

RESEARCH

Open Access



Comparative genomics and transcriptomics of the *Spiroplasma glossinidia* strain sGff reveal insights into host interaction and trypanosome resistance in *Glossina fuscipes fuscipes*

Daniel J. Bruzzese^{1*}, Fabian Gstötenmayer^{1,2*}, Brian L. Weiss¹, Hager Khalil², Robert L. Mach³, Adly M.M. Abd-Alla² and Serap Aksoy¹

Abstract

Tsetse (*Glossina* spp.) are vectors of African trypanosomes, the causative agents of Human and African Animal trypanosomiasis, diseases that remain significant medical and socioeconomic challenges in sub-Saharan Africa. In addition to trypanosomes, tsetse harbor both obligate and facultative symbiotic bacteria that can influence vector competence and reproductive biology. One such facultative symbiont, *Spiroplasma glossinidia*, infects several tsetse species within the *Palpalis* subgroup. In *Glossina fuscipes fuscipes* (*Gff*), the *Spiroplasma glossinidia* strain sGff induces a trypanosome-refractory phenotype and negatively impacts reproductive fitness by reducing female fecundity. However, the mechanisms behind these *Spiroplasma*-derived phenotypes remain poorly understood. Here, we report successful *in vitro* cultivation of sGff and present complete genomes from three sources: *in vitro* cultured sGff and sGff isolated from both laboratory-maintained and wild-caught (Uganda) *Gff* flies. Comparative genomic analyses revealed a high degree of similarity in gene content and synteny among these sGff samples, confirming that they represent isolates of the same strain. Phylogenomic analyses placed sGff within the *Spiroplasma poulsonii* clade. We found the sGff genome to be highly dynamic, containing numerous mobile genetic elements. Additionally, *in silico* annotations indicated that sGff relies on its host for both lipids and carbohydrates and can produce several toxins, all of which could be implicated in the observed trypanosome refractory phenotype. Finally, comparative transcriptomic analysis of sGff from host hemolymph versus *in vitro* culture provided insights into potential factors relevant to host-symbiont interactions. Our findings provide a foundation for understanding the nutritional dialogue between sGff and its host and identify symbiotic products that may contribute to trypanosome resistance. Furthermore, the establishment of an *in vitro* culture system for sGff represents a significant resource for future functional studies with potential implications for vector control.

*Correspondence:

Daniel J. Bruzzese
daniel.bruzzese@yale.edu
Fabian Gstötenmayer
fabian.gstoettenmayer@yale.edu

Full list of author information is available at the end of the article



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

Keywords Spiroplasma, *Glossina fuscipes fuscipes*, Tsetse, Symbiosis, Spiroplasma glossinidia

Background

Tsetse (*Glossina* spp.) transmit African trypanosomes, the causative agents of Human and African Animal Trypanosomiasis (HAT and AAT, respectively). Approximately 60 million people in sub-Saharan Africa live in tsetse-infested areas at risk for HAT, while AAT constrains livestock productivity across much of the continent [1, 2]. With no vaccines available for either disease, vector-control strategies remain paramount for disease management [2–4]. Complementary approaches that block or reduce trypanosome development in the fly have the potential to enhance disease control efforts [5]. Successful transmission of trypanosomes is influenced by a combination of intrinsic factors, including the fly's innate immunity, host and parasite genotypes, as well as host nutritional status at the time of parasite acquisition [6]. In addition to these intrinsic factors, extrinsic factors, including environmental factors and the composition of fly's microbiota, also modulate pathogen transmission efficiency [5, 7]. In this context, modification of tsetse's heritable symbiotic microbes, some of which coexist in close proximity to trypanosomes in the midgut, provide a promising avenue for “paratransgenic” interventions [8].

Tsetse species harbor a complex community of heritable symbionts, each playing distinct roles in host biology. All tsetse species carry the obligate mutualist *Wigglesworthia*, which supplements the fly's nutrient-restricted blood diet with essential vitamins necessary for reproductive success [7, 9], and proper development of the fly's immune system [10–13]. *Wolbachia* infects tsetse species and can induce cytoplasmic incompatibility, potentially influencing population structure and mating compatibility [14]. *Sodalis* also colonizes tsetse species, with its presence correlated with trypanosome infection prevalence in some contexts, depending on geographic location and tsetse species [15]. Finally, some tsetse house *Spiroplasma glossinidia*, which influences several key tsetse processes, including immune modulation, reproduction, and capacity for trypanosome transmission [16–18]. Despite *Spiroplasma*'s substantial impacts on tsetse biology, the underlying molecular mechanisms governing the bacterium's interactions with tsetse remain largely unknown.

Spiroplasma are helical, wall-less Mollicutes, first characterized as plant pathogens [19, 20], but are now known to colonize a wide range of arthropod hosts, most commonly infecting insects [21–27]. In insects, *Spiroplasma* exhibits a broad range of symbiotic phenotypes, including protective effects in different *Drosophila* species against parasitic wasps and nematodes by either producing ribosome-inactivating proteins (RIPs), which

disrupt parasite protein synthesis machinery [28–30], or by competing with the parasite for macronutrients [31]. Conversely, some *Spiroplasma* strains act as reproductive parasites in insects, such as in *Drosophila* [32], *Anisosticta* (ladybugs) [33], and *Danaus* butterflies [34], where they induce selective male-killing. Of note, *Spiroplasma* genomes evolve rapidly, at rates comparable to RNA viruses [35]. This rapid pace of evolution is attributed to the abundance of mobile genetic elements and absence of key DNA mismatch repair genes, which together promote genomic plasticity, rapid diversification, and horizontal gene transfer of these key symbiosis genes [23, 36].

Within *Glossina*, *S. glossinidia* infections are limited to species in the *Palpalis* subgroup, including *Glossina fuscipes fuscipes* (*Gff*), *Glossina palpalis palpalis*, and *G. tachinoides* [14, 37]. In Uganda, the *S. glossinidia* strain infecting *Gff* (*sGff*) is geographically restricted and polymorphic, with prevalence ranging from 5% to 34% in Northwestern populations, while absent from Central and Southern regions. The infection prevalence in Uganda remains relatively stable across time and space, although seasonality can impact infection dynamics [18]. Interestingly, in the *Gff* line reared from the Insect Pest Control laboratory at the International Atomic Energy Agency (IAEA), *sGff* infection is also not fixed, but is stably maintained at approximately 50% prevalence [38]. Laboratory transmission studies in this *Gff* line indicate that *sGff* is maternally inherited with high fidelity, although paternal transmission can also occur [17]. The bacterium colonizes multiple tissues, such as the gonads, gut, and hemolymph [14, 16]. *Gff* infected with *sGff* show altered gene expression in reproductive and gut tissues [16], reduced hemolymph triacylglyceride (TAG) levels, impaired sperm fitness, and reduced female fecundity [17]. In addition, *sGff* infection is negatively correlated with trypanosome infection prevalence in both field and lab populations, suggesting that presence of the bacterium induces a parasite refractory phenotype in tsetse [18].

Whether *sGff* directly (e.g., via the production of anti-trypanosomal factors) or indirectly (e.g., competition for nutrients) confers the observed parasite resistance phenotype to its tsetse host remains unknown. Transcriptomic analyses of *sGff*-infected *Gff* midguts revealed minimal immune stimulation but indicated elevated oxidative stress and impaired lipid biosynthesis [16]. These changes may create a hostile gut environment for parasites by increasing production of trypanocidal nitric oxide (NO) and/or by reducing the availability of key metabolites required for trypanosome survival.

To investigate the mechanisms by which *Spiroplasma* alters tsetse's physiology and impacts vector competence, we leveraged recently developed *in vitro* methods to culture *sGff*. We then sequenced whole genomes of *sGff* from *in vitro* culture, laboratory-reared, and wild-caught *Gff* individuals using the Oxford Nanopore Technology (ONT) platform. Comparative analysis of the three *sGff* genomes revealed subtle isolate-specific genetic variation residing primarily in mobile genetic elements. We also characterized *sGff* gene expression from both *in vitro* *sGff* and *sGff*-infected hosts, and identified core metabolic pathways and candidate toxins that could mediate the tsetse-*sGff* symbiosis. Collectively, this study lays the foundation for dissecting the molecular basis of the *sGff*-*Gff* interactions and for evaluating its potential in trypanosomiasis control strategies.

Methods

Initiation of *sGff in vitro* culture

Establishment of the *sGff in vitro* culture followed the protocol of Masson et al. [39], developed and optimized for the cultivation of *Spiroplasma poulsonii* sMel_MSRO. This protocol is based on Barbour-Stoenner-Kelly H (BSK-H) media without L-Glutamine (Bio&Sell, Germany), originally designed for *Borrelia burgdorferi*, supplemented with rabbit serum, *Gff* fly extract, lipids, antibiotics, and amino acids. Detailed media and cultivation protocols are indicated in Supplementary Document 1. Briefly, hemolymph from female and male *Gff* from a laboratory line with high *sGff* infection prevalence [38] was collected by removing one prothoracic leg and aspirating the exposed hemolymph droplet using a 10 μ l pipette tip. Cultures were established in three biological replicates using 10 μ l hemolymph and incubated at 25 $^{\circ}$ C for 14 days without agitation under microaerobic atmospheric conditions. Routine microscopy checks were performed by fluorescent microscopy on a Leica DMi8 inverted microscope (Leica Microsystems) following staining with SYTO 9 (0.025 mM; ThermoFisher Scientific). *Spiroplasma* presence was further confirmed by PCR-amplification using *Spiroplasma*-specific 16S rRNA primers (Supplementary Table 1). Cultures underwent passaging every 10–14 days by diluting cultures 1:1 with fresh media.

In vitro culture growth kinetics

Samples from *sGff* cultures for growth curve analysis were collected in triplicate biological replicates from day 0 to day 15 following the inoculation of BSK-H-spiro media with *sGff* precultures. Each day, five μ l of culture was diluted in 200 μ l nuclease-free double-distilled water (ddH₂O) in a sterile PCR tube and stored at -20 $^{\circ}$ C. Prior to quantitative PCR (qPCR), samples were lysed via osmotic heat-shock by incubation at 95 $^{\circ}$ C for 10 min.

qPCR reactions were performed in triplicate technical replicates, in a final volume of 15 μ l consisting of 7.5 μ l iQ[™] SYBR[®] Green Supermix (Bio-Rad, CA, USA), 5.5 μ l PCR-grade water, 0.5 μ l each of the forward and reverse *Spiroplasma* qPCR primers (Supplementary Table 1), and 1 μ l heat-shocked culture. To accurately estimate *Spiroplasma* copy numbers, a standard was prepared by purifying *sGff*-specific qPCR amplicons using the Zymo DNA Clean & Concentrator-25 kit (Zymo Research, USA), followed by DNA quantification with a nanodrop spectrophotometer. The purified DNA was then serially diluted 1:10 to create a range of known concentrations. Cycling conditions for the qPCR comprised of an initial denaturation step at 95 $^{\circ}$ C for 2 min, followed by 40 cycles of 95 $^{\circ}$ C for 5 s, and 56 $^{\circ}$ C for 30 s on a Bio-Rad CFX96 Touch Real-Time PCR Detection System (Hercules, CA, USA). Raw Ct values were extracted in the CFX maestro software and analyzed using R Studio v4.4.2 [40, 41]. Briefly, the mean value of the three technical replicates of each sample was calculated and the copy number of each sample was inferred based on the standard curve of known copy numbers [42]. Calculation of culture doubling time was performed according to the formula $\log(2) / \text{slope}$ [43].

Injections of *in vitro* cultured *sGff* into naive *Gff*

Teneral females and males from the *Gff* line were placed in individual cages and their *Spiroplasma* infection status was determined by PCR-amplification of DNA extracted from one mesothoracic leg using *Spiroplasma* 16S rRNA and tsetse tubulin gene primers (Supplementary Table 1). Negative-*sGff* flies were pooled into two treatment groups (12 females and 6 males per group): (1) high dose and (2) low dose. In addition, 12 females and 6 males screened positive for *sGff* were retained as a control group with a natural infection. *Gff* were anesthetized on ice prior to intrathoracic injection using 29-gauge insulin needles on a micrometer syringe unit injector (Burkard Manufacturing, UK). Injected copy numbers were estimated via quantitative PCR compared to a standard curve of known copy numbers: High Dose: 1.1×10^5 copies in Replicate 1 and 4.3×10^5 copies in Replicate 2; Low Dose: 5.7×10^2 copies in Replicate 1 and 1.2×10^3 copies in Replicate 2. Three hours post-injection, flies were offered their first blood meal and subsequently maintained under standard rearing conditions [44] for 14 days, receiving defibrinated bovine blood three times a week via an artificial membrane feeding system [45]. At designated time points (0, 1, 6, 9 and 14 days), two females and one male were collected, flash-frozen, and stored at -20 $^{\circ}$ C. Genomic DNA from individual whole flies was extracted using the Qiagen DNeasy Blood & Tissue kit (Qiagen, Hilden, GER) according to the manufacturer's protocols and diluted to 4 ng/ μ l. The relative abundance of *sGff*

over time was measured by qPCR, with three technical replicates per sample. Ct values for *sGff* were normalized to those of the tsetse housekeeping gene β -tubulin (Supplementary Table 1) and relative fold-change was calculated using the $2^{-(\Delta Ct)}$ method [46]. Primer efficiencies were determined prior to the qPCR experiment and found to be similar for *sGff* and β -tubulin (91.94% and 90.71%, respectively). Statistical analysis was performed in RStudio v4.4.2 [40, 41]. Significant differences in *sGff* fold-change across timepoints between and within each group were identified using the Kruskal-Wallis test and Dunn's post-hoc test with Benjamini-Hochberg correction for multiple comparisons.

sGff sequencing and assembly

Three *Gff Spiroplasma* (*sGff*) whole genome assemblies were generated: (1) *sGff*-IAEA-CC - derived from the *in vitro* culture of *Spiroplasma* established in 2023 from IAEA *Gff* females and males as described above; (2) *sGff*-IAEA-Fly - assembled from a previously sequenced *Spiroplasma*-infected *Gff* female obtained in 2023 from the IAEA colony (NCBI BioSample: SAMN47211303 and NCBI SRA: SRR32737746), and (3) *sGff*-UG-Tol - assembled from a previously sequenced *Spiroplasma*-infected *Gff* female collected in 2019 from Toloyang village in Atiak Sub County, Amuru District, Uganda (3.304883654, 32.37452634) (NCBI BioSample: SAMN47211302 and NCBI SRA: SRR32737747) [47].

For *sGff*-IAEA-CC genome, high density *in vitro Spiroplasma* culture ($\sim 10^6$ cells) was collected at passage nine, pelleted, and washed with PBS. HMW DNA was extracted using the PureGene kit (Qiagen, Hilden, GER). An ONT LSK14 library was prepared following standard protocols and sequenced on a ONT 10.4.1 minION flow cell for 12 h. Generated POD5 signal data files were basecalled with Dorado v0.9.1 (<https://github.com/nanoporetech/dorado>) using the SUP duplex pipeline. Simplex reads were dropped with Samtools v1.18 [48], and remaining sequencing adapters were trimmed using Dorado trim. The produced FASTQ files were quality filtered with nanoq v0.10.0 [49] to have a minimum quality of q10 and minimum 3000 bp read length. To assemble the *sGff*-IAEA-CC genome, Autocycler v0.2.1 (<https://github.com/rrwick/Autocycler>) was used to generate multiple assemblies from four independent subsampled 100X read sets using five assemblers for each read set (Flye v2.9.5 [50], Canu v2.3 [51], miniasm v.3 [52], NECAT [53], and NextDenovo v2.5.2 [54]). A consensus assembly for these 20 assemblies was then determined with Autocycler v0.2.1 and polished with Medaka v2.0.1 using the bacteria model (<https://github.com/nanoporetech/medaka>). The genome was further polished using short reads made from the same HMW DNA used for ONT sequencing. These reads were sequenced on an Illumina

NovaSeq X system (Eurofins Genomics, Germany) and quality trimmed to Q30 with a minimum length of 50 bp with fastp v0.24.0 [55]. Short read polishing was performed with Polypolish v0.6.0 [56] followed by pypolca v0.3.1 [57].

To assemble the *sGff*-IAEA-Fly genome, we mapped raw Nanopore simplex reads generated with the SUP Dorado v0.9.0 model from the *Gff*-IAEA genome assembly (SRR32737746) to the *sGff*-IAEA-CC assembly using minimap2 v2.28 [58]. Unmapped reads were dropped with Samtools. The remaining simplex reads were pre-corrected with Dorado correct to generate near-Hifi quality reads and were assembled using six assemblers with varying parameters (Hifiasm v0.25 [59], Flye, Canu, miniasm, NECAT, and NextDenovo). A consensus assembly was determined with Autocycler v0.2.1.

Finally, to assemble the *sGff*-UG-Tol genome, we mapped raw Nanopore simplex reads generated with the SUP Dorado v0.9.0 model from the *Gff*-IAEA genome assembly (SRR32737747) to the *sGff*-IAEA-CC assembly using minimap2. Unmapped reads were dropped with Samtools. Due to the shorter read-lengths of this dataset we were unable to utilize Dorado correct to pre-correct ONT reads. Instead, we assembled the mapped ONT Simplex reads using five assemblers (Hifiasm (--ont), Flye, miniasm, NECAT, and NextDenovo). A consensus assembly was determined with Autocycler v0.2.1 and was polished with Medaka v2.0.1 using the bacteria model.

All completed *sGff* circular genomes and plasmids were reoriented to start with the *dnaA* or *repA* with Dnaapler v0.7.0 [60]. Assembly stats for the assembled genomes were established using gfastats v1.3.10 [61] and genome completeness was assessed with BUSCO v5.8.3 [62] using the entomoplasmatales_odb10 dataset. Plasmids were clustered with pling v2 [63] to determine similarity and were named in order according to their size. All genomic data were deposited under the NCBI BioProject: PRJNA1235259.

Functional annotations

Assembled genomes were first annotated using the NCBI Prokaryotic Genome Annotation Pipeline [64]. For the reference *sGff*-IAEA-CC genome, additional functional annotations were added with Bakta [65], eggNOG-mapper [66], BlastKOALA [67], and COGclassifier (<https://github.com/moshi4/COGclassifier>). Mobile genetic elements were annotated using PHASTEST [68] for prophages, ICEberg v3 [69] for integrative conjugative elements, and ISEScan v1.7.3 [70] for insertion sequence elements. Metabolic pathways were characterized with the KEGG terms assigned above with eggNOG-mapper [66] and BlastKOALA [67]. Putative proteins were analyzed using SignalP v6 [71] to determine N-terminal signal peptide sequences, indicative of secretion. Finally,

Spiroplasma specific symbiosis genes identified from the literature were manually searched for with BLASTP [72] (Supplementary Table 2). To better characterize the RIP identified in *sGff*, we generated a phylogeny using *Spiroplasma* members of the RIP superfamily from InterPro [73] and our putative RIP sequence (Supplementary Table 3). PRANK [74] aligned amino acid sequences were trimmed to only contain RIP domains and an amino acid phylogeny was constructed with RAxML-NG [75] using a LG + G8 + F model, with 25 parsimony and 25 random starting trees with 1000 bootstraps.

Spiroplasma phylogenomics

To determine *sGff*'s relationship to other *Spiroplasma* strains, we constructed phylogenetic trees at two scales: (1) using *Spiroplasma* strains from several clades, including *Spiroplasma citri*, *Spiroplasma mirum*, *Spiroplasma chrysopicola*, and *Spiroplasma poulsonii*, and (2) using only members from the *S. poulsonii* clade. Reference genomes and their NCBI RefSeq (or if missing, NCBI GenBank) annotations in GBFF format were downloaded using NCBI datasets [76]. These genomes were clustered using RabbitTClust [77], dropping *Spiroplasma* strains that clustered apart from the *S. mirum*, *S. chrysopicola*, *S. citri*, and *S. poulsonii* clades. For the remaining *Spiroplasma* strains (Supplementary Table 4), single-copy orthologous core genes present in >95% of taxa were identified with PPanGGOLiN [78]. From this gene set, DNA sequences were extracted, aligned with MAFFT [79], and concatenated for each *Spiroplasma* taxa. Phylogenetic trees were generated from the concatenated multiple sequence alignment using RAxML-NG [75] with the GTR + G substitution model with 50 random and 50 parsimony-based starting trees, and with 1,000 bootstrap replicates. The resulting draft RAxML tree was used to identify duplicate *Spiroplasma* strains, which were removed before re-running the PPanGGOLiN and RAxML pipeline. Finally, the whole process was repeated at a smaller scale for members of the *Spiroplasma poulsonii* clade only.

Genome comparisons

Pairwise average nucleotide identity was calculated between the three assembled *sGff* genomes and the closest sister taxa, *Spiroplasma. sp.* TU-14 (GCF_001792795.1), using pyani v0.2.13.1 [80]. Whole-genome synteny, SNP, and indel comparisons between *sGff*-IAEA-CC, *sGff*-IAEA-Fly, and *sGff*-UG-Tol were performed with pairwise alignments from NUCmer [81] that were annotated with SyRI [82], and visualized using plotsr v1.1 [83]. For the *sGff*-IAEA-CC, *sGff*-IAEA-Fly, *sGff*-UG-Tol, and sister taxa *S. sp.* TU-14 (GCF_001792795.1), we identified shared and unique gene families by parsing the PPanGGOLiN dataset

generated above for the *Spiroplasma poulsonii* RAxML tree using R [40] with the tidyverse [84] and VennDiagram [85] packages. Unique gene family amino acid sequences were functionally annotated with Bakta [86].

RNA sequencing and analyses

To assess *sGff* transcriptomes, five *sGff* RNA samples were prepared for RNAseq: three from *sGff* cell-culture and two from *Gff* hemolymph. For the *sGff* cell-culture samples, *sGff*-IAEA-CC (~ 10⁶ cells) at passage 12 were pelleted at 12,000 g for 15 min at 4 °C and resuspended in RNeasy lysis buffer (Qiagen, MD, USA) and stored at -80 °C. For the *Gff* hemolymph samples, female IAEA *Gff* flies were screened for *sGff* infections using the above PCR protocol. Positive flies were surface-sterilized in 70% ethanol and hemolymph was collected as described above. Hemolymph was pooled into two samples from 70 to 35 *sGff*-positive females and stored in RNeasy lysis buffer (Qiagen, MD, USA) at -80 °C. Total RNA was extracted from samples by first diluting the samples stored in RNeasy lysis buffer 1:1 with PBS, pelleting at 5000 g for 15 min at 4 °C, and resuspending the pellets in 50 µl PBS. Total RNA was extracted from the PBS washed samples using the Total RNA Miniprep kit (NEB, MA, USA) following standard protocols. Ribodepleted RNA-seq libraries were prepared by the Yale Center for Genomic Analysis and sequenced on the NovaSeq 6000 platform. Raw reads were quality filtered (Phred score >20 and read length >50 bp) and adapter-trimmed using fastp v0.24.0 [55]. Trimming and filtering results were summarized with MultiQC v1.27 [87].

We first used the RNA sequencing data to validate the *in silico* *sGff*-IAEA-CC annotations. We pooled all transcriptomes into a single sample and quantified transcript abundance using Salmon [88] with the *sGff*-IAEA-CC genome as a reference. Transcripts were considered expressed if their transcripts per million (TPM) count was >5. Operons were identified with the OpDetect v1.0 [89] pipeline, which utilized STAR v2.7.11b [90] alignments of the transcriptome to the *sGff*-IAEA-CC genome. We then quantified differential gene expression for the three culture and two *Gff* hemolymph samples, again utilizing Salmon [88] to quantify transcript abundance to the *sGff*-IAEA-CC reference genome. Differential transcript analysis was performed with edgeR [91] using the quasi-likelihood model. To identify the most prominent upregulated and downregulated genes, we utilized conservative parameters to identify significantly differentially expressed genes: a false discovery rate (FDR) <= 0.05 and log₂-fold-change (log₂FC) >= 1.5 or <= -1.5.

Results

Establishment of *sGff* *in vitro* cultures

Following the published protocol for *Spiroplasma poulsonii* sMel_MSRO, we successfully established an *in vitro* culture of the *S. glossinidia* strain *sGff* isolated from the hemolymph of *Gff*. Microscopic examination of SYTO 9-stained cells revealed a homogenous culture, with cells exhibiting the characteristic helical morphology of *Spiroplasma* (Fig. 1 A and B). Growth curve analysis of cultures inoculated with frozen preculture aliquots identified two distinct phases of exponential growth, occurring from days 0 to 4 and days 7 to 11, separated by a plateau phase (Fig. 1 C), which was not observed in subsequent subculturing. The estimated doubling time was 40 h during the first exponential phase and 85 h during the second. Fluorescence microscopy revealed that cultures reached high cell densities within 14 days post inoculation. Remarkably, the *sGff* cultures have remained viable for over 12 months with routine subculturing every 10–14 days, demonstrating the robustness and long-term stability of the *in vitro* system.

sGff cultures can establish infections in naïve *Gff* hosts

To assess whether *in vitro*-cultured *sGff* retains its ability to replicate *in vivo* following prolonged culture, we conducted microinjection experiments using two groups of naïve *Gff* that were confirmed negative for *sGff* by PCR screening of a mesothoracic leg. Teneral *sGff*-negative flies were injected with either a high- or low-dose of cultured *sGff* and infection dynamics were monitored over a 14-day period. The results demonstrate that *in vitro*-cultured *sGff* retains its replicative capacity within naïve *Gff*, validating its potential utility for downstream functional studies. Quantification of *sGff* in transinfected

Gff, alongside naturally infected *Gff* controls, revealed differences in initial *sGff* titers between groups on day 0, with the low-dose group showing the lowest titer (fold-change=0.014), and high-dose and natural infection groups exhibiting similar titers (fold-change=2.3 and 2.451, respectively) (Supplementary Table 5). We observed a dose-dependent increase within groups over time (Supplementary Fig. 1). Statistical analysis confirmed significant differences in *sGff* titers between treatment groups on days 0, 1 and 6 (Kruskal-Wallis: p -values = 0.003, 0.003 and 0.0007, respectively) (Supplementary Table 6) and a significant increase within both high- and low-dose groups over time (Kruskal-Wallis: p -value=0.000044 for low-dose; p -value=0.00045 for high-dose) (Supplementary Table 7), with post-hoc tests indicating that titers on days 9 and 14 were significantly higher compared to days 0 and 1 (Supplementary Tables 8 and 9). In contrast, the increase of *sGff* titers in the naturally infected control group was not statistically significant, likely due to higher variability among individuals (Supplementary Table 5).

sGff complete genome assembly

To generate a high-quality reference genome for *sGff*, we sequenced the new *sGff* culture and obtained 78,639 Nanopore duplex reads (565.8 Mb), over 200x coverage for the primary assembly. Additionally, we obtained 4,166,901 Illumina short reads (1.2 Gb) for polishing. The nanopore reads were split into four subsets to construct 20 independent *sGff* assemblies that were merged into a consensus sequence using Autocycler and subsequently polished with the Illumina data. This approach produced a closed, circular 1.489 Mb reference genome from the *in vitro* cultured *sGff* strain, designated

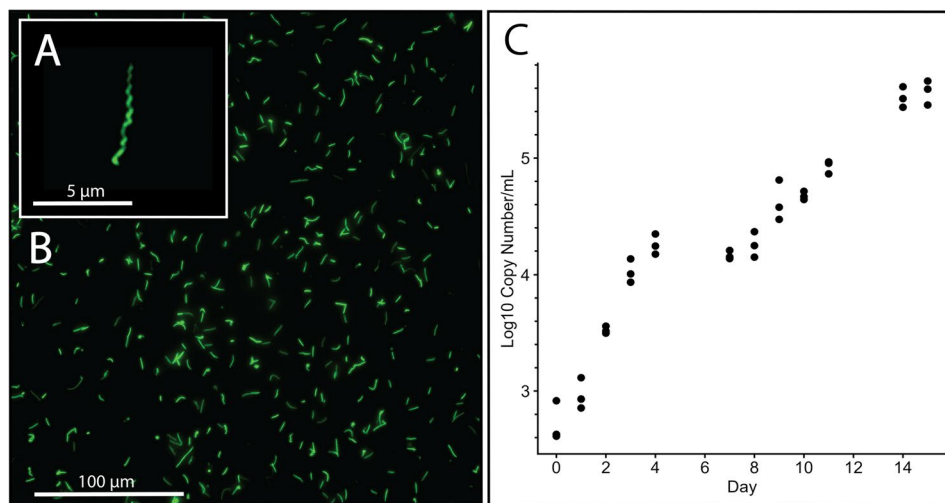


Fig. 1 *In vitro* cultivation of *sGff*. **A**) Fluorescent microscopy of *in vitro* *sGff* cells at passage nine stained with SYTO 9 and imaged using a Leica DMI8 inverted microscope (FITC channel). Scale bar: 5 μ m. **B**) High density *sGff* culture at passage nine. Scale bar: 100 μ m. **C**) Growth kinetics of *sGff* in supplemented BSK-H media at 25°C under microaerobic conditions, assessed by qPCR with threefold technical and biological replication.

sGff-IAEA-CC, which was deposited in NCBI (RefSeq: GCF_049669535.1, BioProject: PRJNA1235259). We also assembled two additional sGff genomes using publicly available NCBI SRA datasets from *Gff* flies (Bioproject: PRJNA1231403): one from a female IAEA colony fly (sGff-IAEA-Fly) and another from a wild-caught female from northwestern Uganda (sGff-UG-Tol).

For all assemblies, BUSCO scores exceeded 97.3% based on the OrthoDB entomoplasmatales_odb10 dataset, indicating high quality, complete genomes (Table 1). Genome size varied slightly among the assemblies. The sGff-IAEA-CC reference genome was 1.489 Mb and included four plasmids ranging in size from 6 kb to 16 kb, while the sGff-IAEA-Fly genome had a slightly smaller chromosome of 1.469 Mb (approximately 20 kb less than the reference sGff-IAEA-CC genome), but retained all four plasmids. Interestingly, the sGff-UG-Tol genome was more reduced, with a 1.418 Mb chromosome, and only three plasmids, missing the 13 kb plasmid, p_sGff2 (Table 1). We manually verified the absence of the plasmid p_sGff2 in sGff-UG-Tol by mapping its reads to the sGff-IAEA-CC assembly, confirming no coverage over the sGff-IAEA-CC p_sGff2 region. Gene counts across assemblies showed moderate variation. The sGff-IAEA-CC genome, including plasmids, contained 1,829 genes, including 1,687 protein-coding genes, 104 pseudogenes,

32 tRNAs, 3 rRNAs, and 3 non-coding RNAs (ncRNAs) (Table 1). In comparison, sGff-IAEA-Fly encoded 1,795 genes (1,653 protein-coding genes), while sGff-UG-Tol encoded 1,695 genes (1,553 protein-coding genes) (Table 1).

sGff clusters within the *Spiroplasma poulsonii* clade

To place the newly assembled sGff genomes within a broader phylogenetic framework, we constructed a RAxML maximum likelihood tree based on 93–101 single-copy core orthologous genes conserved in over 95% of representative *Spiroplasma* taxa from the *S. citri*, *S. mirum*, *S. chrysopicola*, and *S. poulsonii* clades (Fig. 2). The sGff isolates are grouped most closely with two poorly characterized but genetically identical *Spiroplasma* strains: *S. sp.* TU-14 and *S. sp.* NBRC_100390. The *S. sp.* TU-14 strain was originally isolated from a contaminated sample of another Mollicute (*Entomoplasma lucivorax* PIPN-2) and has an unknown host origin [92]. The *S. sp.* NBRC_100390 strain was initially misclassified as *S. atrichopogonis* GNAT3597, a symbiont of biting midges, but has since been recognized as a distinct *Spiroplasma* taxon [93]. Together, the sGff isolates, *S. sp.* TU-14, and *S. sp.* NBRC_100390 form a distinct, well-supported lineage within the *S. poulsonii* clade (Fig. 2). To further resolve their evolutionary relationships, we constructed a higher-resolution RAxML phylogeny using 509–563 single-copy core orthologs shared among >95% of representatives of *Spiroplasma* taxa within the *Poulsonii* clade. This second RAxML tree reinforced the same relationships observed in the previous tree. It also resolved the sGff polytomy, showing that the sGff-IAEA-CC and sGff-IAEA-Fly isolates were genetically identical and formed a monophyletic group, with sGff-UG-Tol positioned as their sister lineage (Supplementary Fig. 2).

sGff genomes are highly similar

Pairwise genome comparisons among the sGff assemblies supported the relationships inferred from the RAxML phylogenies. Average Nucleotide Identity (ANI) was 99.9% between the sGff-IAEA-CC and sGff-IAEA-Fly genomes, and slightly reduced at 99.8% between the sGff-IAEA-CC and sGff-UG-Tol (Fig. 3A). These ANI values are a magnitude higher than those observed between any sGff isolate and its closest *Spiroplasma* outgroup, strain *S. sp.* TU-14, with pairwise ANI values of 98.5% to sGff-IAEA-CC, sGff-IAEA-Fly, and sGff-UG-Tol (Fig. 3A).

Gene content analysis revealed subtle differences between the nearly identical sGff-IAEA isolates and the more divergent sGff-UG-Tol isolate. Specifically, 38 gene families were unique to the sGff-IAEA isolates, while 17 were unique to the sGff-UG-Tol (Fig. 3B). Among the 38 sGff-IAEA-specific gene families, 20 were associated with mobile genetic elements (MGEs) – including twelve from

Table 1 Assembly statistics for the three sGff genomes. BUSCO scores are for the entomoplasmatales_odb10 orthodb dataset.

	sGff-IAEA-CC	sGff-IAEA-Fly	sGff-UG-Tol
Chromosome (bp)	1,489,281	1,469,171	1,418,152
p_sGff1 (bp)	16,389	16,389	16,375
p_sGff2 (bp)	13,621	13,621	N/A
p_sGff3 (bp)	13,208	13,208	12,717
p_sGff4 (bp)	6,466	6,466	6,460
GC content	27.16%	27.21%	27.29%
BioProject	PRJNA1235259	PRJNA1235259	PRJNA1235259
BioSample	SAMN47572768	SAMN48057343	SAMN48057344
Assembly	GCF_049669535.1	GCA_049949145.1	GCA_049949155.1
BUSCO score	C:97.6% [S:97.6%,D:0.0%]	C:97.6% [S:97.6%,D:0.0%]	C:97.3% [S:97.3%,D:0.0%]
genes	1,829	1,795	1,695
protein-coding genes	1,687	1,653	1,553
pseudo-genes	104	104	104
tRNA	32	32	32
rRNA	3	3	3
ncRNA	3	3	3

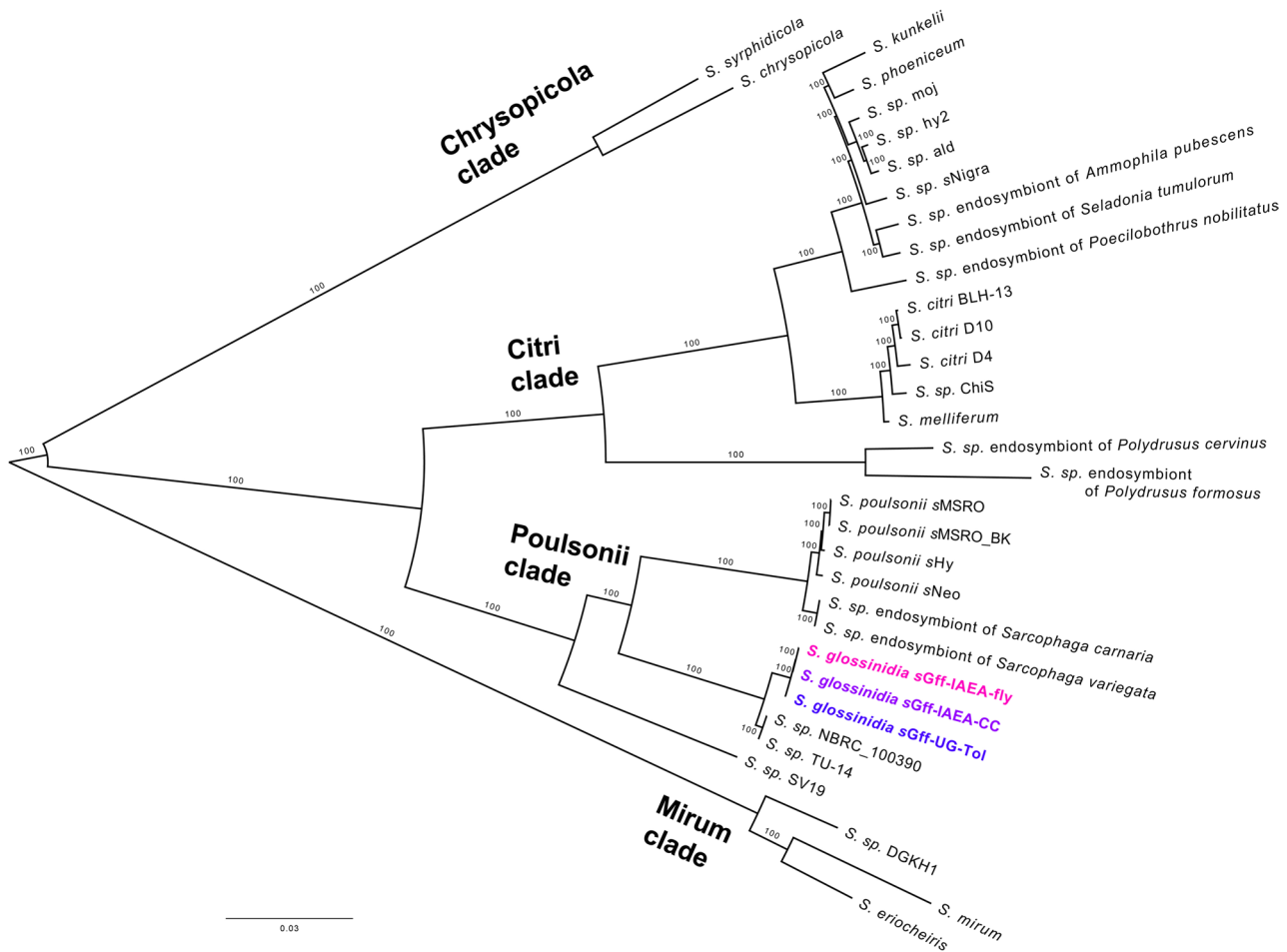


Fig. 2 Phylogenetic placement of *sGff* among representative *Spiroplasma* genomes. Maximum likelihood (RAxML) tree showing the relationship between *sGff* and representative *Spiroplasma* species from the *S. citri*, *S. mirum*, *S. chrysopicola*, and *S. poulsonii* clades (see Supplementary Table 4 for accession details). Bootstrap support values are indicated on branches and the tree is midpoint rooted.

plasmids with six from *p_sGff2* that is absent from *sGff-UG-Tol*, nine were related to metabolism and symbiosis, and the remaining nine encoded hypothetical proteins (Supplementary Table 10). The 17 *sGff-UG-Tol*-specific gene families included four genes located on plasmid *p_sGff3*, three associated with MGEs, two symbiosis-related genes, and eight hypothetical proteins (Supplementary Table 11). When compared to the nearest sister taxon, *S. sp. TU-14*, we identified 310 unique gene families shared among the *sGff* isolates, consisting mostly of MGEs, including *Spiroplasma* prophages, plasmids, and integrative conjugative elements (ICEs) (Supplementary Table 12).

Whole genome alignments further revealed that the two IAEA-derived *sGff* genomes, *sGff-IAEA-CC* and *sGff-IAEA-Fly*, are nearly identical, differing by only three SNPs, 49 insertions/deletions (indels), and one duplication (Fig. 3C and Supplementary Table 13). The three SNPs were located in genes encoding a prophage, the DNA-binding protein *WhiA*, and the fructoselysine

transporter *frlA* (Supplementary Table 14). With no fructose supplemented in the culture medium, the non-synonymous SNP converting threonine to arginine in *frlA* could be the result of relaxed selection. Most indels were located within homopolymeric tracts, suggesting potential assembly artifacts. However, we identified a notable 20 kb duplication involving a *Spiroplasma* prophage region in the *sGff-IAEA-CC* genome, but absent compared to *sGff-IAEA-Fly* (Fig. 3C and Supplementary Table 15). This duplication likely represents a prophage polymorphism in the *sGff-IAEA* isolate that became fixed during the *in vitro* culture process. In contrast, the *sGff-UG-Tol* genome exhibited more extensive divergence, with 550 SNPs, 171 indels, five large duplications, and one translocation relative to *sGff-IAEA-CC* (Fig. 3C and Supplementary Table 16). Of the 550 SNPs, 383 were found within protein coding genes with over 44% associated with MGEs (Supplementary Table 17). While most of the indels were homopolymeric, at least 22 appeared to be genuine sequence differences rather

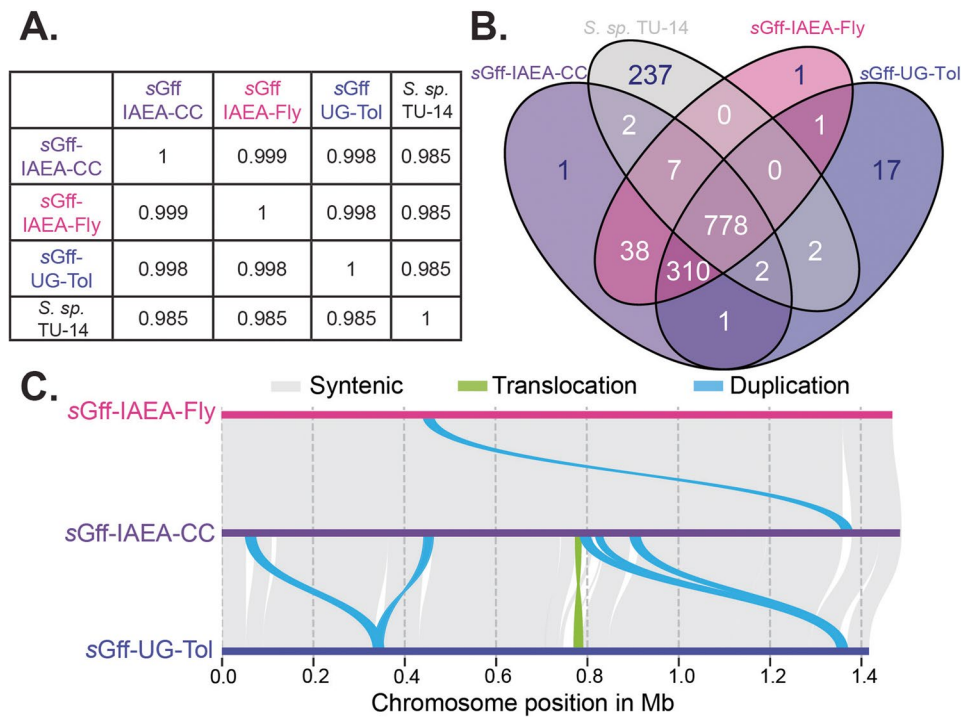


Fig. 3 Genome comparisons between sGff isolates and *S. sp. TU-14*. **A**) Average Nucleotide Identity (ANI) values between sGff isolates and *S. sp. TU-14*, indicating high sequence similarity. **B**) Venn diagram showing shared and unique gene families for sGff isolates and *S. sp. TU-14*. **C**) Genome synteny plot illustrating conserved gene order among sGff isolates, highlighting structural similarities and isolate-specific variations.

than homopolymeric assembly artifacts. Large-scale structural variation between the sGff-IAEA-CC and the sGff-UG-Tol genomes included five duplications and one translocation (Fig. 3C). All these rearrangements were associated with *Spiroplasma* prophages (Supplementary Table 18).

sGff mobile genetic elements

The sGff-IAEA-CC chromosome encodes a total of 1,778 genes, including 1,637 protein-coding genes, of which 800 are predicted to encode hypothetical proteins with unknown functions. In addition, the genome contains 103 pseudogenes, 32 tRNAs, 3 rRNAs, and 3 non-coding RNAs (ncRNAs) (Table 1, Supplementary Table 19). Genes associated with MGEs represented the most prevalent Cluster of Orthologous Groups (COG) functional category, accounting for over 28% of the annotated functions (Fig. 4 and Supplementary Fig. 3). When the 646 proteins with unknown COG annotations (36.33% of the genome) are excluded, the relative proportion of MGE-associated functions increases to 44.35%, highlighting the central role of MGEs in shaping the sGff genome. Other prominent COG functional categories include translation, ribosomal structure and biogenesis, replication, recombination and repair, and carbohydrate transport and metabolism (Supplementary Fig. 3).

The most common MGEs are *Spiroplasma* prophages, specifically from two families:

Plectroviridae (*Plectovirus* prophages) and Microviridae (*Spiromicrovirus* prophages). Using Phastest, we identified 20 supported prophage regions, along with several other putative prophage-related genes scattered across the genome, including within plasmids (Fig. 4). Insertion sequence (IS) elements were also highly prevalent, with 129 identified on the chromosome and many associated with prophage regions. In addition, a single well-supported ICE was identified that encoded a complete Type I restriction-modification system (Fig. 4). Like the prophages, ICE-associated genes were dispersed throughout the genome, including some within prophage regions and plasmids.

sGff metabolism

We reconstructed the metabolic pathways of sGff using PGAP-predicted proteins using BLASTKoala and eggNOG-mapper. This analysis revealed that sGff possesses highly reduced biosynthetic capability and relies primarily on its *Gff* host for energy and essential metabolites (Supplementary Table 19). The bacterium’s energy metabolism centers on glycolysis, with glucose and fructose serving as its main carbohydrate substrates. We identified four phosphotransferase system (PTS) transporters predicted to facilitate uptake of specific sugars: fructose, glucose, 2-(alpha-D-mannosyl)-D-glycerate, and cellobiose/diacetylchitobiose. Notably, the presence of a diacetylchitobiose transporter, together with putative

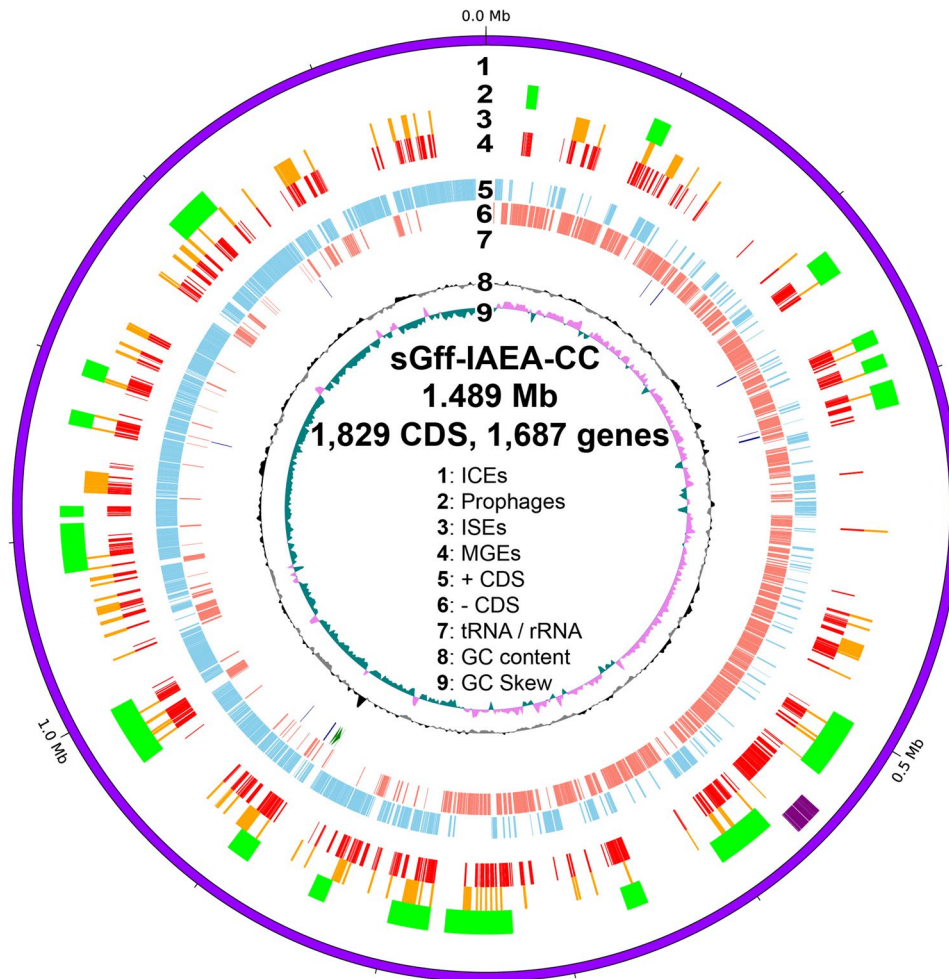


Fig. 4 Circular genome plot of sGff-IAEA-CC showing the distribution of Mobile Genetic Elements (MGE). The genome is represented in nine concentric rings (1–9), each corresponding to a feature labeled in the center of the plot. The figure highlights the high proportion and broad distribution of MGE-associated genes across the chromosome.

chitinases, suggests that sGff may have the capacity to utilize chitin as an alternative energy source. The genome also reveals metabolic versatility through two additional pathways: a functional acetyltransferase-acetate kinase pathway indicating potential acetogenic metabolism, as well as gene-clusters involved with sulfur reduction.

Like other *Spiroplasma* species, sGff exhibits minimal capacity for *de novo* lipid synthesis, possessing functionality limited to the citric acid cycle and fatty acid elongation. Instead, it predominantly incorporates host-derived fatty acids and cholesterol directly into its cell membranes [94]. Nevertheless, sGff retains crucial lipid-modifying capabilities, particularly the complete cardiolipin biosynthesis pathway that converts host-derived diacylglycerols (DAGs) into cardiolipin, a process identical to the experimentally verified pathway in *S. poulsonii* [95] that is essential for bacterial membrane formation [96]. The genome also encodes the non-mevalonate pathway

for terpenoid backbone biosynthesis, providing additional lipid processing capability.

Beyond lipid modification, sGff possesses several other significant biosynthetic capabilities, including a complete folate biosynthesis pathway, supporting one-carbon metabolism essential for nucleotide and amino acid synthesis. Hemolysin-related proteins and ferritin are also produced by sGff potentially facilitating iron sequestration in the host environment (Supplementary Table 19). We also note that sGff exhibits limited amino acid biosynthesis as it is predicted to produce three amino acids *de novo*: aspartic acid, serine, and asparagine, indicating a high degree of dependence on host-derived nutrients.

Symbiosis genes

In addition to its highly specialized metabolism, sGff has several mechanisms that may facilitate its persistence within the tsetse host. Among these are lipoproteins, which play essential roles in symbiosis through

interactions with the host immune system, host cells, and metabolites present in the hemolymph. Of the 1,637 predicted protein-coding genes in the *sGff* genome, 72 encode lipoproteins (Supplementary Table 19). Many of these lipoproteins are associated with *Spiroplasma* prophage regions, suggesting that horizontal gene transfer and the utilization of prophage-derived lipoproteins may be key to *sGff*'s adaptation to the host environment. The most abundant *Spiroplasma* lipoprotein is *Spiralin*, which is a well-characterized protein important in host interactions and vertical transmission [97]. We identified one conserved *Spiralin* gene (WP_424526001.1) as well as two other *Spiralin*-derived copies containing one and two repeats, respectively (WP_424525940.1 and WP_424525799.1). We note six adhesion-related genes: (WP_424527554.1, WP_424527500.1, WP_424527571.1, WP_424527528.1, WP_424527276.1, and WP_424526807.1) four of which are located on plasmids (one per plasmid) and are likely involved in host cell adhesion and invasion. Similar adhesion genes are essential for midgut cell invasion in bees [98] and adhesion to leafhopper cells [99].

We identified one putative ribosome-inactivating protein (RIP) gene in the *sGff* genome (Supplementary Table 19). The gene (WP_424526995.1) encodes a protein with a conserved RIP domain, most similar to *sGff*'s sister taxa *S. sp.* NBRC_100390 (Supplementary Fig. 4). However, the *sGff* RIP lacks a signal peptide, suggesting it is not secreted. Instead, this protein is predicted to be embedded in the outer cell membrane, with the RIP domain exposed to the intracellular environment. Whether this toxin is released via proteolysis by an unidentified peptidase remains unknown. We also identified a more divergent RIP-like protein (WP_424526987.1) that includes a signal peptide, suggesting it may be secreted into the host environment.

We identified two overlapping copies of the *glycerol-3-phosphatase* (*glpO*) gene in the *sGff* genome (WP_424526233.1 and WP_424526234.1) (Supplementary Table 19). *glpO* catalyzes the oxidation of glycerol-3-phosphate, generating reactive oxygen species (ROS) as a byproduct of glycerol metabolism. In *Mycoplasma*, the gene functions as a virulence factor [100] and in *Spiroplasma* it may provide protection against parasitoid wasps [36]. Typically, *glpO* is part of a conserved operon flanked by *glpF* and *glpK*, which encode a glycerol transporter and glycerol kinase, respectively. However, in *sGff*, *glpF* appears truncated due to the insertion of a *Spiroplasma* prophage, with only the first 32 amino acid residues present upstream of the insertion site. Interestingly, in silico analyses suggest that the truncated *glpF* and the two *glpO* copies remain potentially functional.

Transcriptome validation

We performed RNAseq-based transcriptomic analysis utilizing three *in vitro* samples derived from *sGff* culture and two *in vivo* *sGff* samples isolated from *Gff* hemolymph. The transcriptomic dataset was first used to validate in silico annotations of the *sGff* genome. We pooled transcripts from both culture and hemolymph samples, resulting in over 57 million reads mapped to the reference genome. Of the 1,778 annotated genes, 464 showed no detectable expression (TPM < 5) in either condition (Supplementary Table 20). Over 261 of these non-expressed genes were related to MGEs, primarily *Spiroplasma* prophages and suggests that prophage expression may be actively suppressed in *sGff*. Among the 464 non-expressed genes, 52 were annotated as pseudogenes via in silico predictions. Interestingly, 52 of the 104 predicted pseudogenes exhibited transcriptional activity, with 34 annotated as MGEs, including transposases, phage-related proteins, and ICEs, indicating that some may retain regulatory or functional roles, particularly in genome plasticity and host adaptation. Additionally, we confirmed expression for all candidate symbiosis-associated genes identified in our genome analyses (Supplementary Table 20). Transcripts were detected from all four plasmids, indicating that plasmid-encoded genes are expressed under both *in vivo* and *in vitro* conditions.

We identified the most abundant transcripts from the pooled dataset. Among the top 20 most abundant genes, seven were associated with plasmid functions, three were involved in fructose metabolism and five were involved in core transcription and translation processes. The remaining highly expressed genes included two with unknown functions, a malate permease, a rod shape determining protein, and *Spiralin* (Supplementary Table 20). In addition, we used transcriptomic data to predict operon structures in the *sGff* genome, as operons may encode co-regulated clusters, including putative secreted effectors relevant to host interactions or trypanocidal effects. Of the 1,675 protein-coding genes analyzed, 1,471 (87.8%) were predicted to be organized into 297 operons. The mean operon length was 3.95 genes, with the largest operon containing 39 genes (Supplementary Table 21). Operon 64 was particularly interesting as it encoded the conserved RIP, MATE family efflux transporters, and Peroxiredoxin, which is a strong antioxidant.

Transcriptomic differences between *in vivo* and *in vitro* *sGff*

We next investigated *in vivo* vs. *in vitro* *sGff* gene expression by comparing transcriptomes of *sGff* isolated from hemolymph and from *in vitro* culture. In total, we identified 41 differentially expressed genes between the two sample types, using a \log_2 FC threshold of ≥ 1.5 or ≤ -1.5 and a false discovery rate (FDR) ≤ 0.05 (Fig. 5A). Of these, 12 genes were upregulated in *sGff* from



Fig. 5 Differential gene expression between culture- and hemolymph-derived *sGff*. **A** MA plot depicting differentially expressed genes (DEGs) and their expression levels measured in log counts per million (logCPM), with positive log₂FC values representing genes up-regulated in hemolymph-derived *sGff* relative to culture-derived *sGff*, and negative log₂FC values representing down-regulation. Genes were considered differentially expressed if they had a log₂FC >= 1.5 or <=-1.5 with FDR <= 0.05 (Supplementary Table 22). DEGs are numbered and correspond to those in panel B. The dotted horizontal lines represent 1.5 and -1.5 log₂FC values. **B** Functional annotations and genomic locations (chromosome or plasmid) for the DEGs shown in panel A.

hemolymph relative to the culture, while 29 genes were downregulated (Fig. 5B and Supplementary Table 22). Among the 12 upregulated genes in hemolymph relative to culture, 11 were located on the main chromosome, and one was plasmid-encoded. The upregulated main chromosomal genes contained five metabolism-related genes, including a fructose-specific phosphotransferase system (PTS) transporter, two translation-related genes, a detoxifying-related gene, a repair-related gene, a stress-related gene, and a pseudogene. Of the 29 downregulated genes in *sGff* hemolymph compared to culture, 17 were located on the main chromosome and 12 were plasmid-encoded. The genes on the chromosome included five hypothetical proteins, six metabolism-related genes, including a glucose-specific PTS transporter, two prophage-related genes, two lipoproteins, a DNA polymerase, and a serine protease. These differential expression patterns likely reflect environmental differences between the *in vivo* (hemolymph) and *in vitro* (culture) conditions. The upregulation of a fructose-specific PTS transporter and downregulation of a glucose-specific PTS transporter may indicate a shift in available carbohydrate sources

between the hemolymph and the culture medium, as fructose is absent from the culture media (Supplementary Document 1). Reduced expression of prophage genes could also suggest *sGff* can suppress prophage expression in certain environments.

Discussion

This study advances our understanding of the interactions between *Glossina fuscipes fuscipes* (*Gff*) and its symbiont *Spiroplasma glossinidia sGff* through the establishment of an *in vitro* culture system, whole genome sequencing of multiple *sGff* isolates, and transcriptomic comparisons between *sGff* derived from culture and from *Gff* hemolymph.

While *Spiroplasma citri* was first cultured in 1971 [101] and some *Spiroplasma* strains were propagated using both cell-based and cell-free culture systems [102, 103], optimizing growth conditions for many strains remained challenging, likely due to the bacterium’s specialized adaptations to its specific insect host environment [104]. More recently, the use of Barbour-Stoener-Kelly (BSK-H) media, supplemented with specific nutrients, has

facilitated the sustained *in vitro* growth of the previously unculturable *S. poulsonii* sMel_MSRO strain [39], and now *S. glossinidia* sGff. The initial growth pattern of cultured sGff was characterized by two exponential phases separated by a plateau phase, which was not observed in subsequent passages. Culture plateaus at specific density thresholds were also observed in sMel_MSRO [105], however, the sGff culture resumed growth after the plateau phase, possibly reflecting adaptation to the culture environment or modulation of endogenous prophages, which can impact growth dynamics in related strains [106, 107]. The growth rate of sGff was slower than sMel_MSRO, with a doubling time of 40 h compared to 30 h, which may reflect strain-specific metabolic traits [39]. Further optimization of the sGff culture media, informed by insights from its genome-derived metabolic profile [104], may improve these growth rates. Importantly, sGff retained capacity for infection after prolonged *in vitro* cultivation, as demonstrated by successful colonization of naive Gff following experimental infection, with higher inoculum doses resulting in higher sGff titers. Of note, the fold-change of sGff-growth in naive injected Gff was lower than that observed in culture, which is expected due to the limited nutrient availability within the host compared to nutrient-rich culture media. It will be important to establish whether sGff infections can recapitulate the same tissue tropism and transmission dynamics of natural infections. Overall, the availability of a reliable culture system provides a powerful tool for dissecting the molecular dialogue between sGff and its tsetse host, including the potential for genetic manipulation of the symbiont.

Using a hybrid sequencing approach that combined Nanopore and Illumina reads, we assembled a high-quality, closed genome from the sGff culture consisting of a 1.489 Mb chromosome along with four plasmids ranging in size from 6 to 16 kb. This reference genome (sGff-IAEA-CC) was compared to two additional sGff assemblies – one from a colony fly (sGff-IAEA-Fly) and another from a fly collected in Northwest Uganda (sGff-UG-Tol). Comparative analysis confirmed that all three assemblies represent closely related isolates of the same strain, differing by several SNPs, indels, and unique gene families, with very little genomic variation between the sGff-IAEA-CC and sGff-IAEA-Fly genomes. The few genomic differences observed between the two IAEA-fly derived sGff genomes could be attributed to either natural polymorphisms that became fixed during cultivation or perhaps from an elevated mutation rate due to the absence of genes in the mismatch repair pathway, namely *MutS*, *MutL*, and *MutH*, which is a trait shared across the *Spiroplasma* genus [35]. In contrast, more pronounced genomic variation was observed between the IAEA-derived genomes and field-derived sGff-UG-Tol isolate.

This divergence is consistent with the moderate genetic differentiation between their respective Gff hosts, as sGff-UG-Tol isolate was obtained from a Northwestern Ugandan population, while the IAEA Gff colony originated from flies collected in the Central African Republic in 1986 [47, 108]. Nearly all the genomic variation between these isolates was localized to MGEs, including genes associated with *Spiroplasma* prophages. Notably, one plasmid present in the IAEA isolates was absent in the Uganda field isolate. Studies characterizing the range of genetic variation found in sGff across the landscape will be important for understanding the full extent of genetic diversity in this symbiont.

Phylogenetic analysis placed the sGff isolates within the *S. poulsonii* clade, a group that contains several other well-characterized *Spiroplasma* strains known for protecting their hosts against nematodes, parasitoids, and viruses [109–111]. Among these sister strains is *S. poulsonii* sMel_MSRO, whose culture protocol we successfully adopted for sGff and likely explains the ease with which sGff was cultivated. The closest known relatives of sGff are two genetically identical but poorly characterized strains: *S. sp.* TU-14 and *S. sp.* NBRC_100390. These strains share 98.5% genome-wide identity with sGff and have 778 gene families in common. However, sGff possesses over 310 unique gene families, consisting primarily of hypothetical proteins and MGEs. These shared and unique gene families highlight both its close relationship to its sister taxa as well as its substantial repertoire of distinctive genetic features.

The sGff genome consists of a remarkably high proportion of MGEs, including prophages, integrative conjugative elements (ICEs), and insertion sequence elements (ISEs), which together comprise over 28% of the genome. When hypothetical proteins are excluded, MGE-associated content increases to more than 44% of the annotated genome. This high MGE load is consistent with observations in several other *S. poulsonii* and *S. citri* strains [35, 112]. In total, we identified 20 distinct prophage regions, consisting largely of *Spiromicrovirus* and *Plectrovirus* prophages, along with one ICE, and 129 ISEs. We also identified additional prophage- and ICE-related genes dispersed throughout the genome and plasmids – all pointing to a dynamic genome characterized by extensive structural and functional plasticity. Prophages and MGEs can play major roles in the evolution and virulence of their hosts [113–116], acting as a reservoir of adaptive genes that can either be co-opted or horizontally transferred to enhance symbiotic capabilities [116, 117]. For example, in *Wolbachia*, key symbiosis-associated genes, such as the *cif* genes responsible for cytoplasmic incompatibility [118], and *wmk*, a male-killing factor [119], are both derived from and horizontally transferred via prophages [120, 121]. In sGff, as well as other *Spiroplasma*

strains, this pattern holds true [35, 122]. In *sGff*, key symbiosis genes, such as *RIP* and *glpO*, are flanked by prophage sequences. Several prophages in *sGff* encode lipoproteins, which could mediate interactions with host cells, metabolites, and lipids, contributing to symbiotic function [123–125].

The abundance of prophages present in the *sGff* genome may also play a mechanistic role in regulating *Spiroplasma* symbiosis. In *Wolbachia*, the genomic island *octomom* modulates bacterial density to prevent overproliferation [126]. Although *sGff* lacks a clearly defined *octomom* ortholog, prophage induction may serve a similar density-dependent regulatory function, by triggering bacterial lysis in response to stress, such as overcrowding or resource limitation [127, 128], or via a quorum sensing mechanism as observed in *Vibrio cholerae* [129] or *E. coli* [130]. Our transcriptomic comparison between cultured *sGff* and *sGff* isolated from tsetse hemolymph supports this hypothesis, as prophage-associated genes were highly expressed in the dense *in vitro* cultures as compared to hemolymph-derived *sGff*. Additionally, prophage induction can also modulate biofilm formation [131], interbacterial competition [132], or be induced via reactive oxygen species (ROS) [133]. Future studies comparing prophage activity in trypanosome-infected and uninfected tsetse will be key in determining whether *sGff* prophages contribute to trypanosome resistance or influence interactions with other prominent tsetse symbionts, including *Sodalis* and *Wigglesworthia*.

Our metabolic profiling confirms that, like other *Spiroplasma* strains [134, 135], *sGff* has limited biosynthetic capacity and relies on its host for essential metabolites. Unlike its sister taxon *sMel_MSRO* that utilizes only glucose [36], *sGff* appears to use both fructose and glucose as primary carbon sources. Notably, a functional trehalose transporter was absent, which may reduce *sGff* pathogenicity, as trehalose is abundant in tsetse [136]. The inability of *sGff* to utilize trehalose could limit its replication within the host and reduce host resource exploitation, thereby supporting a more commensal relationship. Transcriptomic analysis revealed condition-specific regulation of carbohydrate transporters, where fructose-specific transporters were upregulated in hemolymph-derived *sGff*, while glucose-specific transporters were downregulated. This finding likely reflects differences in nutrient availability, as fructose is absent from the culture medium, and suggests that fructose supplementation may accelerate *in vitro* growth. Nutrient-driven tissue tropism may also influence *sGff* distribution in its *Gff* host. High fructose levels in reproductive tissues [137] could explain the elevated *sGff* titers observed in these organs, while post-blood meal glucose spikes [138] may promote gut colonization and proliferation. We also identified a transporter for diacetylchitobiose, a sugar

derived from chitin. While *sGff* encodes several putative chitinases, it may access chitin breakdown products indirectly, potentially through *Sodalis*-derived secreted chitinase enzymes [139].

Beyond carbohydrates, *sGff* scavenges host-derived fatty acids and cholesterol, and like other *Spiroplasma* strains, can incorporate them directly into its membrane [94]. Pathway reconstructions also indicate that *sGff* can metabolize host diacylglycerols (DAGs) to synthesize cardiolipins, which are key membrane lipids in *Spiroplasma* [94] that could be cytotoxic [36] and in *Drosophila*, serve as important virulence factors by depleting host DAGs [95]. In *Gff* flies, the lipid-depleting phenotype associated with *sGff* is supported by previous work that shows decreased expression of *Gff* genes involved in fatty acid synthesis in *sGff* infected flies, suggesting reduced lipid availability [16]. In addition, triacylglycerol (TAG) levels are decreased in the fat bodies of *sGff*-infected tsetse [17] and since TAGs are synthesized from DAGs, *sGff*'s consumption of host DAGs may be contributing to this decrease. Functional annotations also indicated that *sGff* can produce ferritin and a hemolysin-related protein, both of which are used for iron sequestration. Iron sequestration is an important aspect of nutritional immunity, limiting microbe overproliferation and providing pathogen resistance [111, 140, 141]. In *Drosophila*, the iron-binding protein transferrin is upregulated in the presence of *sMel_MSRO*, which relies on host transferrin-bound iron for survival [142]. Finally, consistent with other *Spiroplasma* species, *sGff* is unable to synthesize most amino acids, and instead must import them from the host, further emphasizing its nutritional dependence on *Gff* [97, 143].

Our metabolic analysis reveals an intriguing potential competitive dynamic between *sGff* and trypanosomes within the tsetse gut, as both microorganisms require the same host-derived resources. In the *Gff* gut, both organisms prefer glucose as their primary carbohydrate source, with trypanosomes later switching to proline [138] – an amino acid that *sGff* also catabolizes. Trypanosomes, like *sGff*, scavenge host lipids and cholesterol [144] and both reduce the expression of *Gff* genes associated with lipid biosynthesis [16] and deplete host TAG reserves [17, 145]. While trypanosomes acquire transferrin-bound iron through a high affinity receptor [146], *sGff* likely acquires transferrin-bound iron in a manner similar to its sister *sMel_MSRO* strain [142]. This potential multifaceted resource competition may contribute to the trypanosome-refractory phenotype observed in *sGff*-infected flies [18]. By limiting parasite access to essential nutrients, *sGff* may delay parasite proliferation, potentially providing the host sufficient time to mount an effective immune response. These competitive dynamics require new studies to characterize *sGff*'s metabolic

demands *in vivo* and to test how resource competition influences host-pathogen dynamics, perhaps contributing to trypanosome resistance in tsetse.

In addition to a nutritional impact, *Spiroplasma* produces a diverse repertoire of toxins that can impact host fitness via reproductive manipulation or protection against parasites and pathogens [109, 110]. In *sGff*, we identified a limited set of potential encoded toxins: two putative Glycerol-3-phosphatase (*glpO*) genes and one ribosome-inactivating-protein (RIP) gene – all of which are expressed according to our transcriptome data. The *glpO* enzyme generates reactive oxygen species (ROS) and is a major virulence factor in *Mycoplasma* [147] and may protect *sMel_MSRO*-infected *Drosophila* against parasitoids [36]. In response to trypanosome infection in the gut, tsetse express nitric-oxide synthase (NOS), which catalyzes the production of trypanocidal nitric oxide (NO) and ROS [16]. These ROS are an essential part of the tsetse innate immune response, such that supplementing infected blood meals with antioxidants significantly enhances trypanosome infection success [148]. Thus, the ROS produced by *glpO* encoded by *sGff* could contribute to *Gff*'s immune response and enhance its trypanocidal effect.

We identified one well-conserved RIP gene in the *sGff* genome, which is transcriptionally active. RIPs act by depurinating the sarcin-ricin loop of 28s rRNA, thereby damaging host ribosomes and causing cell-death [149]. In sister *S. poulsonii* strains, *sMel_MSRO* and *sNeo*, RIPs can protect their hosts from parasitic nematodes and parasitoid wasps [28, 29]. *Gff* is of course host to parasitic trypanosomes, but there are also reports of pupal parasites from the Bombyliidae and Mutillidae families [150]. Determining whether *sGff*-derived RIPs are active against trypanosomes or pupal parasitoids should be a high priority for future research to improve vector control efforts.

Unlike most RIP genes present in the *S. poulsonii* clade, which are typically secreted, the *sGff* RIP gene lacks a signal peptide [109], and is predicted to be anchored in the cell membrane. In dense microenvironments such as plaques, high concentration of the membrane-bound RIPs could exert toxic effects in close proximity, while having limited effects at a distance. Alternatively, the RIP may be proteolytically cleaved from the membrane and released into the extracellular space, enabling it to act on specific targets, or perhaps it is secreted by MATE family efflux transporters encoded within the same operon. Given that secreted RIPs exhibit high substrate specificity, future work is needed to determine the nature of its molecular targets and whether this RIP remains membrane-associated or is secreted. Furthermore, other potential RIP functions need to be evaluated, including enzymatic activity such as chitinase or phosphatase

activity on lipids [151]. RIPs can also impact host tissues and incur fitness costs [28], but such effects may be tolerated or advantageous in contexts where protection against parasites or pathogens improves host survival. This trade-off could explain the polymorphic *sGff* distribution in *Gff* populations in northern Uganda, where it persists at about ~ 30% prevalence [18]. The potential for RIPs to mediate protective benefits in a population-specific manner warrants future research.

Conclusions

Here, we report the first successful *in vitro* cultivation and complete genome assembly of the *S. glossinidia sGff* – the *Spiroplasma* symbiont of *Gff* associated with deleterious effects on host metabolism and a trypanosome-refractory phenotype. We assembled a closed, circular ~ 1.5 Mb genome that clusters within the *S. poulsonii* clade. Comparative genomic analyses of cultured, laboratory- and field-derived *sGff* isolates revealed only minor isolate-specific genetic variations, residing primarily in mobile genetic elements in field-derived *sGff*. Metabolic profiling confirmed that *sGff* has limited biosynthetic capacity and relies on its *Gff* host for essential carbohydrates, lipids, and amino acids. We also identified two putative toxin genes: *RIP* and *glpO* that may contribute to its trypanocidal potential. Our genomic discoveries and the availability of a stable culture system will enable future functional studies to elucidate this symbiosis and identify potential implications for trypanosome transmission control.

Symbionts can protect their hosts through several mechanisms, including immune priming, nutrient supplementation, competitive exclusion of pathogens, or secretion of effector proteins that target pathogens. In the case of *sGff*, previous work found no evidence for activation of the tsetse immune system [16], and our work here identified no key nutrient supplementation apart from folate, which is already produced in excess by *Gff*'s obligate symbiont *Wigglesworthia*. However, we identified several points of potential metabolic competition between *sGff* and trypanosomes, including glucose, cholesterol, fatty acids and iron, which could collectively have an additive effect to limit parasite fitness. In addition, the two toxin genes identified, *glpO* and *RIP*, represent potential candidates for direct trypanocidal activity. These genomic discoveries, along with the establishment of a stable culture system, lay the foundation for functional studies to dissect the mechanisms of trypanosome resistance. Such studies may involve metabolomics, RNAseq, culture media modifications, tsetse transfections, or perhaps *sGff* gene knockouts. Future work also needs to address fundamental aspects of *sGff* biology, including its tissue specific distribution, specific roles of *sGff* within different host compartments, and strategies it

uses for vertical transmission. Expanding transcriptomic analyses to additional tissues, such as the reproductive organs, salivary glands, and midgut, will help reveal whether *sGff* gene expression is spatially regulated and potentially linked to the diverse physiological phenotypes observed in the tsetse host.

Abbreviations

AAT	African Animal Trypanosomiasis
ANI	Average Nucleotide Identity
BLASTP	Basic Local Alignment Search Tool for proteins
BSK-H medium	Barbour-Stoenner-Kelly H medium
BUSCO	Benchmarking Universal Single-Copy Orthologs
CDS	Coding Sequences
COG	Cluster of Orthologous Groups
DAGs	Diacylglycerols
DNA	Deoxyribonucleic acid
<i>E. coli</i>	<i>Escherichia coli</i>
FDR	False-Discovery Rate
GC content	Guanine-Cytosine content
Gff	<i>Glossina fuscipes fuscipes</i>
HAT	Human African Trypanosomiasis
HMW DNA	High Molecular Weight Deoxyribonucleic Acid
IAEA	International Atomic Energy Agency
ICEs	Integrative Conjugative Elements
ISEs	Insertion Sequence Elements
KEGG	Kyoto Encyclopedia of Genes and Genomes
Log ₂ FC	Log ₂ -fold-change
MGEs	Mobile Genetic Elements
ncRNAs	non-coding Ribonucleic Acids
NCBI	National Center for Biotechnology Information
NCBI SRA	National Center for Biotechnology Information Sequence Read Archive
NO	Nitric Oxide
ONT	Oxford Nanopore Technologies
PCR	Polymerase Chain Reaction
<i>p_sGff</i>	Plasmid of <i>Spiroplasma</i> endosymbiont of <i>Glossina fuscipes fuscipes</i>
PTS transporters	Phosphotransferase System Transporter
qPCR	quantitative Polymerase Chain Reaction
RIPs	Ribosome-inactivating Proteins
RNA/rRNA	Ribonucleic Acid/ribosomal Ribonucleic Acid
RNAseq	Ribonucleic Acid Sequencing
ROS	Reactive Oxygen Species
<i>sGff</i>	<i>Spiroplasma</i> endosymbiont of <i>Glossina fuscipes fuscipes</i> /strain <i>sGff</i>
<i>sMel_MSRO</i>	<i>Spiroplasma poulsonii</i> Melanogaster Sex-Ratio Organism
SNPs	Single Nucleotide Polymorphisms
SpAID	<i>Spiroplasma poulsonii</i> Androcinidin
TAG	Triacylglyceride
TPM	Transcripts per Million
tRNA	transfer Ribonucleic Acids

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12864-025-12351-w>.

Supplementary Material 1.

Supplementary Material 2.

Supplementary Material 3.

Acknowledgements

The authors would like to acknowledge all involved personnel at the Joint FAO/IAEA Centre of Nuclear Techniques in Food and Agriculture, Insect Pest Control Subprogram for tsetse rearing and colony management. We are grateful to Florent Masson who provided detailed insights into *in vitro*

cultivation methods. We thank the Yale Center for Research Computing for maintenance and use of the research computing infrastructure.

Authors' contributions

D.J.B., F.G., B.L.W., R.L.M., A.M.A. and S.A. conceived and designed the study. F.G. initiated *in vitro* cultures. D.J.B., F.G. and A.M.A. collected samples and generated sequencing data. F.G. and H.K. performed injection experiments. D.J.B. and F.G. analyzed and interpreted data. D.J.B., F.G., B.L.W. and S.A. wrote the manuscript. All authors contributed to the manuscript with comments and edits. B.L.W., S.A. and A.M.A. provided project supervision. D.J.B. and F.G. are equally contributing first authors.

Funding

This work was generously supported with funding from Ambrose Monell Foundation (to SA), National Institutes of Health (R01AI068932 and R01AI139525 to SA) and National Institutes of Health (R21AI163969 to SA and B.L.W.). This work also was funded by the International Atomic Energy Agency under the coordinated research project D42017 and the regular budget of the Joint FAO/IAEA Centre of Nuclear Techniques in Food and Agriculture for the project 2.1.4.2: Management of transboundary livestock insect pests for sustainable agriculture and rural development. The funders played no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Data availability

All genomic and transcriptomic data generated in this study is available at the NCBI BioProject: PRJNA1235259. Raw data from culture growth kinetics and injection experiments is deposited under <https://doi.org/10.60600/YU/2979VU>.

Declarations

Ethics approval and consent to participate

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

¹Department of Epidemiology of Microbial Diseases, Yale School of Public Health, New Haven, USA

²Insect Pest Control Laboratory, Joint FAO/IAEA Centre of Nuclear Techniques in Food and Agriculture, Vienna, Austria

³Institute of Chemical, Environmental, and Bioscience Engineering, Vienna University of Technology, Vienna, Austria

Received: 5 August 2025 / Accepted: 14 November 2025

Published online: 22 November 2025

References

1. Simarro PP, Diarra A, Ruiz Postigo JA, Franco JR, Jannin JG. The human African trypanosomiasis control and surveillance programme of the world health organization 2000–2009: the way forward. *PLoS Negl Trop Dis*. 2011;5:e1007. <https://doi.org/10.1371/journal.pntd.0001007>.
2. Franco JR, Cecchi G, Priotto G, Paone M, Diarra A, Grout L, et al. Monitoring the elimination of human African trypanosomiasis at continental and country level: update to 2018. *PLoS Negl Trop Dis*. 2020;14:e0008261. <https://doi.org/10.1371/journal.pntd.0008261>.
3. Holt HR, Selby R, Mumba C, Napier GB, Guitian J. Assessment of animal African trypanosomiasis (AAT) vulnerability in cattle-owning communities of sub-Saharan Africa. *Parasit Vectors*. 2016;9:53. <https://doi.org/10.1186/s13071-016-1336-5>.
4. Vreysen MJB, Seck MT, Sall B, Bouyer J. Tsetse flies: their biology and control using area-wide integrated pest management approaches. *J Invertebr Pathol*. 2013;112:S15–25. <https://doi.org/10.1016/j.jip.2012.07.026>.
5. Wang J, Weiss BL, Aksoy S. Tsetse fly microbiota: form and function. *Front Cell Infect Microbiol*. 2013. <https://doi.org/10.3389/fcimb.2013.00069>.
6. Aksoy S, Weiss BL, Attardo GM. Trypanosome transmission dynamics in Tsetse. *Curr Opin Insect Sci*. 2014;3:43–9. <https://doi.org/10.1016/j.cois.2014.07.003>.

7. Bing X, Attardo GM, Vigneron A, Aksoy E, Scolari F, Malacrida A, et al. Unraveling the relationship between the tsetse fly and its obligate symbiont *Wigglesworthia*: transcriptomic and metabolomic landscapes reveal highly integrated physiological networks. *Proceedings of the Royal Society B: Biological Sciences*. 2017;284:20170360. <https://doi.org/10.1098/rspb.2017.0360>.
8. Aksoy S, Attardo G. Paratransgenesis applied for control of Tsetse transmitted sleeping sickness. *Transgenesis and the management of vector-borne disease*. New York, NY: Springer New York; 2008. pp. 35–48. https://doi.org/10.1007/978-0-387-78225-6_3.
9. Michalkova V, Benoit JB, Weiss BL, Attardo GM, Aksoy S. Vitamin B6 generated by obligate symbionts is critical for maintaining proline homeostasis and fecundity in Tsetse flies. *Appl Environ Microbiol*. 2014;80:5844–53. <https://doi.org/10.1128/AEM.01150-14>.
10. Benoit JB, Vigneron A, Broderick NA, Wu Y, Sun JS, Carlson JR, et al. Symbiont-induced odorant binding proteins mediate insect host hematopoiesis. *eLife*. 2017. <https://doi.org/10.7554/eLife.19535>.
11. Weiss BL, Wang J, Aksoy S. Tsetse immune system maturation requires the presence of obligate symbionts in larvae. *PLoS Biol*. 2011;9:e1000619. <https://doi.org/10.1371/journal.pbio.1000619>.
12. Weiss BL, Maltz M, Aksoy S. Obligate symbionts activate immune system development in the Tsetse fly. *J Immunol*. 2012;188:3395–403. <https://doi.org/10.4049/jimmunol.1103691>.
13. Weiss BL, Wang J, Maltz MA, Wu Y, Aksoy S. Trypanosome infection establishment in the Tsetse fly gut is influenced by microbiome-regulated host immune barriers. *PLoS Pathog*. 2013;9:e1003318. <https://doi.org/10.1371/journal.ppat.1003318>.
14. Doudoumis V, Blow F, Saridaki A, Augustinos A, Dyer NA, Goodhead I, et al. Challenging the *Wigglesworthia*, *Sodalis*, *Wolbachia* symbiosis dogma in Tsetse flies: *Spiroplasma* is present in both laboratory and natural populations. *Sci Rep*. 2017;7:4699. <https://doi.org/10.1038/s41598-017-04740-3>.
15. Dieng MM, Dera KM, Moyaba P, Ouedraogo GM, Demirbas-Uzel G, Gstöttenmayer F, et al. Prevalence of *Trypanosoma* and *Sodalis* in wild populations of Tsetse flies and their impact on sterile insect technique programmes for Tsetse eradication. *Sci Rep*. 2022;12:3322. <https://doi.org/10.1038/s41598-022-06699-2>.
16. Awuoch EO, Smallenberger G, Bruzzese DL, Orfano A, Weiss BL, Aksoy S. *Spiroplasma* endosymbiont reduction of host lipid synthesis and Stomoxyn-like peptide contribute to trypanosome resistance in the Tsetse fly *Glossina fuscipes*. *PLoS Pathog*. 2025;21:e1012692. <https://doi.org/10.1371/journal.ppat.1012692>.
17. Son JH, Weiss BL, Schneider DI, Dera KM, Gstöttenmayer F, Opiro R, et al. Infection with endosymbiotic *Spiroplasma* disrupts Tsetse (*Glossina fuscipes fuscipes*) metabolic and reproductive homeostasis. *PLOS Pathog*. 2021;17:e1009539. <https://doi.org/10.1371/journal.ppat.1009539>.
18. Schneider DI, Saarman N, Onyango MG, Hyseni C, Opiro R, Echodu R, et al. Spatio-temporal distribution of *Spiroplasma* infections in the Tsetse fly (*Glossina fuscipes fuscipes*) in Northern Uganda. *PLoS Negl Trop Dis*. 2019;13:e0007340. <https://doi.org/10.1371/journal.pntd.0007340>.
19. Bové JM, Renaudin J, Saillard C, Foissac X, Garnier M. *Spiroplasma citri*, a plant pathogenic Mollicute: relationships with its two hosts, the plant and the leafhopper vector. *Annu Rev Phytopathol*. 2003;41:483–500. <https://doi.org/10.1146/annurev.phyto.41.052102.104034>.
20. Whitcomb RF, Tully JG. *The Mycoplasmas*. United States of America: Academic press; 1979.
21. Anbutu H, Fukatsu T. *Spiroplasma* as a model insect endosymbiont. *Environ Microbiol Rep*. 2011;3:144–53. <https://doi.org/10.1111/j.1758-2229.2010.00240.x>.
22. Fukatsu T, Nikoh N. Endosymbiotic microbiota of the bamboo pseudococcid *Antonina crawii* (Insecta, Homoptera). *Appl Environ Microbiol*. 2000;66:643–50. <https://doi.org/10.1128/AEM.66.2.643-650.2000>.
23. Lo W-S, Chen L-L, Chung W-C, Gasparich GE, Kuo C-H. Comparative genome analysis of *Spiroplasma melliferum* IPMB4A, a honeybee-associated bacterium. *BMC Genomics*. 2013;14:22. <https://doi.org/10.1186/1471-2164-14-22>.
24. Markham PG. *Spiroplasma* in leafhoppers: a review. *Yale J Biol Med*. 1983;56:745–51.
25. Mouches C, Bové JM, Tully JG, Rose DL, McCoy RE, Carle-Junca P, et al. *Spiroplasma apis*, a new species from the honey-bee *Apis mellifera*. *Ann Inst Pasteur Microbiol*. 1983;134:383–97. [https://doi.org/10.1016/S0769-2609\(83\)0063-5](https://doi.org/10.1016/S0769-2609(83)0063-5).
26. Tully JG, Rose DL, Yunker CE, Carle P, Bové JM, Williamson DL, et al. *Spiroplasma ixodetis* sp. nov., a new species from *Ixodes pacificus* ticks collected in Oregon. *Int J Syst Evol Microbiol*. 1995;45:23–8. <https://doi.org/10.1099/0020-7713-45-1-23>.
27. Filee J, Lopez-Villavicencio M, Debat V, Fourdin R, Salazar C, Silva-Brandao K, et al. Genome evolution and between-host transmission of *Spiroplasma* endosymbiont in wild communities of *Morpho* butterflies. Preprint. 2024. <https://doi.org/10.1101/2024.02.22.581604>.
28. 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. <https://doi.org/10.1371/journal.ppat.1006431>.
29. Hamilton PT, Peng F, Boulanger MJ, Perlman SJ. A ribosome-inactivating protein in a *Drosophila* defensive symbiont. *Proc Natl Acad Sci*. 2016;113:350–5. <https://doi.org/10.1073/pnas.1518648113>.
30. Jaenike J, Unckless R, Cockburn SN, Boelio LM, Perlman SJ. Adaptation via symbiosis: recent spread of a *Drosophila* defensive symbiont. *Science*. 2010;329:212–5. <https://doi.org/10.1126/science.1188235>.
31. Paredes JC, Herren JK, Schüpfer F, Lemaître B. The role of lipid competition for endosymbiont-mediated protection against parasitoid wasps in *Drosophila*. *mBio*. 2016;7:e01006-16. <https://doi.org/10.1128/mBio.01006-16>.
32. Harumoto T, Lemaître B. Male-killing toxin in a bacterial symbiont of *Drosophila*. *Nature*. 2018;557:252–5. <https://doi.org/10.1038/s41586-018-0086-2>.
33. Tinsley MC, Majerus MEN. A new male-killing parasitism: *Spiroplasma* bacteria infect the Ladybird beetle *Anisosticta novemdecimpunctata* (Coleoptera: Coccinellidae). *Parasitology*. 2006;132:757–65. <https://doi.org/10.1017/S0031182005009789>.
34. Jiggins FM, Hurst GDD, Jiggins CD, Schulenburg JHG, Majerus MEN. The butterfly *Danaus chrysippus* is infected by a male-killing *Spiroplasma* bacterium. *Parasitology*. 2000;120:439–46. <https://doi.org/10.1017/S0031182099005867>.
35. Gerth M, Martinez-Montoya H, Ramirez P, Masson F, Griffin JS, Aramayo R, et al. Rapid Mol Evol Spiroplasma Symbionts *Drosophila*. 2020. <https://doi.org/10.1101/2020.06.23.165548>.
36. Paredes JC, Herren JK, Schüpfer F, Marin R, Claverol S, Kuo C-H, et al. Genome sequence of the *Drosophila melanogaster* male-killing *Spiroplasma* strain MSRO endosymbiont. *mBio*. 2015;6:e02437-14. <https://doi.org/10.1128/mBio.02437-14>.
37. Dera K-SM, Dieng MM, Moyaba P, Ouedraogo GM, Pagabeleguem S, Njokou F, et al. Prevalence of *Spiroplasma* and interaction with wild *Glossina tachinoides* microbiota. *Parasite*. 2023;30:62. <https://doi.org/10.1051/parasite/2023064>.
38. Dera KM, Barro DT, Kaboré BA, Gstöttenmayer F, Dieng MM, Pagabeleguem S, et al. *Spiroplasma* infection in colonized *Glossina fuscipes fuscipes*: impact on mass rearing and the sterile insect technique. *Insect Sci*. 2025;1744–7917. <https://doi.org/10.1111/1744-7917.70078>.
39. Masson F, Calderon Copete S, Schüpfer F, Garcia-Arreaez G, Lemaître B. *In vitro* culture of the insect endosymbiont *Spiroplasma poulsonii* highlights bacterial genes involved in host-symbiont interaction. *mBio*. 2018;9:e00024-18. <https://doi.org/10.1128/mBio.00024-18>.
40. R Core Team. R: A language and environment for statistical computing. 2024.
41. RStudio Team. RStudio: Integrated Development for R. 2022.
42. Heid CA, Stevens J, Livak KJ, Williams PM. Real time quantitative PCR. *Genome Res*. 1996;6:986–94. <https://doi.org/10.1101/gr.6.10.986>.
43. Madigan MT, Bender KS, Buckley DH, Sattley WM, Stahl DA, Brock TD. *Brock biology of microorganisms*. Sixteenth edition, global edition. Harlow: Pearson Education Limited; 2022.
44. International Atomic Energy Agency. *FAO/IAEA Standard operating procedures for mass-rearing tsetse flies*. 2006.
45. Bauer B, Wetzal H. A new membrane for feeding *Glossina morsitans* Westw. (Diptera: Glossinidae). *Bull Entomol Res*. 1976;65:563–5. <https://doi.org/10.1017/S0007485300006246>.
46. Pfaffl MW. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res*. 2001;29:e45. <https://doi.org/10.1093/nar/29.9.e45>.
47. Bruzzese DJ, Weiss BL, Echodu R, Mireji PO, Abd-Alla AMM, Aksoy S. New Tsetse (*Glossina fuscipes fuscipes*) genomes generated from wild and laboratory-reared specimens. *Insect Sci*. 2025;1744–7917. <https://doi.org/10.1111/1744-7917.70085>.
48. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, et al. Twelve years of samtools and BCFtools. *Gigascience*. 2021;10:giab008. <https://doi.org/10.1093/gigascience/giab008>.
49. Steinig E, Coin L. Nanoq: ultra-fast quality control for nanopore reads. *J Open Source Softw*. 2022;7:2991. <https://doi.org/10.21105/joss.02991>.
50. Kolmogorov M, Bickhart DM, Behsaz B, Gurevich A, Rayko M, Shin SB, et al. MetaFlye: scalable long-read metagenome assembly using repeat graphs. *Nat Methods*. 2020;17:1103–10. <https://doi.org/10.1038/s41592-020-00971-x>.

51. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res.* 2017. <https://doi.org/10.1101/gr.215087.116>. :gr.215087.116.
52. Li H. Minimap and miniasm: fast mapping and de novo assembly for noisy long sequences. *Bioinformatics.* 2016;32:2103–10. <https://doi.org/10.1093/bioinformatics/btw152>.
53. Chen Y, Nie F, Xie S-Q, Zheng Y-F, Dai Q, Bray T, et al. Efficient assembly of nanopore reads via highly accurate and intact error correction. *Nat Commun.* 2021;12:60. <https://doi.org/10.1038/s41467-020-20236-7>.
54. Hu J, Wang Z, Sun Z, Hu B, Ayoola AO, Liang F, et al. NextDenovo: an efficient error correction and accurate assembly tool for noisy long reads. *Genome Biol.* 2024;25:107. <https://doi.org/10.1186/s13059-024-03252-4>.
55. Chen S, Zhou Y, Chen Y, Gu J. Fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics.* 2018;34:i884–90. <https://doi.org/10.1093/bioinformatics/bty560>.
56. Wick RR, Holt KE. Polypolish: short-read polishing of long-read bacterial genome assemblies. *PLoS Comput Biol.* 2022;18:e1009802. <https://doi.org/10.1371/journal.pcbi.1009802>.
57. Bouras G, Judd LM, Edwards RA, Vreugde S, Stinear TP, Wick RR. How low can you go? Short-read polishing of Oxford Nanopore bacterial genome assemblies. *Microb Genom.* 2024;10:001254. <https://doi.org/10.1099/mgen.0.001254>.
58. Li H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics.* 2018;34:3094–100. <https://doi.org/10.1093/bioinformatics/bty191>.
59. Cheng H, Concepcion GT, Feng X, Zhang H, Li H. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nat Methods.* 2021;18:170–5. <https://doi.org/10.1038/s41592-020-01056-5>.
60. Bouras G, Grigson SR, Papudeshi B, Mallawaarachchi V, Roach MJ. Dnaapler: a tool to reorient circular microbial genomes. *J Open Source Softw.* 2024;9:5968. <https://doi.org/10.21105/joss.05968>.
61. Formenti G, Abueg L, Brajuka A, Brajuka N, Gallardo-Alba C, Giani A, et al. Gfastats: conversion, evaluation and manipulation of genome sequences using assembly graphs. *Bioinformatics.* 2022;38:4214–6. <https://doi.org/10.1093/bioinformatics/btac460>.
62. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics.* 2015;31:3210–2. <https://doi.org/10.1093/bioinformatics/btv351>.
63. Frolova D, Lima L, Roberts LW, Bohnenkämper L, Wittler R, Stoye J, et al. Applying rearrangement distances to enable plasmid epidemiology with plying. *Microb Genom.* 2024;10:001300. <https://doi.org/10.1099/mgen.0.001300>.
64. Li W, O'Neill KR, Haft DH, DiCuccio M, Chetvernin V, Badretin A, et al. RefSeq: expanding the prokaryotic genome annotation pipeline reach with protein family model curation. *Nucleic Acids Res.* 2020;49:D1020–8. <https://doi.org/10.1093/nar/gkaa1105>.
65. Schwengers O, Jelonek L, Dieckmann MA, Beyvers S, Blom J, Goesmann A. Bakta: rapid and standardized annotation of bacterial genomes via alignment-free sequence identification. *Microb Genomics.* 2021;7. <https://doi.org/10.1099/mgen.0.000685>.
66. Cantalapiedra CP, Hernández-Plaza A, Letunic I, Bork P, Huerta-Cepas J. eggNOG-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale. *Mol Biol Evol.* 2021;38:5825–9. <https://doi.org/10.1093/molbev/msab293>.
67. Kanehisa M, Sato Y, Morishima K. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. *J Mol Biol.* 2016;428:726–31. <https://doi.org/10.1016/j.jmb.2015.11.006>.
68. Wishart DS, Han S, Saha S, Oler E, Peters H, Grant JR, et al. PHASTEST: faster than PHASTER, better than PHAST. *Nucleic Acids Res.* 2023;51:W443–50. <https://doi.org/10.1093/nar/gkad382>.
69. Wang M, Liu G, Liu M, Tai C, Deng Z, Song J, et al. ICEberg 3.0: functional categorization and analysis of the integrative and conjugative elements in bacteria. *Nucleic Acids Res.* 2024;52:D732–7. <https://doi.org/10.1093/nar/gka4935>.
70. Xie Z, Tang H. ISEScan: automated identification of insertion sequence elements in prokaryotic genomes. *Bioinformatics.* 2017;33:3340–7. <https://doi.org/10.1093/bioinformatics/btx433>.
71. Teufel F, Almagro Armenteros JJ, Johansen AR, Gíslason MH, Pihl SI, Tsirigos KD, et al. SignalP 6.0 predicts all five types of signal peptides using protein language models. *Nat Biotechnol.* 2022;40:1023–5. <https://doi.org/10.1038/s41587-021-01156-3>.
72. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. *BMC Bioinformatics.* 2009;10:421. <http://doi.org/10.1186/1471-2105-10-421>.
73. Blum M, Andreeva A, Florentino LC, Chuguransky SR, Grego T, Hobbs E, et al. InterPro: the protein sequence classification resource in 2025. *Nucleic Acids Res.* 2025;53:D444–56. <https://doi.org/10.1093/nar/gkae1082>.
74. Löytynoja A. Phylogeny-aware alignment with PRANK. In: Russell DJ, editor. Multiple sequence alignment methods. Totowa, NJ: Humana; 2014. pp. 155–70. https://doi.org/10.1007/978-1-62703-646-7_10.
75. Kozlov AM, Darriba D, Flouri T, Morel B, Stamatakis A. RAXML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics.* 2019;35:4453–5. <https://doi.org/10.1093/bioinformatics/btz305>.
76. O'Leary NA, Cox E, Holmes JB, Anderson WR, Falk R, Hem V, et al. Exploring and retrieving sequence and metadata for species across the tree of life with NCBI datasets. *Sci Data.* 2024;11:732. <https://doi.org/10.1038/s41597-024-03571-y>.
77. Xu X, Yin Z, Yan L, Zhang H, Xu B, Wei Y, et al. RabbitTClust: enabling fast clustering analysis of millions of bacteria genomes with MinHash sketches. *Genome Biol.* 2023;24:121. <https://doi.org/10.1186/s13059-023-02961-6>.
78. Gautreau G, Bazin A, Gachet M, Planel R, Burlot L, Dubois M, et al. PPanG-GOLIN: depicting microbial diversity via a partitioned pangenome graph. *PLoS Comput Biol.* 2020;16:e1007732. <https://doi.org/10.1371/journal.pcbi.1007732>.
79. Katoh K, Misawa K, Kuma K, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast fourier transform. *Nucleic Acids Res.* 2002;30:3059–66. <https://doi.org/10.1093/nar/gkf436>.
80. Pritchard L, Glover RH, Humphris S, Elphinstone JG, Toth IK. Genomics and taxonomy in diagnostics for food security: soft-rotting enterobacterial plant pathogens. *Anal Methods.* 2015;8:12–24. <https://doi.org/10.1039/C5AY02550H>.
81. Marçais G, Delcher AL, Phillippy AM, Coston R, Salzberg SL, Zimin A. MUMmer4: a fast and versatile genome alignment system. *PLoS Comput Biol.* 2018;14:e1005944. <https://doi.org/10.1371/journal.pcbi.1005944>.
82. Goel M, Sun H, Jiao W-B, Schneeberger K, SyRI: finding genomic rearrangements and local sequence differences from whole-genome assemblies. *Genome Biol.* 2019;20:277. <https://doi.org/10.1186/s13059-019-1911-0>.
83. Goel M, Schneeberger K. Plotsr: visualizing structural similarities and rearrangements between multiple genomes. *Bioinformatics.* 2022;38:2922–6. <https://doi.org/10.1093/bioinformatics/btac196>.
84. Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R, et al. Welcome to the Tidyverse. *J Open Source Softw.* 2019;4:1686. <https://doi.org/10.21105/joss.01686>.
85. Chen H, Boutros PC. VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R. *BMC Bioinformatics.* 2011;12:35. <https://doi.org/10.1186/1471-2105-12-35>.
86. O S, Ma LJ. Bakta: rapid and standardized annotation of bacterial genomes via alignment-free sequence identification. *Microb Genomics.* 2021;7. <https://doi.org/10.1099/mgen.0.000685>.
87. Ewels P, Magnusson M, Lundin S, Käller M. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics.* 2016;32:3047–8. <https://doi.org/10.1093/bioinformatics/btw354>.
88. Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat Methods.* 2017;14:417–9. <https://doi.org/10.1038/nmeth.4197>.
89. Karaji R, Peña-Castillo L, OpDetect. A convolutional and recurrent neural network classifier for precise and sensitive operon detection from RNA-seq data. 2025;2025.03.24.645056. <https://doi.org/10.1101/2025.03.24.645056>.
90. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics.* 2013;29:15–21. <https://doi.org/10.1093/bioinformatics/bts635>.
91. Chen Y, Chen L, Lun ATL, Baldoni PL, Smyth GK. edgeR v4: powerful differential analysis of sequencing data with expanded functionality and improved support for small counts and larger datasets. *Nucleic Acids Res.* 2025;53:gkaf018. <https://doi.org/10.1093/nar/gkaf018>.
92. Lo W-S, Haryono M, Gasparich GE, Kuo C-H. Complete genome sequence of *Spiroplasma* sp. TU-14. *Genome Announc.* 2017;5:e01465-16. <https://doi.org/10.1128/genomeA.01465-16>.
93. Haryono M, Lo W-S, Gasparich GE, Kuo C-H. Complete genome sequence of *Spiroplasma* sp. NBRC 100390. *Genome Announc.* 2017;5:e00008-17. <https://doi.org/10.1128/genomeA.00008-17>.

94. Freeman BA, Sissenstein R, McManus TT, Woodward JE, Lee IM, Mudd JB. Lipid composition and lipid metabolism of *Spiroplasma citri*. J Bacteriol. 1976;125:946–54. <https://doi.org/10.1128/jb.125.3.946-954.1976>.
95. Herren JK, Paredes JC, Schüpfer F, Arafah K, Bulet P, Lemaître B. Insect endosymbiont proliferation is limited by lipid availability. eLife. 2014;3:e02964. <http://doi.org/10.7554/eLife.02964>.
96. Wingreen NS, Huang KC. Physics of intracellular organization in bacteria. Annu Rev Microbiol. 2015;69(1):361–79. <https://doi.org/10.1146/annurev-micro-091014-104313>.
97. Masson F, Rommelaere S, Schüpfer F, Boquete J-P, Lemaître B. Disproportionate investment in Spiralin B production limits in-host growth and favors the vertical transmission of *Spiroplasma* insect endosymbionts. Proc Natl Acad Sci U S A. 2022;119:e2208461119. <https://doi.org/10.1073/pnas.2208461119>.
98. Zha G-D, Yang D-H, Wang J-J, Yang B, Yu H-S. Infection function of adhesin-like protein ALP609 from *Spiroplasma melliferum* CH-1. Curr Microbiol. 2018;75:701–8. <https://doi.org/10.1007/s00284-018-1435-y>.
99. Béven L, Duret S, Batailler B, Dubrana M-P, Saillard C, Renaudin J, et al. The repetitive domain of SCARP3d triggers entry of *Spiroplasma citri* into cultured cells of the vector *Circulifer haematocaps*. PLoS ONE. 2012;7:e48606. <https://doi.org/10.1371/journal.pone.0048606>.
100. Pilo P, Frey J, Vilei EM. Molecular mechanisms of pathogenicity of *Mycoplasma mycoides* subsp. *mycoides* SC. Vet J. 2007;174:513–21. <https://doi.org/10.1016/j.tvjl.2006.10.016>.
101. Tully JG, Whitcomb RF, Clark HF, Williamson DL. Pathogenic <Emphasis Type="Italic">Mycoplasmas</Emphasis>: cultivation and vertebrate pathogenicity of a new <Emphasis Type="Italic">Spiroplasma</Emphasis>. Science. 1977;195:892–4. <https://doi.org/10.1126/science.841314>.
102. Bell-Sakyi L, Palomar AM, Kazimirova M. Isolation and propagation of a *Spiroplasma* sp. from Slovakian *Ixodes ricinus* ticks in *Ixodes* spp. cell lines. Ticks Tick-borne Dis. 2015;6:601–6. <https://doi.org/10.1016/j.ttbdis.2015.05.002>.
103. Steiner T, McGarrity GJ, Phillips DM. Cultivation and partial characterization of *Spiroplasmas* in cell cultures. Infect Immun. 1982;35:296–304. <https://doi.org/10.1128/iai.35.1.296-304.1982>.
104. Masson F, Lemaître B. Growing ungrowable bacteria: overview and perspectives on insect symbiont culturability. Microbiol Mol Biol Rev. 2020;84:e00089–20. <https://doi.org/10.1128/MMBR.00089-20>.
105. Masson F, Schüpfer F, Jollivet C, Lemaître B. Transformation of the *Drosophila* sex-manipulative endosymbiont *Spiroplasma poulsonii* and persisting hurdles for functional genetic studies. Appl Environ Microbiol. 2020;86:e00835–20. <https://doi.org/10.1128/AEM.00835-20>.
106. Catchpole J, Maynard J, Chang BJ, Payne MS, Beeton ML, Furfuro LL. Minuscule mollicutes: current hurdles to bacteriophage identification. Sustainable Microbiol. 2024;1:qvae019. <https://doi.org/10.1093/sumbio/qvae019>.
107. Bébéar C-M, Aullo P, Bové J-M, Renaudin J. *Spiroplasma citri* virus SpV1: characterization of viral sequences present in the *Spiroplasma* host chromosome. Curr Microbiol. 1996;32:134–40. <https://doi.org/10.1007/s002849900024>.
108. Krafur ES, Marquez JG, Ouma JO. Structure of some East African *Glossina fuscipes fuscipes* populations. Med Vet Entomol. 2008;22:222–7. <https://doi.org/10.1111/j.1365-2915.2008.00739.x>.
109. Moore LD, Ballinger MJ. The toxins of vertically transmitted *Spiroplasma*. Front Microbiol. 2023;14:1148263. <https://doi.org/10.3389/fmicb.2023.1148263>.
110. Ballinger MJ, Perlman SJ. The defensive *Spiroplasma*. Curr Opin Insect Sci. 2019;32:36–41. <https://doi.org/10.1016/j.cois.2018.10.004>.
111. Hrdina A, Serra Canales M, Arias-Rojas A, Frahm D, Iatsenko I. The endosymbiont *Spiroplasma poulsonii* increases *Drosophila melanogaster* resistance to pathogens by enhancing iron sequestration and melanization. MBio. 2024;15:e0093624. <https://doi.org/10.1128/mbio.00936-24>.
112. Rattner R, Thapa SP, Dang T, Osman F, Selvaraj V, Maheshwari Y, et al. Genome analysis of *Spiroplasma citri* strains from different host plants and its leafhopper vectors. BMC Genomics. 2021;22:373. <https://doi.org/10.1186/s12864-021-07637-8>.
113. Wang GH, Sun BF, Xiong TL, Wang YK, Murfin KE, Xiao JH, et al. Bacteriophage WO can mediate horizontal gene transfer in endosymbiotic *Wolbachia* genomes. Front Microbiol. 2016;7:1867. <https://doi.org/10.3389/fmicb.2016.01867>.
114. Ochman H, Lawrence JG, Groisman EA. Lateral gene transfer and the nature of bacterial innovation. Nature. 2000;405:299–304. <https://doi.org/10.1038/35012500>.
115. Arnold BJ, Huang I-T, Hanage WP. Horizontal gene transfer and adaptive evolution in bacteria. Nat Rev Microbiol. 2022;20:206–18. <https://doi.org/10.1038/s41579-021-00650-4>.
116. Fortier L-C, Sekulovic O. Importance of prophages to evolution and virulence of bacterial pathogens. Virulence. 2013;4:354–65. <https://doi.org/10.4161/viru.24498>.
117. Tanaka K, Furukawa S, Nikoh N, Sasaki T, Fukatsu T. Complete WO phage sequences reveal their dynamic evolutionary trajectories and putative functional elements required for integration into the *Wolbachia* genome. Appl Environ Microbiol. 2009;75:5676–86. <https://doi.org/10.1128/AEM.01172-09>.
118. Beckmann JF, Ronau JA, Hochstrasser M. A *Wolbachia* deubiquitylating enzyme induces cytoplasmic incompatibility. Nat Microbiol. 2017;2:17007. <https://doi.org/10.1038/nmicrobiol.2017.7>.
119. Perlmutter JI, Bordenstein SR, Unckless FM, LePage DP, Metcalf JA, Hill T, et al. The phage gene Wmk is a candidate for male killing by a bacterial endosymbiont. PLoS Pathog. 2019;15:e1007936. <https://doi.org/10.1371/journal.ppat.1007936>.
120. McNamara CJ, Ant TH, Harvey-Samuel T, White-Cooper H, Martinez J, Alphey L, et al. Transgenic expression of *Cif* genes from *Wolbachia* strain wAlbB recapitulates cytoplasmic incompatibility in *Aedes aegypti*. Nat Commun. 2024;15:869. <https://doi.org/10.1038/s41467-024-45238-7>.
121. Martinez J, Klasson L, Welch JJ, Jiggins JM. Life and death of selfish genes: comparative genomics reveals the dynamic evolution of cytoplasmic incompatibility. Mol Biol Evol. 2021. <https://doi.org/10.1093/molbev/msaa209>.
122. Arai H, Legeai F, Kageyama D, Sugio A, Simon J-C. Genomic insights into *Spiroplasma* endosymbionts that induce male-killing and protective phenotypes in the pea aphid. FEMS Microbiol Lett. 2024;371:fnae027. <https://doi.org/10.1093/femsle/fnae027>.
123. Knoke LR, Abad Herrera S, Götz K, Justesen BH, Günther Pomorski T, Fritz C, et al. *Agrobacterium tumefaciens* small lipoprotein Atu8019 is involved in selective outer membrane vesicle (OMV) docking to bacterial cells. Front Microbiol. 2020;11:1228. <https://doi.org/10.3389/fmicb.2020.01228>.
124. Jin S, Joe A, Lynett J, Hani EK, Sherman P, Chan LV. JlpA, a novel surface-exposed lipoprotein specific to *Campylobacter jejuni*, mediates adherence to host epithelial cells. Mol Microbiol. 2001;39:1225–36. <https://doi.org/10.1111/j.1365-2958.2001.02294.x>.
125. Speare L, Woo M, Dunn AK, Septon AN. A putative lipoprotein mediates cell-cell contact for type VI secretion system-dependent killing of specific competitors. mBio. 2022;13:e0308521. <https://doi.org/10.1128/mbio.03085-21>.
126. Duarte EH, Carvalho A, López-Madrigal S, Costa J, Teixeira L. Forward genetics in *Wolbachia*: regulation of *Wolbachia* proliferation by the amplification and deletion of an addictive genomic island. PLoS Genet. 2021;17:e1009612. <https://doi.org/10.1371/journal.pgen.1009612>.
127. Mills S, Shanahan F, Stanton C, Hill C, Coffey A, Ross RP. Movers and shakers: influence of bacteriophages in shaping the mammalian gut microbiota. Gut Microbes. 2013;4:4–16. <https://doi.org/10.4161/gmic.22371>.
128. Ofir G, Sorek R. Contemporary phage biology: from classic models to new insights. Cell. 2018;172:1260–70. <https://doi.org/10.1016/j.cell.2017.10.045>.
129. Silpe JE, Bassler BL. A host-produced quorum-sensing autoinducer controls a phage lysis-lysogeny decision. Cell. 2019;176:268–80. <https://doi.org/10.1016/j.cell.2018.10.059>.e13.
130. Laganenka L, Sander T, Lagonenko A, Chen Y, Link H, Sourjik V. Quorum sensing and metabolic state of the host control lysogeny-lysis switch of bacteriophage T1. mBio. 2019. <https://doi.org/10.1128/mbio.01884-19>.
131. Li D, Liang W, Hu Q, Ren J, Xue F, Liu Q, et al. The effect of a spontaneous induction prophage, phi458, on biofilm formation and virulence in avian pathogenic <Emphasis Type="Italic">Escherichia coli</Emphasis>. Front Microbiol. 2022. <https://doi.org/10.3389/fmicb.2022.1049341>.
132. de Sablet T, Chassard C, Bernalier-Donadille A, Varelle M, Gobert AP, Martin C. Human microbiota-secreted factors inhibit shiga toxin synthesis by enterohemorrhagic *Escherichia coli* O157:H7. Infect Immun. 2009;77:783–90. <https://doi.org/10.1128/IAI.01048-08>.
133. Loš JM, Loš M, Węgrzyn A, Węgrzyn G. Hydrogen peroxide-mediated induction of the Shiga toxin-converting lambdoid prophage ST2-8624 in *Escherichia coli* O157:H7. FEMS Immunol Med Microbiol. 2010;58:322–9. <https://doi.org/10.1111/j.1574-695X.2009.00644.x>.
134. Ramirez P, Martinez Montoya H, Aramayo R, Mateos M. Diverse toxin repertoire but limited metabolic capacities inferred from the draft genome assemblies of three *Spiroplasma* (Citri clade) strains associated with *Drosophila*. Microb Genomics. 2025;11. <https://doi.org/10.1099/mgen.0.001408>.
135. Lo W-S, Ku C, Chen L-L, Chang T-H, Kuo C-H. Comparison of metabolic capacities and inference of gene content evolution in mosquito-associated *Spiroplasma diminutum* and *S. taiwanense*. Genome Biol Evol. 2013;5:1512–23. <https://doi.org/10.1093/gbe/evt108>.

136. Geigy R, Huber M, Weinman D, Wyatt GR. Demonstration of trehalose in the vector of African trypanosomiasis: the tsetse fly. *Acta Trop.* 1959;16:255–62.
137. Scolari F, Benoit JB, Michalkova V, Aksoy E, Takac P, Abd-Alla AMM, et al. The spermatophore in *Glossina morsitans morsitans*: insights into male contributions to reproduction. *Sci Rep.* 2016;6:20334. <https://doi.org/10.1038/srep20334>.
138. Naguleswaran A, Fernandes P, Bevkál S, Rehmann R, Nicholson P, Roditi I. Developmental changes and metabolic reprogramming during establishment of infection and progression of *Trypanosoma brucei brucei* through its insect host. *PLoS Negl Trop Dis.* 2021;15:e0009504. <https://doi.org/10.1371/journal.pntd.0009504>.
139. Welburn SC, Arnold K, Maudlin I, Gooday GW. Rickettsia-like organisms and chitinase production in relation to transmission of trypanosomes by Tsetse flies. *Parasitology.* 1993;107:141–5. <https://doi.org/10.1017/S003118200006724X>.
140. Hennigar SR, McClung JP. Nutritional immunity: starving pathogens of trace minerals. *Am J Lifestyle Med.* 2016;10:170–3. <https://doi.org/10.1177/1559827616629117>.
141. Iatsenko I, Marra A, Boquete J-P, Peña J, Lemaitre B. Iron sequestration by transferrin 1 mediates nutritional immunity in *Drosophila melanogaster*. *Proc Natl Acad Sci USA.* 2020;117:7317–25. <https://doi.org/10.1073/pnas.1914830117>.
142. Marra A, Masson F, Lemaitre B. The iron transporter Transferrin 1 mediates homeostasis of the endosymbiotic relationship between *Drosophila melanogaster* and *Spiroplasma poulsonii*. *microLife.* 2021;2:uqab008. <https://doi.org/10.1093/femsm/luqab008>.
143. Patterson A, Stevens C, Cody RM, Gudauskas RT. Differential amino acid utilization by *Spiroplasma* and the effect on growth kinetics. *J Gen Appl Microbiol.* 1985;31:499–505. <https://doi.org/10.2323/jgam.31.499>.
144. de Parreira Aquino G, Mendes Gomes MA, Köpke Salinas R, Laranjeira-Silva MF. Lipid and fatty acid metabolism in trypanosomatids. *Microb Cell.* 2021;8:262–75. <https://doi.org/10.15698/mic2021.11.764>.
145. Atella T, Bittencourt-Cunha PR, Araujo MFC, Silva-Cardoso L, Maya-Monteiro CM, Atella GC. *Trypanosoma cruzi* modulates lipid metabolism and highjacks phospholipids from the midgut of *Rhodnius prolixus*. *Acta Trop.* 2022;233:106552. <https://doi.org/10.1016/j.actatropica.2022.106552>.
146. Steverding D, Stierhof YD, Fuchs H, Tauber R, Overath P. Transferrin-binding protein complex is the receptor for transferrin uptake in *Trypanosoma brucei*. *J Cell Biol.* 1995;131:1173–82. <https://doi.org/10.1083/jcb.131.5.1173>.
147. Pilo P, Vilei EM, Peterhans E, Bonvin-Klotz L, Stoffel MH, Dobbelaere D, et al. A metabolic enzyme as a primary virulence factor of *Mycoplasma mycoides* subsp. *mycoides* small colony. *J Bacteriol.* 2005;187:6824–31. <https://doi.org/10.1128/JB.187.19.6824-6831.2005>.
148. MacLEOD ET, Maudlin I, Darby AC, Welburn SC. Antioxidants promote establishment of trypanosome infections in Tsetse. *Parasitology.* 2007;134:827–31. <https://doi.org/10.1017/S0031182007002247>.
149. Stirpe F. Ribosome-inactivating proteins. *Toxicon.* 2004;44:371–83. <https://doi.org/10.1016/j.toxicon.2004.05.004>.
150. Markham RH. Biological control of tsetse: prospects and progress in the use of pupal parasites. *Int J Trop Insect Sci.* 1986;7:1–4. <https://doi.org/10.1017/S1742758400003015>.
151. De Zaeytjij J, Van Damme EJM. Extensive evolution of cereal ribosome-inactivating proteins translates into unique structural features, activation mechanisms, and physiological roles. *Toxins.* 2017;9:123. <https://doi.org/10.3390/toxins9040123>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.