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

Michael Gerth1,*,†, Humberto Martínez-Montoya2, Paulino Ramirez3, Florent Masson4, Joanne S. Griffin1, Rodolfo Aramayo5, Stefanos Siozios1, Bruno Lemaitre4, Mariana Mateos6 and Gregory D. D. Hurst1

**Abstract**

*Spiroplasma* is a genus 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 sHy within flies, and sMel 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 sMel and sHy (>99.5% sequence identity in shared loci) show extensive structural genomic differences, which potentially indicates a higher degree of host adaptation in sHy, 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.

**DATA SUMMARY**

All novel sequencing data are available through National Center for Biotechnology Information (NCBI) repositories: raw reads for sHy-Tx under BioProject accession PRJNA274591; raw reads for sMel-Br under BioProject accession PRJNA507275; and all other raw reads and all assemblies under BioProject accession PRJNA640980.

Alignments for phylogenetic reconstructions, and a supplementary protocol are available from Zenodo under the DOI 10.5281/zenodo.3903209 (https://zenodo.org/record/3903209#.YBkwNzH7SM8).

Supplementary Material is available under the DOI 10.6084/m9.figshare.13584902.

All *Spiroplasma* genome assemblies used for comparative genomics are available from NCBI repositories, with the accession numbers listed in Table S1.
INTRODUCTION

Many bacterial lineages have evolved to become associates of animal hosts [1]. In arthropods, such associations are the rule, and maternally inherited, endosymbiotic bacteria are especially common and diverse [2, 3]. Prominent examples include obligate nutritional symbionts of blood and sap feeders [4], reproductive manipulators [5], and protective symbionts [6]. 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 [7], protective symbionts [8] and reproductive manipulators [9]. Importantly, these symbiont-conferred traits may be modulated by environmental factors [10] and host genetic background [11].

Spiroplasma are small, helical bacteria that lack cell walls. Like other Mollicutes (Mycoplasma-like organisms), they are exclusively found in host association and may be pathogenic [12]. Among the first discovered spiroplasmas were plant pathogens (reviewed by Saglio and Whitcomb [13]) and an inherited symbiont of Drosophila with the ability to shift sex ratios towards females [14]. Subsequently, Spiroplasma symbionts were found in a number of different Drosophila species [15], as well as many other arthropods [16]. More recently, molecular evidence was brought forward for Spiroplasma symbionts in non-arthropod animals [17, 18]. In insects, two phenotypes induced by inherited Spiroplasma have been studied in detail (reviewed by Anbutsu and Fukatsu [19], and Ballinger and Perlman [20]). Firstly, Spiroplasma can enhance survival rates of its hosts under parasite or parasitoid attack. In Drosophila neotestacea, Spiroplasma inhibits growth and reproduction of the parasitic nematode Howardula and, thus, reverses nematode-induced sterility [21]. Further, Spiroplasma symbionts may greatly enhance Drosophila survival when attacked by parasitoids [22, 23]. Both protective phenotypes are mediated by Spiroplasma-encoded RIP (ribosome-inactivating protein) toxins, which target the attackers' ribosomes [24–26]. These processes may be complementary to protection through competition between Spiroplasma and parasitoids for lipids in the host haemolymph [27]. Secondly, Spiroplasma kills males early in development in several species of Drosophila, planthoppers [28], ladybird beetles [29], lacewings [30] and butterflies [31]. The male-killing phenotype is also linked to a toxin: in Drosophila melanogaster, a Spiroplasma protein containing OTU (ovarian tumour-like deubiquinase) and ankyrin domains was demonstrated to kill male embryos and, thus, termed Spiroplasma poulsonii androcidin (‘Spaid’) [32].

Spiroplasma-induced phenotypes have been commonly observed to be dynamic compared with other symbiont traits, with repeated observation of phenotype change within laboratory culture over relatively short time frames. For example, spontaneous loss of male killing was found in Spiroplasma symbionts of Drosophila nebulosa [33] and Drosophila willistoni [34], and spontaneous emergence of non-male-killing Spiroplasma in D. melanogaster has occurred at least twice in a single culture [32, 35]. 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 [36]. Genomic analysis has further documented dynamic Spiroplasma evolution driven by viral proliferation [37–39] and by extensive transfer of genetic material between plasmids, and from plasmids to chromosomes [40, 41]. Notably, plasmids in Spiroplasma commonly encode the systems that establish their phenotypic effect on their host [42]. Examples include Spaid [32] and RIP genes [26], 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 [43, 44], which are pathogens closely related to Spiroplasma [45].

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 [46]. 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 [47, 48].

In this study, we systematically investigated rates and patterns of Spiroplasma evolution. To this end, we employed two strains of S. poulsonii. (i) sHy is a natural associate of D. hydei [49] and
confers protection against parasitic wasps [22]. We monitored its evolution over ~10 years in fly culture, and determined changes through de novo reference genome sequencing and re-sequencing of multiple isolates. (ii) sMel occurs naturally in D. melanogaster, where it kills male offspring [50] and protects from parasitoids [23]. Using the previously established complete genome [51] and cell-free culture [52], we re-sequenced isolates spanning ~2.5 years of evolution. This approach enabled tracing Spiroplasma evolution on various levels: within-strain (<10 years divergence), between strains, and through comparison with other Spiroplasma genomes, across a whole clade of arthropod symbionts. It further allowed comparison between symbionts evolving in a host and in axenic culture. We observed that rates of molecular evolution and chromosomal rearrangements in S. poulsonii are substantially faster than in other inherited bacteria, and that this has resulted in markedly different genomic organization in sHy and sMel. Our work, thus, highlights S. poulsonii as an amenable, highly evolvable but potentially unpredictable model insect symbiont.

**METHODS**

**Spiroplasma strains and genome sequencing**

We documented the evolution of the S. poulsonii strains sHy and sMel in fly hosts and in cell-free culture, respectively (Table 1). sHy is one of the Spiroplasma symbionts naturally associated with D. hydei [53]. Our evolution experiments covered approximately 10 years and three variants of sHy, all of which all originated from a single isolate (Fig. 1). Over the course of the study, D. hydei hosts were maintained on a standard corn meal agar diet (1% agarose, 8.5% sugar, 6% maize meal, 2% autolysed yeast, 0.25% nipagin), at 25°C, and a 12:12 h light:dark cycle. To isolate Spiroplasma sHy-Liv DNA from D. hydei, we followed the protocol of Paredes et al. [51]. We collected approximately 300 D. hydei virgins and aged them for 4 weeks. The flies were then anesthetized with CO₂ and pricked in the anterior sternopleurum using a sterile needle. Haemolymph was extracted via centrifugation (10000 g, 5 min, at 4°C) through a 0.45 µm filter (Corning Costar Spin-X). To avoid haemolymph coagulation, we only extracted batches of 20 flies at a time and immediately transferred extracted haemolymph into ice-cold PBS solution (1× PBS buffer: 137 mmol NaCl, 2.7 mmol KCl, 10 mmol Na₂HPO₄, 1.8 mmol KH₂PO₄). Extracts of all batches were

| Name used in the current study | Natural host | Establishment in fly or cell-free culture | Source and date of DNA extraction | Also known as |
|--------------------------------|-------------|-----------------------------------------|----------------------------------|--------------|
| sHy                            |             |                                         |                                  |              |
| sHy-Tx12                       | Drosophila hydei | 01/2004 [53]          | Haemolymph extract, 03/2012   | Hap_1, TEN104-106 |
| sHy-Liv18a                     | Drosophila hydei | 06/2008 (copy of sHy-Tx12) | Haemolymph extract, 04/2018 | –             |
| sHy-Liv18b                     | Drosophila hydei | 03/2014 (copy of sHy-Liv18A) | Haemolymph extract, 04/2018 | –             |
| sMel                           |             |                                         |                                  |              |
| sMel-Ug16                      | Drosophila melanogaster | 10/2016 [52]      | Cell-free culture, four time points in 2016–2018 | MSRO-Uganda |
| sMel-Br18                      | Drosophila melanogaster | 07/1997 [54]      | Whole fly, 03/2018            | MSRO-RED42   |

![Fig. 1](image_url) Overview on S. poulsonii strains investigated in this study. Three sHy strains were sampled from D. hydei cultures over a total time span of 10 years. Blue dates denote splitting of D. hydei cultures, red dates mark Spiroplasma extraction for sequencing. For sMel, samples from four points of a time series of axenic culture were sequenced (HL, start of culture; P14, after 14 passages; P57, after 57 passages; P79, after 79 passages). Here, red dates correspond to the date of isolation for sequencing.
combined, and the PBS solution with *Spiroplasma* cells was centrifuged (12000g, 5 min, at 4°C). Supernatant PBS was discarded, and the cell pellet was subjected to DNA extraction using a phenol/chloroform protocol and subsequent ethanol precipitation. DNA was eluted in water, and sequenced at the Centre for Genomics Research at the University of Liverpool (UK). Libraries were prepared using a Nextera XT kit and sequenced as 2×250bp runs on an Illumina MiSeq sequencer.

DNA extraction for the *Spiroplasma* strain sHy-Tx followed the protocol described above, with the following modifications: immediately after the piercing, 35–40 flies were placed into 0.5 ml microcentrifuge tubes previously pierced in the bottom, each of which was placed within a 1.5 ml microcentrifuge tube containing 20 µl PBS, and centrifuged at 4500g for 10s to collect haemolymph. DNA was recovered using a chloroform/ethanol procedure (Supplementary protocol 1) and diluted in AE buffer (Qiagen). A pair-ended library was constructed by Eureka Genomics. DNA was fragmented, end-repaired, A’-tagged, ligated to adaptors, size-selected and enriched with 25 cycles of PCR during which an index was incorporated to the sample. Sample preparation was performed according to Illumina’s Multiplexing Sample Preparation Guide and Eureka Genomics’ proprietary method. The resulting library was subjected to Illumina HiSeq 2500 sequencing at the Texas A&M AgriLife Genomics and Bioinformatics Services Facility (College Station, TX, USA).

sMel is a naturally occurring *Spiroplasma* symbiont of *D. melanogaster* best known for its male-killing phenotype [32, 54]. Cell-free culture of sMel-Ug was recently accomplished [52], and we used a time series of this culture to investigate sMel molecular rates of evolution outside of host tissues. The time series covered 29 months in total, and sequencing was performed at four time points (Fig. 1). Illumina DNA sequencing was performed at MicrobesNG using the ‘standard service’. Culturing, passaging and DNA extraction was performed as described by Masson et al. [52].

We further sequenced a single isolate of sMel-Br deriving from a strain established in 1997 (Table 1). For Illumina sequencing, 1–2g of whole body flies from *Spiroplasma*-positive *D. melanogaster* were collected for CTAB-(N-cetyl-N,N,N-trimethylammonium bromide)/phenol based DNA extraction. Illumina libraries were prepared with an Illumina TruSeq DNA kit and sequenced on Illumina NovaSeq-6000 S2 150 paired-end flow cell by Texas A&M AgriLife Genomics and Bioinformatics Services Facility. For Oxford Nanopore MinION sequencing, heads were first removed by immersing whole flies in liquid nitrogen and vigorous shaking on a metal sieve. The headless specimens were immediately subjected to phenol/chloroform extraction. Nanopore library preparation was performed using an Oxford Nanopore ligation sequencing kit 1D (SQK-LSK108), following precautionary measures to ensure long read lengths. The library was then sequenced on a MinION sequencing device (MIN-101B) with a SpotON Flow Cell Mark 1 (R9.5-FLO-MIN107.1). Base calling was performed using Albacore version 2.0 (Oxford Nanopore) and the reads were processed with Porechop version 0.24 [55] to remove adapter sequences. A draft assembly of sMel-Br was created using Illumina and Nanopore reads and Unicycler version 0.4.7 [56]. The assembly was fragmented, but did contain three circular contigs with sequence similarity to other *Spiroplasma* contigs.

### sHy reference genome sequencing and annotation

In order to reconstruct a high-quality reference genome for sHy, we extracted DNA from 30 *D. hydei* virgin females carrying the symbiont. To limit the impact of potentially heterogeneous *Spiroplasma* populations within the flies, we used F3 individuals from a single isofemale line for DNA extraction. The flies were anesthetized at −20°C, washed with 70% ethanol and briefly dried. The flies were then homogenized in G2 lysis buffer from the Qiagen genomic DNA buffer set using a ml dounce tissue grinder (DWK Life Science), and DNA was extracted from the homogenate using a Qiagen Genomic-tip 20G, following the protocol modifications from Miller et al. [57]. A total of 4 µg DNA was then used in the ‘1D gDNA long reads without BluePippin protocol’ library preparation protocol with a 1D ligation sequencing kit (SQK-LSK108; Oxford Nanopore). Sequencing was performed on a MinION sequencer (Oxford Nanopore) using a single FLO-MIN106 R9 flow cell for 48h, and the raw Nanopore signals were base-called using Albacore version 2.3.1.

All reads passing the quality checks were used to create an assembly with unicycler version 0.4.7 [56]. The resulting assembly contained six circular contigs (~15 kb --- 1.6 Mbp) with high similarity to the previously published sMel genome as determined by BLAST+ version 2.9 searches [58] in Bandage [59]. We also created a hybrid assembly with Unicycler using all long reads and the sHy-Liv18a Illumina reads. The hybrid assembly was more fragmented, but did contain three contigs that spanned almost the entire 1.6 Mbp circular contig of the long-read-only assembly. Because hybrid assemblies are typically more accurate than assemblies based on long reads only [60], we corrected the long-read assembly using the hybrid contigs. This was achieved by calling differences between the assemblies with the ‘dndiff’ function implemented in mummer version 3.18 [61], and correcting the assembly accordingly. Finally, pilon version 1.23 [62] was used to polish the corrected assembly, again using the sHy-Liv18a Illumina reads. After nine rounds of polishing with pilon, no further improvements could be observed and the assembly was considered final.

### Rates of molecular evolution in sHy and sMel

We determined changes in sHy and sMel chromosomes over time by employing the snippy pipeline version 4.1.0 [63]. We used our newly assembled sHy genome and the previously sequenced sMel genome [52] as references, annotated using the prokka pipeline version 1.13 with standard parameters and the *Spiroplasma* genetic code (translation table 4) [64]. sHy and sMel variants determined by snippy were filtered to only include positions that were covered by all sHy or all sMel Illumina libraries with at least 5x coverage, respectively. Further, to compare rates of molecular evolution between
Spiroplasma and other microbial species, we counted the number of changes in third codon positions of coding sequences (CDSs). For both strains, we only considered CDSs that were entirely covered by a sequencing depth of at least 5x in all Illumina libraries (see Table S2 for sequencing and mapping statistics). Rates of molecular evolution were calculated by number of observed changes at third codon positions/time of years over which the changes were observed. For sHy, using sHy-Liv18a as a reference, we calculated two rate estimates, one based on comparison with sHy-Liv18b, and one by comparing sHy-Tx12 to the reference (Fig. 1). For sMel, the haemolymph extract from which the culture was established was used as the reference (HL; Fig. 1), and three rate estimates were calculated based on the changes observed in the three sampled time points of the culture (Fig. 1).

Comparative genomics
We first compared the newly assembled sHy genome with the most closely related, fully sequenced Spiroplasma genome, sMel. Both genomes were annotated using the Kyoto Encyclopedia of Genes and Genomes (KEGG) and BlastKOALA [65, 66], and orthologous protein sequences determined using orthofinder version 2.2.3 [67]. Synteny between the genomes was assessed through aligning the genomes with minimap2 version 2.15 [68]. Genome degradation was investigated by documenting the number of unique KEGG numbers for sHy and sMel, all of which were verified by blast+ searches. For each strain, we considered genes to be likely pseudogenized or truncated if the longest CDS within an orthogroup spanned at most 60% of the CDS length from the other strain.

Secondly, we compared sHy with 17 Spiroplasma genomes from the Citri–Chrysopicola–Mirum clade (see Table S1 for a full list of names and accession numbers). All genomes were annotated using prokka version 1.4.0, and orthology between predicted CDSs was established using orthofinder. We extracted single-copy orthologues present in all strains, and aligned each locus separately with the L-INS-i method implemented in mafft version 7.450 [69]. Recombination was evaluated with pairwise homology index and window sizes of 100, 50 and 20 amino acids for each locus [70], and any locus showing evidence for recombination was excluded. We used iq-TREE version 1.6.10 [71] to reconstruct a core-genome phylogeny from the remaining 96 loci (covering 26019 amino acid positions). Each locus was treated as a separate partition with a distinct evolutionary rate [72], and optimal models and number of partitions were estimated with iq-TREE [73, 74]. Branch support was assessed using 1000 replicates of ultrafast bootstraps [75], and approximate likelihood ratio test [76] in iq-TREE.

For all genomes, insertion sequence elements were annotated with prokka. Prophage regions were predicted with phispy version 3.7.8, which uses typical prophage features (e.g. gene length, strand directionality, AT/GC skews, insertion points) for its predictions [77]. We also used the phaster web server, which uses sequence similarities and other criteria to predict prophages [78]. We screened for the presence of toxin genes implied in protective phenotypes [25, 26] and male killing [32]. To this end, we downloaded UniProt sequence alignments from the Pfam database [79] for the protein families RIP (PF00161) and OTU (PF02338, OTU-like cysteine protease), respectively. We used these alignments as databases for searches with hmmver version 3.2.1 [80], and protein sequences from all genomes as queries. Domain architecture of all matching proteins was then determined using PammScan with an E value of 0.001 [81], signalp 5.0 [82], and TMHMM 2.0 [83]. RIP domains showed a high degree of divergence and, therefore, were aligned to the reference RIP HMM (hidden Markov model) profile using hhmalign from the hmmver software package. From this alignment, positions present in fewer than three sequences were removed, and a phylogeny reconstructed with iq-TREE. OTU domains were aligned using mafft and a phylogeny reconstructed with iq-TREE.

Thirdly, we determined plasmid synteny across S. poulsonii strains. In addition to the plasmids of sMel and sHy, we also included a plasmid of sNeo and all circular contigs from our sMel-Br assembly. Plasmids were annotated using prokka version 1.13 with a protein database composed of all plasmid proteins available from the National Center for Biotechnology Information (NCBI) GenBank. Plasmid alignment and visualization were performed with alitv version 1.06 [84].

For an extended set of 31 Spiroplasma genomes that also included strains from the Apis clade (Table S1), we performed gene tree–species tree reconciliations to investigate the evolutionary history of prophage loci. To this end, we annotated all genomes and determined orthologous groups of loci as described above. Orthogroups were blasted against a database containing all available Spiroplasma virus proteins from NCBI Protein (N=192). Genes were classified as ‘prophage related’ if hits were at least 60% identical over 50% of the length of any viral protein. Reconciliation was performed with the prophage loci using generax version 1.2.2, using maximum-likelihood gene trees determined with iq-TREE as starting trees and the LG+G model for all loci. We predicted CRISPR/Cas systems and CRISPR arrays using the tools crisprcastyper version 1.2.1 [85] and crispridentify [86], respectively.

Data visualization
Figures were prepared in R version 3.6.2 [87] using the packages ‘ape’ [88], ‘cowplot’ [89], ‘ggalluvial’ [90], ‘ggplot2’ [91], ‘ggtree’ [92] and ‘ggridges’ [93]. Phylogenetic trees and domain architecture of toxin loci were visualized with evolview version 3.0 [94].

RESULTS
Rates and patterns of evolutionary change in S. poulsonii
The S. poulsonii sHy reference genome comprises a single circular chromosome (1625797 bp) and five circular contigs
viral pathogens (Fig. 2). The chromosome contains 1584 predicted CDSs, a single rRNA cluster and 30 tRNAs. Overall, genome content and metabolic capacities are highly similar to the previously sequenced, very closely related sMel genome of S. poulsonii [51], and a detailed comparison follows further below.

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

Over ~10 years of evolution, we observed a similar absolute number of variants on the chromosome (~1.6 Mbp) and plasmids (~0.1 Mbp) of sHy, i.e. relatively more changes on the plasmids (Fig. 3, Table S3). Most variants in CDSs affected hypothetical proteins, but were enriched in prophage-associated loci and adhesin-related proteins (Fig. 3). Overall, sHy variants in CDSs were about equally often found to be synonymous as non-synonymous (Fig. 3), and changes were biased towards GC > AT substitutions (Fig. S1). The changes in sMel over ~2.5 years in culture affected only 15 different CDSs in total, of which 4 were adhesin-related proteins and 3 lipoproteins (Table S3). Thus, the rates and patterns of evolutionary change are similar between the axenically cultured sMel and the host-associated sHy.

Comparing the genomes of sHy and sMel revealed a notable contrast between the high degree of nucleotide sequence identity on the one hand, and striking structural and gene content differences on the other hand. In 604 single copy orthologues shared between the genomes, the mean nucleotide identity was > 99.5%. However, compared with sMel, sHy is reduced in gene content (1584 vs 2388 genes), chromosome size (1.61 vs 1.88 Mbp) and coding density (65 vs 82%). Further, sHy contains fewer predicted insertion sequences (9 vs 89 in sMel) and intact prophages as predicted with PHASTER (sMel, 16 such regions spanning 426 kb; sHy, 2 regions, 14 kb), but not with PHISPY (Fig. S2). Both sMel and sHy have a number of missing or truncated (i.e. potentially pseudogenized) genes when compared with each other, but the level of genomic deterioration is higher in sHy, and covers a range of different genes (Fig. 4a, Table S4). Notable pseudogenized or absent loci in sHy include parts of the phosphotransferase system, in particular the loci required for the uptake of N-acetylmuramic acid and N-acetylglycosamine (see Table S3 for a full list).
Further, while recA is truncated in sMel, the copy in sHy appears complete and functional. As suggested by Paredes et al. [51], the loss of recA function in sMel is, therefore, likely very recent.

Strikingly, there is evidence for a history of extensive chromosomal rearrangements since the last common ancestor of sHy and sMel, and genome-wide synteny between the strains is low (Fig. 4b). On average, syntenic blocks between the strains contained fewer predicted CDSs for sHy (on average −0.3 CDSs kb⁻¹; Fig. 4c), in line with a higher degree of genomic deterioration in sHy compared with sMel. A comparison of plasmids across different S. poulsonii strains (sHy, sMel-Ug, sMel-Br and sNeo) revealed similar gene content across the plasmids of different strains, but large differences in arrangement and number (Fig. S3).

**Genome and toxin evolution in the genus Spiroplasma**

Comparing sHy 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 (~1.1–1.9 Mbp), and the S. poulsonii strains sHy, sMel and sNeo are among the strains with the largest main chromosomes (Fig. 5). Plasmid numbers range from 0 to 5, and sHy has the highest number of plasmids of any of the investigated strains, although plasmid number is unclear for some of the genomes with draft status, and Spiroplasma citri may have seven plasmids [95]. 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) [39]. Reconciliation of prophage gene trees with the Spiroplasma species tree revealed that prophage proliferation has likely happened relatively recently, and repeatedly in the Citri and Poulsonii clades (Fig. S4). Prophage loci are entirely absent, or very low in number, for Spiroplasma strains that harbour CRISPR/Cas systems, or remnants thereof (Fig. S4). Further, Spiroplasma genome size correlated with the number of insertion sequences (Figs 5 and S5). The distribution of CDS lengths varies across the investigated genomes (Fig. 5), which may be explained by differences in proportion of prophage regions, level of pseudogenization and assembly quality. Overall, our comparison indicates that sHy may be more degraded than its closest relatives sNeo and sMel, with a smaller chromosome size, fewer predicted CDSs and mobile genetic elements, and the lowest coding density of these three (Fig. 5).

Spiroplasmas’ most striking phenotypes in insects have been mechanistically linked to toxin genes. For example, RIPs may protect Spiroplasma hosts by cleaving ribosomal RNA
of parasites and parasitoids [25]. Five RIP loci are present in sMel (RIP1–5, of which 3–5 are almost identical copies). As expected, the protective sHy also encodes RIPs, but only has a single orthologue for RIP1. However, it contains an additional RIP gene in three copies that appears to be absent in sMel, and one of these copies also contains ankyrin repeats (Fig. S6).

Further, the male-killing phenotype of sMel 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 [32]. hmmmer searches using OTU domain profiles revealed a number of Spiroplasma loci similar to Spaid (Fig. 6) that, however, lack its characteristic domain composition. For example, sHy 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 detected in the symbiotic taxa Rickettsia and Wolbachia. In the phylogenetic reconstruction of OTU domains rooted with the eukaryotic sequences (from the ciliate Stentor coeruleus), Rickettsia and Wolbachia loci are nested within the Spiroplasma sequences. This topology suggests lateral gene transfer of OTU-domain-containing proteins from Spiroplasma to the other intracellular taxa, although differences in genetic code between Spiroplasma and other bacteria likely limit the probability for transfer of functional protein-encoding genes.

**DISCUSSION**

*Spiroplasma, an exceptionally fast evolving symbiont*

The substitution rates we observed in *S. poulsonii* are among the highest reported for any bacteria (Fig. 1). In a study comparing genome-wide evolutionary rates in bacterial human pathogens [96], most taxa showed considerably slower rates, and our estimate of *S. poulsonii* evolutionary rates overlaps with substitution rates of RNA viruses [97]. 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 synthesize them [98]. Obsolete metabolic genes subsequently accumulate now neutral substitutions, resulting in pseudogenization. Secondly, reduced population sizes of symbionts, together with bottlenecks at transmission events, limit purifying selection of deleterious mutations [99] and mobile genetic elements [100] – both in turn create further genomic deterioration. Thirdly, genomes of host-associated microbes are AT rich [101], which makes introduction of errors through replication slippage more likely [102]. Fourthly, many symbionts lack DNA repair genes [103], which may result in increased substitution rates.

All of these points very likely apply for *S. poulsonii*: its metabolic capacities are reduced compared with free-living microbes, it has a high proportion of pseudogenes and prophages, very high AT content, and lacks several DNA
Fig. 5. Genome properties of *Spiroplasma* strains from Citri, Poulsonii, Chrysopicola and Mirum clades. Genome size (Mbp) and plasmid number were taken from the literature (Table S1). Coding density, CDS length and number of IS elements were 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 (26019 amino acid positions). Scale bar corresponds to inferred substitutions per site. All nodes were maximally supported by UFB and aLRT. Asterisks indicate strains with genomes with draft status (i.e. a chromosome assembly split into multiple contigs).

Fig. 6. Spaid-like proteins in *Spiroplasma* and other symbionts. The 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. Scale bar corresponds to inferred substitutions per site. Ank, Ankyrin repeats; OTU, ovarian tumour-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.
repair genes. However, many symbiotic bacteria show similar trends of genome evolution [104] and it is, therefore, somewhat puzzling that S. poulsonii stands out with this exceptionally high substitution rate. Loss of DNA repair genes has been shown to co-occur with elevated mutation rates in intracellular bacteria [105–107], and may explain our observations. The mismatch repair protein-encoding loci mutS and mutL are universally lacking in Spiroplasma, but are present in the slower-evolving symbionts Wolbachia wMel and Buchnera aphidicola (Fig. 7) [108–110]. In Escherichia coli, hypermutator strains that have lost such loci have a ~10–300-fold increase of spontaneous mutation rates [111, 112], which is comparable to the approximately two orders of magnitude difference between substitution rates of Buchnera and Spiroplasma (Fig. 2). Like Spiroplasma, Mycoplasma universally lacks the mismatch repair loci mutS, mutL, and mutH [113, 114]. In line with this observation, elevated evolutionary rates were hypothesized for mycoplasmas [43], and in our comparison, Mycoplasma gallisepticum appears to be the only bacterial taxon with substitution rates similar to S. poulsonii [44]. Absence of the DNA mismatch repair pathway may, thus, be ancestral to Entomoplasmatales (Spiroplasmataceae + Entomoplasmataceae) and contribute to the dynamic genome evolution across this taxon [115, 116].

Alternatively, increased substitutional rates caused by the loss of these loci could have arisen multiple times independently in Entomoplasmatales.

Further to absence of DNA repair genes causing elevated mutation rates, a recent comparative study demonstrated a strong negative correlation between mutation rate and genome size in free-living and endosymbiotic bacteria [117]. This correlation is, however, not apparent in the genomes of endosymbionts we have investigated. For example, the considerably slower evolving Buchnera genomes are much smaller than Spiroplasma, and Wolbachia would be predicted to have much larger genomes if their size were mainly determined by mutational rates. This suggests that substitution rates alone are a poor predictor for the sizes of genomes investigated here. Likely, these genome sizes result from an interplay of multiple factors such as population size, patterns of DNA repair gene absence and mutational rates [118, 119].

Our rate estimate is potentially biased by at least two factors. Firstly, we have only investigated laboratory populations of S. poulsonii. Each vertical transmission event creates symbiont population bottlenecks potentially increasing genetic drift and, thus, substitution rates [46]. Because the number of generations in natural populations of the Spiroplasma host D. hydei is lower compared with laboratory reared hosts, vertical transmission events are rarer under natural conditions and, therefore, substitution rates potentially lower. Further, laboratory strains could experience relaxed selection compared with natural symbiont populations. This may lead to higher substitution rate estimates from laboratory populations compared with natural populations. Secondly, substitution rates often appear larger when estimated over brief time periods [120]. Duchene et al. found that substitution rates measured over 10 years can be up to one order of magnitude larger than those measured over 100 years [96]. In agreement with such bias, we found more variants in our sMel culture after 19 months than for the 29 months isolate (Table S3). The back mutations that likely have happened between 19 and 29 months of our sMel culture would go unnoticed when comparing genomes over larger time scales.

More generally, it is difficult to estimate divergence times between Spiroplasma strains. The substitution rates estimated in sHy and sMel would suggest the two strains have diverged 1120–2260 years ago. However, this estimate is unreliable due to a number of factors that we cannot control for. For example, the number of generations per year for Drosophila hosts of Spiroplasma differ depending on species and location, and is expected to be lower in the wild compared with laboratory strains. Further, Spiroplasma may move between species, for example via ectoparasitic mites [121]. A number of partially sympatric Drosophila species carry similar Spiroplasma strains (D. hydei, D. melanogaster, D. willistoni, D. neotestacea) [15], and with our data it is impossible to determine the number and direction of potential Spiroplasma transfers between these species.

| Genus                | Mismatch repair | Base-excision repair | Nucleotide-excision repair | Homologous recombination |
|----------------------|-----------------|----------------------|---------------------------|-------------------------|
| Arsenophonus nasoniae|                 |                      |                           |                         |
| Wolbachia wMel       |                 |                      |                           |                         |
| Buchnera aphidicola  |                 |                      |                           |                         |
| Mycoplasma gallisepticum |           |                      |                           |                         |
| Spiroplasma poulsonii sMel |   |                      |                           |                         |
| Spiroplasma poulsonii sHy |              |                      |                           |                         |

**Fig. 7.** Absence (light colours) and presence (dark colours) of genes involved in DNA repair in Buchnera aphidicola, Wolbachia wMel, Mycoplasma gallisepticum, and S. poulsonii sHy and sMel. Arsenophonus nasoniae is included as an example of a symbiont with a relatively large genome and complete DNA repair pathways. Genes involved in more than one pathway are listed only once, and those missing in all of the taxa are not displayed. Data is taken from the KEGG pathway database [66], and annotation of sMel and sHy genomes is as described in Methods.
Reductive genome evolution in *Spiroplasma*

According to a widely supported model of genome reduction in symbiotic bacteria [46], the first stages of host restriction involve accumulation of pseudogenes and mobile genetic elements, chromosomal rearrangements, increased substitution rates, and excess deletions. Advanced stages of host association are accompanied by further genome shrinkage, and purging of mobile genetic elements, which overall result in more stable chromosomes. Using these characteristics, different levels of genome reduction are apparent in the investigated *Spiroplasma* genomes, across which genome sizes correlate positively with number of mobile genetic elements (plasmids, prophages, insertion sequences; Figs 5 and S5). On this spectrum, *Spiroplasma syrphidicola* and *Spiroplasma chrysopicala* have the most reduced genomes, without homologues of prophage loci, and high levels of synteny between the two strains [39]. Therefore, it was argued that phages have likely invaded *Spiroplasma* only after the split of the Syrphidicola and Citri-Poulsonii clades [39]. Our prophage gene tree-species tree reconciliations are in line with this hypothesis, but also indicate that prophage proliferation has largely happened independently in different *Spiroplasma* lineages (Fig. S4). CRISPR/Cas systems have multiple origins in *Spiroplasma* [122] and only occur in strains lacking prophages (Fig. S4). While the absence of antiviral systems often coincides with prophage proliferation (e.g. in the Citri clade), several strains with compact, streamlined genomes lack CRISPR/Cas and prophages (e.g. TU-14 and NBRC-100390; Fig. S4). These strains also show other hallmarks of reduced symbiont genomes (small size, high coding density, lack of plasmids and transposons; Fig. 5), which is in line with the model of genome reduction discussed above and suggests prophage regions were purged from these genomes. Alternatively, these strains may never have been exposed to phages.

Using the model of genome reduction during restriction to the host environment introduced above, *S. poulsonii*’s genome characteristics suggest that such restriction has happened relatively recently. Although very closely related (99.5% sequence identity) and found in very similar hosts, *sMel* and *sHy* differ markedly in coding density, genome size and proportion of prophage regions. Interestingly, *phaster* predicted 426 kb of intact phage regions in *sMel*, but only 14 kb for *sHy* (Fig. S2). Both phage proliferation in *sMel* and prophage loss in *sHy* could have contributed to this. However, when using the similarity agnostic tool *phispy*, the predicted prophage regions were similar in size between *sHy* and *sMel* (Fig. S2). This observation is compatible with degradation and/or pseudogenization of prophage regions in *sHy*, which would lead to reduced sequence similarity to viral loci (and, thus, reduced detectability by *phaster*), but not entirely blur prophage characteristics employed by *Phispy* (e.g. gene length, strand directionality, AT/GC skews, insertion points) [77]. Since the split of the lineages, *sHy* has not only lost prophage regions and insertion sequences compared with *sMel*, but also several genes that are often lost in host-restricted bacteria, such as parts of the phosphotransferase system for the uptake of carbohydrates, one tRNA locus, and *uvrD*, which plays a role in mismatch repair and nucleotide excision repair (Table S3).

Using signatures of genomic degradation as a proxy, our findings collectively suggest that *sHy* is in a more advanced stage of host restriction than *sMel*. This may indicate co-adaptation with the host as a result of the fitness benefits associated with *sHy* under parasitoid pressure, and the absence of detectable costs for carrying *sHy* in *D. hydei* [22, 123, 124]. However, the *Spiroplasma* symbiont of *D. neotestacea*, *sNeo*, is also protective, does not cause obvious fitness costs [21], but has a less reduced genome (Fig. 5) [26]. Further, it is also possible that genome reduction in *sHy* was mainly driven by stochastic effects or even by adaptation to laboratory conditions, as we have not investigated contemporary *sHy* from wild *D. hydei* populations.

Implications for *Spiroplasma* evolutionary ecology

*Spiroplasma* evolutionary ecology shows several parallels to that of the most widely distributed arthropod symbiont, *Wolbachia*: both symbionts are found in a range of different hosts [3], have the ability to invade novel hosts [125, 126], may confer protection [8, 22] but also kill males [127], and can spread across host populations swiftly [128, 129]. However, there are also pronounced differences between the symbionts: *Spiroplasma* rarely reaches the high infection frequencies often observed in *Wolbachia* [130, 131], and is arguably found in a more diverse host range that encompasses arthropods and plants [132], molluscs [18], echinoderms [17] and cnidarians [133]. Further, *Wolbachia* show greatly reduced spontaneous mutation rates compared with *Spiroplasma*, likely caused by a more complete set of DNA repair genes (Fig. 7).

In theory, fast evolutionary rates should enable *Spiroplasma* to adapt to novel hosts quickly (i.e. to reduce pathogenicity, and to maximize vertical transmission efficiency), and experimental studies have found high horizontal transmission efficiency of *Spiroplasma* [36, 134, 135]. Consistent with this, we found that genes implicated in host–symbiont compatibility and virulence have evolved especially fast in our evolution experiments. For example, adhesion-related proteins are important in cell invasion in other *Spiroplasma* species [136–138], and are enriched for evolutionary changes in *sHy* and *sMel* (Fig. 2). In addition, we documented dynamic evolution and turnover of toxin loci, which are important for host fitness and symbiont compatibility (Figs 6 and S6). 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 [139]. This may explain the generally low prevalence of *Spiroplasma* symbionts [15], which seems to increase only when carrying the symbiont is associated with a large fitness benefit [21]. In contrast, virtually identical *Wolbachia* strains (‘superspreaders’) are found in many different host species at very high frequencies [140] – 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, S. 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 [32, 35, 36]. However, 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.

Funding information
M. G. has received funding from the European Union’s Horizon 2020 Research and Innovation Program under Marie Skłodowska-Curie grant agreement 703 379, and from EMBO through a long-term fellowship (ALTF 48–2015) co-funded by the European Commission (LTFCO-FUND2013, GA-2013–609409). Sequencing was supported by an award from the University of Liverpool Technology Directorate Voucher Scheme, and by a Texas A&M Genomics Seed Grant.

Author contributions
Designed and conceived project: M. G., G. D. D. H., M. M. Experimental passage of Spiroplasma in vivo: M. G., J. S. G., G. D. D. H., M. M. Experimental passage of Spiroplasma in vitro: F. M., B. L. Symbiont DNA purification and sequencing: M. G., P. R., S. S., R. A., H. M.–M., F. M. Assembly and annotation: M. G., H. M.–M., P. R., R. A. Molecular evolution analyses: M. G. Wrote first draft of paper: M. G., M. M., G. D. D. H. All authors contributed to revising the manuscript, and have read and approved the final manuscript.

Conflicts of interest
The authors declare that there are no conflicts of interest.

References
1. McFall-Ngai M, Hadfield MG, Bosch TCG, Carey HV, Domazet-Lošo T et al. Animals in a bacterial world, a new imperative for the life sciences. Proc Natl Acad Sci USA 2013;110:3229–3236.
2. Weinert LA, Araujo-Jnr EV, Ahmed MZ, Welch JJ. The incidence of bacterial endosymbionts in terrestrial arthropods. Proc Biol Sci 2015;282:20150249.
3. Duron O, Bouchon D, Boutin S, Bellamy L, Zhou L et al. The diversity of reproductive parasites among arthropods: Wolbachia do not walk alone. BMC Biol 2008;6:27.
4. Douglas AE. The microbial dimension in insect nutritional ecology. Funct Ecol 2009;23:38–47.
5. Drew GC, Frost CL, Hurst GDD. Reproductive parasitism and positive fitness effects of heritable microbes. eLS 2019:2.a0028327.
6. Oliver KM, Smith AH, Russell JA. Defensive symbiosis in the real world - advancing ecological studies of heritable, protective bacteria in aphids and beyond. Funct Ecol 2014;28:341–355.
7. Hosokawa T, Koga R, Kikuchi Y, Meng X-Y, Fukatsu T. Wolbachia as a bacteriocyte-associated nutritional mutualist. Proc Natl Acad Sci USA 2010;107:769–774.
8. Teixeira L, Ferreira Á, Ashburner M. The bacterial symbiont Wolbachia induces resistance to RNA viral infections in Drosophila melanogaster. PLoS Biol 2008;6:e1000002.
9. Dyson EA, Kamath MK, Hurst G. Wolbachia infection associated with all-female broods in Hypolimnas bolina (Lepidoptera: Nymphalidae): evidence for horizontal transmission of a butterfly male killer. Heredity 2002;88:166–171.
10. Corbin C, Heyworth ER, Ferrari J, Hurst GDD. Heritable symbionts in a world of varying temperature. Heredity 2017;118:10–20.
11. Hornett EA, Duploye AMR, Davies N, Roderick GK, Wedell N et al. You can’t keep a good parasite down: evolution of a male-killer suppressor uncovers cytoplasmic incompatibility. Evolution 2008;62:1258–1263.
12. Razin S. The genus Mycoplasma and related genera (class Mollicutes). In: Dworkin M, Falkow S, Rosenberg E, Schleifer KH, Stackebrandt E (eds). The Prokaryotes. New York: Springer; 2006. pp. 836–904.
13. Saglio PHM, Whitcomb RF. Diversity of wall-less prokaryotes in plant vascular tissue, fungi, and invertebrate animals. In: Whitcomb RF, Tully JG (eds). The Mycoplasmas. New York: Academic Press; 1979. pp. 1–36.
14. Poulson DF, Sakaguchi B. Nature of ‘sex-ratio’ agent in Drosophila. Science 1961;133:1489–1490.
15. Haselkorn TS. The Spiroplasma heritable bacterial endosymbiont of Drosophila. Fly 2010;4:80–87.
16. Gasparich GE. Spiroplasmas: evolution, adaptation and diversity. Front Biosci 2002;7:d619–d640.
17. He L–S, Zhang P-W, Huang J–M, Zhu F–C, Danchin A et al. The enigmatic genome of an obligate ancient Spiroplasma symbiontholothaplan. Appl Environ Microbiol 2018;84:e01965–17.
18. Duperron S, Pottier M–A, Léger N, Gaudron SM, Puilandre N et al. A tale of two chitons: is habitat specialisation linked to distinct associated bacterial communities? FEMS Microbiol Ecol 2013;83:552–567.
19. Anbutsu H, Fukatsu T. Spiroplasma as a model insect endosymbiont. Environ Microbiol Rep 2011;3:144–153.
20. Ballinger MJ, Perlman SJ. The defensive Spiroplasma. Curr Opin Insect Sci 2019;32:36–41.
21. Jaenike J, Unckless R, Cockburn SN, Boelio LM, Perlman SJ. Adaptation via symbiosis: recent spread of a Drosophila defensive symbiont. Science 2010;329:212–215.
22. Xie JL, Vilchez I, Mateos M. Spiroplasma bacteria enhance survival of Drosophila hydei attacked by the parasitic wasp Leptopilina heterotoma. PLoS One 2010;5:e12149.
23. Xie J, Butler S, Sanchez G, Mateos M. Male killing Spiroplasma protects Drosophila melanogaster against two parasitoid wasps. Heredity 2014;112:399–408.
24. Hamilton PT, Leong JS, Koop BF, Perlman SJ. Transcriptional responses in a Drosophila defensive symbiosis. Mol Ecol 2014;23:1558–1570.
25. Hamilton PT, Peng F, Boulanger MJ, Perlman SJ. A ribosome-inactivating protein in a Drosophila defensive symbiont. Proc Natl Acad Sci USA 2016;113:350–355.
26. Ballinger MJ, Perlman SJ. Generality of toxins in defensive symbiosis: ribosome-inactivating proteins and defense against parasitic wasps in Drosophila. PLoS Pathog 2017;13:e1006431.
27. Paredes JC, Herren JK, Schüpf F, Lemaître B. The role of lipid competition for endosymbiont-mediated protection against parasitoid wasps in Drosophila. mBio 2016;7:e01008–16.
28. Sanada-Morimura S, Matsumura M, Noda H. Male killing caused by a Spiroplasma symbiont in the small brown planthopper, Laodelphax striatellus. J Hered 2013;104:821–829.
29. Tinsley MC, Majerus MEN. A new male-killing parasitism: Spiroplasma bacteria infect the ladybird beetle Anisosticta novemdecimpunctata (Coleoptera: Coccinellidae). Parasitology 2006;132:757–765.
30. Hayashi M, Watanabe M, Yukihito F, Nomura M, Kageyama D. A nightmare for males? A maternally transmitted male-killing bacterium and strong female bias in a green lacewing population. PLoS One 2016;11:e0155794.
31. Jiggins FM, Hurst GDD, Jiggins CD, von der Schulenberg JHG, Majerus MEN. The butterfly Danaus chrysippus is infected by a male-killing Spiroplasma bacterium. Parasitology 2000;120:439–446.
32. Harumoto T, Lemaitre B. Male-killing toxin in a bacterial symbiont of Drosophila. Nature 2018;557:252–255.
33. Yamada M-A, Nawa S, Watanabe TK. A mutant of SR organism (SRo) in Drosophila that does not kill the host males. Jpn J Genet 1982;57:301–305.
34. Ebbert MA. The interaction phenotype in the Drosophila willistoni–Spiraemia symbiosis. Evolution 1991;45:971–988.
35. Masson F, Calderon-Copete S, Schüpfer F, Garcia-Arraiz G, Lemaitre B. In vitro culture of the insect endosymbiont Spiroplasma poulsonii highlights bacterial genes involved in host-symbiont interaction. mBio 2018;9:e00024-18.
36. Mateos M, Castrezana SJ, Nankivil BJ, Estes AM, Markow TA et al. Heritable endosymbionts of Drosophila. Genetics 2006;174:363–376.
37. Montenegro H, Souza WN, Da Silva Leite D, Klcakzo LB. Male-killing selfish cytoplasmic element causes sex-ratio distortion in Drosophila melanogaster. Heredity 2000;85:465–470.
38. Wick RR, Judd LM, Gorrie CL, Holt KE. Completing bacterial genome assemblies with multiplex MinION sequencing. Microb Genom 2017;3:e000132.
39. Wick RR, Judd LM, Gorrie CL, Holt KE. Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. PLoS Comput Biol 2017;13:e1005595.
40. Miller DE, Staber C, Zeitlinger J, Hawley RS. Highly contiguous genome assemblies of 15 Drosophila species generated using nanopore sequencing. G3 2018;8:3131–3141.
41. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J et al. BLAST+: architecture and applications. BMC Bioinformatics 2009;10:421.
42. Wick RR, Schultz MB, Zobel J, Holt KE. Bandage: interactive visualization of de novo genome assemblies. Bioinformatics 2015;31:3350–3352.
43. De Maio N, Shaw LP, Hubbard A, George S, Sanderson ND et al. Comparison of long-read sequencing technologies in the hybrid assembly of complex bacterial genomes. Microb Genom 2019;5:e000294.
44. Kurtz S, Philippy A, Delcher AL, Smoot M, Shumway M et al. Versatile and open software for comparing large genomes. Genome Biol 2004;5:R2.
45. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A et al. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One 2014;9:e112963.
46. Seemann T. Snippy: fast bacterial variant calling from NGS reads; 2015.
47. Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics 2014;30:2068–2069.
48. Kanehisa M, Sato Y, Morishima K. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. J Mol Biol 2016;428:726–731.
49. Kanehisa M, Goto S, Hattori M, Aoki-Kinoshita KF, Itoh M et al. From genomics to chemical genomics: new developments in KEGG. Nucleic Acids Res 2006;34:D354–D357.
50. Emms DM, Kelly S. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. Genome Biol 2015;16:157.
51. Li H. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics 2018;34:3094–3100.
52. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol 2013;30:770–780.
53. Bruen TC, Philippe H, Bryant D. A simple and robust statistical test for detecting the presence of recombination. Genetics 2006;172:2665–2681.
54. Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol 2015;32:268–274.
55. Chernomor O, von Haeseler A, Minh BQ. Terrace aware data structure for phylogenomic inference from supermatrices. Syst Biol 2016;65:997–1008.
56. Lanfear R, Calcott B, Ho SYW, Guindon S. Partitionfinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. Mol Biol Evol 2012;29:1695–1701.
74. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermii LS. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods 2017;14:587–589.

75. Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. UFBioT2: improving the ultrafast bootstrap approximation. Mol Biol Evol 2018;35:518–522.

76. Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W et al. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol 2010;59:307–321.

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

78. Arndt D, Grant JR, Marcu A, Sajed T, Pon A et al. PHASTER: a better, faster version of the PHAST phage search tool. Nucleic Acids Res 2016;44:W6–W21.

79. Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J et al. The Pfam protein families database: towards a more sustainable future. Nucleic Acids Res 2016;44:D279–D285.

80. Eddy SR. Accelerated profile HMM searches. PLoS Comput Biol 2011;7:e1002195.

81. Madeira F, Park YM, Lee J, Buso N, Gur T et al. The EMBL-EBI search and sequence analysis tools APIs in 2019. Nucleic Acids Res 2019;47:W636–W641.

82. Almagro Armenteros JJ, Tsisrigos KD, Senderby CK, Petersen TN, Winther Ø et al. SignaP 5.0 improves signal peptide predictions using deep neural networks. Nat Biotechnol 2019;37:420–423.

83. Krogh A, Larsson B, von Heijne G, Sonnhammer EL. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol 2001;305:567–580.

84. Ankenbrand MJ, Hohlfeld P. Predictive visualization of whole genome comparisons. PeerJ Computer Science 2017;3:e116.

85. Russell J, Pinilla-Redondo R, Mayo-Muñoz D, Shah SA, Sørensen SJ. CRISPRCasTyper: automated identification, annotation, and classification of CRISPR-Cas loci. CRISPR J 2020;3:462–469.

86. Mitrofanov A, Alkhbash O, Shmakov SA, Makarova KS, Koonin EV et al. CRISPRidentify: identification of CRISPR arrays using machine learning approach. Nucleic Acids Res 2020 (Epub ahead of print 08 Dec 2020).

87. R Core Team. R: A Language and Environment for Statistical Computing. Vienna: R Foundation for Statistical Computing; 2016.

88. Paradis E, Schliep K. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics 2019;35:526–528.

89. Wilke CO. cowplot: streamlined plot theme and plot annotations for ‘ggplot2’. 2019. https://CRAN.R-project.org/package=cowplot

90. Brunson JC. ggluvial: alluvial plots in ‘ggplot2’. 2019. https://CRAN.R-project.org/package=ggluvial

91. Wickham H. ggplot2: Elegant Graphics for Data Analysis. New York: Springer; 2009.

92. Yu G, Smith DK, Zhu H, Guan Y, Lam TT. ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. Methods Ecol Evol 2017;8:28–36.

93. Wilke CO. ggrridges: ridgeline plots in ‘ggplot2’. 2020. https://CRAN.R-project.org/package=ggrridges

94. Subramanian B, Gao S, Lercher MJ, Hu S, Chen W-H. Evolviev v3: a webserver for visualization, annotation, and management of phylogenetic trees. Nucleic Acids Res 2019;47:W270–W275.

95. Saillard C, Carle P, Duret-Nurbel S, Henri R, Killiny N et al. The abundant extrachromosomal DNA content of the Spiroplasma citri 3X genome. BMC Genomics 2008;9:195.

96. Duchène S, Holt KE, Weill F-X, Le Hello S, Hawkey J et al. Genome-scale rates of evolutionary change in bacteria. Microb Genom 2016;2:e000094.

97. Holmes EC. The Evolution and Emergence of RNA Viruses. Oxford: Oxford University Press; 2009.

98. Moran NA. Tracing the evolution of gene loss in obligate bacterial symbionts. Curr Opin Microbiol 2003;6:512–518.

99. Moran NA. Accelerated evolution and Muller’s ratchet in endosymbiotic bacteria. Proc Natl Acad Sci USA 1996;93:2873–2878.

100. Lynch M. Streamlining and simplification of microbial genome architecture. Annu Rev Microbiol 2006;60:327–349.

101. Rocha EPC, Danchin A. Base composition bias might result from competition for metabolic resources. Trends Gen 2002;18:291–294.

102. Moran NA, McLaughlin HJ, Sorek R. The dynamics and time scale of ongoing genomic erosion in symbiotic bacteria. Science 2009;323:379–382.

103. Moran NA, McCutcheon JP, Nakabachi A. Genomics and evolution of heritable bacterial symbionts. Annu Rev Genet 2008;42:165–190.

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

105. Itoh T, Martin W, Nei M. Acceleration of genomic evolution caused by enhanced mutation rate in endosymbiotic genomes. Proc Natl Acad Sci USA 2002;99:1294a–1294b.

106. Massey SE. The proteomic constraint and its role in molecular evolution. Mol Biol Evol 2008;25:2557–2565.

107. Koonin EV, Mushhegan AR, Rudd KE. Sequencing and analysis of bacterial genomes. Curr Biol 1996;6:404–416.

108. Richardson MF, Weinert LA, Welch JJ, Linheiro RS, Magwire MM et al. Population genomics of the Wolbachia endosymbiont in Drosophila melanogaster. PLoS Genet 2012;8:e1003129.

109. Wu M, Sun LV, Yamathewen J, Riegler M, Deboy R et al. Phylogenomics of the reproductive parasite Wolbachia pipientis wMel: a streamlined genome overun by mobile genetic elements. PLoS Biol 2004;2:e69.

110. Shigenobu S, Watanabe H, Hattori M, Sakaki Y, Ishikawa H. Genome sequence of the endosymbiotic bacterial symbiont of aphids Buchnera sp. APS. Nature 2000;407:81–86.

111. LeClerc JE, Li B, Payne WL, Cebula TH. High mutation frequencies among Escherichia coli and Salmonella pathogens. Science 1996;274:1208–1211.

112. Lee H, Popodi E, Tang H, Foster PL. Rate and molecular spectrum of spontaneous mutations in the bacterium Escherichia coli as determined by whole-genome sequencing. Proc Natl Acad Sci USA 2012;109:E2774–E2783.

113. Carvalho FM, Fonseca MM, Batistuzzo De Medeiros S, Scortecchi KC, Bliha CAG et al. DNA repair in reduced genome: the Mycoplasma model. Gene 2005;360:111–119.

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

115. Rocha EPC, Blanchard A. Genomic repeats, genome plasticity and the dynamics of Mycoplasma evolution. Nucleic Acids Res 2002;30:2031–2042.

116. Lo W-S, Huang Y-Y, Kuo C-H. Winding paths to simplicity: genome evolution in facultative insect symbionts. FEMS Microbiol Rev 2016;40:855–874.

117. Bourguignon T, Kinjo Y, Villa-Martín P, Coleman NV, Tang Q et al. Increased mutation rate is linked to genome reduction in prokaryotes. Curr Biol 2020;30:3848–3855.

118. Kuo C-H, Moran NA, Ochman H. The consequences of genetic drift for bacterial genome complexity. Genome Res 2009;19:1450–1454.

119. Marais GAB, Batut B, Daubin V. Genome evolution: mutation is the main driver of genome size in prokaryotes. Curr Biol 2020;30:R1083–R1085.
120. Ho SYW, Phillips MJ, Cooper A, Drummond AJ. Time dependency of molecular rate estimates and systematic overestimation of recent divergence times. Mol Biol Evol 2005;22:1561–1568.

121. Jaenike J, Polak M, Fiskin A, Helou M, Minhas M. Interspecific transmission of endosymbiotic Spiroplasma by mites. Biol Lett 2007;3:23–25.

122. Ipoutcha T, Tsarmopoulos I, Talenton V, Gaspin C, Moisan A et al. Multiple origins and specific evolution of CRISPR/Cas9 systems in minimal bacteria (Mollicutes). Front Microbiol 2019;10:2701.

123. Xie J, Winter C, Winter L, Mateos M. Rapid spread of the defensive endosymbiont S. robusta in Drosophila hydei under high parasitoid wasp pressure. FEMS Microbiol Ecol 2015;91:fu017.

124. Osaka R, Ichizono T, Kageyama D, Nomura M, Watada M. Natural variation in population densities and vertical transmission rates of a Spiroplasma endosymbiont in Drosophila hydei. Symbiosis 2013;60:73–78.

125. Haselkorn TS, Markow TA, Moran NA. Multiple introductions of the Spiroplasma bacterial endosymbiont into Drosophila. Mol Ecol 2009;18:1294–1305.

126. Conner WR, Blaxter ML, Anfora G, Ometto L, Rota-Stabelli O et al. Genome comparisons indicate recent transfer of wRi-like Wolbachia between sister species Drosophila suzukii and D. subpulchrela. Ecol Evol 2017;7:9391–9404.

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

128. Turelli M, Hoffmann AA. Rapid spread of an inherited incompatibility factor in California Drosophila. Nature 1991;353:440–442.

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

130. Watts T, Haselkorn TS, Moran NA, Markow TA. Variable incidence of Spiroplasma infections in natural populations of Drosophila species. PLoS One 2009;4:e5703.

131. Hilgenboecker K, Hammerstein P, Schlattmann P, Telschow A, Warren JH. How many species are infected with Wolbachia? – a statistical analysis of current data. FEMS Microbiol Lett 2008;281:215–220.

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

133. Viver T, Orellana LH, Watt JK, Urdain M, Diaz S et al. The low diverse gastric microbiome of the jellyfish Cotylorhiza tuberculata is dominated by four novel taxa. Environ Microbiol 2017;19:3039–3058.

134. Haselkorn TS, Cockburn SN, Hamilton PT, Perlman SJ, Jaenike J. Infectious adaptation: potential host range of a defensive endosymbiont in Drosophila. Evolution 2013;67:934–945.

135. Haselkorn TS, Jaenike J. Macroevolutionary persistence of heritable endosymbionts: acquisition, retention and expression of adaptive phenotypes in Spiroplasma. Mol Ecol 2015;24:3752–3765.

136. Hou L, Liu Y, Gao Q, Xu X, Ning M et al. Spiroplasma eriocheiris adhesin-like protein (ALP) interacts with epidermal growth factor (EGF) domain proteins to facilitate infection. Front Cell Infect Microbiol 2017;7:13.

137. Béven L, Duret S, Batailler B, Dubrama M-P, Saillard C et al. The repetitive domain of ScARP3d triggers entry of Spiroplasma citri into cultured cells of the vector Circulifer haematoceps. PLoS One 2012;7:e48606.

138. Dubrama MP, Béven L, Arricau-Bouvery N, Duret S, Claverol S et al. Differential expression of Spiroplasma citri surface protein genes in the plant and insect hosts. BMC Microbiol 2016;16:53.

139. Jaenike J. Population genetics of beneficial heritable symbionts. Trends Ecol Evol 2012;27:226–232.

140. Turelli M, Cooper BS, Richardson KM, Ginsberg PS, Peckenpaugh B et al. Rapid global spread of Wolbachia across multiple Drosophila. Curr Biol 2018;28:963–971.

141. Duffy S, Shackelton LA, Holmes EC. Rates of evolutionary change in viruses: patterns and determinants. Nat Rev Genet 2008;9:267–276.

142. Degnan PH, Lazarus AB, Brock CD, Wernegreen JJ. Host-symbiont stability and fast evolutionary rates in an antibacterium association: copeciation of Campotonus species and their endosymbionts, Candidatus Blochmannia. Syst Biol 2004;53:95–110.

Five reasons to publish your next article with a Microbiology Society journal
1. The Microbiology Society is a not-for-profit organization.
2. We offer fast and rigorous peer review – average time to first decision is 4–6 weeks.
3. Our journals have a global readership with subscriptions held in research institutions around the world.
4. 80% of our authors rate our submission process as ‘excellent’ or ‘very good’.
5. Your article will be published on an interactive journal platform with advanced metrics.

Find out more and submit your article at microbiologyresearch.org.