTU) of this gene are necessary to induce male killing (Harumoto and Lemaitre, 2018).
382 HMMER searches using OTU domain profiles revealed a number of *Spiroplasma* loci similar
383 to Spaid (Fig. 6) which however lack its characteristic domain composition. For example,
384 *s*Hy encodes three loci similar to Spaid: one lacks a signal peptide, one has no ankyrin
385 repeats, and another encodes an epsilon-toxin like domain in addition to OTU (Fig. 6). Other
386 bacterial loci with notable similarities to the *Spiroplasma* OTU could only be detected in the
387 symbiotic taxa *Rickettsia* and *Wolbachia*. In the phylogenetic reconstruction of OTU domains
388 rooted with the eukaryotic sequences (from the ciliate *Stentor coeruleus*), *Rickettsia* and
389 *Wolbachia* loci are nested within the *Spiroplasma* sequences. This topology suggests lateral
390 gene transfer of OTU domain containing proteins from *Spiroplasma* to the other intracellular
391 taxa, although differences in genetic code between *Spiroplasma* and other bacteria likely limit
392 the probability for transfer of functional protein coding genes.

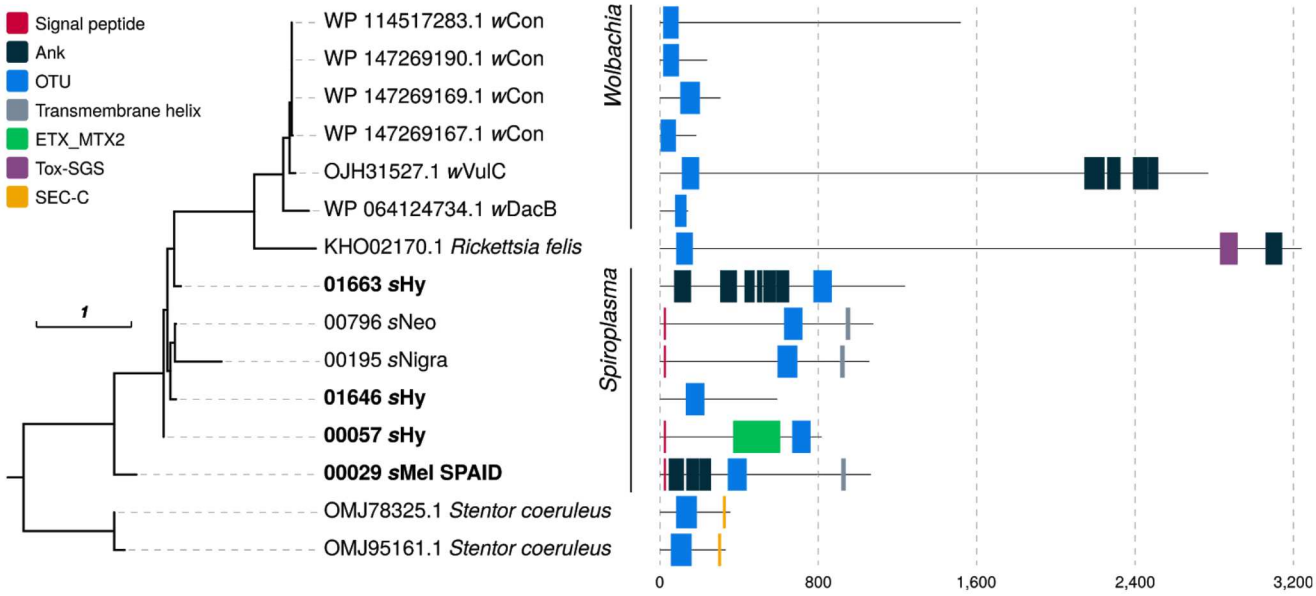


Fig. 6: Spaid-like proteins in *Spiroplasma* and other symbionts. Maximum likelihood tree was reconstructed from an alignment of OTU domains (147 amino acid positions), and domain predictions are based on PfamScan, SignalP, and TMHMM. Abbreviations (corresponding to PFAM domains): Ank - ankyrin repeats,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

Discussion

Spiroplasma, an exceptionally fast evolving symbiont

The substitution rates we observed in *Spiroplasma* symbionts are among the highest reported for any bacteria (Fig. 1). In a study comparing genome-wide evolutionary rates in bacterial human pathogens (Duchêne et al., 2016), most taxa showed considerably slower rates, and our estimate of *Spiroplasma* evolutionary rates overlaps with substitution rates of RNA viruses (Holmes, 2009). A number of reasons have been proposed for elevated substitution rates in bacterial symbionts and pathogens with small genome sizes: Firstly, host-associated microbes may acquire essential metabolic intermediates from their hosts and thus need not synthesise them (Moran, 2003). Obsolete metabolic genes subsequently accumulate now neutral substitutions, resulting in pseudogenization. Secondly, reduced population sizes of

410 symbionts, together with bottlenecks at transmission events limit purifying selection of
411 deleterious mutations (Moran, 1996) and mobile genetic elements (Lynch, 2006) – both in
412 turn create further genomic deterioration. Thirdly, genomes of host-associated microbes are
413 AT rich (Rocha and Danchin, 2002), which makes introduction of errors through replication
414 slippage more likely (Moran et al., 2009). Fourthly, many symbionts lack DNA repair genes
415 (Moran et al., 2008), which may result in increased substitution rates.

416 All of these points very likely apply for *Spiroplasma poulsonii*: its metabolic capacities are
417 reduced compared with free living microbes, it has a high proportion of pseudogenes and
418 prophages, very high AT content, and lacks several DNA repair genes. However, many
419 symbiotic bacteria show similar trends of genome evolution (Burger and Lang, 2003), and it
420 is therefore somewhat puzzling that *Spiroplasma* stands out with this exceptionally high
421 substitution rate. Loss of DNA repair genes has been shown to co-occur with elevated
422 mutation rates in intracellular bacteria (Itoh et al., 2002; Koonin et al., 1996; Massey, 2008),
423 and may explain our observations. The loci encoding the mismatch repair proteins *mutS* and
424 *mutL* are universally lacking in *Spiroplasma*, but are present in the slower evolving symbionts
425 *Wolbachia w*Mel, and *Buchnera aphidicola* (Richardson et al., 2012; Fig. S6, supplementary
426 material, Shigenobu et al., 2000; Wu et al., 2004). In *E. coli*, hypermutator strains that have
427 lost such loci have a ~10–300 fold increase of spontaneous mutation rates (LeClerc et al.,
428 1996; Lee et al., 2012), which is comparable to the ~two orders of magnitude difference
429 between substitution rates of *Buchnera* and *Spiroplasma* (Fig. 2). Like *Spiroplasma*,
430 *Mycoplasma* universally lacks the mismatch repair loci *mutS*, *mutL*, and *mutH* (Carvalho et
431 al., 2005; Chen et al., 2012). In line with this observation, elevated evolutionary rates were
432 hypothesized for mycoplasmas (Woese et al., 1984), and in our comparison, *M. gallisepticum*
433 appears to be the only bacterial taxon with substitution rates similar to *Spiroplasma* (Delaney

434 et al., 2012). Absence of DNA mismatch repair pathway may thus be ancestral to
435 Entomoplasmatales (Spiroplasmatacea + Entomoplasmataceae) and contribute to the dynamic
436 genome evolution across this taxon (Lo et al., 2016; Rocha and Blanchard, 2002).

437 Notably, substitution rates often appear larger when estimated over brief time periods (Ho et
438 al., 2005), and this may have biased our estimates. Duchene et al. (2016) found that
439 substitution rates measured over 10 years can be up to one order of magnitude larger than
440 those measured over 100 years. In agreement with such bias, we found more variants in our
441 *s*Mel culture after 19 months than for the 29 months isolate (Table S2, supplementary
442 material). The back mutations that likely have happened between 19 and 29 months of our
443 *s*Mel culture would go unnoticed when comparing genomes over larger time scales. More
444 generally, it is difficult to estimate divergence times between *Spiroplasma* strains. The
445 substitution rates estimated in *s*Hy and *s*Mel would suggest the two strains have diverged
446 1,120–2,260 years ago. However, this estimate is unreliable due to a number of factors that
447 we cannot control for. For example, the number of generations per year for *Drosophila* hosts
448 of *Spiroplasma* differ depending on species and location, and is expected to be lower in the
449 wild compared with laboratory strains. Further, *Spiroplasma* may move between species, for
450 example via ectoparasitic mites (Jaenike et al., 2007). A number of partially sympatric
451 *Drosophila* species carry similar *Spiroplasma* strains (*D. hydei*, *D. melanogaster*, *D.*
452 *willistoni*, *D. neotestacea*) (Haselkorn, 2010), and with our data it is impossible to determine
453 the number and direction of potential *Spiroplasma* transfers between these species.

454 Reductive genome evolution in *Spiroplasma*

455 According to a widely supported model of genome reduction in symbiotic bacteria
456 (McCutcheon and Moran, 2011), the first stages of host restriction involve accumulation of

457 pseudogenes and mobile genetic elements, chromosomal rearrangements, increased
458 substitution rates, and excess deletions. Advanced stages of adaptation to a host are
459 accompanied by further genome shrinkage, and purging of mobile genetic elements, which
460 overall result in more stable chromosomes. Using these characteristics, different levels of
461 genome reduction are apparent in the investigated *Spiroplasma* genomes, across which
462 genome sizes correlate positively with number of mobile genetic elements (plasmids,
463 prophages, insertion sequences, Fig. 5). On this spectrum, *S. syrphidicola* and *S. chrysopicola*
464 have the most reduced genomes, without homologues of prophage loci, and high levels of
465 synteny between the two strains (Ku et al., 2013). It was therefore argued that phages have
466 likely invaded *Spiroplasma* only after the split of the Syrphidicola and Citri+Poulsonii clades
467 (Ku et al., 2013). However, several observations support the conclusion that the reduced
468 *Spiroplasma* genomes have lost their prophage regions, rather than never having been
469 susceptible to viruses: 1) Genomes lacking prophage loci are also found within the
470 Citri+Poulsonii clade (TU-14 & NBRC100-390, Fig. 5), illustrating that prophage loss after
471 viral invasion is possible in *Spiroplasma*. 2) Strains lacking prophages show other hallmarks
472 of reduced symbiont genomes (small size, high coding density, lack of plasmids and
473 transposons, Fig. 5), which is in line with the model of genome reduction discussed above. 3)
474 Prophage prediction using an algorithm that does not mainly rely on sequence similarities to
475 known viral sequences (PhiSpy, Akhter et al., 2012) identified at least very small prophage
476 regions in all of the investigated genomes. In summary, while the genomic instability in
477 *Spiroplasma* genomes is likely driven by prophages (Ku et al., 2013; Lo et al., 2013; Ye et al.,
478 1996), several lineages display reduced genomes from which the phages have largely been
479 purged.

480 Using the model of genome reduction during host adaptation introduced above, *Spiroplasma*
481 *poulsonii*'s genome characteristics suggest that host restriction has happened relatively
482 recently. Although very closely related (99.5% sequence identity) and found in very similar
483 hosts, *s*Mel and *s*Hy differ markedly in coding density, genome size, and proportion of
484 prophage regions. Interestingly, PHASTER predicted 426 Kb of intact phage regions in *s*Mel,
485 but only 14 Kb for *s*Hy (Fig. S2, supplementary material). Both phage proliferation in *s*Mel,
486 and prophage loss in *s*Hy could have contributed to this. However, when using the similarity
487 agnostic tool PhiSpy, the differences in predicted prophage regions were less pronounced
488 (Fig. 5). This observation is compatible with degradation and/or pseudogenization of
489 prophage regions in *s*Hy, which would lead to reduced sequence similarity to viral loci (and
490 thus reduced detectability by PHASTER), but not entirely blur prophage characteristics
491 employed by PhiSpy (e.g., gene length, strand directionality, AT/GC skews, insertion points,
492 Akhter et al., 2012). Since the split of the lineages, *s*Hy has not only lost prophage regions
493 and insertion sequences compared with *s*Mel, but also several genes that are often lost in
494 adaptation to a host, such as parts of the phosphotransferase system for the uptake of
495 carbohydrates, one tRNA locus, and *uvrD* which plays a role in mismatch repair, and
496 nucleotide excision repair (Table S2, supplementary material). Using genome degradation as
497 a proxy for host adaptation, our findings collectively suggest that host adaptation is more
498 advanced in *s*Hy than in *s*Mel. This may be a result of the fitness benefits associated with *s*Hy
499 under parasitoid pressure, and the absence of detectable costs of *s*Hy infection in *Drosophila*
500 *hydei* (Osaka et al., 2013; Jialei Xie et al., 2014; Xie et al., 2010). However, the *Spiroplasma*
501 symbiont of *Drosophila neotestacea* *s*Neo is also protective, does not cause obvious fitness
502 costs (Jaenike et al., 2010), but has a less reduced genome (Fig.5, Ballinger and Perlman,
503 2017). Further, we cannot exclude a pathway in which the genome reduction in *s*Hy was

504 mainly driven by stochastic effects or even by adaptation to laboratory conditions, as we have
505 not investigated contemporary *s*Hy from wild *D. hydei* populations.

506 Implications for *Spiroplasma* evolutionary ecology

507 *Spiroplasma* evolutionary ecology shows several parallels to that of the most widely
508 distributed arthropod symbiont *Wolbachia*: both symbionts are found in a range of different
509 hosts (Duron et al., 2008), have the ability to invade novel hosts (Conner et al., 2017;
510 Haselkorn et al., 2009), may confer protection (Teixeira et al., 2008; Xie et al., 2010) but also
511 kill males (Hurst and Jiggins, 2000), and can spread across host populations swiftly
512 (Cockburn et al., 2013; Turelli and Hoffmann, 1991). However, there are also pronounced
513 differences between the symbionts: *Spiroplasma* rarely reaches the high infection frequencies
514 often observed in *Wolbachia* (Hilgenboecker et al., 2008; Watts et al., 2009), and is arguably
515 found in a more diverse host range that encompasses arthropods and plants (Regassa and
516 Gasparich, 2006), but also molluscs (Duperron et al., 2013), echinoderms (He et al., 2018),
517 and cnidarians (Viver et al., 2017). Further, *Wolbachia* show greatly reduced spontaneous
518 mutation rates compared with *Spiroplasma*, likely caused by a more complete set of DNA
519 repair genes (Fig. S5, supplementary material).

520 In theory, fast evolutionary rates should enable *Spiroplasma* to adapt to novel hosts quickly
521 (i.e., to reduce pathogenicity, and to maximise vertical transmission efficiency), and
522 experimental studies have found high horizontal transmission efficiency of *Spiroplasma*
523 (Haselkorn et al., 2013; Haselkorn and Jaenike, 2015; Nakayama et al., 2015). Consistent
524 with this, we found that genes implicated in host-symbiont compatibility and virulence (e.g.,
525 adhesion-related proteins) have evolved especially fast in our evolution experiments (Fig. 2).
526 In addition, we documented dynamic evolution and turnover of toxin loci, which are

important for host fitness and symbiont compatibility (Fig. 6, Fig. S4, supplementary material). Such 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 abundance 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.

547 **Author contributions**

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

554 **Acknowledgements**

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

561 **Data accessibility**

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

References

Akhter S, Aziz RK, Edwards RA. 2012. PhiSpy: a novel algorithm for finding prophages in bacterial genomes that combines similarity- and composition-based strategies. *Nucleic Acids Res* **40**:e126.

Almagro Armenteros JJ, Tsirigos KD, Sønderby CK, Petersen TN, Winther O, Brunak S, von Heijne G, Nielsen H. 2019. SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat Biotechnol* **37**:420–423.

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

Ankenbrand MJ, Hohlfeld S, Hackl T, Förster F. 2017. AliTV—interactive visualization of whole genome comparisons. *PeerJ Computer Science* **3**:e116.

Arndt D, Grant JR, Marcu A, Sajed T, Pon A, Liang Y, Wishart DS. 2016. PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Res* **44**:W16–21.

Ballinger MJ, Perlman SJ. 2019. The defensive *Spiroplasma*. *Curr Opin Insect Sci* **32**:36–41.

Ballinger MJ, Perlman SJ. 2017. Generality of toxins in defensive symbiosis: Ribosome-inactivating proteins and defense against parasitic wasps in *Drosophila*. *PLoS Pathog* **13**:e1006431.

Berho N, Duret S, Danet J-L, Renaudin J. 2006. Plasmid pSci6 from *Spiroplasma citri* GII-3 confers insect transmissibility to the non-transmissible strain *S. citri* 44. *Microbiology* **152**:2703–2716.

Bruen TC, Philippe H, Bryant D. 2006. A simple and robust statistical test for detecting the presence of recombination. *Genetics* **172**:2665–2681.

Brunson JC. 2019. ggalluvial: Alluvial Plots in "ggplot2."

Burger G, Lang BF. 2003. Parallels in genome evolution in mitochondria and bacterial symbionts. *IUBMB Life* **55**:205–212.

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. *BMC Bioinformatics* **10**:421.

Carle P, Saillard C, Carrère N, Carrère S, Duret S, Eveillard S, Gaurivaud P, Gourgues G, Gouzy J, Salar P, Verdin E, Breton M, Blanchard A, Laigret F, Bové J-M, Renaudin J, Foissac X. 2010. Partial chromosome sequence of *Spiroplasma citri* reveals extensive viral invasion and important gene decay. *Appl Environ Microbiol* **76**:3420–3426.

Carvalho FM, Fonseca MM, Batistuzzo De Medeiros S, Scortecci KC, Blaha CAG, Agnez-Lima LF. 2005. DNA repair in reduced genome: the *Mycoplasma* model. *Gene* **360**:111–119.

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

Chernomor O, von Haeseler A, Minh BQ. 2016. Terrace Aware Data Structure for Phylogenomic Inference from Supermatrices. *Syst Biol* **65**:997–1008.

Clark TB, Whitcomb RF. 1984. Pathogenicity of mollicutes for insects: possible use in biological control. *Ann Microbiol* **135A**:141–150.

Cockburn SN, Haselkorn TS, Hamilton PT, Landzberg E, Jaenike J, Perlman SJ. 2013. Dynamics of the continent-wide spread of a *Drosophila* defensive symbiont. *Ecol Lett* **16**:609–616.

Conner WR, Blaxter ML, Anfora G, Ometto L, Rota-Stabelli O, Turelli M. 2017. Genome comparisons indicate recent transfer of *w*Ri-like *Wolbachia* between sister species *Drosophila suzukii* and *D. subpulchrella*. *Ecol Evol* **7**:9391–9404.

Corbin C, Heyworth ER, Ferrari J, Hurst GDD. 2017. Heritable symbionts in a world of varying temperature. *Heredity* **118**:10–20.

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.
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.
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.
Haselkorn TS. 2010. The *Spiroplasma* heritable bacterial endosymbiont of *Drosophila*. *Fly* **4**:80–87.
Haselkorn TS, Cockburn SN, Hamilton PT, Perlman SJ, Jaenike J. 2013. Infectious adaptation: potential host range of a defensive endosymbiont in *Drosophila*. *Evolution* **67**:934–945.
Haselkorn TS, Jaenike J. 2015. Macroevolutionary persistence of heritable endosymbionts: acquisition, retention, and expression of adaptive phenotypes in *Spiroplasma*. *Mol Ecol* **14**:3752–3765.
Haselkorn TS, Markow TA, Moran NA. 2009. Multiple introductions of the *Spiroplasma* bacterial endosymbiont into *Drosophila*. *Molecular Ecology* **18**:1294–1305.
Hayashi M, Watanabe M, Yukuhiro F, Nomura M, Kageyama D. 2016. A Nightmare for Males? A Maternally Transmitted Male-Killing Bacterium and Strong Female Bias in a Green Lacewing Population. *PLoS ONE* **11**:e0155794.
He L-S, Zhang P-W, Huang J-M, Zhu F-C, Danchin A, Wang Y. 2018. The Enigmatic Genome of an Obligate Ancient *Spiroplasma* Symbiont in a Hadal Holothurian. *Appl Environ Microbiol* **84**:e01965–17.

Hilgenboecker K, Hammerstein P, Schlattmann P, Telschow A, Werren JH. 2008. How many species are infected with *Wolbachia*?- a statistical analysis of current data. *FEMS Microbiol Lett* **281**:215–220.
Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. 2018. UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Mol Biol Evol* **35**:518–522.
Holmes EC. 2009. The Evolution and Emergence of RNA Viruses. Oxford University Press.
Hornett EA, Duplouy AMR, Davies N, Roderick GK, Wedell N, Hurst GDD, Charlat S. 2008. You Can't Keep a Good Parasite Down: Evolution of a Male-Killer Suppressor Uncovers Cytoplasmic Incompatibility. *Evolution* **62**:1258–1263.
Hosokawa T, Koga R, Kikuchi Y, Meng X-Y, Fukatsu T. 2010. *Wolbachia* as a bacteriocyte-associated nutritional mutualist. *Proc Natl Acad Sci U S A* **107**:769–774.
Ho SYW, Phillips MJ, Cooper A, Drummond AJ. 2005. Time dependency of molecular rate estimates and systematic overestimation of recent divergence times. *Mol Biol Evol* **22**:1561–1568.
Hurst GDD, Jiggins FM. 2000. Male-killing bacteria in insects: mechanisms, incidence, and implications. *Emerg Infect Dis* **6**:329–336.
Itoh T, Martin W, Nei M. 2002. Acceleration of genomic evolution caused by enhanced mutation rate in endocellular symbionts. *Proc Natl Acad Sci U S A* **99**:12944–12948.
Jaenike J. 2012. Population genetics of beneficial heritable symbionts. *Trends Ecol Evol* **27**:226–232.
Jaenike J, Polak M, Fiskin A, Helou M, Minhas M. 2007. Interspecific transmission of endosymbiotic *Spiroplasma* by mites. *Biol Lett* **3**:23–25.
Jaenike J, Unckless R, Cockburn SN, Boelio LM, Perlman SJ. 2010. Adaptation via Symbiosis: Recent Spread of a *Drosophila* Defensive Symbiont. *Science* **329**:212–215.
Jiggins FM, Hurst GDD, Jiggins CD, von der Schulenburg JHG, Majerus MEN. 2000. The butterfly *Danaus chrysippus* is infected by a male-killing *Spiroplasma* bacterium. *Parasitology* **120**:439–446.
Joshi BD, Berg M, Rogers J, Fletcher J, Melcher U. 2005. Sequence comparisons of plasmids pBJS-O of *Spiroplasma citri* and pSKU146 of *S. kunkelii*: implications for plasmid evolution. *BMC Genomics* **6**:175.
Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods* **14**:587–589.
Kanehisa M, Goto S, Hattori M, Aoki-Kinoshita KF, Itoh M, Kawashima S, Katayama T, Araki M, Hirakawa M. 2006. From genomics to chemical genomics: new developments in KEGG. *Nucleic Acids Res* **34**:D354–7.
Kanehisa M, Sato Y, Morishima K. 2016. BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences. *J Mol Biol* **428**:726–731.
Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* **30**:772–780.
Koonin EV, Mushegian AR, Rudd KE. 1996. Sequencing and analysis of bacterial genomes. *Current Biology* **6**:404–416.
Krogh A, Larsson B, von Heijne G, Sonnhammer EL. 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol* **305**:567–580.
Ku C, Lo W-S, Chen L-L, Kuo C-H. 2013. Complete genomes of two dipteran-associated spiroplasmas provided insights into the origin, dynamics, and impacts of viral invasion in spiroplasma. *Genome Biol Evol* **5**:1151–1164.
Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, Salzberg SL. 2004. Versatile and open software for comparing large genomes. *Genome Biol* **5**:R12.
Lanfear R, Calcott B, Ho SYW, Guindon S. 2012. Partitionfinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Mol Biol Evol* **29**:1695–1701.
LeClerc JE, Li B, Payne WL, Cebula TA. 1996. High mutation frequencies among *Escherichia coli* and *Salmonella* pathogens. *Science* **274**:1208–1211.

Lee H, Popodi E, Tang H, Foster PL. 2012. Rate and molecular spectrum of spontaneous mutations in the bacterium *Escherichia coli* as determined by whole-genome sequencing. *Proc Natl Acad Sci U S A* **109**:E2774–83.
Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**:3094–3100.
Lo W-S, Chen L-L, Chung W-C, Gasparich GE, Kuo C-H. 2013. Comparative genome analysis of Spiroplasma melliferum IPMB4A, a honeybee-associated bacterium. *BMC Genomics* **14**:22.
Lo W-S, Huang Y-Y, Kuo C-H. 2016. Winding paths to simplicity: genome evolution in facultative insect symbionts. *FEMS Microbiol Rev* **40**:855–874.
Lynch M. 2006. Streamlining and simplification of microbial genome architecture. *Annu Rev Microbiol* **60**:327–349.
Madeira F, Park YM, Lee J, Buso N, Gur T, Madhusoodanan N, Basutkar P, Tivey ARN, Potter SC, Finn RD, Lopez R. 2019. The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res* **47**:W636–W641.
Massey SE. 2008. The proteomic constraint and its role in molecular evolution. *Mol Biol Evol* **25**:2557–2565.
Masson F, Calderon-Copete S, Schüpfer F, Vigneron A, Rommelaere S, Garcia-Arraez MG, Paredes JC, Lemaitre B. 2020. Blind killing of both male and female *Drosophila* embryos by a natural variant of the endosymbiotic bacterium *Spiroplasma poulsonii*. *Cell Microbiol* **22**:e13156.
Masson F, Copete SC, Schüpfer F, Garcia-Arraez G, Lemaitre B. 2018. In Vitro Culture of the Insect Endosymbiont *Spiroplasma poulsonii* Highlights Bacterial Genes Involved in Host-Symbiont Interaction. *mBio* **9**. doi:10.1128/mbio.00024-18
Mateos M, Castrezana SJ, Nankivell BJ, Estes AM, Markow T a., Moran NA. 2006. Heritable endosymbionts of *Drosophila*. *Genetics* **174**:363–376.
McCutcheon JP, Moran NA. 2011. Extreme genome reduction in symbiotic bacteria. *Nat Rev Microbiol* **10**:13–26.
McFall-Ngai M, Hadfield MG, Bosch TCG, Carey HV, Domazet-Lošo T, Douglas AE, Dubilier N, Eberl G, Fukami T, Gilbert SF, Hentschel U, King N, Kjelleberg S, Knoll AH, Kremer N, Mazmanian SK, Metcalf JL, Nealson K, Pierce NE, Rawls JF, Reid A, Ruby EG, Rumpho M, Sanders JG, Tautz D, Wernegreen JJ. 2013. Animals in a bacterial world, a new imperative for the life sciences. *Proc Natl Acad Sci U S A* **110**:3229–3236.
Miller DE, Staber C, Zeitlinger J, Hawley RS. 2018. Highly Contiguous Genome Assemblies of 15 Species Generated Using Nanopore Sequencing. *G3- Genes Genom Genet* **8**:3131–3141.
Montenegro H, Solferini VN, Klaczko LB, Hurst GDD. 2005. Male-killing *Spiroplasma* naturally infecting *Drosophila melanogaster*. *Insect Mol Biol* **14**:281–287.
Montenegro H, Souza WN, da Silva Leite D, Klaczko LB. 2000. Male-killing selfish cytoplasmic element causes sex-ratio distortion in *Drosophila melanogaster*. *Heredity* **85 Pt 5**:465–470.
Moran NA. 2003. Tracing the evolution of gene loss in obligate bacterial symbionts. *Curr Opin Microbiol* **6**:512–518.
Moran NA. 1996. Accelerated evolution and Muller's rachet in endosymbiotic bacteria. *Proc Natl Acad Sci U S A* **93**:2873–2878.
Moran NA, McCutcheon JP, Nakabachi A. 2008. Genomics and evolution of heritable bacterial symbionts. *Annu Rev Genet* **42**:165–190.
Moran NA, McLaughlin HJ, Sorek R. 2009. The dynamics and time scale of ongoing genomic erosion in symbiotic bacteria. *Science* **323**:379–382.
Mouches C, Barroso G, Gadeau A, Bové JM. 1984. Characterization of two cryptic plasmids from *Spiroplasma citri* and occurrence of their DNA sequences among various spiroplasmas. *Ann Inst Pasteur Microbiol* **135**:17–24.
Nakayama S, Parratt SR, Hutchence KJ, Lewis Z, Price TAR, Hurst GDD. 2015. Can maternally inherited endosymbionts adapt to a novel host? Direct costs of *Spiroplasma* infection, but not vertical transmission efficiency, evolve rapidly after horizontal transfer into *D. melanogaster*. *Heredity* **114**:539–543.

Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* **32**:268–274.
Oliver KM, Smith AH, Russell JA. 2014. Defensive symbiosis in the real world - advancing ecological studies of heritable, protective bacteria in aphids and beyond. *Funct Ecol* **28**:341–355.
Osaka R, Ichizono T, Kageyama D, Nomura M, Watada M. 2013. Natural variation in population densities and vertical transmission rates of a *Spiroplasma* endosymbiont in *Drosophila hydei*. *Symbiosis* **60**:73–78.
Ota T, Kawabe M, Oishi K, Poulson DF. 1979. Non-male-killing spiroplasmas in *Drosophila hydei*. *J Hered* **70**:211–213.
Paradis E, Schliep K. 2019. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* **35**:526–528.
Paredes JC, Herren JK, Schüpfer F, Lemaitre B. 2016. The Role of Lipid Competition for Endosymbiont-Mediated Protection against Parasitoid Wasps in *Drosophila*. *mBio* **7**:e01006–16.
Paredes JC, Herren JK, Schüpfer F, Marin R, Claverol S, Kuo C-H, Lemaitre B, Béven L. 2015. Genome Sequence of the *Drosophila melanogaster* Male-Killing *Spiroplasma* Strain MSRO Endosymbiont. *MBio* **6**:e02437–14.
Poulson DF, Sakaguchi B. 1961. Nature of “sex-ratio” agent in *Drosophila*. *Science* **133**:1489–1490.
Razin S. 2006. The Genus Mycoplasma and Related Genera (Class Mollicutes) In: Dworkin M., Falkow S., Rosenberg E., Schleifer KH., Stackebrandt E., editor. The Prokaryotes. New York: Springer. pp. 836–904.
R Core Team. 2016. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
Regassa LB, Gasparich GE. 2006. Spiroplasmas: evolutionary relationships and biodiversity. *Front Biosci* **11**:2983–3002.
Richardson MF, Weinert LA, Welch JJ, Linheiro RS, Magwire MM, Jiggins FM, Bergman CM. 2012. Population Genomics of the *Wolbachia* Endosymbiont in *Drosophila melanogaster*. *PLoS Genet* **8**:e1003129.
Rocha EPC, Blanchard A. 2002. Genomic repeats, genome plasticity and the dynamics of *Mycoplasma* evolution. *Nucleic Acids Res* **30**:2031–2042.
Rocha EPC, Danchin A. 2002. Base composition bias might result from competition for metabolic resources. *Trends Gene* **18**:291–294.
Saglio PHM, Whitcomb RF. 1979. Diversity of wall-less prokaryotes in plant vascular tissue, fungi, and invertebrate animals In: Whitcomb RF, Tully JG, editors. The Mycoplasmas. New York: Academic Press. pp. 1–36.
Saillard C, Carle P, Duret-Nurbel S, Henri R, Killiny N, Carrère S, Gouzy J, Bové J-M, Renaudin J, Foissac X. 2008. The abundant extrachromosomal DNA content of the *Spiroplasma* citri GII3-3X genome. *BMC Genomics* **9**:195.
Sanada-Morimura S, Matsumura M, Noda H. 2013. Male Killing Caused by a *Spiroplasma* Symbiont in the Small Brown Planthopper, *Laodelphax striatellus*. *J Hered* **104**:821–829.
Schneider DI, Saarman N, Onyango MG, Hyseni C, Opiro R, Echodu R, O’Neill M, Bloch D, Vigneron A, Johnson TJ, Dion K, Weiss BL, Opiyo E, Caccone A, Aksoy S. 2019. Spatio-temporal distribution of *Spiroplasma* infections in the tsetse fly (*Glossina fuscipes fuscipes*) in northern Uganda. *PLoS Neglect Trop D* **13**:e0007340.
Seemann T. 2015. snippy: fast bacterial variant calling from NGS reads.
Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**:2068–2069.
Shigenobu S, Watanabe H, Hattori M, Sakaki Y, Ishikawa H. 2000. Genome sequence of the endocellular bacterial symbiont of aphids *Buchnera* sp. APS. *Nature* **407**:81–86.
Subramanian B, Gao S, Lercher MJ, Hu S, Chen W-H. 2019. Evolview v3: a webserver for visualization, annotation, and management of phylogenetic trees. *Nucleic Acids Res* **47**:W270–W275.
Teixeira L, Ferreira A, Ashburner M. 2008. The bacterial symbiont *Wolbachia* induces resistance to

RNA viral infections in *Drosophila melanogaster*. *PLoS Biol* **6**:e1000002.
Tinsley MC, Majerus MEN. 2006. A new male-killing parasitism: *Spiroplasma* bacteria infect the ladybird beetle *Anisosticta novemdecimpunctata* (Coleoptera: Coccinellidae). *Parasitology* **132**:757–765.
Turelli M, Cooper BS, Richardson KM, Ginsberg PS, Peckenpaugh B, Antelope CX, Kim KJ, May MR, Abrieux A, Wilson DA, Bronski MJ, Moore BR, Gao J-J, Eisen MB, Chiu JC, Conner WR, Hoffmann AA. 2018. Rapid Global Spread of *w*Ri-like *Wolbachia* across Multiple *Drosophila*. *Curr Biol* **28**:963–971.e8.
Turelli M, Hoffmann AA. 1991. Rapid spread of an inherited incompatibility factor in California *Drosophila*. *Nature* **353**:440–442.
Viver T, Orellana LH, Hatt JK, Urdiain M, Díaz S, Richter M, Antón J, Avian M, Amann R, Konstantinidis KT, Rosselló-Móra R. 2017. The low diverse gastric microbiome of the jellyfish *Cotylorhiza tuberculata* is dominated by four novel taxa. *Environ Microbiol* **19**:3039–3058.
Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, Earl AM. 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* **9**:e112963.
Watts T, Haselkorn TS, Moran NA, Markow TA. 2009. Variable Incidence of *Spiroplasma* Infections in Natural Populations of *Drosophila* Species. *PLoS One* **4**:e5703.
Weinert LA, Araujo-Jnr EV, Ahmed MZ, Welch JJ. 2015. The incidence of bacterial endosymbionts in terrestrial arthropods. *Proc R Soc B* **282**:20150249.
Wickham H. 2009. ggplot2: Elegant Graphics for Data Analysis. Springer Science & Business Media.
Wick RR, Judd LM, Gorrie CL, Holt KE. 2017a. Completing bacterial genome assemblies with multiplex MinION sequencing. *Microb Genom* **3**:e000132.
Wick RR, Judd LM, Gorrie CL, Holt KE. 2017b. Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Comput Biol* **13**:e1005595.
Wick RR, Schultz MB, Zobel J, Holt KE. 2015. Bandage: interactive visualization of de novo genome assemblies. *Bioinformatics* **31**:3350–3352.
Wilke CO. 2020. ggridges: Ridgeline Plots in “ggplot2.”
Wilke CO. 2019. cowplot: Streamlined Plot Theme and Plot Annotations for “ggplot2.”
Woese CR, Stackebrandt E, Ludwig W. 1984. What are mycoplasmas: the relationship of tempo and mode in bacterial evolution. *J Mol Evol* **21**:305–316.
Wu M, Sun LV, Vamathevan J, Riegler M, Deboy R, Brownlie JC, McGraw EA, Martin W, Esser C, Ahmadinejad N, Wiegand C, Madupu R, Beanan MJ, Brinkac LM, Daugherty SC, Durkin AS, Kolonay JF, Nelson WC, Mohamoud Y, Lee P, Berry K, Young MB, Utterback T, Weidman J, Nierman WC, Paulsen IT, Nelson KE, Tettelin H, O’Neill SL, Eisen JA. 2004. Phylogenomics of the reproductive parasite *Wolbachia pipientis w*Mel: a streamlined genome overrun by mobile genetic elements. *PLoS Biol* **2**:e69.
Xie J, Butler S, Sanchez G, Mateos M. 2014. Male killing *Spiroplasma* protects *Drosophila melanogaster* against two parasitoid wasps. *Heredity* **112**:399–408.
Xie JL, Vilchez I, Mateos M. 2010. *Spiroplasma* Bacteria Enhance Survival of *Drosophila hydei* Attacked by the Parasitic Wasp *Leptopilina heterotoma*. *PLoS One* **5**. doi:10.1371/journal.pone.0012149
Xie J, Winter C, Winter L, Mateos M. 2014. Rapid spread of the defensive endosymbiont *Spiroplasma* in *Drosophila hydei* under high parasitoid was pressure. *FEMS Microbiol Ecol* **91**:1–11.
Yamada M-A, Nawa S, Watanabe TK. 1982. A mutant of SR organism (SRO) in *Drosophila* that does not kill the host males. *Jpn J Genet* **57**:301–305.
Ye F, Melcher U, Rascoe JE, Fletcher J. 1996. Extensive chromosome aberrations in *Spiroplasma citri* strain BR3. *Biochem Genet* **34**:269–286.
Yu G, Smith DK, Zhu H, Guan Y, Lam TT. 2017. ggtree : an r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol*

870 *Evol* **8**:28–36.

871 **Supplementary material**

872 **Figure S1.** Mutational bias in *s*Hy evolution. Rates were calculated by counting the
873 corresponding numbers of snps and dividing by the total number of AT or GC positions in the
874 *s*Hy genome.

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

877 **Figure S3.** Plasmid synteny in *Spiroplasma poulsonii* strains. Numbers on ticks correspond to
878 nucleotide position on plasmids in kb. Sequence similarities (>=60% and spanning >= 2kb)
879 between plasmids are indicated by ribbons. Abbreviations: ARP - adhesion related protein,
880 OTU - ovarian tumor domain containing protein (in *s*Mel: Spaid), ParA - plasmid partitioning
881 protein like, RIP - ribosome inactivating domain containing protein, Spiralin - spiralin like
882 protein. Figure was created with AliTV.

883 **Figure S4.** RIP loci in different *Spiroplasma* genomes. Maximum likelihood tree was
884 calculated with IQ-TREE based on an alignment of RIP domains (288 amino acid positions),
885 created using the hmmalign function of the HMMER software, and manually trimmed to
886 exclude positions present in < 3 sequences. Domain prediction is based on PfamScan,
887 SignalP, and TMHMM. Clades identified by Ballinger & Perlmann (2017) are indicated, and
888 sequences from *s*Hy highlighted in bold font. UFB Bootstrap values ≥90 are shown on nodes.
889 Abbreviations: SP - signal peptide, TM - transmembrane helix, Ank - ankyrin repeats, RIP -
890 ribosome inactivating domain containing protein

891 **Figure S5.** Absence (light colors) and presence (dark colors) of genes involved in DNA
892 repair in *Buchnera aphidicola*, *Wolbachia w*Mel, *Mycoplasma gallisepticum*, and
893 *Spiroplasma poulsonii s*Hy and *s*Mel. *Arsenophonus nasoniae* is included as an example for a
894 symbiont with a relatively large genome and complete DNA repair pathways. Genes involved
895 in more than one pathway are listed only once, and those missing in all of the taxa are not
896 displayed. Data is taken from the KEGG pathway database (Kanehisa et al., 2006), and
897 annotation of *s*Mel and *s*Hy genomes as described in Materials and Methods section.

898 **Table S1.** *Spiroplasma* genomes included in comparative genomic analyses, with accession
899 numbers, genome properties, and phenotypes induced in host.

900 **Table S2.** List of changes observed in *s*Hy and *s*Mel over the course of the evolution
901 experiments.

902 **Table S3.** List of missing and truncated genes in *s*Hy and *s*Mel with KEGG annotation.