Wild communities of *Morpho* butterflies reveal *Spiroplasma* endosymbiont with inflated genome size and peculiar evolution.

Jonathan Filée [*,1], Manuela López-Villavicencio[2,6], Vincent Debat[2,6], Rachel Fourdin[7], Camilo Salazar[3], Karina Lucas Silva-Brandão[4], Patrick Blandin[2], Andre Victor Lucci Freitas[5], Carolina Pardo-Diaz[3] and Violaine Llaurens[6,2]

[1] Laboratoire Evolution, Génomes, Comportement, Ecologie, CNRS Université Paris-Saclay UMR9191, IRD UMR247, IDEEV, 12 route 128, 91198 Gif sur Yvette, France
[2] Institut de Systématique, Evolution et Biodiversité (UMR 7205 CNRS/MNHN/SU/EPHE/UA), Muséum National d'Histoire Naturelle - CP50, 45 rue Buffon, 75005 PARIS, France
[3] Department of Biology, Faculty of Natural Sciences, Universidad del Rosario. Carrera 24 # 63C-69, Bogota D.C., 111221, Colombia
[4] Leibniz Institute for the Analysis of Biodiversity Change, Museum of Nature Hamburg, Martin-Luther-King-Platz 3, 20146 Hamburg, Germany
[5] Departamento de Biologia Animal, Instituto de Biologia, Universidade Estadual de Campinas, Campinas/SP, Brazil
[6] Centre Interdisciplinaire de Recherche en Biologie (UMR 7241 CNRS/INSERM/Collège de France) 11, place Marcelin Berthelot, 75005 Paris, France
[7] GeT-PlaGe, Bât G2, INRAe, 31326 Castanet-Tolosan Cedex, France

* Corresponding authors: E-mail: jonathan.filee@universite-paris-saclay.fr

Keywords: Lepidoptera, endosymbiont, toxin genes, speciation, insertion sequences

Abstract (229 words)

The evolution of endosymbiont genomes is likely influenced by the ecological interactions with their hosts. Here, we studied the evolution of *Spiroplasma* genomes detected within *Morpho* butterflies sampled in the wild. *Spiroplasma* was detected in 4 out of the 11 *Morpho* species studied and displayed a 3 times larger genome size as compared to *Spiroplasma* genomes documented in other hosts. This inflation in genome size is caused by massive and recent expansion of various mobile genetic elements and by the acquisition of new genes stemming from prophages. In particular, we documented the peculiar evolution of the toxin genes in plasmids that may enhance host resistance to parasites. Phylogenetic comparisons with *Spiroplasma* extracted from other host point at a unique origin of *Spiroplasma* in *Morpho*, and strong divergence from *Spiroplasma* found in other Lepidoptera. Resequencing data obtained for multiple populations of the two sister-species *M. helenor* and *M. achilles* living in sympatry over the majority of their distribution revealed a opposite prevalence (97% in *M. achilles* and 3% in *M. helenor*), suggesting contrasted ecological interactions with these two host-species. Reconciliation analysis of the phylogenetic relationships of *Morpho* mitochondrial genomes and *Spiroplasma* genomes was then consistent with a predominant vertical transfer of the endosymbiont. Altogether, our results suggest a key role of ecological interactions with the host in the evolution of endosymbiont genomes and point at a putative interaction of *Spiroplasma* with reproductive isolation between sympatric species of butterflies.

Introduction

Ecological relationships between species generate selective pressures acting on their genomes. In turn, the evolution of genes in interacting species can modify their ecological interactions. Intracellular bacteria, or endosymbionts, offer the opportunity to investigate feedbacks between genome evolution and ecological interactions. Heritable microbial endosymbionts profoundly impact several life history traits of their hosts, affecting both their survival and reproductive success (Hurst 2017) as sex-ratio distortion (Harumoto and Lemaitre 2018) or protection against pathogens (Ballinger and Perlman 2019). They can therefore play a crucial role in population dynamics and diversification of the host species, especially when they induce cytoplasmic incompatibilities (Werren 1998). At the same time, the endosymbiont lifestyle is often associated with changes in its own genome (Wernegreen 2017): the prevalence of genetic drift in endosymbiont populations and the hyper-specialization to their host induce fast and irreversible genome erosion and progressive loss of metabolic functions. The mutation accumulation and genome decay through Muller's ratchet is indeed documented as a specific feature of endosymbiotic bacteria (Moran 1996). Such reduction in the number of functional genes may in turn increase the extinction risk of endosymbiont populations (Bennett and Moran 2015), and as a consequence, symbiont replacement is commonly observed(Manzano-Marín et al. 2023). However, long-term persistence of some endosymbionts has also been documented, raising the question of how ecological interactions with the hosts limit genome decay (Naito and Pawlowska 2016).

Insect endosymbionts offer prominent examples of the diversity of ecological interactions (Drew et al. 2019), from positive effects as nutritional providers (Sudakaran et al. 2017) or protective agents against pathogens (Ballinger and Perlman 2019; King 2019) to negative ones acting as sex-ratio distorters or male-killing agents (Stevens et al. 2001). While the effects of endosymbionts on host survival and reproduction have been largely explored in insect model species like *Drosophila*, their prevalence and ecological impacts in wild communities are still largely unknown in most insects. Recent publications of large genomic datasets in insects now allow to better characterize the prevalence of these endosymbionts throughout arthropods (Medina et al. 2023) but, also, to investigate the evolution of these endosymbionts and the diversity of their ecological relationships with different hosts in the wild.

In Lepidoptera, the diversity and the impact of cytoplasmic endosymbionts on host phenotypes have been scarcely studied. *Spiroplasma* and *Wolbachia* are the most frequently reported endosymbionts, with various effects on host fitness (Duplouy and Hornett 2018). Both produce male killing and sex ratio distortion in population of different Lepidoptera such as *Acraea encedon*, *Hypolimnas bolina* or *Danaus chrysippus* (Nymphalidae) (Duplouy and Hornett 2018; Jiggins et al. 2000). However, the transmission of these endosymbionts across species is largely unknown, as well as the diversity of their impact on host phenotypes.

Spiroplasma are associated with a large variety of hosts, and their genomes appear highly eroded with reduced metabolic capacities, high proportion of pseudogenes, and elevated evolutionary rates(Gerth et al. 2021; Liu et al. 2022). In *Drosophila*, the prevalence of *Spiroplasma* among natural population is generally low (Watts et al. 2009; Haselkorn 2010), but in some cases, the fitness advantages brought to their hosts, such as protection against parasitic nematodes, can make it more common (Jaenike et al. 2010). Experimental infections show that *Spiroplasma* has high horizontal transmission efficiency (Nakayama et al. 2015) but

this ability is constrained by the phylogenetic distance between different hosts (Tinsley and Majerus 2007). However, *tempo* and patterns of *Spiroplasma* transmission in natural populations remain to be investigated. Therefore, studying the evolution of *Spiroplasma* genomes in natural populations of insects can now shed light on the feedbacks between bacterial genome evolution and ecological interactions with their hosts.

Here, we focus on the *Spiroplasma* of nymphalid butterflies of the genus *Morpho* to characterize their level of ecological specialization, as well as their transmission mode. The genus *Morpho* is composed of emblematic species from the Neotropical rainforests, where up to ten different species can be observed in sympatry in Amazonian lowlands and the Guiana shield (Blandin and Purser 2013). Studying endosymbiont genomes found in *Morpho* butterflies from the wild allows to test (1) how much endosymbionts are shared across closely *vs.* distantly related host species, (2) how their genomes, and more specifically their toxin genes, evolved in different hosts, and (3) how endosymbionts are transmitted within and among sympatric host species. We used whole-genome sequencing data from 11 *Morpho* species to study the evolution of endosymbiont genomes in closely-related hosts. We then investigated the prevalence of endosymbionts in different populations of two sister-species of *Morpho* living in sympatry to characterize the transmission of the endosymbionts within and between species.

Materials and Methods

Genus dataset

To identify the diversity of endosymbiont genomes present in different *Morpho* species, we analyzed the sequencing data obtained using the PacBio HiFi methodology applied to specimens from 11 Morpho species: 9 Amazonian species (*M. marcus*, *M. eugenia*, *M. telemachus*, *M. hecuba*, *M. rhetenor*, *M. menelaus*, *M. deidamia*, *M. helenor*, *M. achilles*), and 2 sympatric species ranging from western Ecuador to Central America (*M. amathonte*, *M. granadensis)*. Note that within *M. telemachus* there are two sympatric morphs (with either blue or yellow wings), so we analyzed one individual per morph. This dataset including all sampled species is referred to as the *genus dataset*.

For each individual included in the *genus dataset*, the DNA extraction was carried out from the thorax muscles of a male individual using the Qiagen Genomic-tip 100/G kit, following supplier instructions. After DNA extraction, the sequencing library was prepared following the manufacturer's instructions "Procedure and Checklist Preparing HiFi SMRTbell Libraries Using SMRTbell Express Template Prep Kit 2.0." for *M. helenor*, *M. achilles* and *M. deidamia* and "Procedure and Checklist – Preparing whole genome and metagenome libraries using SMRTbell® prep kit 3.0" for the other species. Libraries were sequenced on several PacBio Sequel II SMRT cells with the adaptive loading method or by diffusion loading on a SequelII instrument (for additional details see (Bastide et al. 2023). The reads were assembled into contigs using Hifiasm (Cheng et al. 2021) using the option no purge (-l0) to avoid eventual over-purging symbiont sequences. The mitochondrial genome for all *Morpho* species was assembled directly from the PacBio Hifi reads with Rebaler (https://github.com/rrwick/Rebaler). For the assembly of mitochondrial genomes of *M. helenor*, *M. achilles* and *M. deidamia* the mitochondrial genome of the closely related

species *Pararge aegeria* was used as a reference (Bastide et al. 2023) while for the other eight species, we used the genome of *M. helenor* as a reference.

Endosymbiont metagenomic assembly

To detect the presence of endosymbiont genomes in the *genus dataset*, we used Blobtools(Laetsch and Blaxter 2017) with Diamond as search engine (Buchfink et al. 2015) against the UniProt database using a local copy of the NCBI TaxID file for the taxonomic assignation of the best hits. Minimap2 (Li 2018) was used for read mapping with the options -ax map-hifi. Endosymbiont contigs were extracted using seqtk (available at https://github.com/lh3/seqtk) and processed through the Dfast workflow (Tanizawa et al. 2018) to estimate statistics and taxonomic assignation. The completeness of the detected endosymbiont genomes was estimated with CheckM (Parks et al. 2015) with the corresponding gene sets. Endosymbiont genome annotations were carried out using PROKKA (Seemann 2014) with standard parameters, and the corresponding genetic codes. Whole genome alignments were created using the nucmer utility of the Mummer package (Marçais et al. 2018) with standard options. Structural variations were visualized using D-GENIES (Cabanettes and Klopp 2018).

Gene content in Spiroplasma *genomes*

We aimed at distinguishing orthologous genes shared with previously published *Spiroplasma* genomes found in other hosts, from genes specific to *Spiroplasma* in *Morpho*. We downloaded a set of 62 *Spiroplasma* genome assemblies with comparable genome metrics (N50>100kb) from the NCBI Refseq Genomes FTP server 12/02/2022 version (ftp://ftp.ncbi.nlm.nih.gov/genomes/refseq). We used using Ortho-Finder 2.5.4 (Emms and Kelly 2015) to infer orthologous genes in the 62 *Spiroplasma* genomes as well as the *M. achilles Spiroplasma* genome sAch identified in this study. Then, functional annotation of the orthologous was inferred using the BlastKOALA tool against the KEGG database (Kanehisa et al. 2016).

We then studied putative toxin genes either implied in host-protective phenotypes or host male killing in insects. We used the HMMER software (Mistry et al. 2013) seeded with protein sequences of each symbiont genome and the Pfam domain sequence alignments (Finn et al. 2014) corresponding to the OTU (PF02338 and OTU-like cysteine protease) and RIP (PF00161) as databases. Domain architectures of all matching proteins were then computed using PfamScan (Mistry et al. 2007) and SIGNALP 6.0 (Teufel et al. 2022). Phylogenies of the toxin proteins were built by extracting and aligning the corresponding OTU domains. The phylogenetic trees were inferred using IQ-TREE v2.1.3 (Nguyen et al. 2015) by estimating the best substitution models using ModelFinder (Kalyaanamoorthy et al. 2017). Branch support was then assessed by performing 1000 replicates using UltraFast boostraps (Hoang et al. 2018).

Phylogeny of Spiroplasma

A previously published set of 96 single-copy, non-recombinant orthologs from the *Spiroplasma* genomes (Gerth et al. 2021) was used to assess the phylogenetic relationships of these endosymbionts. Orthologs were identified using the best reciprocal BLASTP hits of

each of the 96 protein sequences using the *Spiroplasma poulsonii* sMel gene sequences as seeds and lead to a dataset of 72 gene sequences present in all of the *Morpho Spiroplasma* genomes. Alignments were computed using MAFFT (Katoh et al. 2002) and manually corrected to exclude ambiguous regions and taxa with sequence similarity >99% were removed. Phylogenic analyses were carried out using IQ-TREE v2.1.3 (Nguyen et al. 2015), and genes were partitioned to estimate the best substitution models using ModelFinder (Kalyaanamoorthy et al. 2017). Branches supports were assessed by performing 1000 replicates using UltraFast boostraps(Hoang et al. 2018). The resulting trees were rooted using the *Spiroplasma* sequences belonging to the *ixodetis* clade in accordance with the literature.

Prediction of mobile genetic elements in Spiroplasma *genomes*

Inserted sequences (ISs) were identified by querying the ISFinder database (Siguier et al. 2006) with protein sequences of each endosymbiont genome assemblies using BLAST with *e-value* ≤10e-10 (Altschul et al. 1990).
Plasmid sequences were identified using the Plasflow software (Krawczyk et al. 2018) and prophage regions were found using two methods: (1) a sequence-similarity search using PHASTER (Arndt et al. 2016), and (2) a *de novo* prediction using PhiSpy (Akhter et al. 2012). Predictions gathered from the two methods were then merged in a single file. Comparative genomics with other prokaryotic genomes were then computed using a set of 25,674 genome sequences with comparable genome metrics (N50>100kb) downloaded from the NCBI Refseq Genomes FTP server 12/02/2022 version (ftp://ftp.ncbi.nlm.nih.gov/genomes/refseq). The corresponding proteomes were then downloaded and ISs were identified using the same procedure as described above for *Morpho* endosymbionts (using the 62 *Spiroplasma* genomes for which plasmids and phage regions were identified with the same previous workflow).

Sister-species dataset

To study vertical and horizontal transmission of *Spiroplasma*, we focused on multiple populations of *M. achilles* and *M. helenor*, two sister-species living in sympatry across most of their distribution (Blandin and Purser 2013). We analyzed re-sequencing data obtained from 43 males of *M. helenor* and 33 individuals (20 males and 3 females) of *M. achilles* (Supplementary Table 1). This second dataset is referred to as the *sister-species dataset*.

DNA extractions and genome sequencing of M. helenor and M. achilles

For the *sister-species dataset*, DNA for each individual was extracted from thorax muscle using the DNeasy Blood & Tissue Kit following the producer instructions. In most cases, DNA was extracted from SNAP-frozen individuals or samples preserved in DMSO, but we also used 13 samples of dried pinned *M. achilles* from the personal collection of Patrick Blandin (Supplementary Table 1).

Sequencing was then performed at the GeT-PlaGe core facility of INRAE. DNA-seq libraries were prepared using the Illumina TruSeq Nano DNA LT Library Prep Kit, following supplier instructions. Briefly, DNA was fragmented by sonication and adaptors were ligated. Eight cycles of PCR were then applied to amplify libraries. Library quality was assessed using an

Advanced Analytical Fragment Analyzer and quantified by QPCR using the Kapa Library Quantification Kit. Sequencing was performed on an Illumina Novaseq 6000, using a paired-end read length of 2x150 pb on a S4 Flowcell.

Population genomics of Spiroplasma *in the sister-species dataset*

Adaptors were removed from the reads with cutadapt (Martin 2011) and the reads of each individual were mapped against the original *Spiroplasma* genome found in the reference genome of *M. achilles* (referred to as *sAch* hereafter) using BWA-mem v0.7.17 (Li 2013). Then we used Samtools v1.10 (Li et al. 2009) to sort the resulting SAM and BAM files and recover the fasta sequences from *Spiroplasma* endosymbionts, which were then assembled using Megahit (Li et al. 2015) with default parameters. To assess the presence/absence of *Spiroplasma* in each individual, we blasted each contig against the *sAch* genome using BLASTN with e-value cuttoff = 10e-50 and identity percentage >90%. Matching contigs were extracted and the presence/absence of the endosymbiont in a sample was classified as follows: (i) 'presence' when more than 100kb of aligned matching contigs were obtained, (ii) 'ambiguous' when 5-100kb of aligned contigs were obtained, (iii) 'absence' when less than 5kb of aligned sequences were obtained. To validate this classification, we identified the 16S rDNA gene from each sample using BLASTN against the 16S rDNA gene from the reference genome *sAch*, with e-value cutoff of 10e-50 and identity percentage >90%. Then, we estimated a phylogenetic tree for *Spiroplasma* using the 16S rDNA sequences retrieved from the different butterfly samples, following the same procedure stated earlier, and using the *Spiroplasma mellifera* KC3 sequence (NCBI accession CP029202) as outgroup.

Finally, RIP/Spaid toxin genes were also retrieved from each butterfly sample by applying a TBLASTN using the RIP/Spaid toxin gene identified in the sAch genome as seed, with e-value cuttoff = 10e-10.

Tree reconciliation analysis in order to predict host lateral switches

Spiroplasma, as maternally transmitted endosymbionts, are inherited together with the mitochondrial genome of the host. Therefore, we used ecceTERA (Jacox et al. 2016) to reconcile the mtDNA phylogeny of *Morpho* with the *Spiroplasma* 16S rDNA phylogeny previously obtained, and for which identical sequences were removed to eliminate tree polytomies. SylvX (Chevenet et al. 2016) was used to visualize and interpret the reconciliation tree.

The *Morpho* phylogeny was estimated with whole mitochondrial genomes, for which we combined those obtained in the genus-dataset and the sister-species datasets. Mitochondrial genomes for all individuals in the sister-species dataset were extracted directly from Illumina reads with GetOrganelle v1.7.5.3 (Jin et al. 2020) and the parameters -R 10 -k 21,45,65,85,105 -F animal_mt. Alignments were generated using MAFFT (Katoh et al. 2002) and manually curated to exclude ambiguous regions. The phylogeny was obtained with IQ-TREE v2.1.3 (Nguyen et al. 2015), estimating the best substitution model with ModelFinder (Kalyaanamoorthy et al. 2017), and assesing branch support with 1000 UltraFast boostrap replicates (Hoang et al. 2018). We used the *Heliconius melpomene* sequence (NCBI accession HE579083) as outgroup.

Results

Morpho butterflies are sporadically associated with Spiroplasma symbionts

We sequenced and surveyed genomes of 11 species of *Morpho* butterflies (including two different morphs of *M. telemachus*), for the presence of endosymbiotic bacteria (so-called *genus dataset*). Contigs were binned using their GC%, read coverage and taxonomic assignation (Figure 1). The resulting blob-plots indicate the presence of a limited number of symbionts: the genomes of *M. achilles*, *M. amathonte* and *M. rhetenor* had contigs with very low GC% that match with *Spiroplasma*, whereas the genomes of *M. hecuba* and *M. helenor* had contigs associated with *Wolbachia* and *Enteroccocus*. In contrast, the genomes of the remaining six species of *Morpho* (i.e. *M. marcus*, *M. eugenia*, *M. telemachus*, *M. menelaus, M. deidamia* and *M. granadensis)* do not seem to harbor symbiont sequences (Supplementary Figure 1).

We recovered and assembled three *Spiroplasma* genomes (*sAma*, *sAch* and *sRhe)* from the genomes of *M. amathonte*, *M. achilles* and *M. rhetenor*, respectively. The *sAch* assembly was the most complete, containing 98% of a set of lineage-specific, single-copy, *Spiroplasma* marker genes (Table 1). The other two (*sAma* and *sRhe*) had lower completeness (60% and 78%, respectively, Table 1), indicating that only a fraction of the corresponding genomes was captured.

Table 1. Assembly statistics of the endosymbiont genomes identified in genomes of *Morpho* butterflies

Name	Host	Taxonomic assignment	Size (kb)	N50 (kb)	Completness (%)	Contigs (#)	Coding ratio (%)	GC (%)
sAch	*M. achilles*	*Spiroplasma sp.*	4075	2667	98	46	46	24
sAma	*M. amathonte*	*Spiroplasma sp.*	3144	2570	60	16	45	24
sRhe	*M. rhetenor*	*Spiroplasma sp.*	2860	157	78	44	46	24
wHec	*M. hecuba*	*Wolbachia*	1444	1444	100	1	84	34
eHel	*M. helenor*	*Enterrococcus faecalis*	4618	2787	100	38	86	36

Inflation of genome size of Spiroplasma *found in* Morpho *butterflies*

The assemblies of *Spiroplasma* retrieved from *Morpho* genomes display a considerable larger genome size (2,9Mb to 4,1Mb) than the 62 previously-published genomes of this endosymbiont (1.1 Mb to 1.9 Mb; Supplementary Figure 2). The sAch and sAma assemblies contained a large contig of 2,7 Mb and 2,5 Mb, respectively, with low levels of synteny (Supplementary Figure 3). The sAch and sAma assemblies were also composed of 46 and 15 small contigs ranging from 17kb to 74kb, while the assembly of sRhe is composed of 44 contigs ranging from 14kb to 320kb (Figure 1). The 46 small contigs of the sAch assembly fall into 4 clusters based on sequence alignments, but all of them differ in size and/or in nucleotide similarity (Supplementary Figure 3).

To assess how much new genes contributed to the expansion of genome size in the *Spiroplasma* associated with *Morpho,* we searched for orthologous genes in sAch, which is the most complete assembly. Similar to other *Spiroplasma*, sAch contains a number of conserved orthologous group of genes that ranges from 600 to 1000 ortho-groups, but has an unusually large number of species-specific genes (>350 singletons; Supplementary Figure 4).

Spiroplasma found in Morpho *are divergent from* Spiroplasma *in other Lepidoptera*

To investigate the evolutionary origin of the *Spiroplasma* detected in *Morpho*, we built a phylogeny of this endosymbiont using a set of 72 concatenated single copy genes present in all *Spiroplasma* genomes available in the database, with recognizable homologs in the sAch complete genome. The assemblies *sAma*, *sAch* and *sRhe* retrieved from *Morpho* were monophyletic and are included within the *citri* clade (Figure 2), which includes diverse plant pathogens and endosymbionts of insects such as Hemiptera (*Spirolasma kunkelli*), Diptera (*S. sp. s*Nigra) and Hymenoptera (*S. melliferum*). The 16S rDNA phylogeny that includes a broader taxonomic dataset confirms this observation (Supplementary Figure 6). Intriguingly, the *Spiroplasma* recovered from *Morpho* are highly divergent from those found in other Lepidoptera such as the moth *Homona magnanima* (*S. ixodetis* sHM) or the nymphalid butterfly *Danaus chrysippus* (*S. sp. Danaus chrysipus*), both in the *ixodetis* clade (Figure 2) .

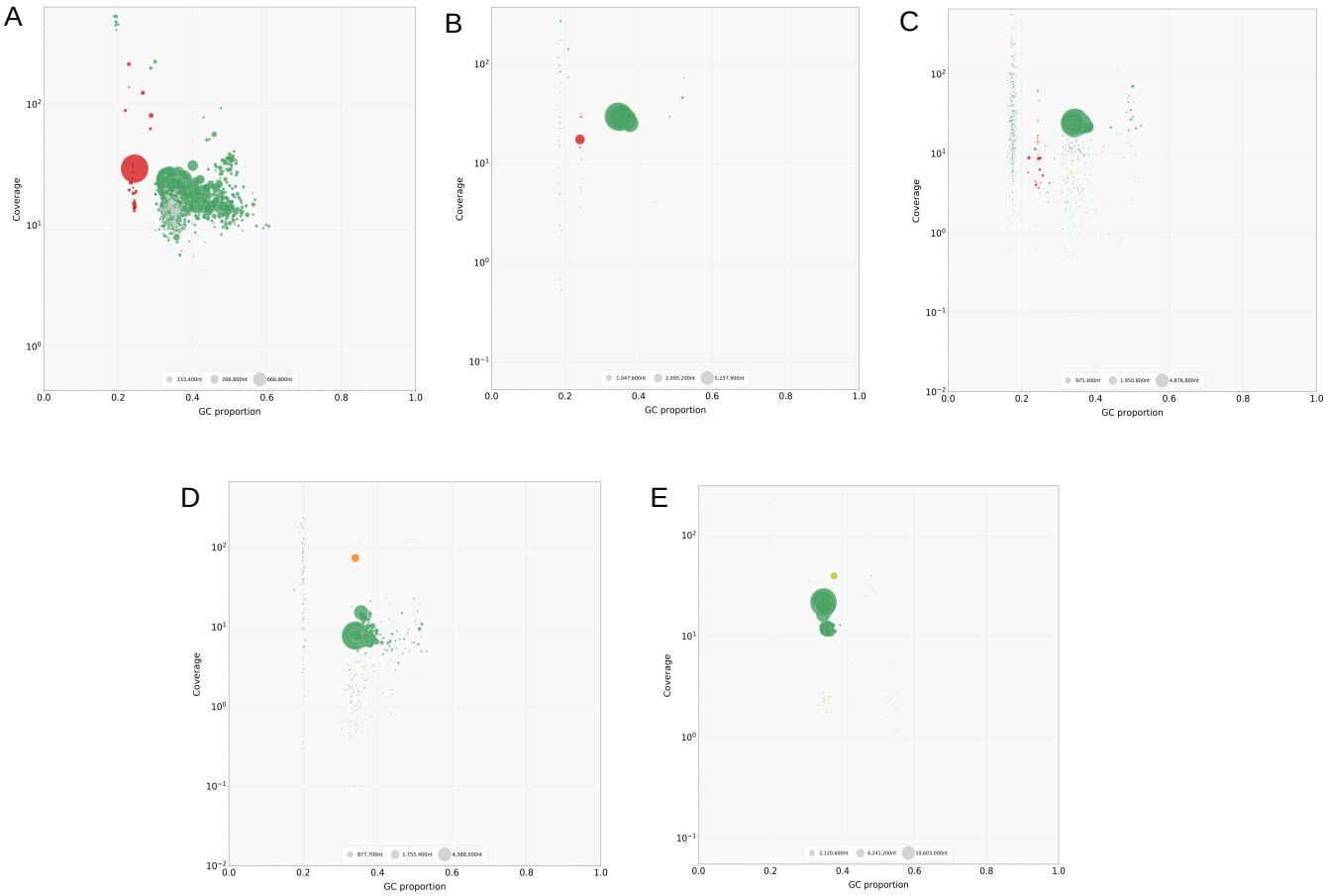

Figure 1: Detection of symbionts within genomic sequences of *Morpho*: *M. achilles* (A), *M. amathonte* (B), *M. rhetenor* (C), *M. hecuba* (D) and *M. helenor* (E). Contigs represented as circles were binned based on their GC%, read coverage, and taxonomic assignation. Dark green contigs matched arthropod sequences, red contigs matched Mollicutes (*Spiroplasma*), orange contigs matched Proteobacteria (*Wolbachia*), and light green contigs matched Firmicutes (*Enterococcus*). The size of the circle is proportional to the size of the contigs.

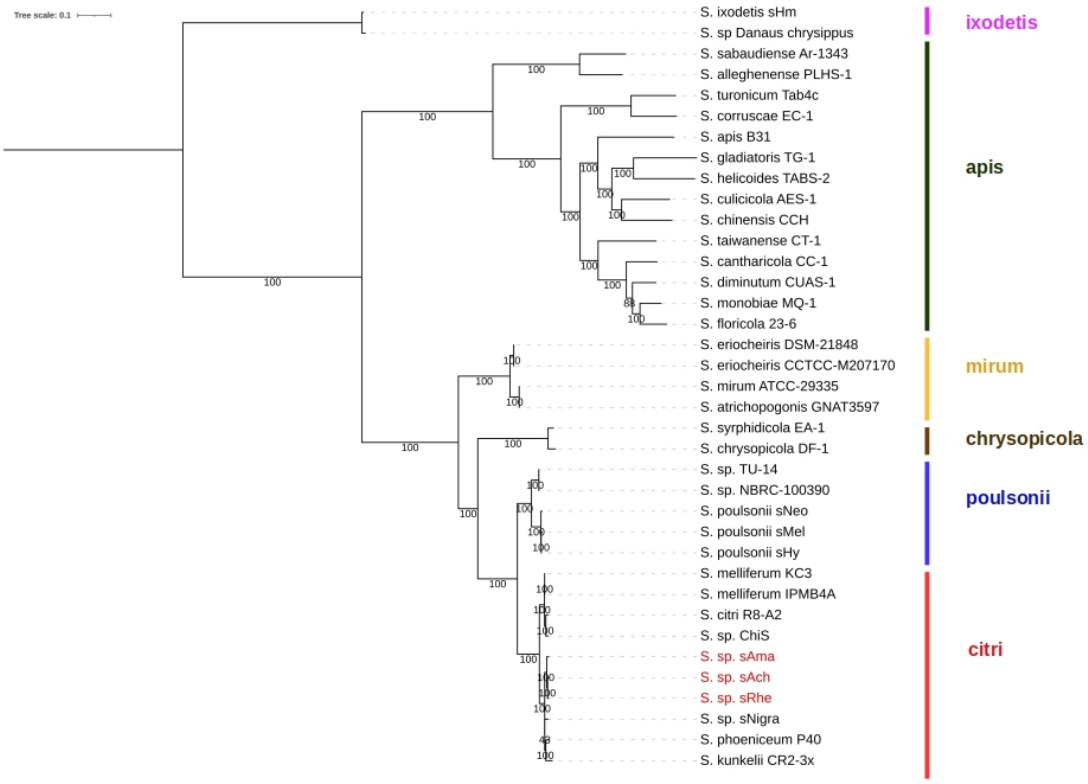

Figure 2: Phylogenetic reconstruction of *Spiroplasma* from different hosts based on the conserved set of 72 single copy orthologous genes. Recognized *Spiroplasma* clades are indicated in colors (right), and the *Spiroplasma* detected in *M. amathonte*, *M. achilles* and *M. rhetenor* (sAma, SAch and sRhe, respectively) are highlighted in red. Phylogeny was constructed using 1000 bootstrap replicates. The phylogeny made with an extended gene dataset of 96 orthologs including only sAch produces the same topology (Supplementary Figure 5). Note that the *Spiroplasma* documented in other Lepidoptera (the butterfly *Danaus chrysippus* and the moth *Homona magnanima*) fall in the distantly-related ixodetis clade (pink).

Proliferation of mobile genetic elements in Spiroplasma found in *Morpho*

We found multiple mobile genetic elements (i.e., prokaryotic transposons, ISs, plasmids and prophages) integrated in the large genomes of the *Spiroplasma* retrieved from *Morpho.* We observed that ISs are unusually abundant in these endosymbiont genomes, reaching a record-level in prokaryotes, that ranges from 398 to 885 copies (Figure 3A). The sAch assembly suggests this proliferation is associated with a surprisingly low number of IS families (Figure 3B). Indeed, only four IS families have expanded: an IS3-like family with 68 complete copies, an IS30 family with 96 complete copies, and a IS481 family with 72 intact copies. The fourth group is a 22kb composite transposon that we named Tn_sAch. This transposon has two IS3 copies at the tips and 24 conserved ORFs in the middle. In the assembly sAch, we observed Tn_sAch elements being especially common in the small contigs (constituting ~54% of their length) and much less frequent in the large 2,7Mb contig (accounting for only 9% of its lenght). The strong structural conservation of the backbone of the 8 intact copies of Tn_sAch suggests *en bloc* successive transpositions in the genome.

Sequence similarities of the different transposon copies analyzed as a proxy of the age of the different transposition events, indicates that most of them are identical or nearly identical, suggesting very recent transpositions (Figure 3C), but the copies of Tn_sAch have higher sequence divergence suggesting the presence of older copies.

The IS30 and IS481 transposons have generated non-autonomous Miniature Inverted repeat Transposable Elements (MITEs) by internal deletion leading to smaller transposons that represent 25% and 32% of the size of the parental elements (Figure 3D). In contrast, most of the complete autonomous IS copies are 100% full-length and presumably intact, showing few truncated copies. Such high level of complete and identical IS copies strongly suggests that these families have recently expanded in the genomes of the *Spiroplasma* found in *Morpho*.

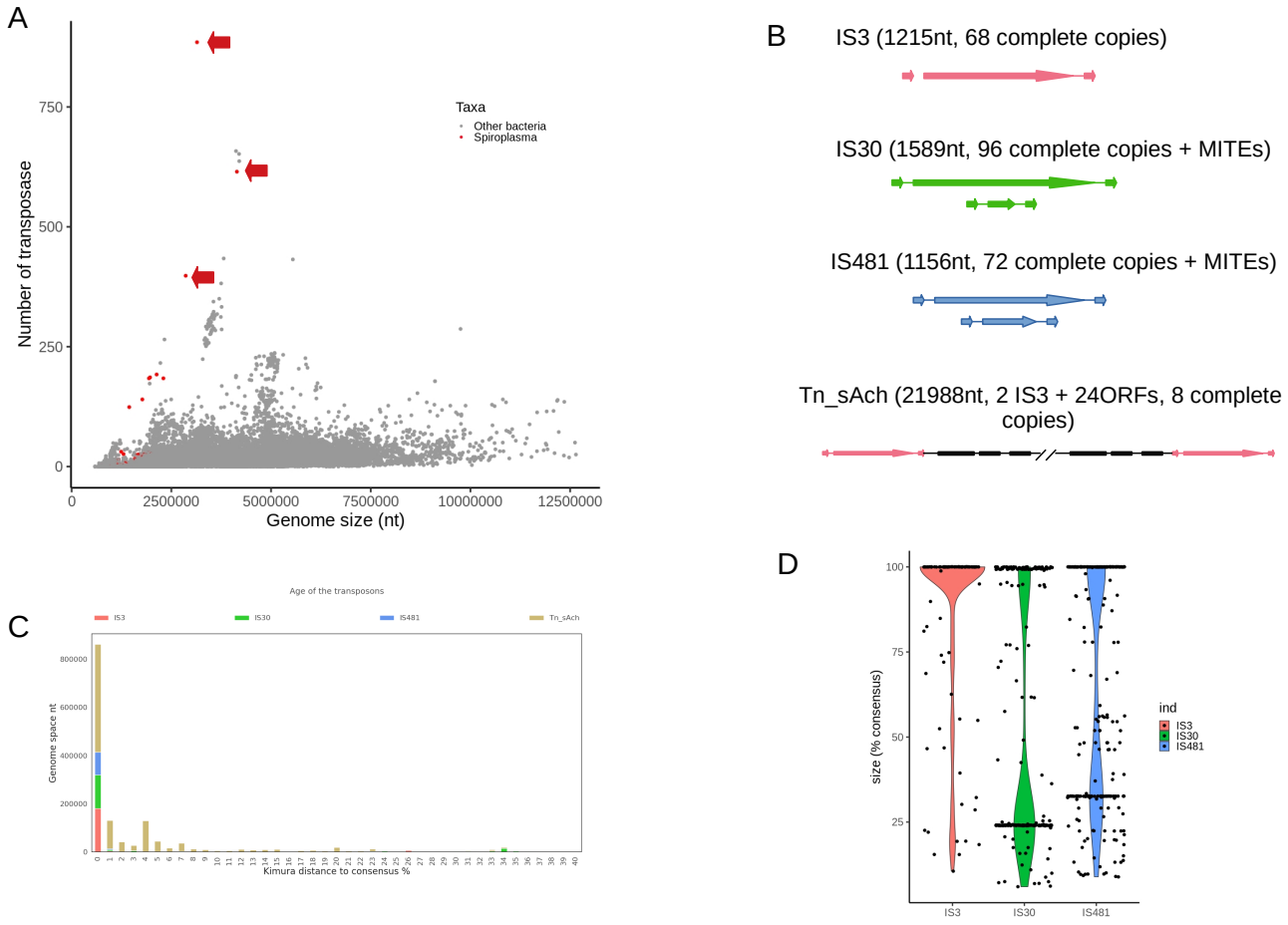

Figure 3: Insertion Sequences (ISs) found in the genomes of *Spiroplasma* found in *Morpho*. A: Number of transposase encoding genes found in a set of 25,675 prokaryotic genomes that include 62 *Spiroplasma* genomes (red dots) and three *Morpho Spiroplasma* (red dots with arrows) plotted against their genome sizes. **B**: Structure of the IS families found in the complete sAch genome and their main properties. Each colors represent a distinct IS families, arrows correspond to transposase genes and their internally deletted derivatives (MITEs), black rectangles indicate passenger genes of the compostite transposon. **C**: Analysis of the age of the IS copies in the sAch genome using the Kimura 2-parameter distance between the consensus sequence of a given family and all the individual copies that compose the family. The results are ordered based on the total amount of nucleotides. **D**: Analysis of the completeness of the different IS copies found in the sAch genome estimated as the percentage of the total length of the corresponding consensus sequences.

Our analyses of phage sequence invasion reveal the 28 to 32 integrated prophages in the genomes of *Spiroplasma* in *Morpho,* accounting for 418 kb in sAma, 482 kb in sRhe, and 538

kb in sAch (Supplementary Figure 7). The density of phage-derived elements (7.8/Mb) is thus larger than that in the most prophage-rich bacterial genome known to date (6.9/Mb), (Touchon et al. 2016; Frost et al. 2020). Interestingly, more than 80% of the sAch species-specific genes (singletons) are located in phage regions (Supplementary Figure 4). Functional analysis of these genes indicated that they are enriched by two KEGG functional categories: “genetic information processing” and “signaling and cellular processes” (Supplementary Figure 4). Therefore, the genome size expansion in *Spiroplasma* of *M. achilles* is associated with the accumulation of new genes acquired through interactions with phages.

Furthermore, the sAch, sAma and sRhe assemblies encode for six, four and five different plasmids respectively. They are all characterized by substantial higher level of read coverage than the genome contigs (Figure 1) suggesting the presence of multiple identical copies per bacterial cells.

Toxin genes identified in the Spiroplasma *genomes of healthy* Morpho *males*

As insects *Spiroplasma* are known to induce striking phenotypes in their host, such as male-killing promoted by the Spaid toxin or protection against parasites (RIP-like enzyme) (Harumoto and Lemaitre 2018; Ballinger and Perlman 2019), we specifically searched for these genes. One of the plasmids detected in both the sAch and sAma assemblies encodes for an ORF that combines both a RIP locus and a complete and structurally conserved Spaid gene (Figure 4B). This apparent bi-functional gene encodes for two RIP proteins in the 5’ end and a Spaid protein in 3’ end. The latter includes both ankyrin repeats (Ank) and a deubiquitinase domain (OTU), which are known to occur in the *Spiroplasma* strain sMel (Figure 4B) and induce male-killing in *Drosophila melanogaster* embryo. We sporadically observed ankyrin repeats, the OTU domain, and RIP domains in other genomes of *Spiroplasma* (Figure 4A). In particular, four homologous copies of RIP/OTU/Ank domains were also present in the sRhe assembly, and all of them are located on four different contigs. Although RIP-encoding genes are present in various *Spiroplasma* genomes (Figure 4A), the fusion of the RIP domain with the Spaid domain is an original feature found in all of the *Morpho Spiroplasma* genomes (Figure 4B). These features open the possibility that *Spiroplasma* endosymbionts may induce some peculiar phenotype in their *Morpho* butterfly hosts.

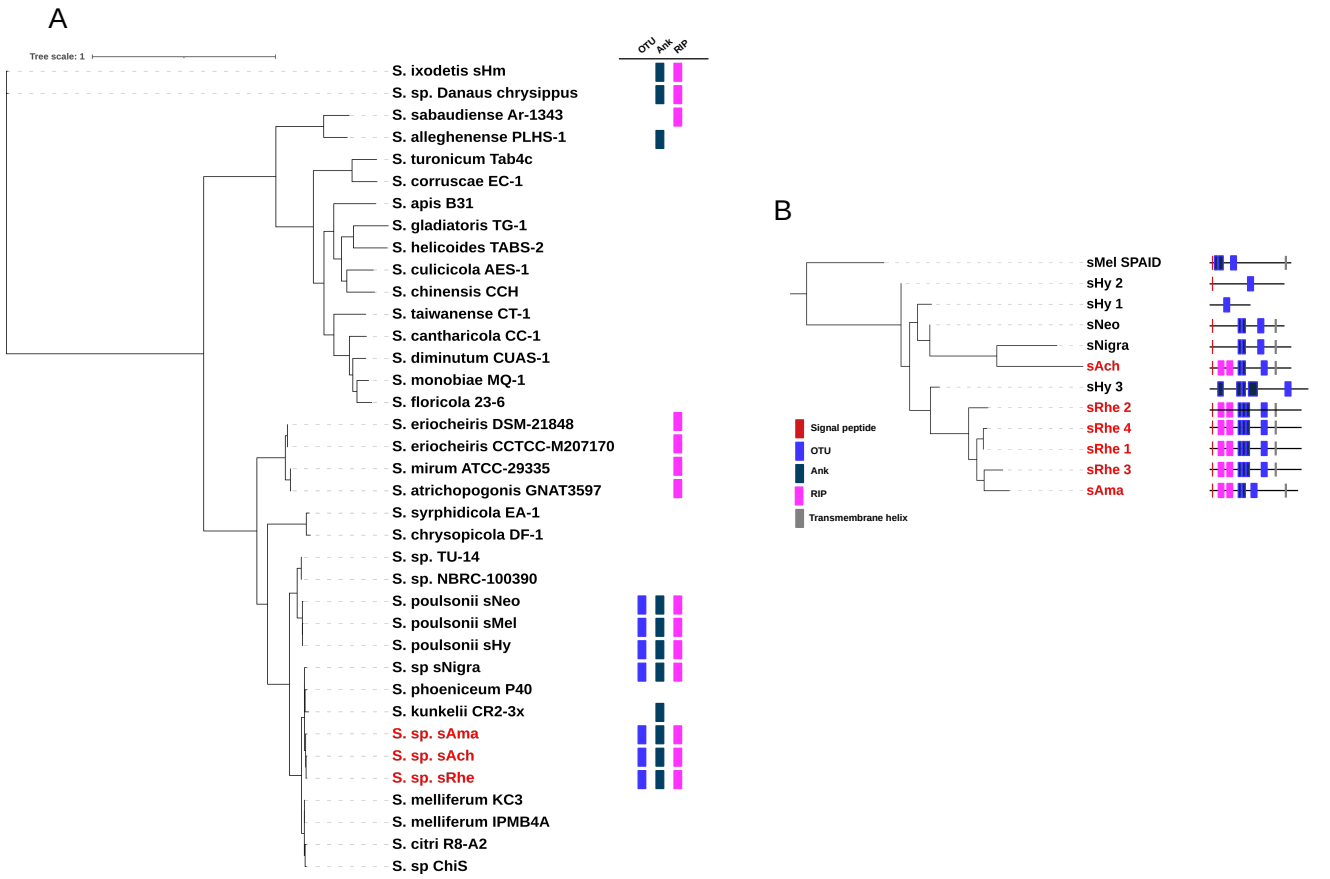

Figure 4: Distribution and organization of the toxin genes found in genomes of *Spiroplasma*. A: Distribution of the OTU (blue), Ankyrin (grey) and RIP (fucsia) encoding domains across the *Spiroplasma* whole-genome phylogeny. **B**. Phylogeny based on the OTU domain alignment of the Spaid-like proteins. Domain prediction based on Pfam similarity with known domains. The position of the *Morpho Spiroplasma* is highlighted in red in both trees.

Horizontal and vertical transfer of Spiroplasma *in* Morpho

To estimate the prevalence of *Spiroplasma* within species of *Morpho* and test for horizontal vs. vertical transfer of this endosymbionts, we searched for the presence of *Spiroplasma* in different populations of *M. achilles* and its sister and sympatric species *M. helenor.* We detected genomes of *Spiroplasma* with assembly size >100 kb and highly matching the sAch assembly in 26 out of 33 individuals of *M. achilles*; only 6 individuals had few contigs matching sAch (with assembly size <100kb), and a single one lacked any genomic trace of it (Figure 5 and Supplementary Table 1). Thus, all populations of *M. achilles* accross south-America had *Spiroplasma*, except for one population in Peru represented by a single individual in our study (Figure 5). By contrast, among the 43 *M. helenor* individuals from 27 populations we investigated, only a single individual had *Spiroplasma*. Therefore, although *M. achilles* and *M helenor* are sympatric species throughout the Amazonian basin and are closely-related species (3,6 millions years of divergence (Chazot et al. 2021), they display completely opposite patterns of infection by *Spiroplasma.*

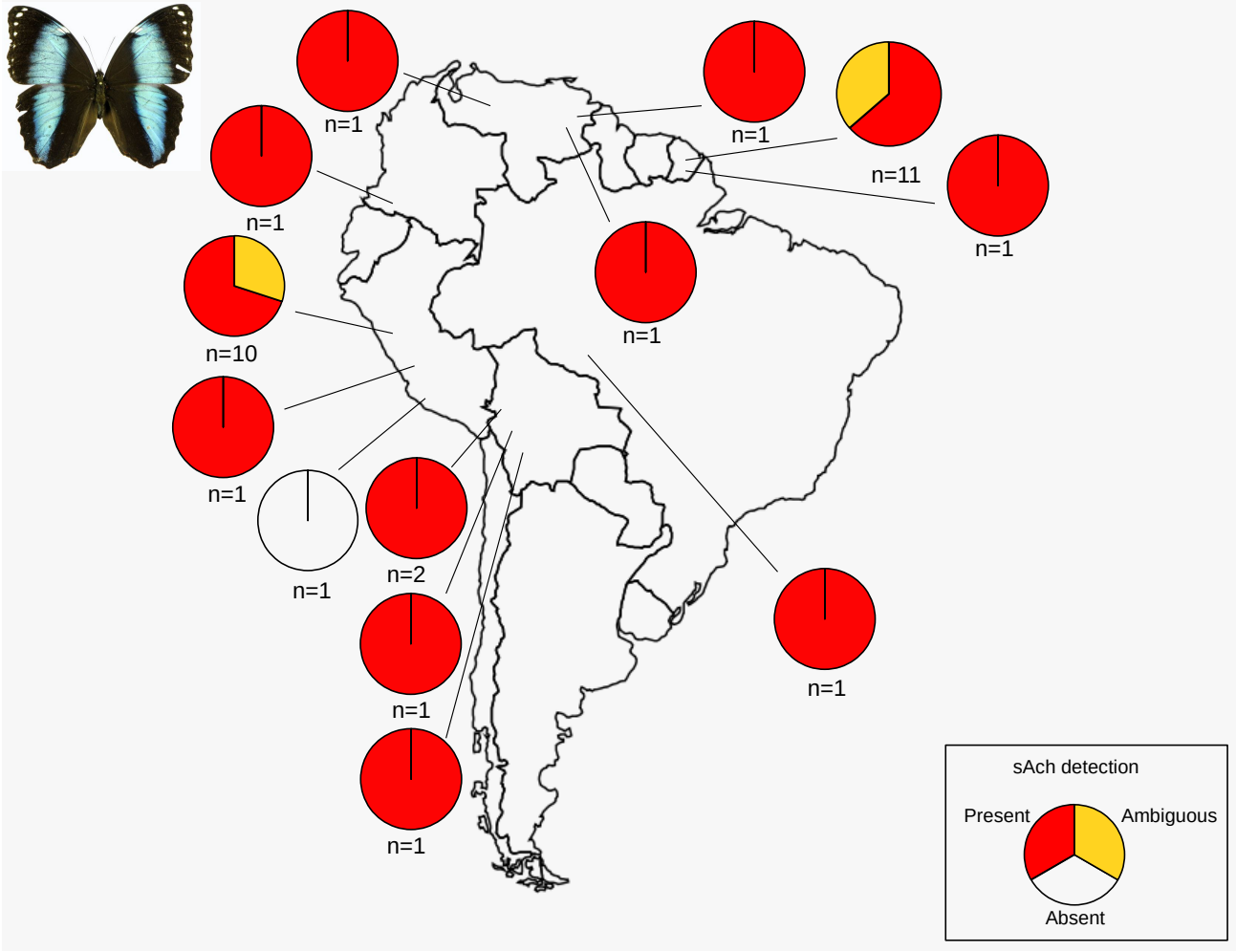

Figure 5: Geographic distribution of *Spiroplasma* in populations of *Morpho achilles*. Each circle represents a population of *M. achilles* and the number of individuals sampled per population is indicated on the bottom. The presence of *Spiroplasma* in each population is color coded as red (presence), yellow (ambiguous) and white (absence). See also Supplementary Figure 8.

Most of the *Spiroplasma* identified here had large genome assemblies (>1 Mb) albeit the use of short-read sequencing technology that generally failed to assemble highly repeated regions. Moreover, some assemblies reached sizes comparable to the reference sAch assembly (>3 Mb; Supplementary Figure 8). The 16S rDNA gene was present in most of these *Spiroplasma* assemblies, and homologous RIP and Spaid toxin genes were also found almost universally (Supplementary Figure 8). In the 16s rDNA phylogeny, the only *Spiroplasma* genome retrieved from *M. helenor* and the sRhe *Spiroplasma* from *M. rhetenor* both appear well nested into the sAch clades from *M. achilles* (Figure 6 and Supplementary Figure 8). This suggests a putative horizontal transmission of *Spiroplasma* between *M. helenor*, *M. achilles* and *M. rhetenor* living in sympatry. The reconciliation analysis between the *Spiroplasma* tree (16S rDNA) and the *Morpho* tree (whole mitochondrial genomes) are highly congruent suggesting that both the endosymbiont and mtDNA are maternally inherited (Figure 6), except for sHel and sRhe. This finding strongly suggests lateral exchange of *Spiroplasma* between *M. achilles*, *M. rhetenor* and *M. helenor* (Figure 6). The high congruence between the *Spiroplasma* and the mitochondrial tree within *M. achilles* agrees with a predominant vertical maternal transmission.

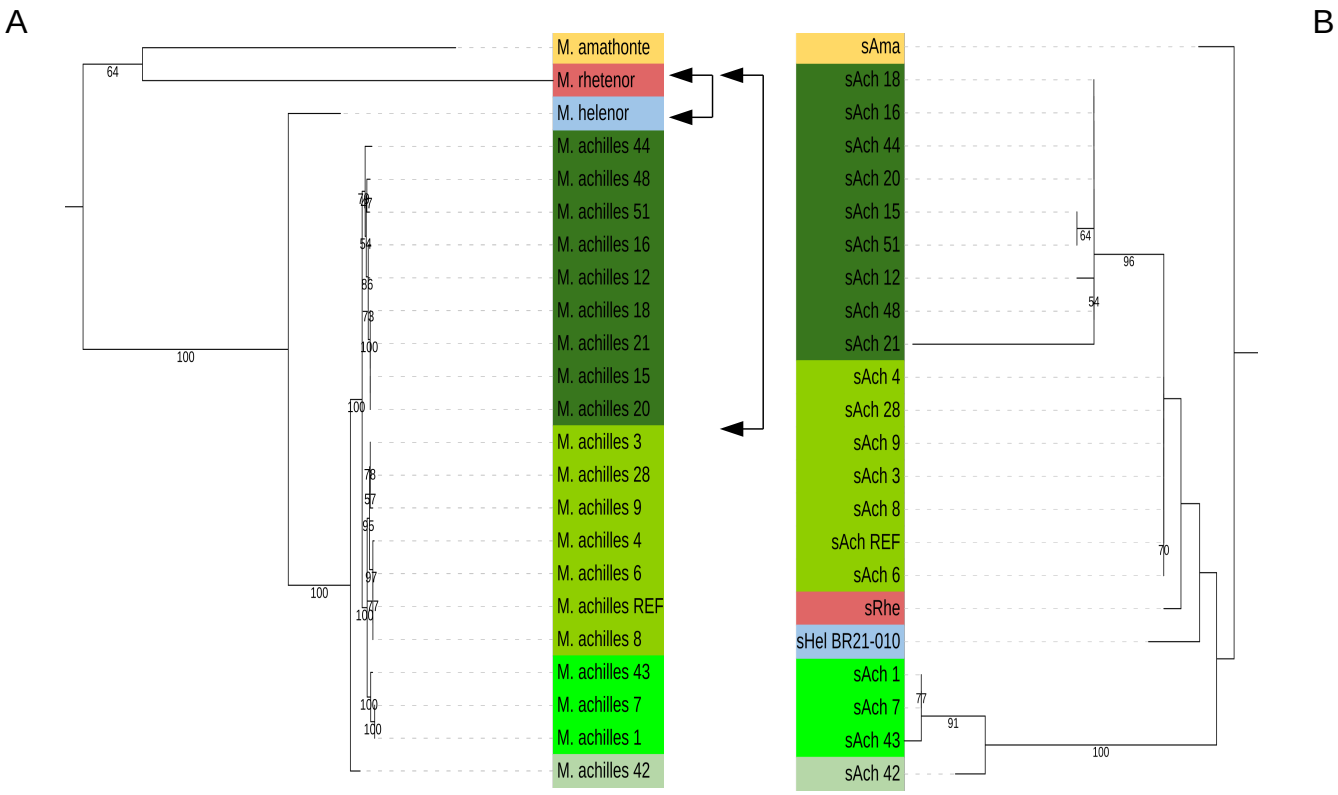

Figure 6: Reconciliation analysis between the host tree (A) contrasted against the endosymbiont tree (B) with predicted horizontal host switches of *Morpho Spiroplasma.* Whole mitochondrial genome tree for *Morpho* against the 16S rDNA tree for *Spiroplasma.* Ultra-fast boostraps values are indicated on each branch. Color blocs correspond to the main phylogenetic clusters identified in the species phylogeny. Double-arrows highlight the possible horizontal host switches.

Discussion

Peculiar evolution of Spiroplasma *genomes in* Morpho *butterflies*

We documented the presence of the bacterial endosymbiont *Spiroplasma* in four out of 11 species of *Morpho* butterflies studied here, highlighting that the association with this endosymbiont greatly varies across closely related host species, even when they live in sympatry. Surprisingly, the genomes of *Spiroplasma* retrieved from *Morpho* sharply differ from those in other Lepidoptera, suggesting horizontal transfer among distantly-related host species. The phylogenetic discrepancies between *Spiroplasma* genes and mitochondrial genes across *Morpho* species also suggests that horizontal transfer between species living in sympatry might occur. The circulation of *Spiroplasma* in the hemolymph is thought to facilitate horizontal transfer across sympatric species, for instance through the consumption of hemolymph by mites (Jaenike et al. 2010). Because the *Spiroplasma* found in *Morpho* butterflies is closely related to strain documented as a plant pathogen, the inter-specific transmission of this symbionts could also be enabled by the shared consumption of host plants by caterpillars. Remarkably, the three *Morpho* species presenting close *Spiroplasma* (*M. achilles*, *M. helenor* and *M. rhetenor)* share several hostplants in French Guyana, supporting this hypothesis (Anon 2017). The presence of *Spiroplasma* in *Morpho* butterflies might thus stem from interactions with other insects or through host plant consumption.

However, population genomics of *Spiroplasma* within the species *M. achilles* species indicate widespread distribution and predominance of vertical transmission at the species level. Interestingly, we have evidenced the presence of *Spiroplasma* symbiont in *M. amathonte* and *M. achilles*, two species that have diverged from more than 17 My (Chazot et al. 2021) and that currently occupying non-overlapping geographical ranges: *M. amathonte* is found in central America and in the western slopes of the Andes while *M. achilles* inhabits the Amazonian basin. The weak genome synteny of the *Spiroplasma* genomes of these two species and the congruence of host and symbiont phylogenies support a vertical transmission from a common ancestor. Thus, the ubiquitous presence of *Spiroplasma* in *M. achilles* populations across south-America and the predominant vertical transmission of the symbiont at various evolutionary time-scale point at a long-term association between *Spiroplasma* symbiont and *Morpho* butterflies. The absence of *Spiroplasma* from some *Morpho* species (e.g. *M. menelaus*, closely related to *M. amathonte*) would then suggest a secondary loss.

Massive genome size promoted by large expansion in diverse mobile genetic elements.

While endosymbionts as *Spiroplasma* display streamlined genomes (Gerth et al. 2021), the genome size of the *Spiroplasma* observed in our study is surprisingly large. Our analyses indicate that recent and massive expansion of diverse mobile genetic elements (MGE) are responsible for this striking inflation in genome size. Record-level of Insertion Sequences (accounting for 1550 kb), integrated prophages (538kb) and plasmids (278 kb) represent nearly 60% of the total genome size of the *Spiroplasma* found in *M. achilles* (accounting for 2366 kb on a total genome size of 4075 kb). All *Spiroplasma* genomes detected in *Morpho* butterflies display such extreme expansion of MGE, in sharp contrast with the paucity of MGE generally found in *Spiroplasma* (Gerth et al. 2021) or *Wolbachia* genomes (Cerveau et al. 2011). Moreover, most of the IS copies found in *Spiroplasma* of *Morpho* butterflies are recently-transposed elements, indicating an ongoing and continuous accumulation. Such MGE proliferation has also been sporadically observed in *Orientia* symbionts, a widespread *Rickettsia*-like, intra-cellular bacteria associated with mites (Batty et al. 2018) and in *Mycoplasma* endosymbionts, associated with diverse fungi (Naito and Pawlowska 2016). In addition, integrated phage genomes provide numerous new genes and functions in *Spiroplasma* associated with *Morpho* butterflies. The massive expansion of MGEs leads to an inflated genome size but also provides a source of new genes and functions expanding the diversity of the genomic repertoire of *Spiroplasma* symbionts infecting *Morpho* butterflies.
Recombination induced by MGEs and the gene flow provided by phage genome integration can explain the lack of genome erosion in the S*piroplasma* of *Morpho*, that contrasts with the important genome size reduction observed in most endosymbionts. Such rapid evolution might stem from peculiar adaptation in the symbionts of *Morpho*, allowing long-term association and high prevalence in some *Morpho* species. For instance, MGE-encoded toxin genes might have contributed to increase the symbiont persistence in some *Morpho* butterfly populations.

Evolution of toxin genes and putative protective effect

By detecting *Spiroplasma* in adult *Morpho* males sampled in the wild, our study suggests that the presence of *Spiroplasma* does not prevent the development of males in *Morpho* butterflies, in sharp discrepancy with the male-killing effects reported in the butterfly *Danaus chrysippus* (Jiggins et al. 2000), but similar to the results obtained with *Wolbachia* in Neotropical Acraeini (Nymphalidae) (Silva-Brandão et al. 2021). More specifically, the toxin

gene *Spaid* found in *Spiroplasma poulsonii* and documented to trigger male killing in *Drosophila* (Harumoto and Lemaitre 2018) does not have the male-killing (MK) effect in *Morpho*. It is thus possible that *Morpho* butterflies have developed resistance to the sex-ratio distortion effect of the *Spiroplasma Spaid* toxins. Such MK-suppression have indeed been observed in plant-hopper (Yoshida et al. 2021) and in lacewing (Hayashi et al. 2018). Alternatively, as *Spaid* toxins target the gene dosage compensation system that increases the transcription of genes on the male single X chromosome in *Drosophila* (Harumoto & Lemaitre 2018), it is also possible that this toxin is ineffective in ZW sex-determination system for which the female is heterogametic. Supporting this view, MK-inducing *Spiroplasma* in Lepidoptera lack the *Spaid* toxins genes, but the genetic determinant(s) of the MK phenotype are unknown (Arai et al. 2022). However, strong conservation of the *Spaid* genes among *Morpho* species and populations favors the idea that it provides a selective advantage.

Interestingly, our study also reveals the evolution of specific architecture of toxin genes in the *Spiroplasma* of *Morpho* butterflies, including *RIP* physical linkage with the *Spaid* gene. While the functional implication of this evolution cannot be inferred from our current results, the high conservation of this specific architecture in the *Morpho* genomes is consistent with an adaptive role. Moreover, the localization of these genes on plasmids also suggests that they may spread across bacteria, and their persistence might have been promoted by natural selection, either because they act as selfish elements or because of positive impact on host fitness. In *Drosophila*, RIP proteins produced by *Spiroplasma poulsonii* have indeed been documented to induce positive effects on host survival, through their protective effect against nematods (Stevens et al. 2001), as well as parasitoid wasps (Ballinger and Perlman 2019). Such defensive effect of *Spiroplasma* could have a positive impact on the fitness of *Morpho* butterflies and might explain their high prevalence in *M. achilles.*

Contrasted prevalence of Spiroplasma *in sympatric sister species: do* Spiroplasma *impact reproductive isolation?*

The evolution of the *Spaid* gene might have resulted in a change of function in *Morpho* butterflies, disabling the male-killing mechanism. Alternatively, resistance to male-killing effect might have evolved in *Morpho* butterflies. The evolution of MK-suppression has been documented in natural populations of the butterfly *Hypolimnas bolina* (Nymphalidae) infected by *Wolbachia* (Hornett et al. 2022). Such evolution of resistance is likely to be under strong positive selection given the high fitness costs for the hosts induced by male-killing genes (Hornett et al. 2022). The *Spiroplasma* is highly prevalent in *M. achilles* and quite rare in the sympatric species *M. helenor* despite the similarity of the *Spaid* gene in both species, this might suggest that resistance to the deleterious effect of *Spiroplasma* could be restricted to *M. achilles*.

Alternatively, the presence of the symbiont might trigger cytoplasmic incompatibilities (CI), explaining the huge difference in its prevalence between these two sister-species living in sympatry. *Spiroplasma*-induced cytoplasmic incompatibilities have been recently documented in the wasp *Lariophagus distinguendus* (Pollmann et al. 2022). In case of CI, crosses between infected males and uninfected females generally do not produce offspring; CI could thus limit genetic exchange between the two sympatric species.

Altogether, our current results on the contrasted *Spiroplasma* prevalence in these sister-species therefore raises the question of the potential impact of this endosymbiont as barrier to gene flow between these sympatric species. Endosymbionts like *Spiroplasma* and *Wolbachia* have indeed been suggested to generate post-zygotic barriers to gene flow (see (Duplouy and Hornett 2018) for a review), but their role in initiating *vs.* reinforcing speciation remains

largely uncovered. Our study therefore points at the needs to investigate the effect of *Spiroplasma* on reproductive isolation and its significance for species diversification and co-existence in sympatry.

Conclusion

Studies on heritable symbionts in natural populations generally support a dynamic model of gene gain and loss shaped by the ecological interactions with their host. On the other hand, heritable symbiont lifestyle induces genetic isolation and population bottlenecks that lead to mutational decay and genome streamlining. In comparison with *Spiroplasma* genomes retrieved from other insects, *Morpho Spiroplasma* genomes display a massive expansion of diverse mobile genetic elements as transposable elements, prophages or plasmids. In addition, we documented a strong conservation of toxin RIP and Spaid-encoding genes in the *Spiroplasma* of *Morpho* species, that might enhance protection of the butterflies against parasites. The study of *Spiroplasma* symbionts in natural population of diverse *Morpho* butterfly species support a stable association in *Morpho achilles* populations across south-America, whereas *Spiroplasma* appears almost absent in sympatric *M. helenor* populations. This contrasted symbiont distribution among sympatric *Morpho* species is associated with a global predominant vertical transmission of the symbiont supporting a model in which the symbiont provides fitness advantages to the butterfly. Indeed, *Morpho Spiroplasma* genomes display a remarkable resistance to genome erosion by the mean of massive expansions of diverse mobile genetic elements as transposable elements, prophages or plasmids. In addition, the strong conservation of toxin RIP and Spaid-encoding genes might be the key-drivers of this lasting association either by conferring host protection against parasite and/or by limiting hybridization with symbiont-free sympatric *Morpho* species. Our study calls for additional investigations of the phenotypic effects of the *Spiroplasma* on their butterfly hosts to better understand how their ecological interactions shapes – and is shaped by – their evolution.

Acknowledgments

We would like to thank Heloïse Bastide and David Ogereau for their helps at the initial phase of this work, Josephine Ledamoisel, Rémi Mauxion and Owen MacMillan for their support during fieldwork in Panama, as well as Mathieu Chouteau and Mélanie McClure during fieldwork in Peru and French Guiana. Some of the analyses and simulations were performed on the the Plateforme de Calcul Intensif et Algorithmique PCIA (MNHN/CNRS), on the MeSU platform at Sorbonne-Université and on the Genotoul bioinformatics platform Toulouse Occitanie (Bioinfo Genotoul).

Data Availability

PacBio *(genus dataset*) and Illumina (*sister-species dataset*) read sequences and genome assemblies generated in this study were deposited in the NCBI database under the bioprojects PRJNA1069011 and PRJNA1063620. Primary raw data for each analysis can be downloaded at https://doi.org/10.6084/m9.figshare.24582867.v1

Funding

This work was supported by the European Research Council (ERC) Consolidator Grant 101088089 “OUTOFTHEBLUE”, the ANR-tremplin and the CNRS MITI project ECODEMO.

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 Research* 40:e126.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J. Mol. Biol.* 215:403–410.

Anon. 2017. Les Morpho – Distribution, diversification, comportement - Les Lépidoptéristes de France. Available from: https://www.lepidofrance.fr/monographies/morpho-distribution-diversification-comportement/

Arai H, Inoue MN, Kageyama D. 2022. Male-killing mechanisms vary between Spiroplasma species. *Frontiers in Microbiology* [Internet] 13. Available from: https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2022.1075199

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 Research* 44:W16–W21.
Ballinger MJ, Perlman SJ. 2019. The defensive *Spiroplasma*. *Current Opinion in Insect Science* 32:36–41.

Bastide H, López-Villavicencio M, Ogereau D, Lledo J, Dutrillaux A-M, Debat V, Llaurens V. 2023. Genome assembly of 3 Amazonian Morpho butterfly species reveals Z-chromosome rearrangements between closely related species living in sympatry. *GigaScience* 12:giad033.

Batty EM, Chaemchuen S, Blacksell S, Richards AL, Paris D, Bowden R, Chan C, Lachumanan R, Day N, Donnelly P, et al. 2018. Long-read whole genome sequencing and comparative analysis of six strains of the human pathogen Orientia tsutsugamushi. *PLOS Neglected Tropical Diseases* 12:e0006566.

Bennett GM, Moran NA. 2015. Heritable symbiosis: The advantages and perils of an evolutionary rabbit hole. *Proceedings of the National Academy of Sciences* 112:10169–10176.

Blandin P, Purser B. 2013. Evolution and diversification of Neotropical butterflies: Insights from the biogeography and phylogeny of the genus Morpho Fabricius, 1807 (Nymphalidae: Morphinae), with a review of the geodynamics of South AAmerica. *Tropical Lepidoptera Research*:62–85.

Buchfink B, Xie C, Huson DH. 2015. Fast and sensitive protein alignment using DIAMOND. *Nat Methods* 12:59–60.

Cabanettes F, Klopp C. 2018. D-GENIES: dot plot large genomes in an interactive, efficient and simple way. *PeerJ* 6:e4958.

Cerveau N, Leclercq S, Leroy E, Bouchon D, Cordaux R. 2011. Short- and Long-term Evolutionary Dynamics of Bacterial Insertion Sequences: Insights from Wolbachia Endosymbionts. *Genome Biology and Evolution* 3:1175–1186.

Chazot N, Blandin P, Debat V, Elias M, Condamine FL. 2021. Punctuational ecological changes rather than global factors drive species diversification and the evolution of wing phenotypes in Morpho butterflies. *Journal of Evolutionary Biology* 34:1592–1607.

Cheng H, Concepcion GT, Feng X, Zhang H, Li H. 2021. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nat Methods* 18:170–175.

Chevenet F, Doyon J-P, Scornavacca C, Jacox E, Jousselin E, Berry V. 2016. SylvX: a viewer for phylogenetic tree reconciliations. *Bioinformatics* 32:608–610.

Drew GC, Frost CL, Hurst GD. 2019. Reproductive Parasitism and Positive Fitness Effects of Heritable Microbes. In: Encyclopedia of Life Sciences. 1st ed. Wiley. p. 1–8. Available from: https://onlinelibrary.wiley.com/doi/10.1002/9780470015902.a0028327

Duplouy A, Hornett EA. 2018. Uncovering the hidden players in Lepidoptera biology: the heritable microbial endosymbionts. *PeerJ* 6:e4629.

Emms DM, Kelly S. 2015. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biology* 16:157.

Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, Heger A, Hetherington K, Holm L, Mistry J, et al. 2014. Pfam: the protein families database. *Nucleic Acids Research* 42:D222–D230.

Frost CL, Siozios S, Nadal-Jimenez P, Brockhurst MA, King KC, Darby AC, Hurst GDD. 2020. The Hypercomplex Genome of an Insect Reproductive Parasite Highlights the Importance of Lateral Gene Transfer in Symbiont Biology. *mBio* 11:10.1128/mbio.02590-19.

Gerth M, Martinez-Montoya H, Ramirez P, Masson F, Griffin JS, Aramayo R, Siozios S, Lemaitre B, Mateos M, Hurst GDD. 2021. Rapid molecular evolution of Spiroplasma symbionts of Drosophila. *Microbial Genomics* [Internet] 7. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8208695/

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.

Hayashi M, Nomura M, Kageyama D. 2018. Rapid comeback of males: evolution of male-killer suppression in a green lacewing population. *Proceedings of the Royal Society B: Biological Sciences* 285:20180369.

Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. 2018. UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Molecular Biology and Evolution* 35:518–522.

Hornett EA, Kageyama D, Hurst GDD. 2022. Sex determination systems as the interface between male-killing bacteria and their hosts. *Proceedings of the Royal Society B: Biological Sciences* 289:20212781.

Hurst GDD. 2017. Extended genomes: symbiosis and evolution. *Interface Focus* 7:20170001.
Jacox E, Chauve C, Szöllősi GJ, Ponty Y, Scornavacca C. 2016. ecceTERA: comprehensive gene tree-species tree reconciliation using parsimony. *Bioinformatics* 32:2056–2058.

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, Schulenburg JHG v d, Majerus MEN. 2000. The butterfly Danaus chrysippus is infected by a male-killing Spiroplasma bacterium. *Parasitology* 120:439–446.

Jin J-J, Yu W-B, Yang J-B, Song Y, dePamphilis CW, Yi T-S, Li D-Z. 2020. GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. *Genome Biol* 21:241.

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, Sato Y, Morishima K. 2016. BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences. *Journal of Molecular Biology* 428:726–731.

Katoh K, Misawa K, Kuma K, Miyata T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* 30:3059–3066.

King KC. 2019. Defensive symbionts. *Curr Biol* 29:R78–R80.

Krawczyk PS, Lipinski L, Dziembowski A. 2018. PlasFlow: predicting plasmid sequences in metagenomic data using genome signatures. *Nucleic Acids Research* 46:e35.

Laetsch DR, Blaxter ML. 2017. BlobTools: Interrogation of genome assemblies. Available from: https://f1000research.com/articles/6-1287

Li D, Liu C-M, Luo R, Sadakane K, Lam T-W. 2015. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* 31:1674–1676.

Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Available from: http://arxiv.org/abs/1303.3997

Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 34:3094–3100.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25:2078–2079.

Liu P, Li Yuxin, Ye Y, Chen J, Li R, Zhang Q, Li Yuan, Wang W, Meng Q, Ou J, et al. 2022. The genome and antigen proteome analysis of Spiroplasma mirum. *Frontiers in Microbiology* [Internet] 13. Available from: https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2022.996938

Manzano-Marín A, Coeur d'acier A, Clamens A-L, Cruaud C, Barbe V, Jousselin E. 2023. Co-obligate symbioses have repeatedly evolved across aphids, but partner identity and nutritional contributions vary across lineages. *Peer Community Journal* [Internet] 3. Available from: https://peercommunityjournal.org/articles/10.24072/pcjournal.278/

Marçais G, Delcher AL, Phillippy AM, Coston R, Salzberg SL, Zimin A. 2018. MUMmer4: A fast and versatile genome alignment system. *PLOS Computational Biology* 14:e1005944.

Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* 17:10–12.

Medina P, Russell SL, Corbett-Detig R. 2023. Deep data mining reveals variable abundance and distribution of microbial reproductive manipulators within and among diverse host species. *PLOS ONE* 18:e0288261.

Mistry J, Bateman A, Finn RD. 2007. Predicting active site residue annotations in the Pfam database. *BMC Bioinformatics* 8:298.

Mistry J, Finn RD, Eddy SR, Bateman A, Punta M. 2013. Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions. *Nucleic Acids Research* 41:e121.

Moran NA. 1996. Accelerated evolution and Muller's rachet in endosymbiotic bacteria. *Proceedings of the National Academy of Sciences* 93:2873–2878.

Naito M, Pawlowska TE. 2016. Defying Muller's Ratchet: Ancient Heritable Endobacteria Escape Extinction through Retention of Recombination and Genome Plasticity. *mBio* 7:10.1128/mbio.02057-15.

Nakayama S, Parratt SR, Hutchence KJ, Lewis Z, Price T a. R, 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. *Molecular Biology and Evolution* 32:268–274.

Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. 2015. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* 25:1043–1055.

Pollmann M, Moore LD, Krimmer E, D'Alvise P, Hasselmann M, Perlman SJ, Ballinger MJ, Steidle JLM, Gottlieb Y. 2022. Highly transmissible cytoplasmic incompatibility by the extracellular insect symbiont Spiroplasma. *iScience* 25:104335.

Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30:2068–2069.

Siguier P, Perochon J, Lestrade L, Mahillon J, Chandler M. 2006. ISfinder: the reference centre for bacterial insertion sequences. *Nucleic Acids Res.* 34:D32-36.

Silva-Brandão KL, Cirino M, Magaldi LDM, Gueratto PE, Mattos RG, Freitas AVL. 2021. Subspecies limits and hidden Wolbachia diversity in Actinote pellenea butterflies. *Systematics and Biodiversity* 19:1012–1025.

STEVENS L, GIORDANO R, FIALHO RF. 2001. Male-killing, nematode infections, bacteriophage infection, and virulence of cytoplasmic bacteria in the genus Wolbachia. *Annu. rev. ecol. syst* 32:519–545.

Sudakaran S, Kost C, Kaltenpoth M. 2017. Symbiont Acquisition and Replacement as a Source of Ecological Innovation. *Trends in Microbiology* 25:375–390.

Tanizawa Y, Fujisawa T, Nakamura Y. 2018. DFAST: a flexible prokaryotic genome annotation pipeline for faster genome publication. *Bioinformatics* 34:1037–1039.

Teufel F, Almagro Armenteros JJ, Johansen AR, Gíslason MH, Pihl SI, Tsirigos KD, Winther O, Brunak S, von Heijne G, Nielsen H. 2022. SignalP 6.0 predicts all five types of signal peptides using protein language models. *Nat Biotechnol* 40:1023–1025.

Tinsley MC, Majerus ME. 2007. Small steps or giant leaps for male-killers? Phylogenetic constraints to male-killer host shifts. *BMC Evolutionary Biology* 7:238.

Touchon M, Bernheim A, Rocha EPC. 2016. Genetic and life-history traits associated with the distribution of prophages in bacteria. *The ISME Journal* 10:2744–2754.

Watts T, Haselkorn TS, Moran NA, Markow TA. 2009. Variable Incidence of Spiroplasma Infections in Natural Populations of Drosophila Species. *PLOS ONE* 4:e5703.

Wernegreen JJ. 2017. In it for the long haul: evolutionary consequences of persistent endosymbiosis. *Current Opinion in Genetics & Development* 47:83–90.

Werren JH. 1998. Wolbachia and Speciation. In: Howard DJ, Berlocher SH, editors. Endless Forms: Species and Speciation. Oxford University Press. p. 0. Available from: https://doi.org/10.1093/oso/9780195109009.003.0018

Yoshida K, Sanada-Morimura S, Huang S-H, Tokuda M. 2021. Silence of the killers: discovery of male-killing suppression in a rearing strain of the small brown planthopper, Laodelphax striatellus. *Proceedings of the Royal Society B: Biological Sciences* 288:20202125.

679