frontiers | Research Topics

Spiroplasma, Mycoplasma, Phytoplasma, and other genome-reduced and wall-less mollicutes: Their genetics, genomics, mechanics, interactions and symbiosis with insects, other animals and plants

Edited by
Takema Fukatsu, Toshiyuki Harumoto, Shigeyuki Kakizawa, Chih-Horng Kuo and Akiko Sugio

Published in
Frontiers in Microbiology

ISSN 1664-8714
ISBN 978-2-8325-5418-0
DOI 10.3389/978-2-8325-5418-0

About Frontiers

Frontiers is more than just an open access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers journal series

The Frontiers journal series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the *Frontiers journal series* operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews. Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the *Frontiers journals series*: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area.

Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers editorial office: frontiersin.org/about/contact

Spiroplasma, Mycoplasma, Phytoplasma, and other genome-reduced and wall-less mollicutes: Their genetics, genomics, mechanics, interactions and symbiosis with insects, other animals and plants

Topic editors

Takema Fukatsu — National Institute of Advanced Industrial Science and Technology (AIST), Japan

Toshiyuki Harumoto — Kyoto University, Japan

Shigeyuki Kakizawa — National Institute of Advanced Industrial Science and Technology (AIST), Japan

Chih-Horng Kuo — Institute of Plant and Microbial Biology, Academia Sinica, Taiwan

Akiko Sugio — Institut de Génétique, Environnement et Protection des Plantes, Institut National de la Recherche Agronomique, France

Citation

Fukatsu, T., Harumoto, T., Kakizawa, S., Kuo, C.-H., Sugio, A., eds. (2024). *Spiroplasma, Mycoplasma, Phytoplasma, and other genome-reduced and wall-less mollicutes: Their genetics, genomics, mechanics, interactions and symbiosis with insects, other animals and plants*. Lausanne: Frontiers Media SA.

doi: 10.3389/978-2-8325-5418-0

Table of contents

frontiers | Frontiers in Microbiology

TYPE Editorial
PUBLISHED 30 August 2024
DOI 10.3389/fmicb.2024.1477536

OPEN ACCESS

EDITED AND REVIEWED BY
Louis S. Tisa,
University of New Hampshire, United States

*CORRESPONDENCE
Takema Fukatsu
✉ t-fukatsu@aist.go.jp

RECEIVED 08 August 2024
ACCEPTED 15 August 2024
PUBLISHED 30 August 2024

CITATION
Fukatsu T, Kakizawa S, Harumoto T, Sugio A and Kuo C-H (2024) Editorial: *Spiroplasma*, *Mycoplasma*, Phytoplasma, and other genome-reduced and wall-less mollicutes: their genetics, genomics, mechanics, interactions and symbiosis with insects, other animals and plants.
Front. Microbiol. 15:1477536.
doi: 10.3389/fmicb.2024.1477536

Editorial: *Spiroplasma*, *Mycoplasma*, Phytoplasma, and other genome-reduced and wall-less mollicutes: their genetics, genomics, mechanics, interactions and symbiosis with insects, other animals and plants

Takema Fukatsu[1,2,3]*, Shigeyuki Kakizawa[1], Toshiyuki Harumoto[4,5], Akiko Sugio[6] and Chih-Horng Kuo[7]

[1]Bioproduction Research Institute, National Institute of Advanced Industrial Science and Technology, Tsukuba, Japan, [2]Graduate School of Life and Environmental Sciences, University of Tsukuba, Tsukuba, Japan, [3]Department of Biological Sciences, Graduate School of Science, The University of Tokyo, Tokyo, Japan, [4]Hakubi Center for Advanced Research, Kyoto University, Kyoto, Japan, [5]Graduate School of Biostudies, Kyoto University, Kyoto, Japan, [6]IGEPP, INRAE, Institut Agro, University of Rennes, Le Rheu, France, [7]Institute of Plant and Microbial Biology, Academia Sinica, Taipei, Taiwan

KEYWORDS

***Spiroplasma, Mycoplasma*, Phytoplasma, symbiont, pathogen, genome, insect, plant**

Editorial on the Research Topic
Spiroplasma, *Mycoplasma*, Phytoplasma, and other genome-reduced and wall-less mollicutes: their genetics, genomics, mechanics, interactions and symbiosis with insects, other animals and plants

Genome-reduced, wall-less, and fastidious bacteria of the genera *Spiroplasma*, *Mycoplasma*, "*Candidatus* Phytoplasma" and allies belonging to the class Mollicutes, are known for a number of unique microbiological features, which have prompted researchers to investigate their basic, applied, and medical aspects (Brown et al., 2018). They are mostly parasitic or symbiotic to a variety of animals or plants, living on or within the eukaryotic cells. Spiroplasmas, recognized by their characteristic helical shapes and active twitching motility, are associated with diverse arthropods and plants (Gasparich et al., 2020), and have been developed as models for the study of facultative symbionts (Anbutsu and Fukatsu, 2011; Lo et al., 2016). Some strains of *Spiroplasma poulsonii* and *Spiroplasma ixodetis* cause a remarkable reproductive phenotype, called male-killing, of their insect hosts (Hurst and Frost, 2015). In contrast, some other insect-associated spiroplasmas protect their hosts from natural enemies, including parasitoid wasps, nematodes, and pathogenic fungi (Ballinger and Perlman, 2019). *Spiroplasma citri* and *Spiroplasma kunkelii* are notorious as devastating pathogens of citrus and maize, respectively (Gasparich et al., 2020). Mycoplasmas are not only medically important as human or animal pathogens like *Mycoplasma pneumoniae* (Waites and Talkington, 2004) and *Mycoplasma mycoides* (Teodoro et al., 2020), but also intensively investigated as minimal-genome bacterial

models (Yus et al., 2009). Microbial genome synthesis and engineering technologies have been developed mainly on *M. mycoides* and *Mycoplasma capricolum* (Venter et al., 2022). Some mycoplasmas are known for their capability of unique gliding motility (Miyata et al., 2020). Phytoplasmas are obligatorily parasitic to plant phloem tissues and vectored by plant-sucking insects, often causing spectacular plant morphological changes like phyllody, virescence, witches' bloom, etc. (Hogenhout et al., 2008; Bertaccini et al., 2022).

While the conventional studies have revealed fascinating aspects of this bizarre bacterial group, the whole picture of their diversity and versatility has long been elusive mainly due to their reluctance to axenic cultivation. However, owing to the recent development and availability of high-throughput DNA sequencing technologies, our knowledge on the diversity of such fastidious microbes in a variety of environments has been growing rapidly. In this context, the Research Topic "*Spiroplasma, Mycoplasma, Phytoplasma, and other genome-reduced and wall-less mollicutes: their genetics, genomics, mechanics, interactions and symbiosis with insects, other animals and plants*" is aimed to provide an opportunity to compile the new information emerging in this research field. In total, 10 articles and one mini review were published, of which five, three, and three articles are on *Spiroplasma*, *Mycoplasma*, Phytoplasma, respectively, thereby covering this research area in a balanced manner.

As for *Spiroplasma*, unique articles are contributed to this Research Topic. Mizutani, Omori et al. reported successful cloning of the whole 1.12 Mbp genome of *Spiroplasma chrysopicola*, which was originally isolated from a deer fly *Chrysops* sp., into the yeast *Saccharomyces cerevisiae*. Now the *S. chrysopicola* genome is retained in yeast cells, can be genetically engineered using sophisticated genetic tools available for *S. cerevisiae*, and can be distributed to anybody and utilized for further research. While a series of elaborate synthetic biological technologies have been developed using *Mycoplasma* spp. (Venter et al., 2022), this study serves as an initial step toward the synthetic biological approaches to *Spiroplasma* spp. Following two articles reported the characterization of newly obtained genomes of *S. ixodetis* strains. Arai et al. analyzed the genome sequence of a male-killing *S. ixodetis* symbiont of *Homona magnanima*. The genome harbored a number of putative virulence-associated genes like ankyrin domain containing genes, while no homologous sequence of *spaid*, a male-killing gene of *S. poulsonii* in *Drosophila* (Harumoto and Lemaitre, 2018), was identified. This implicates the diverged mechanism of male killing evolved in distinct *Spiroplasma* symbionts. The authors also showed bacterial tropism toward host somatic tissues and successful proliferation in various inset cell lines, which conform to the fact that *S. ixodetis* strains have a broad host range. Moore and Ballinger presented the complete genome sequence of a defensive *S. ixodetis* symbiont of *D. atripex* and revealed a set of toxins and virulence genes containing ribosome-inactivating protein toxin (RIP), OTU-like cysteine protease, ankyrin, and other bacterial toxin domains. They also performed a genus-wide comparative analysis of toxin/virulence-related domains between vertically transmitted and non-vertically transmitted strains, then identified a conserved core of toxin domains that is specific to the vertically transmitted strains. Kakizawa et al. highlighted the diversity of *Spiroplasma* as a group of facultative symbiotic bacteria of insects. The authors surveyed diverse stinkbugs representing 13 families, 69 genera, 97 species and 468 individuals, and detected *Spiroplasma* infection from four families (30.8%), seven genera (10.1%), 11 species (11.3%) and 21 individuals (4.5%). Phylogenetically, the stinkbug-associated *Spiroplasma* symbionts were placed in three distinct clades in the Spiroplasmataceae, confirming multiple and dynamic evolutionary trajectories of the stinkbug-*Spiroplasma* associations.

As for *Mycoplasma*, Mizutani, Sasajima et al. contributed a sophisticated biophysical work on the bacterial unique motility. *Mycoplasma pneumoniae*, known as the causative agent of mycoplasma pneumonia, binds to sialylated oligosaccharides and glides on host cell surface, which is essential for initiating the infection process (Miyata and Hamaguchi, 2016). The authors measured the stall force and the gliding speed of each *M. pneumoniae* cell carrying a bead that was manipulatable using optical tweezers. From the measurements of the faster strain M129 and the slower strain FH, detailed parameters of the gliding motility were measured. These results provide fundamental parameters underlying the bacterial gliding movement. Ras et al. conducted comparative genomic survey of *Mycoplasma* spp. for biosynthesis pathway genes of coenzyme A. The authors showed that most *Mycoplasma* genomes retain the genetic capacity to synthesize coenzyme A, but there was a differentiated prevalence of these genes across species. The final enzyme gene in the biosynthesis pathway encoding dephospho-coenzyme A kinase was found to be the most common among the studied *Mycoplasma* genomes. Yang et al. contributed a mini review about a comprehensive summary of lncRNAs and the responses of host cells associated with *M. pneumoniae* infections. With increasing literature on this topic, the review provided insights into the protective roles of lncRNAs against various forms of *M. pneumoniae* infections. They also discussed the involvement of lncRNAs in cardiovascular diseases, neurological disorders, cancers, and diabetes. Despite identifying key lncRNAs linked to *M. pneumoniae* pneumonia, their biological roles and mechanisms remain largely unknown. The authors emphasize the importance of understanding the molecular mechanisms of these lncRNAs.

As for Phytoplasma, two research articles provided technological advancements to facilitate the whole genome sequencing of these uncultivated pathogens from the infected plant samples. Tan et al. described an immunoprecipitation-based method for enriching the phytoplasma cells prior to DNA extraction. Zhang et al. described another enrichment method for phytoplasma cells based on serial filtration, as well as removal of host DNA using DNase I prior to the lysis of bacterial cells. By combining short read sequencing based on the Illumina platform and long read sequencing based on the Oxford Nanopore Technologies platform, both methods are effective for obtaining complete assemblies of these repeat-rich and compositionally biased genomes from metagenomic sequencing. It is worth noting that another enrichment protocol using iodixanol density gradients was developed recently and also shown to be effective (Jardim et al., 2022). In the third phytoplasma research article in this Research Topic, Huang et al.

reported the first complete genome sequence of "*Candidatus* Phytoplasma luffae" and found that a pair of 75 kb repeats and at least 13 potential mobile units (PMUs) account for ∼25% of this 769 kb chromosome. PMUs are phytoplasma-specific mobile genetic elements that often associate with effector genes (Bai et al., 2006, 2009) and likely contribute to the horizontal transfer of these virulence factors (Chung et al., 2013). In this new work, a genus-wide analysis of PMUs established a classification scheme of these mobile elements and identified strong correlations between PMU abundance and genome size at both within- and between-species levels.

In conclusion, the Research Topic presents a valuable overview of the current research coverage on *Spiroplasma*, *Mycoplasma*, and Phytoplasma, which encompasses such diverse areas as microbial diversity, genomics, reproductive manipulation, defense, toxins, biophysics and synthetic biology. These contributions highlight a variety of interesting biological phenomena observed with this wall-less, fastidious, and host-associated bacterial group. We hope that this Research Topic would provide some insight into what directions are promising in our future studies.

Author contributions

TF: Writing – original draft, Writing – review & editing. SK: Writing – review & editing. TH: Writing – review & editing. AS: Writing – review & editing. C-HK: Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. TF, SK, and TH were funded by the Japan Science and Technology Agency (JST) ERATO grant no. JPMJER1902. AS was funded by Plant2Pro-2022-R2V2 and ANR-23-CE20-0040-01 Greenpeas. C-HK was funded by Academia Sinica. AS was funded by INRAE and Agence Nationale de la Recherche (ANR).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Publisher's note

References

Anbutsu, H., and Fukatsu, T. (2011). *Spiroplasma* as a model insect endosymbiont. *Environ. Microbiol. Rep.* 3, 144–153. doi: 10.1111/j.1758-2229.2010.00240.x

Bai, X., Correa, V. R., Toruño, T. Y., Ammar, el-D., Kamoun, S., and Hogenhout, S. A. (2009). AY-WB phytoplasma secretes a protein that targets plant cell nuclei. *Mol. Plant Microbe Interact.* 22, 18–30. doi: 10.1094/MPMI-22-1-0018

Bai, X., Zhang, J., Ewing, A., Miller, S. A., Jancso Radek, A., Shevchenko, D. V., et al. (2006). Living with genome instability: the adaptation of phytoplasmas to diverse environments of their insect and plant hosts. *J. Bacteriol.* 188, 2682–3696. doi: 10.1128/JB.188.10.3682-3696.2006

Ballinger, M. J., and Perlman, S. J. (2019). The defensive *Spiroplasma*. *Curr. Opin. Insect Sci.* 32, 36–41. doi: 10.1016/j.cois.2018.10.004

Bertaccini, A., Arocha-Rosete, Y., Contaldo, N., Duduk, B., Fiore, N., Montano, H. G., et al. (2022). Revision of the '*Candidatus* Phytoplasma' species description guidelines. *Int J. Syst. Evol. Microbiol.* 72:5353. doi: 10.1099/ijsem.0.005353

Brown, D. R., May, M., Bradbury, J. M., and Johansson K. E. (2018). *Mollicutes. Bergey's Manual of Systematics of Archaea and Bacteria.* Hoboken, NJ: Wiley. doi: 10.1002/9781118960608.cbm00048.pub2

Chung, W. C., Chen, L. L., Lo, W. S., Lin, C. P., and Kuo, C. H. (2013). Comparative analysis of the peanut witches'-broom phytoplasma genome reveals horizontal transfer of potential mobile units and effectors. *PLoS ONE* 8:e62770. doi: 10.1371/journal.pone.0062770

Gasparich, G. E., Kuo, C. H., and Foissac, X. (2020). *Spiroplasma. Bergey's Manual of Systematics of Archaea and Bacteria. Hoboken, NJ: Wiley.* doi: 10.1002/9781118960608.gbm01262.pub2a

Harumoto, T., and Lemaitre, B. (2018). Male-killing toxin in a bacterial symbiont of *Drosophila*. *Nature* 557, 252–255. doi: 10.1038/s41586-018-0086-2

Hogenhout, S. A., Oshima, K., Ammar, E.-D., Kakizawa, S., Kingdom, H. N., and Namba, S. (2008). Phytoplasmas: bacteria that manipulate plants and insects. *Mol. Plant Pathol.* 9, 403–423. doi: 10.1111/j.1364-3703.2008.00472.x

Hurst, G. D. D., and Frost, C. L. (2015). Reproductive parasitism: maternally inherited symbionts in a biparental world. *Cold Spring Harb. Perspect. Biol.* 7:a017699. doi: 10.1101/cshperspect.a017699

Jardim, B. R., Tran-Nguyen, L. T. T., Gambley, C., Rodoni, B., and Constable, F. E. (2022). Iodixanol density gradients as an effective phytoplasma enrichment approach to improve genome sequencing. *Front. Microbiol.* 13:937648. doi: 10.3389/fmicb.2022.937648

Lo, W. S., Huang, Y. Y., and Kuo, C. H. (2016). Winding paths to simplicity: genome evolution in facultative insect symbionts. *FEMS Microbiol. Rev.* 40, 855–874. doi: 10.1093/femsre/fuw028

Miyata, M., and Hamaguchi, T. (2016). Integrated information and prospects for gliding mechanism of the pathogenic bacterium *Mycoplasma pneumoniae*. *Front. Microbiol.* 7:960. doi: 10.3389/fmicb.2016.00960

Miyata, M., Robinson, R. C., Uyeda, T. Q. P., Fukumori, Y., Fukushima, S. I., Haruta, S., et al. (2020). Tree of motility - a proposed history of motility systems in the tree of life. *Genes Cells* 25, 6–21. doi: 10.1111/gtc.12737

Teodoro, G. D., Marruchella, G., Di Provvido, A., D'Angelo, A. R., Orsini, G., Di Giuseppe, P., et al. (2020). Contagious bovine pleuropneumonia: a comprehensive overview. *Vet. Pathol.* 57, 476–489. doi: 10.1177/0309858209 21818

Venter, J. C., Glass, J. I., Hutchison, C. A. 3rd, and Vashee, S. (2022). Synthetic chromosomes, genomes, viruses, and cells. *Cell* 185, 2708–2724. doi: 10.1016/j.cell.2022.06.046

Waites, K. B., and Talkington, D. F. (2004). *Mycoplasma pneumoniae* and its role as a human pathogen. *Clin. Microbiol. Rev.* 17, 697–728. doi: 10.1128/cmr.17.4.697-728.2004

Yus, E., Maier, T., Michalodimitrakis, K., van Noort, V., Yamada, T., Chen, W. H., et al. (2009). Impact of genome reduction on bacterial metabolism and its regulation. *Science* 326, 1263–1268. doi: 10.1126/science.1177263

frontiers in Microbiology

ORIGINAL RESEARCH
published: 24 September 2021
doi: 10.3389/fmicb.2021.747905

Force and Stepwise Movements of Gliding Motility in Human Pathogenic Bacterium *Mycoplasma pneumoniae*

Masaki Mizutani[1†], Yuya Sasajima[1] and Makoto Miyata[1,2]*

[1] Graduate School of Science, Osaka City University, Osaka, Japan, [2] The OCU Advanced Research Institute for Natural Science and Technology (OCARINA), Osaka City University, Osaka, Japan

Mycoplasma pneumoniae, a human pathogenic bacterium, binds to sialylated oligosaccharides and glides on host cell surfaces *via* a unique mechanism. Gliding motility is essential for initiating the infectious process. In the present study, we measured the stall force of an *M. pneumoniae* cell carrying a bead that was manipulated using optical tweezers on two strains. The stall forces of M129 and FH strains were averaged to be 23.7 and 19.7 pN, respectively, much weaker than those of other bacterial surface motilities. The binding activity and gliding speed of the M129 strain on sialylated oligosaccharides were eight and two times higher than those of the FH strain, respectively, showing that binding activity is not linked to gliding force. Gliding speed decreased when cell binding was reduced by addition of free sialylated oligosaccharides, indicating the existence of a drag force during gliding. We detected stepwise movements, likely caused by a single leg under 0.2-0.3 mM free sialylated oligosaccharides. A step size of 14-19 nm showed that 25-35 propulsion steps per second are required to achieve the usual gliding speed. The step size was reduced to less than half with the load applied using optical tweezers, showing that a 2.5 pN force from a cell is exerted on a leg. The work performed in this step was 16-30% of the free energy of the hydrolysis of ATP molecules, suggesting that this step is linked to the elementary process of *M. pneumoniae* gliding. We discuss a model to explain the gliding mechanism, based on the information currently available.

Keywords: motility, optical tweezers, class *Mollicutes*, infection, sialic acid

OPEN ACCESS

Edited by:
Chih-Horng Kuo, Academia Sinica, Taiwan

Reviewed by:
Masayoshi Nishiyama, Kindai University, Japan
Ramanujam Srinivasan, National Institute of Science Education and Research (NISER), India

****Correspondence:***
Makoto Miyata
miyata@osaka-cu.ac.jp

†Present address:
Masaki Mizutani, Bioproduction Research Institute, National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba, Japan

Specialty section:
This article was submitted to Microbial Symbioses, a section of the journal Frontiers in Microbiology

Received: *27 July 2021*
Accepted: *24 August 2021*
Published: *24 September 2021*

Citation:
Mizutani M, Sasajima Y and Miyata M (2021) Force and Stepwise Movements of Gliding Motility in Human Pathogenic Bacterium Mycoplasma pneumoniae. Front. Microbiol. 12:747905. doi: 10.3389/fmicb.2021.747905

INTRODUCTION

Members of the bacterial class *Mollicutes*, which includes the genus *Mycoplasma*, are parasitic and occasionally commensal bacteria that are characterized by small cells and genomes and by the absence of a peptidoglycan layer (Razin et al., 1998; Razin and Hayflick, 2010). *Mycoplasma* species bind to host cell surfaces and exhibit gliding motility to spread the infectious area. Interestingly, *Mycoplasma* gliding does not involve flagella or pili and is completely unrelated to other bacterial motility systems or conventional motor proteins that are common in eukaryotic motility (Miyata et al., 2020). The gliding motility of *Mycoplasma* is divided into two types, *Mycoplasma pneumoniae* and *Mycoplasma mobile*, which share no homology in component proteins, indicating independent mechanisms (Miyata and Hamaguchi, 2016a,b).

Mycoplasma pneumoniae is a human pathogen that causes respiratory diseases, including bronchitis and atypical pneumonia (Saraya, 2017). *M. pneumoniae* cells bind to human epithelial surfaces through sialylated oligosaccharides (SOs), which are major structures of animal cell surfaces related to cell-cell recognition, and the binding targets of many pathogens and toxins (Nagai and Miyata, 2006; Varki, 2008; Kasai et al., 2013; Williams et al., 2018). The cells show unidirectional gliding motility at a speed of up to 1 μm/s on SO-coated glass surfaces (**Figure 1A**; Nakane and Miyata, 2009; Miyata and Hamaguchi, 2016a), which is known to be essential for their infection (Prince et al., 2014). The gliding machinery, called the "attachment organelle," is localized at a cell pole (Seto et al., 2001). The attachment organelle is divided into two parts: internal and surface structures. The internal structure is composed of an internal core complex and a surrounding translucent area. The internal core comprises three parts: a terminal button, paired plates, and a bowl complex from the front side of the cell (**Figure 1B**; Nakane et al., 2015; Kawamoto et al., 2016; Miyata and Hamaguchi, 2016a). The major surface structure, called "P1 adhesin complex" or "genitalium and pneumoniae cytoadhesin (GPCA)," is composed of P1 adhesin and P40/P90 proteins and aligned around the internal structure, which plays a dual role as the adhesin to bind to SOs and as the leg for gliding (**Figure 1B**; Nakane et al., 2011; Aparicio et al., 2018, 2020; Vizarraga et al., 2020, 2021). The model for gliding called the "Inchworm model" or "Double-spring hybrid ratchet model" is proposed, in which cells repeat the extensions and contractions of the attachment organelle based on the energy from ATP hydrolysis to enable smooth gliding (Miyata, 2008; Kawamoto et al., 2016; Seybert et al., 2018; Mizutani and Miyata, 2019). Generally, the mechanical characteristics and detailed analysis of movements are essential for creating and completing a detailed model for the motility mechanism (Schnitzer and Block, 1997; Veigel et al., 1999; Tanaka et al., 2002; Mallik et al., 2004; Sowa et al., 2005). However, to date, no information is available about the force for gliding.

In the present study, we measured the stall forces of the two strains and discuss the relationship between binding and force. Furthermore, we succeeded in detecting and measuring stepwise movements that are likely linked to the elementary process of the gliding reaction.

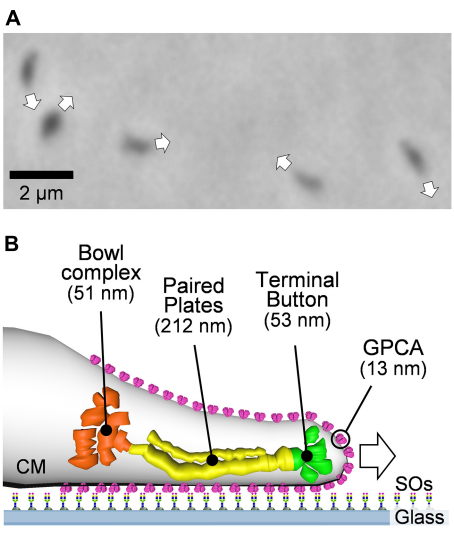


FIGURE 1 | Gliding motility of *Mycoplasma pneumoniae*. **(A)** Phase-contrast micrograph of *M. pneumoniae* cells. The gliding direction of cells are indicated by white arrows. **(B)** Illustration of gliding machinery. The gliding machinery is composed of bowl complex, paired plates, terminal button, and GPCA. The lengths along the cell axis were summarized from previous studies (Nakane et al., 2015; Vizarraga et al., 2020). GPCA, working as a leg is anchored to the cell membrane (CM) and catches sialylated oligosaccharides (SOs) fixed on solid surfaces.

MATERIALS AND METHODS

Strains and Cultivation

Mycoplasma pneumoniae M129 (ATCC29342) and FH strains were grown in SP-4 medium at 37°C in tissue culture flasks (TPP Techno Plastic Products AG, Trasadingen, Switzerland), as described previously (Tully, 1983; Nakane and Miyata, 2009). The FH strain was kindly provided by Tsuyoshi Kenri at the National Institute of Infectious Diseases, Tokyo, Japan.

Optical Tweezers System

An inverted microscope (IX71; Olympus, Tokyo, Japan) was equipped with a Nd:YAG laser (ASF1JE01; Furukawa Electric, Tokyo, Japan) to construct the optical tweezers. The microscope stage was replaced by a piezoelectric stage controlled by a stage controller (MDR14-CA-2.5; SIGMAKOKI, Tokyo, Japan) and a joystick (JS-300; SIGMAKOKI). The irradiated laser beam was concentrated as a finite optical system using a plano-convex lens supported by "optical cage system" (SIGMAKOKI). The concentrated laser beam was inserted into the microscope and focused by an objective lens (CFI Apochromat TIRF 100XC Oil; Nikon, Tokyo, Japan). The actual laser power was measured using a power meter (FieldMaxII; COHERENT, Santa Clara, CA, United States) without the objective lens.

Force Measurements

The cell suspension was mixed with 0.5 mM Sulfo-NHS-LC-LC-biotin (Thermo Fisher Scientific, Waltham, MA, United States) as the final concentration and incubated for 15 min at room temperature (RT). The cell suspension was centrifuged at 12,000 × *g* for 10 min, washed with 10 mM HEPES buffer (pH 7.4) containing 150 mM NaCl, centrifuged again, washed with HEPES buffer containing 10% non-heat-inactivated horse serum (Gibco; Thermo Fisher Scientific) and 20 mM glucose, passed through a 0.45-μm pore size filter and incubated for 15 min at RT. The cell suspension was inserted into a tunnel chamber, which was assembled by taping coverslips cleaned with saturated ethanolic KOH and precoated with 100% non-heat-inactivated horse serum for 60 min and 10 mg/ml bovine serum albumin (Sigma-Aldrich, St. Louis, MO, United States) in HEPES buffer for 60 min at RT. The tunnel chamber was washed with HEPES buffer containing 20 mM glucose and

incubated at 37°C on optical tweezers equipped with a thermo plate (MATS-OTOR-MV; Tokai Hit, Shizuoka, Japan) and a lens heater (MATS-LH; Tokai Hit). Avidin-conjugated beads in the HEPES buffer containing 20 mM glucose and 0.2-0.3 mM of 3′-*N*-acetylneuraminyllactose (SL) were sonicated and inserted into the tunnel chamber. The avidin-conjugated beads were prepared as previously described (Mizutani and Miyata, 2017). The bead movements were recorded using a charge-coupled device (CCD) camera (LRH2500XE-1; DigiMo, Tokyo, Japan) at 30 frames per second and analyzed by displacement of up to 200 nm from the trap center (the linear range of the laser trap) using ImageJ 1.43u[1] and IGOR Pro 6.33 J and 8.02 J (WaveMetrics, Portland, OR, United States) (Tanaka et al., 2016; Mizutani and Miyata, 2017; Mizutani et al., 2018). Measurements were performed using at least five individual cultures.

Binding and Gliding Analyses

Cultured cells were washed with buffer in a culture flask, and then washed with HEPES buffer containing 10% non-heat-inactivated horse serum (Gibco; Thermo Fisher Scientific) and 20 mM glucose. Cultured cells were scraped off the culture flask, passed through a 0.45-μm pore size filter, and incubated for 15 min at RT. The cell suspension was then inserted into the tunnel chamber. The tunnel chamber was washed with HEPES buffer containing 20 mM glucose, incubated at 37°C on an inverted microscope (IX83; Olympus) equipped with a thermo plate and lens heater, observed by phase-contrast microscopy at 37°C, and recorded with a CCD camera (DMK 33UX174; The Imaging Source Asia Co., Ltd., Taipei City, Taiwan). Then, the tunnel chamber was washed with HEPES buffer containing 20 mM glucose and various concentrations of SL. Video data were analyzed using ImageJ 1.43u and IGOR Pro 6.33 J. Measurements were performed using at least three individual cultures.

Genome Sequencing and Variant Analyses

Frozen stocks of cells were plated on Aluotto medium and isolated as previously described (Tulum et al., 2014). Genomic DNA was isolated and sequenced using MiSeq (Illumina, San Diego, CA, United States), as previously described (Mizutani et al., 2018). Sequence read mapping and variant detection were performed using the CLC Genomics Workbench (QIAGEN, Hilden, Germany).

RESULTS

Differences in Amino Acid Sequences of Gliding Proteins Between Two Strains

In this study, mainly we focused on a type strain "M129-B7," and also another major strain called "FH." The FH strain used in this study has not been genomically analyzed. Therefore, we sequenced the genomes of both strains using MiSeq (**Supplementary Data Sheet 1**) and analyzed the sequences of 14 genes that have been reported to be involved in binding and gliding (Miyata and Hamaguchi, 2016a). Only one amino acid was substituted in the M129 strain genome from the M129-B7 genome (GenBank accession no. CP003913), which was V196A in the HMW3 protein, which is positioned at the terminal button in the attachment organelle (**Figure 1B**). Our FH strain had 161 and 155 variations from the reported FH and FH2009 strains, respectively (GenBank accession no. CP010546 and CP017327), indicating that the FH strain is distant from the genome strains. The differences between the two strains analyzed in the present study in terms of the 14 genes were as follows: (i) P1 adhesin showed many differences, including 87 single amino acid substitutions, 18 amino acid insertions, and four amino acid deletions, known differences of FH from M129 strains (**Supplementary Figure 1**). (ii) The P40/P90 protein showed 96 single amino acid substitutions, one amino acid insertion, and 68 amino acid deletions (**Supplementary Figure 2**). (iii) The other 12 genes showed 0-5 mutations in each gene, as summarized in **Supplementary Table 1**.

Stall Force Measurement Using Optical Tweezers

Optical tweezers are commonly used to measure the stall force generated by pili in bacterial motility or motor proteins in eukaryotic motility (Kojima et al., 1997; Merz et al., 2000; Takagi et al., 2006; Gennerich et al., 2007). The stall force is defined as the force needed to stop movements and is equal to the maximal propulsion force for locomotion. Previously, we measured the stall force for the gliding motility of *M. mobile*, based on a mechanism unrelated to *M. pneumoniae* gliding, using optical tweezers (Miyata et al., 2002; Mizutani et al., 2018). In the present study, we applied this method to *M. pneumoniae* gliding. *M. pneumoniae* cells were biotinylated and inserted into a tunnel chamber, which was assembled using two glass plates and double-sided tape (Mizutani and Miyata, 2017). An avidin-conjugated polystyrene bead was trapped by a highly focused laser beam and attached to a gliding cell at the back end of the cell body by exploiting the avidin-biotin interaction. The cells pulled the bead from the trap center with gliding and then stalled (**Figures 2A,B** and **Supplementary Video 1**). The force was calculated by measuring the distance between the centers of the bead and the laser trap, which was multiplied by the trap stiffness; the force acting on the bead increased linearly with the displacement from the trap center (Kojima et al., 1997). Starting from 0 s, the pulling force increased and reached a plateau in 120 s (**Figure 2C**). The maximal value of the force averaged over 1 s was determined as the stall force. The stall force of M129 strain cells was 23.7 ± 6.3 pN (**Figure 2E**). We also measured the stall force of the FH strain. The cells of the FH strain pulled the bead in a manner similar to that of M129 cells (**Supplementary Video 2**). The stall force of the FH strain was 19.7 ± 5.3 pN, significantly weaker than that of the M129 strain (**Figures 2D,E**).

[1] http://rsb.info.nih.gov/ij/

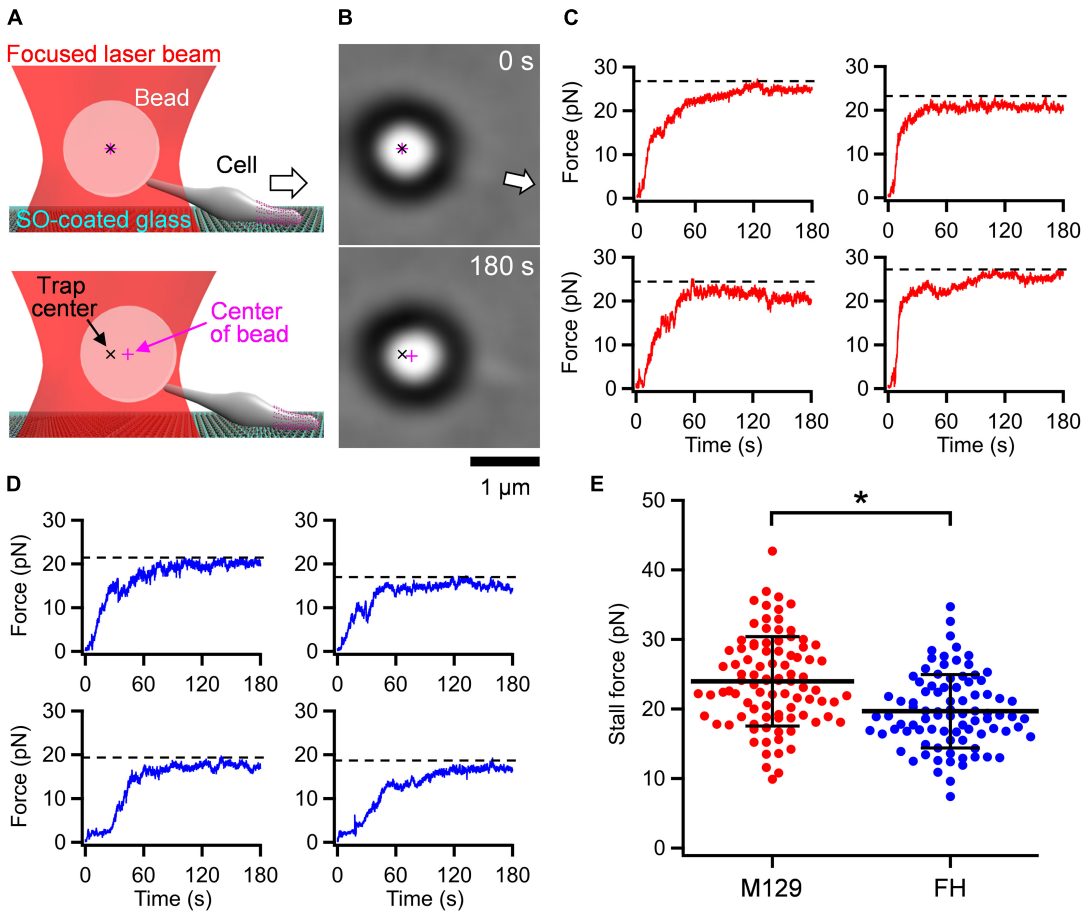


FIGURE 2 | Stall force measurements. **(A)** Illustrations of experimental design for force measurements. The white arrow indicates the gliding direction. **(B)** Optical micrographs of trapped cell. The cell attached with a bead (large black ring with white center) glided in the direction of white arrow from 0 and stalled in 180 s. The cell image can be observed near the base of arrow. **(C)** Four representative time courses of force increments for M129 strain. The broken lines indicate the values of stall force. **(D)** Four representative time courses of force increments for FH strain. The broken lines indicate the values of stall force. **(E)** Scatter dot plot of stall force (n = 92 and 88 in M129 and FH, respectively) shown with averages (thick lines) and standard deviations (thin lines). *$p = 2.4 \times 10^{-6}$ by Student's t-test.

Effect of Free Sialyllactose on Binding and Gliding

Mycoplasma pneumoniae cells have hundreds of legs and glide smoothly on solid surfaces. To trace the behavior of a single propulsion event, the number of working legs should be reduced by adding a free form of SO, which is the binding target for gliding (Kasai et al., 2013; Kinosita et al., 2014; Mizutani et al., 2018). To quantitatively analyze the effects of free SO on the binding activity and gliding under our conditions, we measured the bound ratio of cells and the gliding speed of the M129 strain under various concentrations of the SO, 3′-*N*-acetylneuraminyllactose (SL). *M. pneumoniae* cells suspended in the buffer were inserted into a tunnel chamber. Then, the cell behavior in 0–0.5 mM SL solutions was analyzed. The addition of free SL slowed down and then stopped gliding or released the gliding cells from the glass surfaces, but did not release non-gliding cells, indicating that release requires glass binding by GPCA with displacements (**Figure 3A**). The number of gliding cells and the gliding speed relative to the initial speed were decreased by 0.1-0.5 mM with SL treatments from 55 ± 11% to 26 ± 16% and from 0.33 ± 0.06 to 0.19 ± 0.07 μm/s, respectively (**Figures 3B,C**). We then decided to use 0.2-0.3 mM concentrations for further experiments because the binding and gliding were partially inhibited under these conditions, which were advantageous for observing single propulsion events caused by a single leg during gliding.

Stepwise Gliding Movements Observed Under Free SL

Next, we traced cells under 0.2-0.3 mM free SL conditions as performed in the stall force measurements. The gliding cells slowly pulled the beads from the trap center. Half of the tested cells stalled in 120 s, but the other cells repeated creeping movements and detachment from the glass surfaces (**Figure 4A**). The stall force of the cells that reached a plateau in 120 s was 16.2 ± 4.5 pN (**Figure 4B**), which was significantly weaker than that without SL ($p = 1.1 \times 10^{-5}$ by Student's t-test). In creeping movements, cells occasionally showed discontinuous

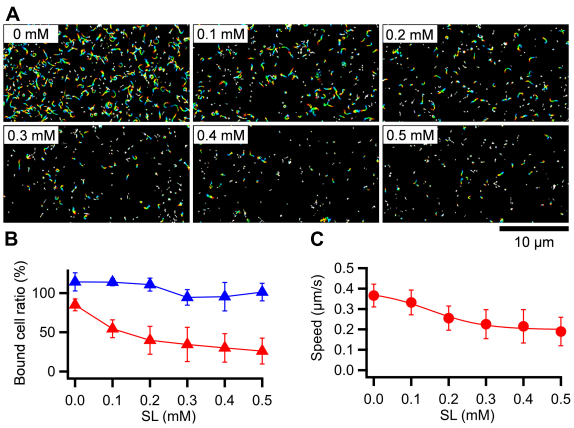


FIGURE 3 | Effects of SL on binding and gliding. **(A)** Gliding cells were treated with free SL. The cell trajectories after SL treatment are presented as a stack for 20 s, changing color from red to blue. SL concentrations are shown at left upper of each panel. **(B)** Number of cells bound to the glass at 5 min after the treatment are plotted relative to the initial cell number. The averages for gliding (red triangles) and non-gliding (blue triangles) cells are shown with the standard deviations (n = 10, 10, 10, 10, 9, 10 for 0, 0.1, 0.2, 0.3, 0.4, 0.5 mM, respectively). **(C)** Averaged gliding speeds under various SL concentrations are plotted with the standard deviations and fitted by a sigmoidal curve (n = 50, 50, 50, 50, 56, 56 for 0, 0.1, 0.2, 0.3, 0.4, 0.5 mM, respectively).

displacements, which were mostly stepwise (**Figures 4C,D**). Individual displacements in a stepwise time course were analyzed using the pairwise distance function (Kinosita et al., 2014; Mizutani et al., 2018). The displacements shown in **Figure 4D** were uniformly distributed at about 17-, 14-, and 16-nm intervals under 0.15, 0.20, and 0.17 pN/nm of trap stiffness, corresponding to 2.55, 2.80, and 2.66 pN of propulsion force, respectively (**Figures 4D,E**). Note that force increments are generally calculated from the trap stiffness × displacement (Kojima et al., 1997; Mizutani et al., 2018).

Next, to examine the load dependency of step sizes in detail, we measured step sizes under 0.07–0.26 pN/nm of trap stiffness. Ninety-seven steps in 36 cell trajectories, including at least two continuous steps, were detected. The step sizes under 0.13 to 0.26 pN/nm of trap stiffness linearly decreased from 21.0 to 8.0 nm with trap stiffness (**Figures 4F,H**). In contrast, the step sizes from 0.07 to 0.12 pN/nm of trap stiffness were mostly constant and distributed at 14-19 nm with an average of 16.2 ± 2.7 nm (n = 21) (**Figures 4G-I**).

From 0.07 to 0.12 pN/nm of trap stiffness, the force increments increased with the trap stiffness (**Figure 4I**). In the cases where the force values increase with trap stiffness, the calculated force does not reflect the actual force because the load is too small to influence the movements. Therefore, we focused on the force increments measured under 0.13–0.26 pN/nm of trap stiffness. The force exerted in a single propulsion step was concluded to be 2.5 ± 0.3 pN (n = 76) (**Figure 4J**).

Binding Activity and Gliding Motility

To discuss about the relationship between the binding activity and gliding speed, we analyzed them for the M129 and FH strains. *M. pneumoniae* cells were suspended in HEPES buffer containing 20 mM glucose to obtain an optical density at 595 nm of 0.07, then inserted into tunnel chambers. After incubation for 5–30 min, the tunnel chambers were washed and observed by phase-contrast microscopy. The number of bound cells in the M129 strain increased with time from 90 ± 16 to 308 ± 20 cells in 100 μm × 100 μm area from 5 to 30 min (**Figures 5A,B**). These values are consistent with the results of a previous report (Kasai et al., 2013). In contrast, the number of bound cells in the FH strain increased from 12 ± 2 to 36 ± 8 cells from 5 to 30 min (**Figures 5A,B**). The number of bound cells in the FH strain was 7.2-8.7-fold smaller than that of the M129 strain at all time points (**Figure 5B**). When 10 × concentrated cell suspension was examined, the bound cell numbers of the FH strain increased from 117 ± 10 to 386 ± 24 cells from 5 to 30 min, 1.2-1.3 times that of M129 at all time points (**Figures 5A,B**). These results indicate that the FH strain has approximately eightfold lower binding activity to SO-coated glass surfaces than the M129 strain. Considering that the stall force of FH was only 1.2-fold smaller than that of the M129 strain, the force is unlikely to be linked to binding activity.

Most of the bound cells showed gliding motility on the SO-coated glass (**Figure 5C** and **Supplementary Videos 3, 4**). To characterize gliding motility, the proportion of gliding cells and gliding speed were analyzed. The proportion of gliding cells to all bound cells was 78.1% (n_{total} = 2,284, n_{glide} = 1785) and 60.0% (n_{total} = 2,271, n_{glide} = 1,362) in the M129 and FH strains, respectively (**Figure 5C**). The gliding speeds averaged for 20 s at 1-s intervals were 0.47 ± 0.08 and 0.21 ± 0.07 μm/s for M129 and FH strains, respectively (**Figure 5D**). These results suggest that binding activity is not directly linked to gliding speed.

DISCUSSION

Mycoplasma pneumoniae Has Weak Stall Force

In the present study, we measured the stall force in *M. pneumoniae* gliding using optical tweezers. Previously, the stall force was measured for some bacterial surface motilities. *Neisseria gonorrhoeae* and *Myxococcus xanthus* show surface motility driven by the retraction of type IV pili (Pelicic, 2008). The stall forces of single-pilus retraction in *N. gonorrhoeae* and *M. xanthus* were measured by optical tweezers to be approximately 80 and 150 pN, respectively (Merz et al., 2000; Clausen et al., 2009). *M. mobile* glides up to 4.0 μm/s by a mechanism unrelated to that of *M. pneumoniae*. Previously, we measured the stall force of *M. mobile* gliding using optical tweezers to be approximately 113 pN (Mizutani et al., 2018). The stall forces of *M. pneumoniae* gliding were 23.7 ± 6.3 and 19.7 ± 5.3 pN in M129 and FH strains, respectively (**Figure 2E**), much weaker than those of other bacterial surface motilities. *M. pneumoniae* has a streamlined cell body about 0.2 μm in diameter (Hatchel and Balish, 2008). This shape and cell size may be beneficial for gliding in human tissues with weak forces.

In liquid culture, *M. pneumoniae* cells bind to the bottom surface of the tissue culture flask. This is distinct from the

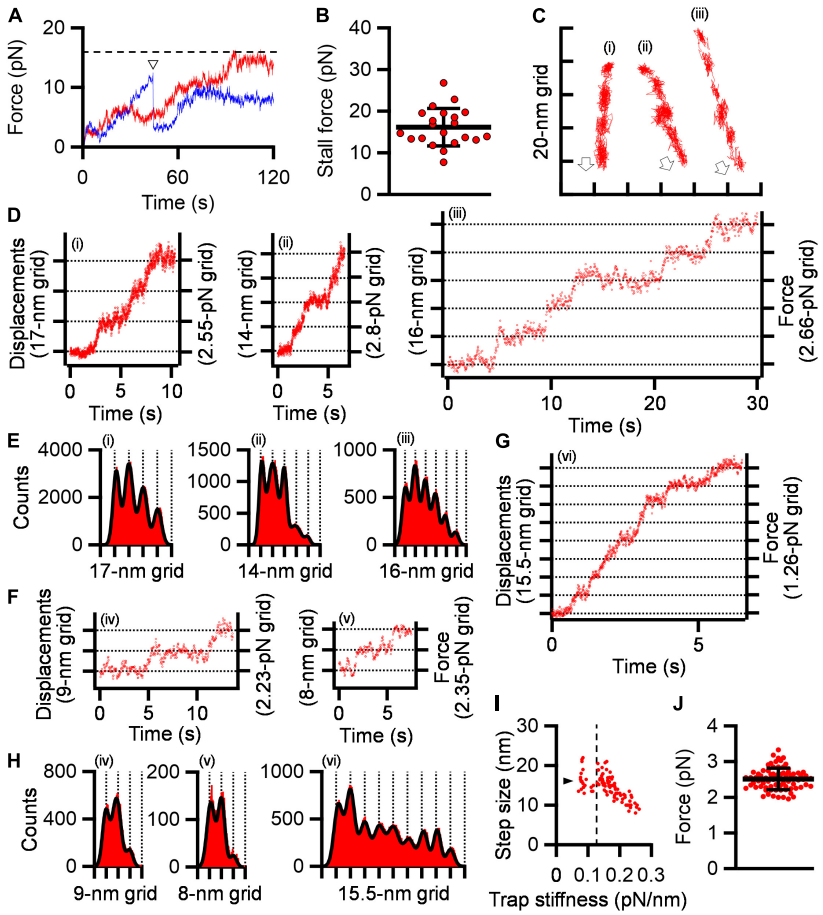


FIGURE 4 | Stepwise movements under SL. **(A)** Representative force traces with time under 0.2–0.3 mM SL are shown for stalled (red) and detached (blue) cells. The broken line indicates the stall force. The time point of detachment was marked by a triangle. **(B)** Scatter dot plot of stall force shown with averages (thick line) and standard deviations (thin lines) (n = 21). **(C)** Three cell trajectories with stepwise movements are shown in a field. Open arrows indicate the gliding direction. **(D)** Displacement and force of cells whose trajectories are shown in panel **(C)**. **(E)** Histograms of pairwise distance function (PDF) analysis of panel **(D)** were fitted by the sum of Gaussian curves. **(F)** Displacement and force under high trap stiffness. **(G)** Displacement and force under low trap stiffness. **(H)** Histograms of PDF analysis of panels **(F,G)**. **(I)** Distribution of step size under various trap stiffnesses. Step sizes were plotted as a function of trap stiffness (n = 97). The position at 0.12 pN/nm of trap stiffness was marked by a dotted line. The average of step size under 0.07–0.12 pN/nm of trap stiffness was marked by a triangle. **(J)** Scatter dot plot of force increments under 0.13–0.26 pN/nm of trap stiffness shown with averages (thick line) and standard deviations (thin lines) (n = 76).

case of *M. mobile*, in which most cells float in the medium. Therefore, *M. pneumoniae* is expected to have a stronger force for gliding than *M. mobile*. However, we found that *M. pneumoniae* had a much weaker stall force than *M. mobile* (**Figure 2E**). When comparing the two strains of *M. pneumoniae*, the FH strain showed much less active binding than, but a stall force comparable to, that of M129 (**Figures 2E**, **3B**). These facts indicate that the binding activity of *M. pneumoniae* cells is not determined by the gliding force, which is represented by the stall force.

Drag Force in Gliding

The gliding speed decreased when cell binding was partially inhibited by the addition of free SL (**Figure 3**). This observation is consistent with previous data; that is, the inhibition of binding by monoclonal antibodies decreased the gliding speed of *M. pneumoniae* (Seto et al., 2005) and the inhibition by SL decreased the speed of *Mycoplasma gallisepticum*, coinciding with the common mechanism with *M. pneumoniae* (Mizutani and Miyata, 2019). The decrease in speed was probably caused by the drag force generated from the substrate surface, because the friction force exerted from water is estimated to be more than 5,000 times smaller than the stall force of 24 pN (**Figure 2E**; Rosengarten et al., 1988; Uenoyama et al., 2004). As the cause of the drag force, two possibilities are considered: GPCA and others. If some proportion of GPCA molecules are not involved in gliding, they are not released from SOs by inhibitory factors, resulting in speed reduction. Interestingly, these observations and explanations are similar to the case of *M. mobile* gliding, even though they do not share the same structure of machinery (Uenoyama et al., 2004, 2009; Kasai et al., 2013; Miyata and Hamaguchi, 2016b; Nishikawa et al., 2019). This scheme may be advantageous for gliding on SOs based on ATP energy, which is common in the both gliding mechanisms.

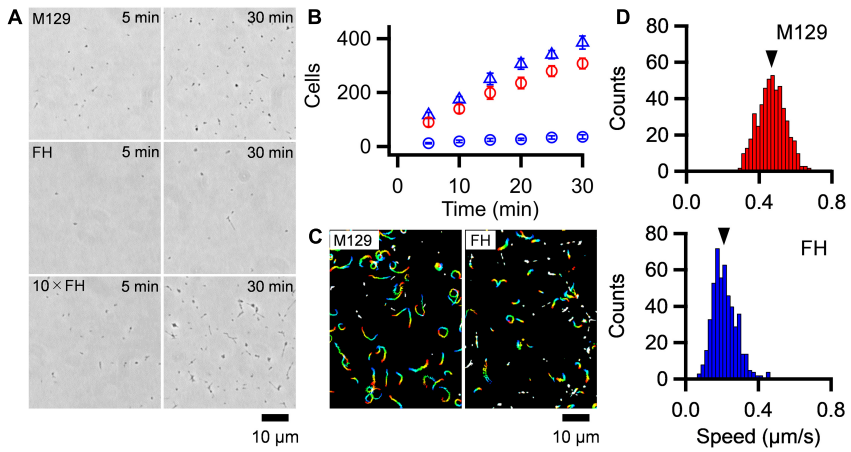


FIGURE 5 | Binding activity and gliding motility. **(A)** Phase-contrast micrographs of cells on SO-coated glass surfaces. M129, FH, and 10 × concentrated FH are shown after 5 and 30 min of incubations. **(B)** Averaged bound cell numbers in 100 μm × 100 μm plotted with standard deviations (n = 10 in each). M129 (Red open circle), FH (Blue open circle), and 10 × concentrated FH (Blue open triangle) are presented. **(C)** Cell trajectories for 20 s, changing color with time from red to blue. **(D)** Distributions of gliding speeds averaged for 20 s at 1-s intervals (n = 500 in each). The mean values are marked by black triangles.

Stepwise Movement as Elementary Process

We succeeded in detecting the stepwise movements of *M. pneumoniae* gliding. *M. pneumoniae* cells glided at a speed of 0.47 μm/s (**Figure 5D**), and the step size in the load-free condition was 14-19 nm (**Figure 4I**), suggesting 25-35 steps per second. The step size of *M. pneumoniae* is shorter than that of *M. mobile* around 70 nm (Kinosita et al., 2014, 2018; Mizutani et al., 2018). This difference is likely related to the lengths of leg complex, 13 nm in *M. pneumoniae* (Kenri et al., 2019; Aparicio et al., 2020; Vizarraga et al., 2021) and 97 nm in *M. mobile* (Miyata and Petersen, 2004; Adan-Kubo et al., 2006).

Stepwise movements are well-studied in ATP-driven eukaryotic motor proteins including myosin, dynein, and kinesin (Bustamante et al., 2004; Mallik et al., 2004). The force and displacement of a single step in myosin II, cytoplasmic dynein, kinesin-1, and myosin V have been reported as 3–5, 7–8, 8, and 2–3 pN and 5.3, 8, 8, and 36 nm, respectively (Kojima et al., 1997; Schnitzer and Block, 1997; Takagi et al., 2006; Gennerich et al., 2007; Fujita et al., 2012; Park and Lee, 2013). Stepwise movements are also present in bacterial motility. A flagellar motor shows 14 degrees of revolution as a step (Sowa et al., 2005; Nakamura et al., 2010). The gliding of *M. mobile* shows stepwise movements of 1.5 pN force and 70 nm length (Kinosita et al., 2014; Mizutani et al., 2018). Generally, stepwise movements are thought to correspond to the elementary process of a motility event.

Previously, we showed that *M. pneumoniae*-type gliding motility is driven by energy from ATP hydrolysis (Mizutani and Miyata, 2019). Motilities driven by ATP energy can be divided into elementary processes that are directly coupled with ATP hydrolysis, such as stepwise movements. The elementary processes observed as steps require a smaller amount of work than the energy produced by ATP hydrolysis, which is ∼80 pN nm (Yasuda et al., 1998). Therefore, we estimated the work done in the stepwise movements of *M. pneumoniae* gliding to determine this possibility. The work was estimated to be 18.2 ± 5.4 pN nm from the equation W_{step} = 0.5 × spring constant × displacement^2, under 0.13–0.26 pN/nm of trap stiffness, where the stiffness is large enough to determine the force (**Supplementary Figure 5**; Mizutani et al., 2018). This value is 16-30% of the free energy of the hydrolysis of ATP molecules, suggesting that we detected the elementary process of *M. pneumoniae* gliding. The energy conversion efficiencies of stepwise movements are 12-40, 10, and 40-60% for myosin II, cytoplasmic dynein, and kinesin, respectively (Bustamante et al., 2004; Mallik et al., 2004). Previously, we estimated it to be 10-40% for *M. mobile* gliding (Mizutani et al., 2018). These facts suggest that the energy efficiency of stepwise movements of *M. pneumoniae* is in a similar range of myosin II and *M. mobile* gliding, although we should be careful that here we measured a cell including many leg units, which is different from motor proteins even if we decreased the working numbers by adding SL.

Load-Dependent Step Size

M. pneumoniae cells showed different step sizes depending on the load provided by optical tweezers continuously (**Figure 4I**), suggesting that the step size of *M. pneumoniae* is load-dependent. In the infectious process, *M. pneumoniae* cells glide to the deep positions of respiratory systems and experience a large load at these positions (Prince et al., 2014). The load-dependent stepping behavior would be useful to glide against large loads because a load-independent stepping motor, kinesin, shows frequent back steps under large loads (Carter and Cross, 2006; Toleikis et al., 2020). Different step sizes under different loads have also been observed in *M. mobile* gliding (Kinosita et al., 2014; Mizutani et al., 2018).

Suggestion for Gliding Mechanism

How can we image the gliding mechanism? The GPCA probably plays a critical role in *M. pneumoniae* gliding, because antibodies

against the P1 adhesin, a component of GPCA, decreased the gliding speed and ultimately replaced the *M. pneumoniae* cells on the glass surface (Seto et al., 2005). Recently, the detailed structure of GPCA has been solved (Aparicio et al., 2018, 2020; Vizarraga et al., 2020, 2021). It is a mushroom structure composed of two P1 adhesins and two P40/P90 molecules. The C-terminal regions of the four molecules are bundled and anchored to the cell membrane. P40 and P90 proteins are synthesized as a single protein and processed into two proteins that are not observed in *M. genitalium*. Although P1 adhesin is believed to be the receptor for SOs, the binding site exists at the distal end of P40/P90. This complex is thought to undergo conformational changes between open and closed with respect to the binding pockets, which are likely involved in the gliding mechanism (Aparicio et al., 2020; Vizarraga et al., 2021).

The attachment organelle responsible for gliding can be divided into a surface structure including GPCA and an internal rod structure (Henderson and Jensen, 2006; Seybert et al., 2006, 2018; Kawamoto et al., 2016; Krause et al., 2018). Briefly, the force for gliding is likely generated at the bowl complex because a few proteins essential for gliding and not binding are localized there (Hasselbring et al., 2005; Jordan et al., 2007; Kawakita et al., 2016). The paired plates and elastic components play a role in force transmission, because the gliding speed decreases severely in a deletion mutant (Garcia-Morales et al., 2016). The terminal button likely connects the rod front to the cell membrane, because an end component, the P30 protein, features transmembrane segments (Chang et al., 2011; Relich and Balish, 2011).

Here, we focus on the following observations to construct the working model for the gliding scheme: (1) The force is probably generated around the bowl complex, transmits through the internal structures including paired plates, and reaches GPCAs. (2) Inhibition of cell binding decreases gliding speed. (3) GPCA has open and closed conformations. (4) Gliding can be divided into steps because binding and force are not tightly coupled. (5) The gliding movement can be divided into 14-19 nm steps with a 2.5 pN force.

Gliding Scheme to Explain Stepwise Movements

The model is composed of repeated cycle of four stages (i)-(iv) connected by four steps: "Power stroke," "Release," "Displacement," and "Catch" (**Figure 6**). (i) The GPCA in the open state catches an SO on the glass surface. (ii) Contraction of the internal structure pulls the cell body with 2.5 pN and a step size of 14-19 nm. This step is possibly linked to energy from an ATP molecule. (iii) GPCA switches into a closed state by triggering the pulling force transmitted from other working GPCAs, resulting in the release of after-stroke GPCA from the SO. (iv) The released GPCA returns to the open state and displaces another SO in the next position by the extension of the internal structure. (i) The GPCA captures the next SO. The full cycle was then repeated. This scheme explains the relationship between the binding and force of different strains, if we assume that release is enhanced and power stroke and catch are reduced in FH compared with those in M129, because force is determined by power stroke. The slow gliding speed in FH also can be explained by the reduction of catch step. These differences in the steps may be caused by the structural differences of the 14 proteins involved in binding and gliding. The characterization of defined genetic mutants may solve these problems. The observation that shorter steps occur under load can be explained if we consider that the power stroke is shortened by the load (**Figure 3I**).

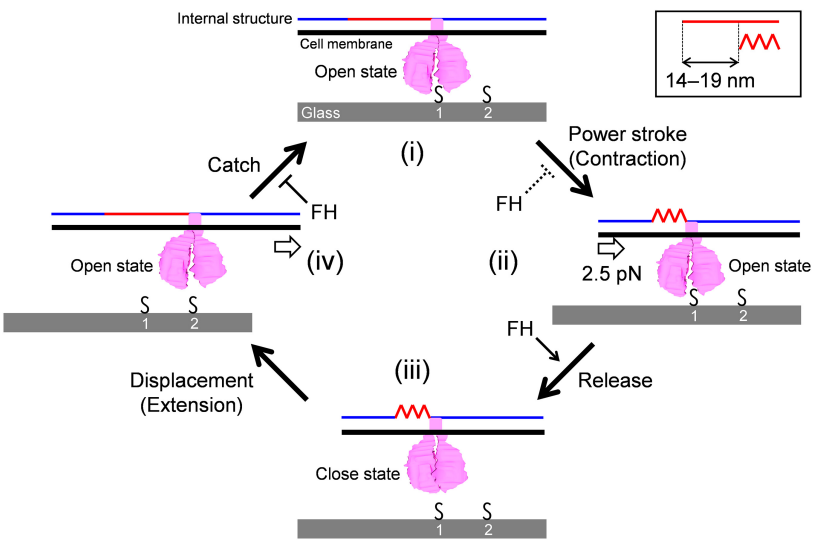


FIGURE 6 | Schematic illustration of leg behaviors in stepwise movements. Internal structure presented by blue and red lines is responsible for force transmission. It repeats contraction and extension cooperatively with the surface structures. Conformational change of internal structure and step size are shown in a right-upper box. GPCA is presented in pink. Sialylated oligosaccharides are shown by "S." Force exerted in a step is shown by an arrow in panel (ii). Displacement of cell body forward is shown by an arrow in panel (iv). The gliding occurs through stages (i)–(iv). See the text for details.

Previously, our group suggested a working model for *M. pneumoniae* gliding, focusing mainly on the information of the internal structure with regard to the attachment organelle (Kawamoto et al., 2016; Miyata and Hamaguchi, 2016a). The previous model suggested "directed detachment of feet" (GPCA), due to the lack of information about step, force, and foot structure. In this study, we succeeded in adding new information and completing an updated model.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in the **Supplementary Material**.

AUTHOR CONTRIBUTIONS

MMZ designed the study. MMZ and YS analyzed the genomes. MMZ performed all other experiments. MMZ and MMY wrote the manuscript, and all the authors completed it. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by Grants-in-Aid for Scientific Research: (A) (MEXT KAKENHI Grant Number JP17H01544) and by JST CREST (Grant Number JPMJCR19S5, Japan). MMZ and YS are the recipients of a Research Fellowship of the Japan Society for the Promotion of Science (18J15362, 21J15218).

ACKNOWLEDGMENTS

We would like to thank Tsuyoshi Kenri at the National Institute of Infectious Diseases and Ikuko Fujiwara at Osaka City University for helpful discussions, and Shigeyuki Kakizawa at the National Institute of Advanced Industrial Science and Technology (AIST) for supporting the genome sequence analyses.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2021.747905/full#supplementary-material

Supplementary Figure 1 | Multiple sequence alignments for P1 adhesin of M129 and FH strains. The symbols "*" ":" "." indicate fully conserved residue, conservation between groups of strongly similar properties, and conservation between groups of weakly similar properties, respectively.

Supplementary Figure 2 | Multiple sequence alignments for P40/P90 of M129 and FH strains. The symbols "*" ":" "." indicate fully conserved residue, conservation between groups of strongly similar properties, and conservation between groups of weakly similar properties, respectively.

Supplementary Figure 3 | Work performed by stepwise movements. The scatter dot plot of works calculated from individual steps is shown with average (thick line) and standard deviation (thin lines).

Supplementary Table 1 | Amino acid variations of gliding related proteins in M129 and FH strains.

Supplementary Video 1 | Stall force measurement in M129 strain. A polystyrene bead was attached to the back end of cell body. The cell pulled the bead from trap center of optical tweezers. The video was played at 5 × speed.

Supplementary Video 2 | Stall force measurement in FH strain. A polystyrene bead was attached to the back end of cell body. The cell pulled the bead from trap center of optical tweezers. The video was played at 5 × speed.

Supplementary Video 3 | Gliding movement of M129 strain cells. Cells bound to the SO-coated glass surface were observed by phase-contrast microscopy. The video was played at 5 × speed.

Supplementary Video 4 | Gliding movement of FH strain cells. Cells bound to the SO-coated glass surface were observed by phase-contrast microscopy. The video was played at 5 × speed.

REFERENCES

Adan-Kubo, J., Uenoyama, A., Arata, T., and Miyata, M. (2006). Morphology of isolated Gli349, a leg protein responsible for *Mycoplasma mobile* gliding via glass binding, revealed by rotary shadowing electron microscopy. *J. Bacteriol.* 188, 2821–2828. doi: 10.1128/JB.188.8.2821-2828.2006

Aparicio, D., Scheffer, M. P., Marcos-Silva, M., Vizarraga, D., Sprankel, L., Ratera, M., et al. (2020). Structure and mechanism of the Nap adhesion complex from the human pathogen *Mycoplasma genitalium*. *Nat. Commun.* 11:2877. doi: 10.1038/s41467-020-16511-2

Aparicio, D., Torres-Puig, S., Ratera, M., Querol, E., Pinol, J., Pich, O. Q., et al. (2018). *Mycoplasma genitalium* adhesin P110 binds sialic-acid human receptors. *Nat. Commun.* 9:4471. doi: 10.1038/s41467-018-06963-y

Bustamante, C., Chemla, Y. R., Forde, N. R., and Izhaky, D. (2004). Mechanical processes in biochemistry. *Annu. Rev. Biochem.* 73, 705–748. doi: 10.1146/annurev.biochem.72.121801.161542

Carter, N. J., and Cross, R. A. (2006). Kinesin's moonwalk. *Curr. Opin. Cell Biol.* 18, 61–67. doi: 10.1016/j.ceb.2005.12.009

Chang, H. Y., Jordan, J. L., and Krause, D. C. (2011). Domain analysis of protein P30 in *Mycoplasma pneumoniae* cytadherence and gliding motility. *J. Bacteriol.* 193, 1726–1733. doi: 10.1128/JB.01228-10

Clausen, M., Jakovljevic, V., Sogaard-Andersen, L., and Maier, B. (2009). High-force generation is a conserved property of type IV pilus systems. *J. Bacteriol.* 191, 4633–4638. doi: 10.1128/JB.00396-09

Fujita, K., Iwaki, M., Iwane, A. H., Marcucci, L., and Yanagida, T. (2012). Switching of myosin-V motion between the lever-arm swing and brownian search-and-catch. *Nat. Commun.* 3:956. doi: 10.1038/ncomms1934

Garcia-Morales, L., Gonzalez-Gonzalez, L., Querol, E., and Pinol, J. (2016). A minimized motile machinery for *Mycoplasma genitalium*. *Mol. Microbiol.* 100, 125–138. doi: 10.1111/mmi.13305

Gennerich, A., Carter, A. P., Reck-Peterson, S. L., and Vale, R. D. (2007). Force-induced bidirectional stepping of cytoplasmic dynein. *Cell* 131, 952–965. doi: 10.1016/j.cell.2007.10.016

Hasselbring, B. M., Jordan, J. L., and Krause, D. C. (2005). Mutant analysis reveals a specific requirement for protein P30 in *Mycoplasma pneumoniae* gliding motility. *J. Bacteriol.* 187, 6281–6289. doi: 10.1128/JB.187.18.6281-6289.2005

Hatchel, J. M., and Balish, M. F. (2008). Attachment organelle ultrastructure correlates with phylogeny, not gliding motility properties, in *Mycoplasma pneumoniae* relatives. *Microbiology* 154, 286–295. doi: 10.1099/mic.0.2007/012765-0

Henderson, G. P., and Jensen, G. J. (2006). Three-dimensional structure of *Mycoplasma pneumoniae*'s attachment organelle and a model for its role in

gliding motility. *Mol. Microbiol.* 60, 376–385. doi: 10.1111/j.1365-2958.2006.05113.x

Jordan, J. L., Chang, H. Y., Balish, M. F., Holt, L. S., Bose, S. R., Hasselbring, B. M., et al. (2007). Protein P200 is dispensable for *Mycoplasma pneumoniae* hemadsorption but not gliding motility or colonization of differentiated bronchial epithelium. *Infect. Immun.* 75, 518–522. doi: 10.1128/IAI.01344-06

Kasai, T., Nakane, D., Ishida, H., Ando, H., Kiso, M., and Miyata, M. (2013). Role of binding in *Mycoplasma mobile* and *Mycoplasma pneumoniae* gliding analyzed through inhibition by synthesized sialylated compounds. *J. Bacteriol.* 195, 429–435. doi: 10.1128/JB.01141-12

Kawakita, Y., Kinoshita, M., Furukawa, Y., Tulum, I., Tahara, Y. O., Katayama, E., et al. (2016). Structural study of MPN387, an essential protein for gliding motility of a human-pathogenic bacterium, *Mycoplasma pneumoniae*. *J. Bacteriol.* 198, 2352–2359. doi: 10.1128/JB.00160-16

Kawamoto, A., Matsuo, L., Kato, T., Yamamoto, H., Namba, K., and Miyata, M. (2016). Periodicity in attachment organelle revealed by electron cryotomography suggests conformational changes in gliding mechanism of *Mycoplasma pneumoniae*. *MBio* 7, e00243–16. doi: 10.1128/mBio.00243-16

Kenri, T., Kawakita, Y., Kudo, H., Matsumoto, U., Mori, S., Furukawa, Y., et al. (2019). Production and characterization of recombinant P1 adhesin essential for adhesion, gliding, and antigenic variation in the human pathogenic bacterium, *Mycoplasma pneumoniae*. *Biochem. Biophys. Res. Commun.* 508, 1050–1055. doi: 10.1016/j.bbrc.2018.11.132

Kinosita, Y., Miyata, M., and Nishizaka, T. (2018). Linear motor driven-rotary motion of a membrane-permeabilized ghost in *Mycoplasma mobile*. *Sci. Rep.* 8:11513. doi: 10.1038/s41598-018-29875-9

Kinosita, Y., Nakane, D., Sugawa, M., Masaike, T., Mizutani, K., Miyata, M., et al. (2014). Unitary step of gliding machinery in *Mycoplasma mobile*. *Proc. Natl. Acad. Sci. U. S. A.* 111, 8601–8606. doi: 10.1073/pnas.1310355111

Kojima, H., Muto, E., Higuchi, H., and Yanagida, T. (1997). Mechanics of single kinesin molecules measured by optical trapping nanometry. *Biophys. J.* 73, 2012–2022. doi: 10.1016/S0006-3495(97)78231-6

Krause, D. C., Chen, S., Shi, J., Jensen, A. J., Sheppard, E. S., and Jensen, G. J. (2018). Electron cryotomography of *Mycoplasma pneumoniae* mutants correlates terminal organelle architectural features and function. *Mol. Microbiol.* 108, 306–318. doi: 10.1111/mmi.13937

Mallik, R., Carter, B. C., Lex, S. A., King, S. J., and Gross, S. P. (2004). Cytoplasmic dynein functions as a gear in response to load. *Nature* 427, 649–652. doi: 10.1038/nature02293

Merz, A. J., So, M., and Sheetz, M. P. (2000). Pilus retraction powers bacterial twitching motility. *Nature* 407, 98–102. doi: 10.1038/35024105

Miyata, M. (2008). Centipede and inchworm models to explain *Mycoplasma* gliding. *Trends Microbiol.* 16, 6–12. doi: 10.1016/j.tim.2007.11.002

Miyata, M., and Hamaguchi, T. (2016a). Integrated information and prospects for gliding mechanism of the pathogenic bacterium *Mycoplasma pneumoniae*. *Front. Microbiol.* 7:960. doi: 10.3389/fmicb.2016.00960

Miyata, M., and Hamaguchi, T. (2016b). Prospects for the gliding mechanism of *Mycoplasma mobile*. *Curr. Opin. Microbiol.* 29, 15–21. doi: 10.1016/j.mib.2015.08.010

Miyata, M., and Petersen, J. D. (2004). Spike structure at the interface between gliding *Mycoplasma mobile* cells and glass surfaces visualized by rapid-freeze-and-fracture electron microscopy. *J. Bacteriol.* 186, 4382–4386. doi: 10.1128/JB.186.13.4382-4386.2004

Miyata, M., Robinson, R. C., Uyeda, T. Q. P., Fukumori, Y., Fukushima, S. I., Haruta, S., et al. (2020). Tree of motility - A proposed history of motility systems in the tree of life. *Genes Cells* 25, 6–21. doi: 10.1111/gtc.12737

Miyata, M., Ryu, W. S., and Berg, H. C. (2002). Force and velocity of *Mycoplasma mobile* gliding. *J. Bacteriol.* 184, 1827–1831. doi: 10.1128/JB.184.7.1827-1831.2002

Mizutani, M., and Miyata, M. (2017). Force measurement on *Mycoplasma mobile* gliding using optical tweezers. *Bio Protoc.* 7:e2127. doi: 10.21769/BioProtoc.2127

Mizutani, M., and Miyata, M. (2019). Behaviors and energy source of *Mycoplasma gallisepticum* gliding. *J. Bacteriol.* 201, e00397–19. doi: 10.1128/JB.00397-19

Mizutani, M., Tulum, I., Kinosita, Y., Nishizaka, T., and Miyata, M. (2018). Detailed analyses of stall force generation in *Mycoplasma mobile* gliding. *Biophys. J.* 114, 1411–1419. doi: 10.1016/j.bpj.2018.01.029

Nagai, R., and Miyata, M. (2006). Gliding motility of *Mycoplasma mobile* can occur by repeated binding to *N*-acetylneuraminyllactose (sialyllactose) fixed on solid surfaces. *J. Bacteriol.* 188, 6469–6475. doi: 10.1128/JB.00754-06

Nakamura, S., Kami-Ike, N., Yokota, J. P., Minamino, T., and Namba, K. (2010). Evidence for symmetry in the elementary process of bidirectional torque generation by the bacterial flagellar motor. *Proc Natl Acad Sci U. S. A.* 107, 17616–17620. doi: 10.1073/pnas.1007448107

Nakane, D., Adan-Kubo, J., Kenri, T., and Miyata, M. (2011). Isolation and characterization of P1 adhesin, a leg protein of the gliding bacterium *Mycoplasma pneumoniae*. *J. Bacteriol.* 193, 715–722. doi: 10.1128/JB.00796-10

Nakane, D., Kenri, T., Matsuo, L., and Miyata, M. (2015). Systematic structural analyses of attachment organelle in *Mycoplasma pneumoniae*. *PLoS Pathog.* 11:e1005299. doi: 10.1371/journal.ppat.1005299

Nakane, D., and Miyata, M. (2009). Cytoskeletal asymmetrical dumbbell structure of a gliding mycoplasma, *Mycoplasma gallisepticum*, revealed by negative-staining electron microscopy. *J. Bacteriol.* 191, 3256–3264. doi: 10.1128/JB.01823-08

Nishikawa, M. S., Nakane, D., Toyonaga, T., Kawamoto, A., Kato, T., Namba, K., et al. (2019). Refined Mechanism of *Mycoplasma mobile* gliding based on structure, ATPase activity, and sialic acid binding of machinery. *MBio* 10, e02846–19. doi: 10.1128/mBio.02846-19

Park, P. J., and Lee, K. J. (2013). A modified active Brownian dynamics model using asymmetric energy conversion and its application to the molecular motor system. *J. Biol. Phys.* 39, 439–452. doi: 10.1007/s10867-013-9300-5

Pelicic, V. (2008). Type IV pili: *e pluribus unum*? *Mol. Microbiol.* 68, 827–837. doi: 10.1111/j.1365-2958.2008.06197.x

Prince, O. A., Krunkosky, T. M., and Krause, D. C. (2014). In vitro spatial and temporal analysis of *Mycoplasma pneumoniae* colonization of human airway epithelium. *Infect. Immun.* 82, 579–586. doi: 10.1128/IAI.01036-13

Razin, S., and Hayflick, L. (2010). Highlights of mycoplasma research–an historical perspective. *Biologicals* 38, 183–190. doi: 10.1016/j.biologicals.2009.11.008

Razin, S., Yogev, D., and Naot, Y. (1998). Molecular biology and pathogenicity of mycoplasmas. *Microbiol. Mol. Biol. Rev.* 62, 1094–1156. doi: 10.1128/MMBR.62.4.1094-1156.1998

Relich, R. F., and Balish, M. F. (2011). Insights into the function of *Mycoplasma pneumoniae* protein P30 from orthologous gene replacement. *Microbiology* 157, 2862–2870. doi: 10.1099/mic.0.052464-0

Rosengarten, R., Fischer, M., Kirchhoff, H., Kerlen, G., and Seack, K. H. (1988). Transport of erythrocytes by gliding cells of *Mycoplasma mobile* 163K. *Curr. Microbiol.* 16, 253–257. doi: 10.1007/BF01568687

Saraya, T. (2017). *Mycoplasma pneumoniae* infection: basics. *J. Gen. Fam. Med.* 18, 118–125. doi: 10.1002/jgf2.15

Schnitzer, M. J., and Block, S. M. (1997). Kinesin hydrolyses one ATP per 8-nm step. *Nature* 388, 386–390. doi: 10.1038/41111

Seto, S., Kenri, T., Tomiyama, T., and Miyata, M. (2005). Involvement of P1 adhesin in gliding motility of *Mycoplasma pneumoniae* as revealed by the inhibitory effects of antibody under optimized gliding conditions. *J. Bacteriol.* 187, 1875–1877. doi: 10.1128/JB.187.5.1875-1877.2005

Seto, S., Layh-Schmitt, G., Kenri, T., and Miyata, M. (2001). Visualization of the attachment organelle and cytadherence proteins of *Mycoplasma pneumoniae* by immunofluorescence microscopy. *J. Bacteriol.* 183, 1621–1630. doi: 10.1128/JB.183.5.1621-1630.2001

Seybert, A., Gonzalez-Gonzalez, L., Scheffer, M. P., Lluch-Senar, M., Mariscal, A. M., Querol, E., et al. (2018). Cryo-electron tomography analyses of terminal organelle mutants suggest the motility mechanism of *Mycoplasma genitalium*. *Mol. Microbiol.* 108, 319–329. doi: 10.1111/mmi.13938

Seybert, A., Herrmann, R., and Frangakis, A. S. (2006). Structural analysis of *Mycoplasma pneumoniae* by cryo-electron tomography. *J. Struct. Biol.* 156, 342–354. doi: 10.1016/j.jsb.2006.04.010

Sowa, Y., Rowe, A. D., Leake, M. C., Yakushi, T., Homma, M., Ishijima, A., et al. (2005). Direct observation of steps in rotation of the bacterial flagellar motor. *Nature* 437, 916–919. doi: 10.1038/nature04003

Takagi, Y., Homsher, E. E., Goldman, Y. E., and Shuman, H. (2006). Force generation in single conventional actomyosin complexes under high dynamic load. *Biophys. J.* 90, 1295–1307. doi: 10.1529/biophysj.105.068429

Tanaka, A., Nakane, D., Mizutani, M., Nishizaka, T., and Miyata, M. (2016). Directed binding of gliding bacterium, *Mycoplasma mobile*, shown by

detachment force and bond lifetime. *MBio* 7, e00455–16. doi: 10.1128/mBio.00455-16

Tanaka, H., Homma, K., Iwane, A. H., Katayama, E., Ikebe, R., Saito, J., et al. (2002). The motor domain determines the large step of myosin-V. *Nature* 415, 192–195. doi: 10.1038/415192a

Toleikis, A., Carter, N. J., and Cross, R. A. (2020). Backstepping mechanism of kinesin-1. *Biophys. J.* 119, 1984–1994. doi: 10.1016/j.bpj.2020.09.034

Tully, J. G. (1983). New laboratory techniques for isolation of *Mycoplasma pneumoniae*. *Yale J Biol. Med.* 56, 511–515.

Tulum, I., Yabe, M., Uenoyama, A., and Miyata, M. (2014). Localization of P42 and F_1-ATPase alpha-subunit homolog of the gliding machinery in *Mycoplasma mobile* revealed by newly developed gene manipulation and fluorescent protein tagging. *J. Bacteriol.* 196, 1815–1824. doi: 10.1128/JB.01418-13

Uenoyama, A., Kusumoto, A., and Miyata, M. (2004). Identification of a 349-kilodalton protein (Gli349) responsible for cytadherence and glass binding during gliding of *Mycoplasma mobile*. *J. Bacteriol.* 186, 1537–1545. doi: 10.1128/JB.186.5.1537-1545.2004

Uenoyama, A., Seto, S., Nakane, D., and Miyata, M. (2009). Regions on Gli349 and Gli521 protein molecules directly involved in movements of *Mycoplasma mobile* gliding machinery, suggested by use of inhibitory antibodies and mutants. *J. Bacteriol.* 191, 1982–1985. doi: 10.1128/JB.01012-08

Varki, A. (2008). Sialic acids in human health and disease. *Trends Mol. Med.* 14, 351–360. doi: 10.1016/j.molmed.2008.06.002

Veigel, C., Coluccio, L. M., Jontes, J. D., Sparrow, J. C., Milligan, R. A., and Molloy, J. E. (1999). The motor protein myosin-I produces its working stroke in two steps. *Nature* 398, 530–533. doi: 10.1038/19104

Vizarraga, D., Kawamoto, A., Matsumoto, U., Illanes, R., Perez-Luque, R., Martin, J., et al. (2020). Immunodominant proteins P1 and P40/P90 from human pathogen *Mycoplasma pneumoniae*. *Nat. Commun.* 11:5188. doi: 10.1038/s41467-020-18777-y

Vizarraga, D., Torres-Puig, S., Aparicio, D., and Pich, O. Q. (2021). The sialoglycan binding adhesins of *Mycoplasma genitalium* and *Mycoplasma pneumoniae*. *Trends Microbiol.* 29, 477–481. doi: 10.1016/j.tim.2021.01.011

Williams, C. R., Chen, L., Driver, A. D., Arnold, E. A., Sheppard, E. S., Locklin, J., et al. (2018). Sialylated receptor setting influences *Mycoplasma pneumoniae* attachment and gliding motility. *Mol. Microbiol.* 109, 735–744. doi: 10.1111/mmi.13997

Yasuda, R., Noji, H., Kinosita, K. Jr., and Yoshida, M. (1998). F_1-ATPase is a highly efficient molecular motor that rotates with discrete 120 degree steps. *Cell* 93, 1117–1124. doi: 10.1016/s0092-8674(00)81456-7

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 11 November 2021
doi: 10.3389/fmicb.2021.766221

Accelerating Complete Phytoplasma Genome Assembly by Immunoprecipitation-Based Enrichment and MinION-Based DNA Sequencing for Comparative Analyses

Choon Meng Tan[1†], Yu-Chen Lin[2†], Jian-Rong Li[3], Yuan-Yu Chien[1], Chien-Jui Wang[1], Lin Chou[2], Cheng-Wei Wang[1], Yi-Ching Chiu[1,4], Chih-Horng Kuo[2,4,5] and Jun-Yi Yang[1,4,5,6]**

[1] Institute of Biochemistry, National Chung Hsing University, Taichung, Taiwan, [2] Institute of Plant and Microbial Biology, Academia Sinica, Taipei, Taiwan, [3] Institute of Genomics and Bioinformatics, National Chung Hsing University, Taichung, Taiwan, [4] Ph.D. Program in Microbial Genomics, National Chung Hsing University and Academia Sinica, Taichung, Taiwan, [5] Institute of Biotechnology, National Chung Hsing University, Taichung, Taiwan, [6] Advanced Plant Biotechnology Center, National Chung Hsing University, Taichung, Taiwan

Phytoplasmas are uncultivated plant-pathogenic bacteria with agricultural importance. Those belonging to the 16SrII group, represented by '*Candidatus* P. aurantifolia', have a wide range of plant hosts and cause significant yield losses in valuable crops, such as pear, sweet potato, peanut, and soybean. In this study, a method that combines immunoprecipitation-based enrichment and MinION long-read DNA sequencing was developed to solve the challenge of phytoplasma genome studies. This approach produced long reads with high mapping rates and high genomic coverage that can be combined with Illumina reads to produce complete genome assemblies with high accuracy. We applied this method to strain NCHU2014 and determined its complete genome sequence, which consists of one circular chromosome with 635,584 bp and one plasmid with 4,224 bp. Although '*Ca*. P. aurantifolia' NCHU2014 has a small chromosome with only 471 protein-coding genes, it contains 33 transporter genes and 27 putative effector genes, which may contribute to obtaining nutrients from hosts and manipulating host developments for their survival and multiplication. Two effectors, the homologs of SAP11 and SAP54/PHYL1 identified in '*Ca*. P. aurantifolia' NCHU2014, have the biochemical activities in destabilizing host transcription factors, which can explain the disease symptoms observed in infected plants. Taken together, this study provides the first complete genome available for the 16SrII phytoplasmas and contributes to the understanding of phytoplasma pathogenicity.

Keywords: phytoplasma, Nanopore, effector, SAP11, genome sequencing, potential mobile unit

OPEN ACCESS

Edited by:
David William Waite, Ministry for Primary Industries, New Zealand

Reviewed by:
Amit Yadav, National Centre for Cell Science, India
Kenro Oshima, Hosei University, Japan

****Correspondence:***
Chih-Horng Kuo
chk@gate.sinica.edu.tw
Jun-Yi Yang
jyang@nchu.edu.tw

† These authors have contributed equally to this work

Specialty section:
This article was submitted to Microbial Symbioses, a section of the journal Frontiers in Microbiology

Received: *28 August 2021*
Accepted: *11 October 2021*
Published: *11 November 2021*

Citation:
Tan CM, Lin Y-C, Li J-R, Chien Y-Y, Wang C-J, Chou L, Wang C-W, Chiu Y-C, Kuo C-H and Yang J-Y (2021) Accelerating Complete Phytoplasma Genome Assembly by Immunoprecipitation-Based Enrichment and MinION-Based DNA Sequencing for Comparative Analyses. Front. Microbiol. 12:766221. doi: 10.3389/fmicb.2021.766221

INTRODUCTION

Phytoplasmas are wall-less bacterial pathogens that are known to infect numerous plant species and lead to significant agricultural losses (Gurr et al., 2016; Kumari et al., 2019; Pierro et al., 2019). They are parasitic bacteria multiplying exclusively in phloem sieve elements and are transmitted between plants by phloem-feeding insects (Lee et al., 2000; Hogenhout et al., 2008). Plants infected by phytoplasmas exhibit a range of symptoms, including witches' broom, phyllody, virescence, purple top, stunting, yellowing, and general decline (Christensen et al., 2005). To date, more than 40 *Candidatus* Phytoplasma (abbreviated as *Ca.* P.) species have been described (Kumari et al., 2019). However, as an obligate intracellular parasite, phytoplasmas remain some of the most challenging plant pathogens to characterize due to the lack of an axenic culture.

Beginning with the important step in whole-genome sequencing of phytoplasmas, many aspects have been intensively studied, and a better understanding of the molecular interactions between phytoplasmas and their hosts is revealed (Oshima et al., 2004; Namba, 2019). Overall, phytoplasmas have relatively small genomes (i.e., < 1,000 kb) compared with other bacteria, and many metabolic genes required for biosynthetic pathways of indispensable compounds are missing, including those for the biosynthesis of amino acids, nucleotides and fatty acids, the tricarboxylic acid cycle, and ATP synthases (Oshima et al., 2004, 2013; Namba, 2019). As a result, phytoplasmas rely on an enriched environment and their transporters to obtain nutrients from hosts for their survival and multiplication. At the same time, pathogenicity factors (i.e., effectors), such as TENGU, SAP11, and SAP54/PHYL1, are secreted by phytoplasmas to interfere with plant developmental processes as well as plant hormone homeostasis (Hoshi et al., 2009; Sugio et al., 2011a; MacLean et al., 2014; Maejima et al., 2014; Tan et al., 2016; Chang et al., 2018). The morphological and physiological changes of host plants not only associate with symptom development, but also improve the fitness of phytoplasmas and their insect vectors, resulting in the spread of phytoplasma diseases (Tomkins et al., 2018; Dermastia, 2019).

The spreading of phytoplasmas might also rely on the interactions between phytoplasma immunodominant membrane proteins (IDPs) and host proteins (Konnerth et al., 2016). IDPs are highly abundant proteins in the cell membrane of phytoplasmas, which can be grouped into three classes, including immunodominant membrane protein (Imp), immunodominant membrane protein A (IdpA), and antigenic membrane protein (Amp) (Kakizawa et al., 2006b). These proteins are highly variable in the amino acid sequences due to the selective pressure from interactions with the environment and hosts (Kakizawa et al., 2006a, 2009). Previous studies show that the Imp of '*Ca.* P. mali' can interact with the actin of plant hosts (Boonrod et al., 2012); the Amp of '*Ca.* P. asteris' can interact with the actin of insect vectors and form the Amp-microfilament complex together with myosin (Suzuki et al., 2006; Galetto et al., 2011). These interactions seem to play important roles in determining the mobility of phytoplasma within plant hosts and the transmissibility of phytoplasma by insect vectors (Konnerth et al., 2016).

Despite having small genomes compared with other bacteria, phytoplasmas contain high numbers of repetitive genes lying within potential mobile unit (PMU) or sequence-variable mosaic (SVM) regions, which are proposed to be remnants of transposons or prophage (Bai et al., 2006; Wei et al., 2008). These regions are flanked by inverted repeats and contain genes involved in DNA recombination (*tra5*, *ssb*, *himA*) and replication (*dnaG*, *dnaB*), suggesting that they have the ability to transpose within the genome and be horizontally transferred among phytoplasmas (Toruno et al., 2010; Chung et al., 2013; Ku et al., 2013a; Cho et al., 2019). The repeat-rich nature of phytoplasma genomes is proposed to contribute to the frequent recombination events, leading to the considerable variation in genome sizes among phytoplasmas (Kube et al., 2008; Andersen et al., 2013; Marcone, 2014; Seruga Music et al., 2019). Moreover, as putative pathogenicity islands, these regions often carry effector genes and are proposed to contribute to the adaptation of phytoplasma switching in plant and insect hosts (MacLean et al., 2011; Oshima et al., 2011; Chung et al., 2013).

The availability of complete genomes not only provides a better chance to identify potential genes involved in phytoplasma–host interactions, but it also contributes to a better understanding of the phytoplasma genome organization and evolution (Hogenhout and Seruga Music, 2009; Kube et al., 2012; Marcone, 2014). In the past, phytoplasma DNA was enriched via procedures, such as cesium chloride density gradient centrifugation and pulse-field gel electrophoresis (PFGE) (Hogenhout and Seruga Music, 2009). However, the density gradient centrifugation method may also enrich plant chloroplast DNA and reduce DNA integrity, and the PFGE method typically has low yields. Given the difficulties in isolating phytoplasma DNA from an infected host, together with the fact that phytoplasmas have AT- and repeat-rich genomes, only seven complete genome sequences of phytoplasmas were released over 15 years since the first Onion Yellows phytoplasma strain M (OY-M) was completed in 2004 (Oshima et al., 2004; Bai et al., 2006; Kube et al., 2008; Tran-Nguyen et al., 2008; Andersen et al., 2013; Orlovskis et al., 2017; Wang J. et al., 2018). These genome sequences represent only four 16S rRNA gene RFLP (16Sr) groups, including 16SrI, 16SrX, 16SrXII, and 16SrV.

Recently, with the introduction of Illumina sequencing technology and decreasing cost, more than 30 phytoplasma draft genome sequences were published. Those draft genomes were determined without further enrichment of phytoplasma DNA and utilized selective exclusion of the host reads by using the healthy plant genome as a reference or selective inclusion of phytoplasma reads using available phytoplasma genome sequences as references (Chung et al., 2013; Polano and Firrao, 2018). Although draft genome sequences provide insights into phytoplasma biology, the fragmented nature of draft genomes limits the type of comparative genomics analysis that can be conducted (Ricker et al., 2012). Previously, a draft genome of '*Ca.* P. aurantifolia' NCHU2014 (16SrII group) associated with *Echinacea purpurea* witches' broom (EpWB) disease was obtained based on Illumina paired-end sequencing, which contains 28

contigs with a combined size of 545,427 bp and encodes 433 protein-coding genes (Chang et al., 2015). In this study, an integrated solution that combined immunoprecipitation-based enrichment of phytoplasma cells prior to DNA extraction and Oxford Nanopore Technologies (ONT) MinION long-read DNA sequencing was developed to obtain a complete genome sequence of '*Ca*. P. aurantifolia' NCHU2014. Furthermore, with the complete genome sequence of this strain determined, we conducted comparative analysis with other phytoplasmas and experimental characterization of its effectors.

MATERIALS AND METHODS

Polyclonal Antibody Production

A codon-optimized DNA fragment encoding the Imp of '*Ca*. P. aurantifolia' NCHU2014 without the transmembrane domain was subcloned into the SUMO-pET vector and introduced into *Escherichia coli* BL21 (DE3). The N-terminal His-SUMO tagged ImpΔN protein was produced at 24°C by isopropyl β-D-1-thiogalactopyranoside induction and purified by Ni^{2+}-NTA resin (Qiagen) according to the manufacturer's instructions. The purified protein was cleaved with Ubiquitin-like-specific protease 1 and then reapplied to Ni^{2+}-NTA resin for removing the cleaved His-SUMO tag and uncleaved His-SUMO-ImpΔN. The recombinant ImpΔN was obtained in the flowthrough and prepared for polyclonal antibody production in rabbits.

Affinity Purification of '*Ca*. P. aurantifolia' NCHU2014

The antibody-based purification was developed to enrich '*Ca*. P. aurantifolia' NCHU2014, which was maintained in periwinkle (*Catharanthus roseus*) by grafting. The stems and leaf veins of symptomatic periwinkle were sampled and gently homogenized with liquid nitrogen using mortar and pestle. Subsequently, 3 g of grinded tissues were suspended with 6 ml PBS buffer (0.137 M NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, pH 6.8) and filtrated with a cell strainer (100 μm) to remove unfragmented tissues. Meanwhile, the anti-Imp antibody-coated beads were prepared by incubating antiserum (containing polyclonal anti-Imp antibodies) and Novex® Dynabeads Protein A (superparamagnetic beads with recombinant Protein A covalently coupled to the surface). The flowthrough containing homogenized cells was incubated with antibody-coated beads (3 mg) for 30 min at 4°C, and then the unwanted substances were separated from the beads by magnet. The remaining materials containing phytoplasma cells were washed extensively with PBS buffer and collected for genomic DNA extraction.

MinION-Based DNA Sequencing

After the affinity-purified cells enriched from 3 g of grinded tissues were lysed, about 6 μg genomic DNA was obtained from extraction by the Plant Genomic DNA Purification Kit (Gene Mark, Taiwan) according to the manufacturer's instructions. After further cleaning by the Quick-DNA Plant/Seed Miniprep Kit (ZYMO RESEARCH), approximately 4 μg of DNA was

TABLE 1 | Raw reads statistics.

Run	Read count	Sequencing output (bp)	Max. length (bp)	Av. length (bp)	Reads mapped to phytoplasma genome	% Reads mapped to phytoplasma genome	Reads mapped to chloroplast genome	% Reads mapped to chloroplast genome	Reads mapped to plant nuclear genome	% Reads mapped to plant nuclear genome	% Unmapped reads	Phytoplasma genome coverage (bp)	Phytoplasma genome sequencing depth (fold)
A01	148,631	554,291,700	46,543	3,729	53,487	36.0	11,474	7.7	78,134	52.6	3.7	639,808	209
A02	88,111	347,096,923	41,777	3,939	30,963	35.1	6,565	7.5	47,564	54.0	3.4	639,808	126
A03	58,938	243,448,556	40,264	4,131	20,511	34.8	4,179	7.1	32,044	54.4	3.7	639,808	87
A04	14,718	61,459,870	46,555	4,176	5,165	35.1	1,092	7.4	7,834	53.2	4.3	639,808	22
A05	23,978	95,643,639	34,810	3,989	8,495	35.4	1,678	7.0	12,802	53.4	4.2	639,808	34
A06	533,668	2,653,344,966	55,606	4,972	119,855	22.5	44,280	8.3	353,160	66.2	3.1	639,808	605
A07	292,158	1,537,572,472	52,917	5,263	74,261	25.4	24,101	8.2	185,853	63.6	2.7	639,808	384
A08	161,139	771,282,841	39,480	4,786	48,419	30.0	13,476	8.4	94,699	58.8	2.8	639,808	244
A09	36,062	156,720,727	41,438	4,346	11,348	31.5	3,358	9.3	20,183	56.0	3.3	639,808	51
A10	28,469	119,536,595	44,746	4,199	9,168	32.2	2,721	9.6	15,516	54.5	3.7	639,808	41
Subtotal	1,385,872	6,540,398,289			381,672	27.5	112,924	8.1	847,789	61.2	3.1		1,802
B01	207,087	509,103,092	61,271	2,458	34,235	16.5	8,341	4.0	142,745	68.9	10.5	602,845	47
B02	25,499	127,421,361	44,713	4,997	2,043	8.0	1,001	3.9	19,672	77.1	10.9	613,315	8
B03	475,713	2,194,510,560	91,125	4,613	110,450	23.2	16,647	3.5	314,627	66.1	7.1	619,020	194
Subtotal	708,299	2,831,035,013			146,728	20.7	25,989	3.7	477,044	67.4	8.3		249
C01	34,321,466	10,206,843,846	300	297	5,477,697	16.0	2,265,456	6.6	26,462,971	77.1	0.3	637,986	2,117

Runs A01–A10 were based on the ONT sequencing with the immunoprecipitation enrichment of phytoplasma cells; four biological samples for runs A01–A02, A03–A05, A06–07, and A08–A10, respectively. Runs B01–B03 were based on the ONT sequencing without the enrichment; each used a different biological sample. Run C01 was based on the Illumina paired-end sequencing from our previous study (DOI: 10.1128/genomeA.01398-1); no enrichment was performed.

obtained from one sample with 3 g of starting plant materials. A total of 2 μg genomic DNA was used for each library construction by the KAPA Hyper Prep Kit optimization for Nanopore 1D (Oxford Nanopore Technologies, Oxford, United Kingdom) according to the manufacturer's instructions. To perform long-read sequencing, DNA libraries were loaded on MinION flow cells (R9.4), and the Nanopore reads were base-called from FAST5 files using the ONT Albacore Software (version 2.0.1).

Genome Assembly and Annotation

The procedures for genome assembly and annotation were based on those described in our previous studies (Chung et al., 2013; Cho et al., 2020). All bioinformatics tools were used with the default settings unless stated otherwise. Briefly, the MinION reads were mapped to a previously published draft genome (Chang et al., 2015) using Minimap2 v2.15 (Li, 2018). Visual inspection of the mapping result using IGV v2.3.57 (Robinson et al., 2011) produced one circular scaffold representing the chromosome and one circular contig representing the plasmid. Next, an iterative process was used to complete and validate the assembly. In each iteration, the MinION reads were mapped as described, and the Illumina reads from our previous study (Chang et al., 2015) were mapped to the assembly using BWA v0.7.12 (Li and Durbin, 2009). The raw read mapping results were programmatically checked using SAMtools v1.2 (Li et al., 2009) and manually inspected using IGV v2.3.57 (Robinson et al., 2011). The iterative process was repeated until all polymorphic sites were resolved and a complete assembly was obtained. For the assembly process, only the MinION reads from runs A01 to A10 (**Table 1**) were used. The MinION reads from runs B01 to B03 were only used for checking coverage after the complete assembly was validated.

For annotation, RNAmmer v1.2 (Lagesen et al., 2007), tRNAscan-SE v1.3.1 (Lowe and Eddy, 1997), and Prodigal v2.6.3 (Hyatt et al., 2010) were used for gene prediction. The homologous genes in other representative phytoplasma genomes (**Table 2**) were identified using OrthoMCL (Li et al., 2003) to provide gene name and product description. Gene encoding putative secreted proteins were identified based on two established procedures, including one that uses SignalP v4.1 (Petersen et al., 2011) as described in Garcion et al. (2021) and another that uses SignalP v5.0 (Armenteros et al., 2019) as described in Cho et al. (2019). Additionally, BlastKOALA (Kanehisa et al., 2016) and GenBank (Benson et al., 2018) sequence similarity searches were used for manual curation of the annotation results.

For quality check, the raw reads that were not mapped to the phytoplasma genome were collected and mapped to the host (*Catharanthus roseus*) genome in two steps using Minimap2 v2.15 (Li, 2018). The first step mapped to the chloroplast genome (accession KC561139) (Ku et al., 2013b), and the second step mapped to the host nuclear genome (accession JQHZ01000000) (Kellner et al., 2015).

Co-expression Assays

A codon-optimized DNA fragment encoding SAP11 of '*Ca.* P. aurantifolia' NCHU2014 without the signal peptide was subcloned into a binary vector pBA002 for expression under the control of the *CaMV* 35S promoter. Plasmids for expression of the N-terminal FLAG-tagged *Arabidopsis* TCP transcription factors (SFP-AtT) were obtained as previously described (Chang et al., 2018). *Nicotiana benthamiana* grown at 26°C was used for transient co-expression assays. A mixture of *A. tumefaciens* strain ABI carrying the desired constructs of *35S:SAP11* and *35S:SFP-AtTCPs* was introduced into *N. benthamiana* leaves by agroinfiltration. After 2 days, samples prepared from two infiltrated leaves (the third and fourth leaves counting from the top of 4- to 5-week-old plants) were collected for Western blotting analysis.

Western Blotting

Collected samples were ground into powder using liquid nitrogen. About 0.1 g sample powder was added to 0.2 ml 2.5X SDS sample buffer (5 mM EDTA, 5% SDS, 0.3 M Tris-HCl, pH 6.8, 20% glycerol, 1% β-mercaptoethanol, and bromophenyl blue) and heated in a boiling water bath for 5 min. After centrifugation, supernatants were obtained as total cell extracts, and proteins were separated by SDS-PAGE. Antibodies against Imp (polyclonal), SAP11 (polyclonal), and FLAG^{TM} tag (monoclonal) were used to monitor protein amounts. Chemiluminescence signals generated by Amersham

TABLE 2 | List of the representative phytoplasma genome sequences analyzed.

Strain	16Sr Group	Accession	Chromosome	Size (bp)	G + C Content (%)	Coding density (%)	Coding sequences	Pseudogenes	tRNA genes	rRNA genes
'*Ca.* P. aurantifolia' NCHU2014	II	CP040925	Circular	635,584	24.5	66.3	471	35	24	6
'*Ca.* P. luffae' NCHU2019	VIII	CP054393	Circular	769,143	23.3	80.3	725	13	31	6
'*Ca.* P. ziziphi' Jwb-nky	V	CP025121	Circular	750,803	23.2	75.4	640	31	32	6
'*Ca.* P. mali' AT	X	CU469464	Linear	601,943	21.4	76.7	482	15	32	6
'*Ca.* P. australiense' PAa	XII	AM422018	Circular	879,959	27.4	64.1	684	155	35	6
'*Ca.* P. asteris' AYWB	I-A	CP000061	Circular	706,569	26.9	73.5	671	0	31	6
'*Ca.* P. asteris' OY-M	I-B	AP006628	Circular	853,092	27.8	73.0	752	0	32	6

The pseudogene counts are based on those annotated in the standard format (i.e., with the "/pseudo" tag). For 'Ca. P. asteris' OY-M, 46 coding sequences are annotated as "possible pseudogene" in the "/note" field.

ECL reagents were captured using the ImageQuant LAS 4000 mini (GE Healthcare).

RESULTS

Immunoprecipitation-Based Enrichment of '*Ca*. P. aurantifolia' NCHU2014

The '*Ca*. P. aurantifolia' NCHU2014 was originally collected from purple coneflower (*Echinacea purpurea*) in Taiwan and transferred to periwinkle by dodder (Chang et al., 2015). The diseased plants infected with '*Ca*. P. aurantifolia' NCHU2014 exhibited phyllody, virescence, and witches' broom phenotypes in which the development of leaf-like structures with the loss of flower pigment and the proliferation of shoots in place of carpels were observed (**Figures 1A–C**).

As an obligate intracellular parasite, it is challenging to obtain a complete genome sequence of phytoplasma due to the host DNA contamination. To enrich '*Ca*. P. aurantifolia' NCHU2014 for whole genome sequencing, immunoprecipitation-based purification of phytoplasma cells was conducted using the polyclonal antibody raised against the Imp of '*Ca*. P. aurantifolia' NCHU2014. The antibody was generated in rabbit using the recombinant protein of Imp without N-terminal transmembrane domain (**Figure 1D**), which specifically recognized the 16SrII group '*Ca*. P. aurantifolia' NCHU2014 in symptomatic periwinkle but not the 16SrVIII group '*Ca*. P. luffae' NCHU2019 associated with loofah witches' broom disease (**Figure 1E** and **Supplementary Figure 1**). For affinity purification of '*Ca*. P. aurantifolia' NCHU2014, the stems and leaf veins from the symptomatic periwinkle displaying chlorosis, witches' broom, phyllody, and virescence were used (**Figure 1F**). Genomic DNA extracted from the immunoprecipitated fraction containing '*Ca*. P. aurantifolia' NCHU2014 was obtained for PCR examination. The PCR products specific for EpWB_v2c2400, EpWB_v2c2070, EpWB_v2c5130, and EpWB_v2c1780 genes of '*Ca*. P. aurantifolia' NCHU2014 encoding SAP54/phyllogen, putative secreted protein, Mn/Zn-binding protein, and 50S ribosomal protein L17, respectively, were amplified only with the DNA sample extracted from symptomatic plants but not healthy plants (**Figure 1G**). The extracted DNA was then further processed for high-throughput DNA sequencing.

MinION-Based Long-Read DNA Sequencing and Genome Assembly

To improve the draft genome assembly of '*Ca*. P. aurantifolia' NCHU2014 (Chang et al., 2015), we used ONT MinION to generate long sequencing reads. A total of 1,385,872 reads containing 6,540,398,289 bp were obtained from the libraries constructed using four DNA samples from our enrichment procedure (**Table 1**). These reads were obtained from 10 sequencing runs using two MinION flow cells. The first flow cell was used for two libraries (runs A01–A02 and A03–A05, respectively), producing 334,376 reads totaling 1,301,940,688 bp. The second flow cell was used for two additional libraries (runs A06–A07 and A08–A10, respectively), producing 1,051,496 reads totaling 5,238,457,601 bp. The average read lengths ranged from 3.7 to 5.2 kb with a maximum read length of 55.6 kb. Using these newly generated long reads, combined with our previous Illumina data set (Chang et al., 2015), we obtained a complete genome assembly for '*Ca*. P. aurantifolia' NCHU2014. This assembly contains one circular chromosome with 635,584 bp and one circular plasmid with 4,224 bp (**Figure 2**). To the best of our knowledge, this is the first complete genome available for 16SrII phytoplasma.

Compared with the previous Illumina-only draft assembly of the same strain (Chang et al., 2015), 23 chromosomal regions ranging from 214 to 32,404 bp in size are newly assembled; the plasmid is unchanged. One 193-bp region (positions 354,755–354,947 of the chromosome) is shown as a gap based on the ONT reads but is supported by the Illumina reads. One possible explanation is that this region was deleted between our sample collections for Illumina (in 2014) and ONT sequencing (in 2019). Additionally, by examining the mapping results of Illumina reads to this complete assembly, we found that 15 chromosomal regions (mostly 36–128 bp in size, totaling 1,822 bp) are imperfect repeats and have no Illumina reads mapped.

As a quality evaluation, we also performed MinION sequencing to generate 708,299 reads (B01–B03; three runs totaling 2,831,035,013 bp using one MinION flowcell) using the libraries constructed with three DNA samples obtained without the immunoprecipitation-based enrichment. Compared with the enrichment protocol from which 22.5 to 36.0% of the raw reads were from phytoplasma, only 8.0 to 23.2% of the raw reads were from phytoplasma without the enrichment (**Table 1**). Importantly, the reads obtained with the enrichment procedure can cover all base pairs of the assembled phytoplasma genome even when only 14,718 reads were obtained in run A04 (**Table 1**). In contrast, up to approximately 37 kb of the phytoplasma genome were not covered by reads obtain from unenriched samples (B01). Based on these results, our enrichment procedure provides substantial improvements applicable to phytoplasma genome research.

Genome Comparison Between Two Closely Related '*Ca*. P. aurantifolia' Strains

A phylogenetic tree based on 16S ribosomal RNA (16S rRNA) genes of representative phytoplasmas was inferred to illustrate the evolutionary relationships (**Figure 3**). As the first complete genome sequence available for 16SrII group phytoplasma, '*Ca*. P. aurantifolia' NCHU2014 shares 100% sequence identity of the 16S rRNA gene with '*Ca*. P. aurantifolia' NTU2011. '*Ca*. P. aurantifolia' NTU2011 was originally collected from infected peanut plants associated with peanut witches' broom (PnWB) disease in Taiwan. So far, only a draft genome was available for '*Ca*. P. aurantifolia' NTU2011, which contains 13 chromosomal contigs with a combined size of 562,473 bp and a plasmid with 4,221 bp (Chung et al., 2013). Comparative genomic analysis between two closely related strains revealed that the average nucleotide identity (ANI) is 99.72% across 90.79% chromosomal segments. This result reflects a high genomic

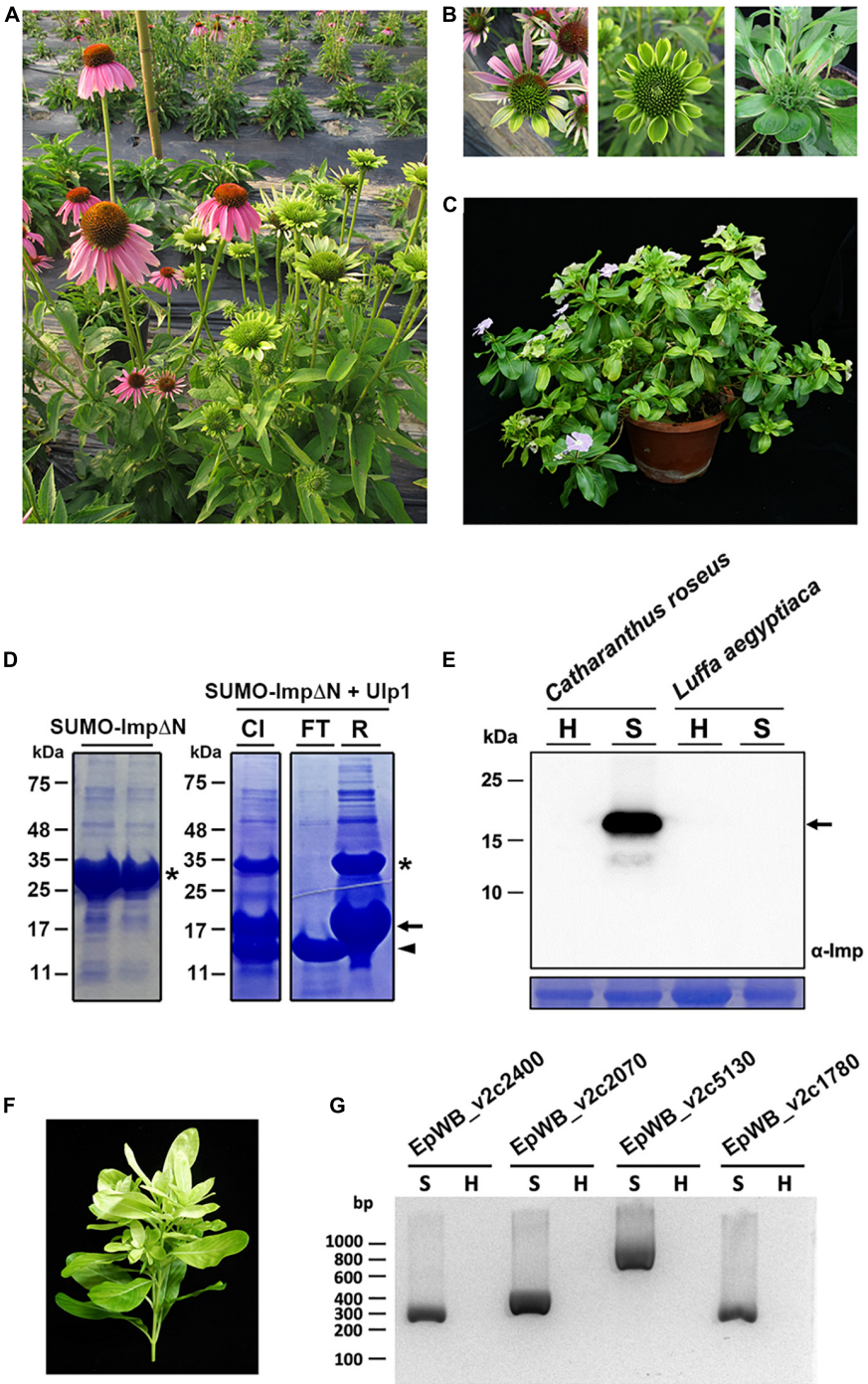


FIGURE 1 | Immunoprecipitation-based enrichment of '*Ca*. P. aurantifolia' NCHU2014. **(A)** Phenotypic comparisons between healthy (left) and '*Ca*. P. aurantifolia' NCHU2014-infected (right) purple coneflower (*Echinacea purpurea*). The diseased plants exhibited phyllody and witches' broom symptoms. **(B)** Enlarged images of the abnormal flowers carrying phyllody, virescence, and bud proliferation symptoms in '*Ca*. P. aurantifolia' NCHU2014-infected purple coneflower. **(C)** Symptomatic periwinkle (*Catharanthus roseus*) associated with '*Ca*. P. aurantifolia' NCHU2014 transmitted by dodder. **(D)** The His-SUMO tagged ImpΔN was expressed in *E. coli* and purified by Ni^{2+}-NTA resin (left panel). After cleavage of the His-SUMO tag by Ulp1 (CI), the reaction mixture was applied on the Ni^{2+}-NTA column. Arrowhead indicates the purified ImpΔN in flowthrough (FT); asterisk and arrow indicate the uncleaved His-SUMO-ImpΔN and His-SUMO, respectively, in the resin (R). **(E)** Total cell extracts prepared from healthy (H) and symptomatic (S) leaves were examined by Western blotting using specific antibody against Imp of '*Ca*. P. aurantifolia' NCHU2014. The specific signal of Imp was only detected in symptomatic *C. roseus* infected with '*Ca*. P. aurantifolia' NCHU2014, but not symptomatic loofah (*Luffa aegyptiaca*) infected with '*Ca*. P. luffae' NCHU2019 (upper panel). As a loading control, the large subunit of Rubisco was visualized with Coomassie Brilliant Blue staining (lower panel). Arrow indicates the 19 kDa Imp. **(F)** The '*Ca*. P. aurantifolia' NCHU2014-infected periwinkle used for immunoprecipitation. **(G)** PCR examination of DNA extracted from the immunoprecipitated fraction containing '*Ca*. P. aurantifolia' NCHU2014.

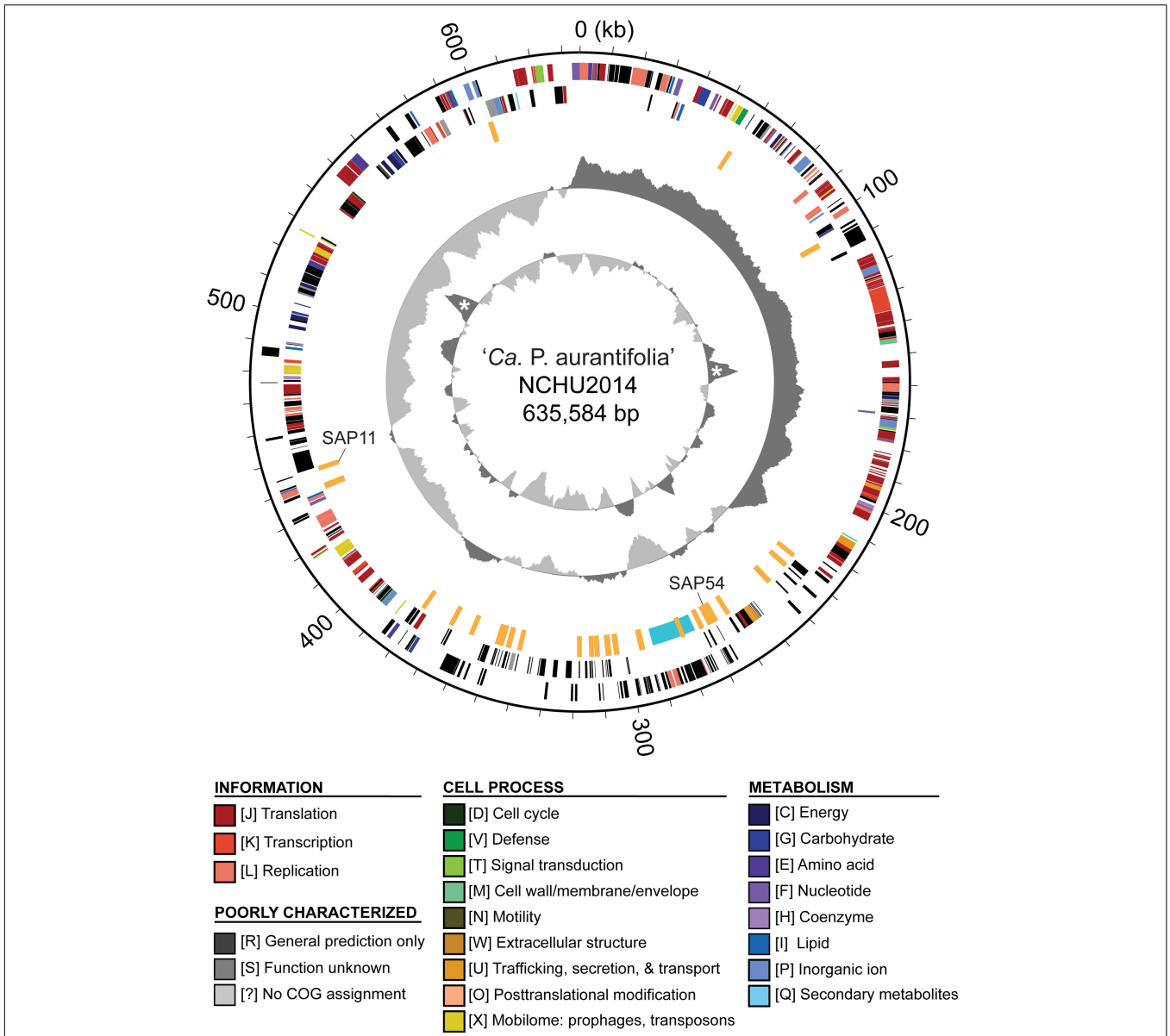


FIGURE 2 | Genome map of '*Ca*. P. aurantifolia' NCHU2014. Rings from outside in: (1) scale marks; (2 and 3) coding sequences on the forward and reverse strand, respectively (color-coded by functional categories); (4) putative effectors (orange) and putative mobile unit (light blue); (5) GC skew (positive: dark gray; negative: light gray); (6) GC content (above average: dark gray; below average: light gray), two peaks corresponding to the rRNA operons are marked by "*".

similarity between these two strains in Taiwan. However, the draft genome of NTU2011 lacks two large segments corresponding to positions 232,370–246,973 and 313,259–348,010 of the NCHU2014 chromosome (**Supplementary Figure 2**). Further analysis by the pairwise genome alignment revealed that the chromosomes of these two strains are largely collinear although some rearrangements, including inversions and translocations, were observed (**Figure 4A**). The rearranged regions were confirmed experimentally by PCR analysis (**Figure 4B**).

The plasmids found in these two strains are nearly identical with 99.13% sequence identity. Only four protein-coding genes were found, including those corresponding to one replication protein, one DNA primase, one threonine synthase, and one hypothetical protein. Our BLASTN search against the NCBI Nucleotide Collection (nt) database did not find any similar sequence available for comparative analysis.

Comparative Analysis of Gene Content of '*Ca*. P. aurantifolia' NCHU2014

The '*Ca*. P. aurantifolia' NCHU2014 genome consists of one circular chromosome with 635,584 bp and one plasmid with 4,224 bp. Based on the annotation, the chromosome encodes six rRNA genes, 24 tRNA genes, 471 protein-coding genes, and 35 pseudogenes (**Table 2**). Notably, '*Ca*. P. aurantifolia'

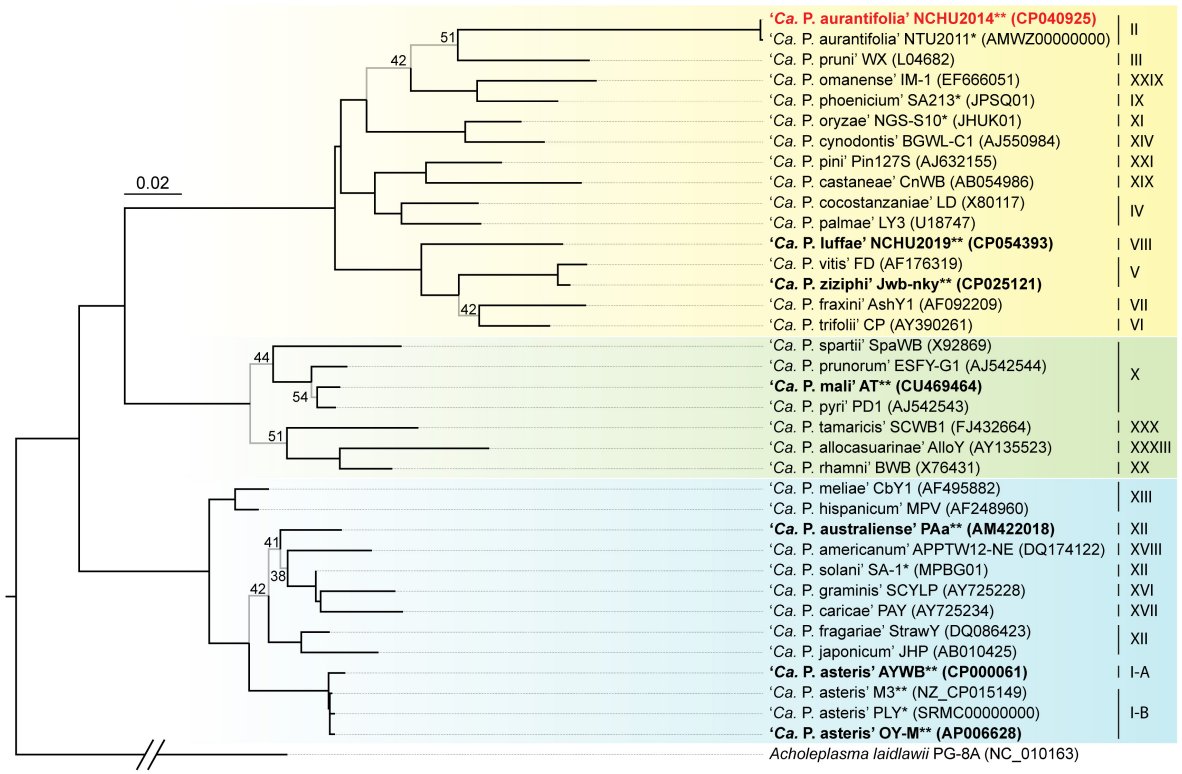


FIGURE 3 | Maximum likelihood phylogeny of representative phytoplasmas based on the 16S rRNA genes. The internal branches with < 60% bootstrap support are drawn in light gray, and the support values are labeled. Strains with genome sequences available are highlighted ("*", draft; "**", complete), those included in the genome comparisons are in bold. The focal strain of this study, '*Ca*. P. aurantifolia' NCHU2014, is highlighted in red. Sequence accession numbers are provided in parentheses following the strain names. For genome sequences with two 16S rRNA gene sequences available, the first one (i.e., the one with a smaller numerical value in locus tag) was selected. The 16Sr group assignments are provided on the right. Background colors are used to highlight the three phylogenetic clusters of phytoplasmas (I: blue; II; green; III: yellow). *Acholeplasma laidlawii* is included as the outgroup.

NCHU2014 has a low coding density (66.3%) with the lowest numbers of tRNA and protein-coding genes compared with other phytoplasmas with complete genome sequences available (**Table 2**). Among the protein-coding genes, only 352 (75%) genes were assigned to COG categories with specific functions (**Figure 2**). There were 119 (25%) genes annotated as hypothetical proteins without COG functional category assignments. Similar to other phytoplasmas with complete genome information, '*Ca*. P. aurantifolia' NCHU2014 has a small chromosome with low GC content (24.5%) and lacks the genes for many metabolic pathways although it has 33 genes annotated as transporters. The comparative analysis further revealed that '*Ca*. P. aurantifolia' NCHU2014 share 284–313 homologous gene clusters with other lineages surveyed (**Figure 5A**). Among them, a conserved core of 204 homologous gene clusters was found in all lineages surveyed (**Figure 5B**). This is consistent with other obligate parasitic bacteria that pose a high level of genomic plasticity (Andersen et al., 2013).

Effectors (Virulence-Related Factors) and Potential Mobile Units

Phytoplasmas possess the Sec secretion system for transportation of effectors into the host cell cytoplasm (Sugio et al., 2011b; Oshima et al., 2013). Based on the prediction of N-terminal signal peptide by SignalP-5.0, 28 putative secreted proteins were identified (**Supplementary Table 1**). Surprisingly, when SignalP-4.1 was used, one of these 28 (EPWB_v2c3230) was excluded, and 70 additional putative secreted proteins were identified. Our manual inspection of these prediction results found that many of the putative secreted proteins identified by SignalP-4.1 are likely to be false positives (e.g., ribosomal proteins). In comparison, only one of the 28 candidates identified by SignalP-5.0 was an obvious false positive (EPWB_v2c2520; ATP-dependent Zn protease). Based on these findings, 27 putative secreted proteins predicted by SignalP-5.0 (excluding EPWB_v2c2520) were used for downstream analysis.

Among these 27 candidates, only one (EpWB_v2c2530) with uncharacterized function was found in PMU (**Figure 6**), which is different from the previous report that the many of the 56 secreted AYWB protein (SAP) genes in '*Ca*. P. asteris' were associated with PMUs (Bai et al., 2009). This result may be explained by the fact that only one PMU was found in '*Ca*. P. aurantifolia' NCHU2014, which contains PMU-associated genes (*dnaG*, *dnaB*, *tmk*, *smc*, *hflB*, *himA*, *ssb, and rpoD*) and is closely related to those of '*Ca*. P. luffae' NCHU2019 and '*Ca*. P. ziziphi' Jwb-nky in gene content and organization (**Figure 6**).

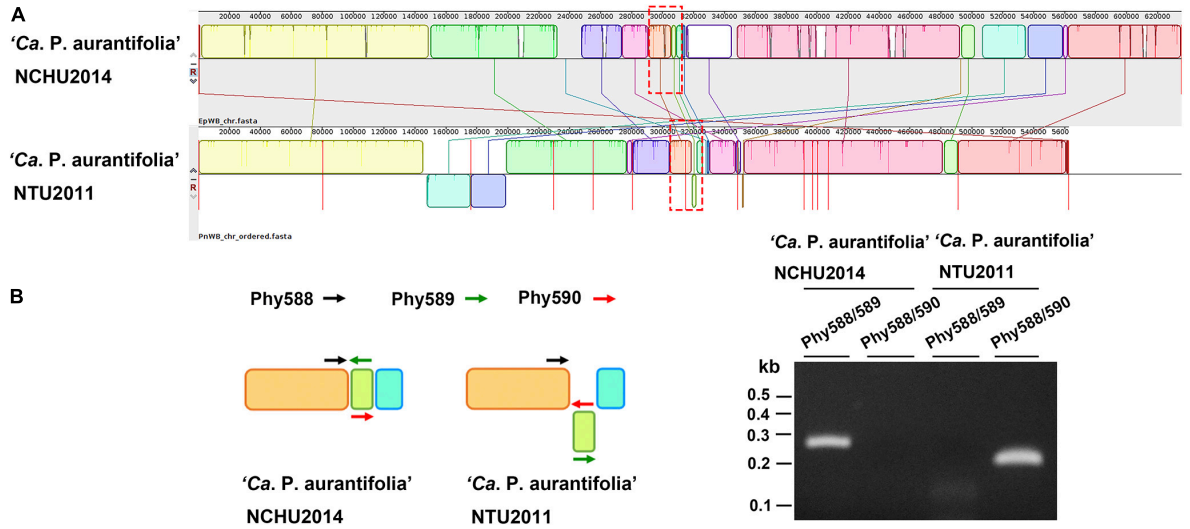
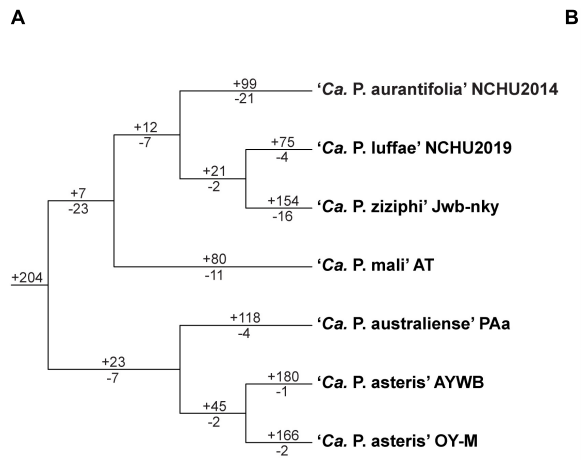


FIGURE 4 | Pairwise alignment of linearized genomes of '*Ca*. P. aurantifolia' strains NCHU2014 and NTU2011. **(A)** The complete genome of '*Ca*. P. aurantifolia' NCHU2014 was aligned with the draft genome of '*Ca*. P. aurantifolia' NTU2011. **(B)** Schematic diagram represented the inversion rearrangements indicated by red-dashed rectangles in **(A)** and the specific primers designed for PCR assays (left). The DNA fragments corresponding to the breakpoint of inverted structures in both chromosomal genomes were amplified (right). Primers: Phy588 (5′-GTTACTTTCGTGATTAATTCC-3′), Phy589 (5′-GAATTTACAACAAGCTCAATTGG-3′), Phy560 (5′-CTTGACTAATTTACTTTCGTCGC-3′).



B

	NCHU 2014	NCHU 2019	Jwb-nky	AT	PAa	AYWB	OY-M
NCHU 2014	446	313	293	284	300	306	303
NCHU 2019		462	340	313	322	319	319
Jwb-nky			537	301	310	311	312
AT				431	312	311	316
PAa					517	368	365
AYWB						620	418
OY-M							610

FIGURE 5 | Distribution patterns of homologous gene clusters. **(A)** Phylogenetic distribution. The cladogram was based on a maximum likelihood phylogeny inferred using 204 shared single-copy genes; the concatenated alignment contains 72,653 aligned amino acid sites. Numbers above a branch and preceded by a " + " sign indicate the number of clusters that are uniquely present in all daughter lineages; numbers below a branch and preceded by a "–" sign indicate the number of clusters that are uniquely absent. **(B)** Numbers on the diagonal indicate the counts of homologous gene clusters found in each of the *Candidatus* species genomes; numbers above the diagonal indicate the counts of homologous gene clusters shared in pairwise comparisons.

In comparison, the chromosome of '*Ca*. P. asteris' AY-WB harbors at least five PMU regions (Bai et al., 2006, 2009). Nevertheless, two thirds of putative secreted protein genes found in '*Ca*. P. aurantifolia' NCHU2014 formed clusters and closely located on both sides of PMU (**Figure 2**). In addition, 12 out of the 13 SAP homologs in the NCHU2014 genome are located in these clusters. The only exception is the homolog of SAP11 ($SAP11_{EpWB}$; EPWB_v2c3970), which is located far from these clusters. Comparison of putative effector gene content revealed that '*Ca*. P. aurantifolia' NCHU2014 harbors SAP11 (Bai et al., 2009) and SAP54 (MacLean et al., 2011) homologs while lacking homologs o SAP05 (Huang et al., 2021) or TENGU (Hoshi et al., 2009; **Figure 7** and **Supplementary Table 2**).

Phytoplasma $SAP11_{EpWB}$ Destabilizes *Arabidopsis* Class II CYC/TB1-TCPs

Previously, it has been demonstrated that SAP11 has the ability to destabilize CYC/TB1 (CYCLOIDEA/TEOSINTE-BRANCHED1)-TCPs (TCP12 and TCP18), leading to the proliferation of axillary meristems (Chang et al., 2018;

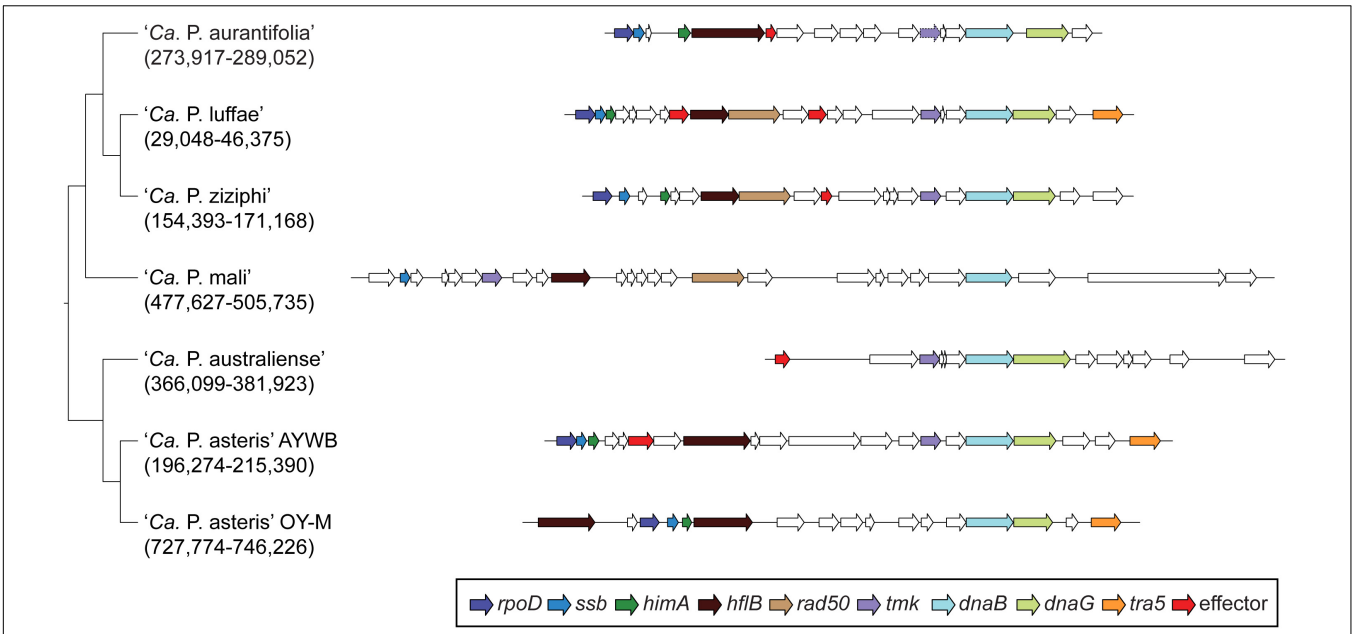


FIGURE 6 | Gene organization of PMUs. One representative PMU was selected from each genome for visualization, the chromosomal locations are labeled below the *Candidatus* species names. PMU-associated genes are color-coded according to annotation. Additionally, putative effector genes are colored in red. Intact genes are drawn with solid borders, putative pseudogenes are drawn with dotted borders.

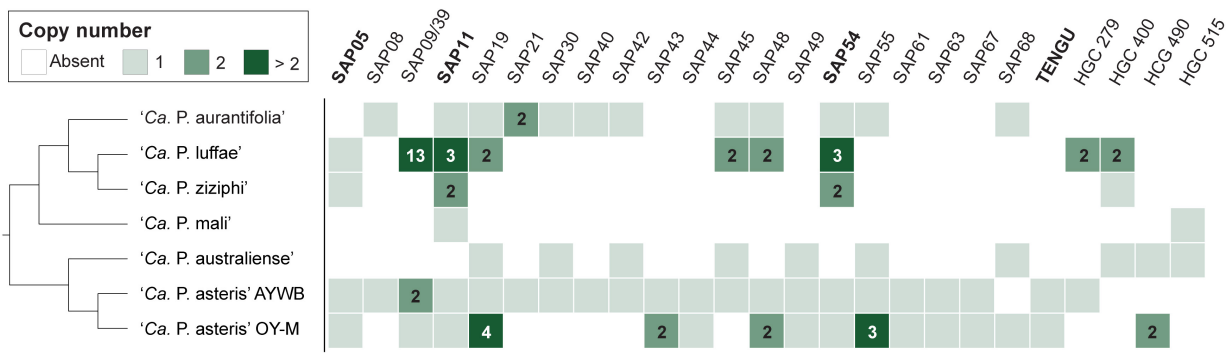
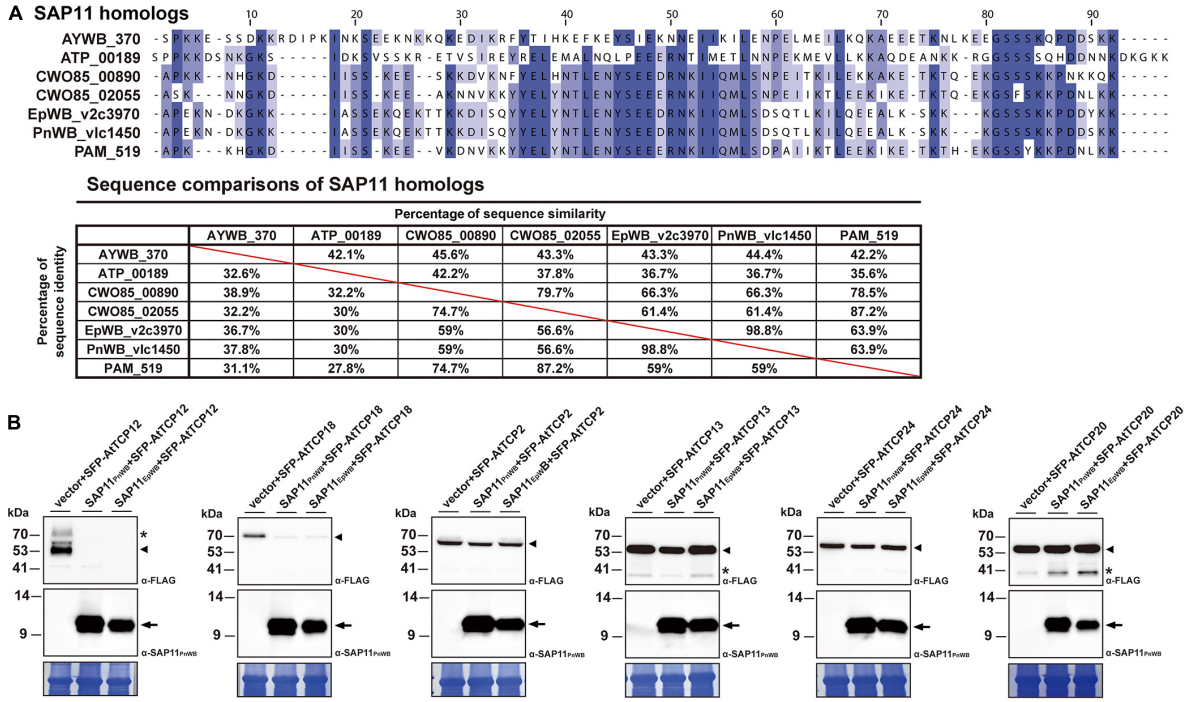


FIGURE 7 | Distribution of genes that encode putative secreted proteins among representative phytoplasmas. Only those found in at least two genomes are presented, the gene copy numbers are presented in the form of a heat map. The gene names are labeled according to the '*Ca*. P. asteris' isolates AYWB or OY-M annotation when available; the remaining ones are labeled according the homologous gene cluster (HGC) IDs of this work. Those effectors genes that have been experimentally characterized (i.e., SAP05, SAP11, SAP54, and TENGU) are highlighted in bold. '*Ca*. P. aurantifolia' NCHU2014 has two divergent SAP05 homologs (locus tags EPWB_v2c2740 and EPWB_v2c3230) that do not meet the criteria for inclusion in the figure.

Wang N. et al., 2018; Pecher et al., 2019). The $SAP11_{EpWB}$ identified in '*Ca*. P. aurantifolia' NCHU2014 shares a high degree of amino acid sequence identity (98.8%) with $SAP11_{PnWB}$ but only has 36.7% sequence identity with $SAP11_{AYWB}$ (**Figure 8A**). To understand the ability of $SAP11_{EpWB}$ in destabilizing TCP transcription factors, co-expression of $SAP11_{EpWB}$ and FLAG-tagged *A. thaliana* TCPs were conducted in *N. benthamiana* using agroinfiltration. Similar to $SAP11_{PnWB}$, $SAP11_{EpWB}$ exhibited a strong ability to destabilize class II CYC/TB1-TCPs (AtTCP12 and AtTCP18) but did not destabilize class II CIN-TCPs (AtTCP2, AtTCP13, and AtTCP24) as well as class I PCF-TCP (AtTCP20) (**Figure 8B**). As a control, the AtTCP12 and AtTCP18 were not decreased in abundance with the presence of the vector alone. Despite the fact that $SAP11_{EpWB}$ display one amino acid difference with $SAP11_{PnWB}$ in the C-terminal end, the potential biochemical activities in destabilizing CYC/TB1-TCP transcription factors were similar to each other. This is consistent with the fact that SAP11 lacking C-terminus (SAP11ΔC) still has the abilities for TCP binding and destabilization (Sugio et al., 2014). Because another phytoplasma effector, TENGU, with the ability to enhance the proliferation of axillary meristems, was not found in '*Ca*. P. aurantifolia' NCHU2014, $SAP11_{EpWB}$ would be the core virulence factor to induce witches' broom symptom in diseased plants (**Figures 1A–C**).



Percentage of sequence identity	AYWB_370	ATP_00189	CWO85_00890	CWO85_02055	EpWB_v2c3970	PnWB_vlc1450	PAM_519
AYWB_370		42.1%	45.6%	43.3%	43.3%	44.4%	42.2%
ATP_00189	32.6%		42.2%	37.8%	36.7%	36.7%	35.6%
CWO85_00890	38.9%	32.2%		79.7%	66.3%	66.3%	78.5%
CWO85_02055	32.2%	30%	74.7%		61.4%	61.4%	87.2%
EpWB_v2c3970	36.7%	30%	59%	56.6%		98.8%	63.9%
PnWB_vlc1450	37.8%	30%	59%	56.6%	98.8%		63.9%
PAM_519	31.1%	27.8%	74.7%	87.2%	59%	59%	

FIGURE 8 | Destabilization of *Arabidopsis* CYC/TB1-TCPs transcription factors by SAP11 of '*Ca*. P. aurantifolia' NCHU2014. **(A)** Sequence comparison of SAP11 homologs identified in '*Ca*. P. asteris' AYWB (AYWB_370), '*Ca*. P. mali' AT (ATP_00189), '*Ca*. P. ziziphi' Jwb-nky (CWO85_00890), '*Ca*. P. ziziphi' Jwb-nky (CWO85_02055), '*Ca*. P. aurantifoli' NCHU2014 (EpWB_v2c3970), '*Ca*. P. aurantifolia' NTU2011 (PnWB_v1c1450), and '*Ca*. P. asteris' OY-M (PAM_519). SAP11 homologs without the signal peptide were aligned by MEGA 7.0 using ClustalW. The sequence alignment was then edited by Jalview software in which identical residues are shaded in blue (upper panel). The color gradient indicates the level of sequence conservation at each position. The sequence identity and sequence similarity between SAP11 homologs are presented on the lower panel. **(B)** After transient co-expression in *N. benthamiana*, the relative abundance levels of FLAG-tagged PCF-TCP (AtTCP20), CIN-TCPs (AtTCP2, AtTCP13, and AtTCP24), and CYC/TB1-TCPs (AtTCP12 and AtTCP18) were examined in the presence of SAP11 effectors through Western blotting. Monoclonal antibody against the FLAG-tagged TCPs (upper panel) and polyclonal antibody against SAP11 effectors (middle panel) were used. As a loading control, the large subunit of Rubisco was visualized with Coomassie Brilliant Blue staining (lower panel). Non-specific bandings recognized by antibodies are indicated by asterisks.

DISCUSSION

In general, phytoplasmas contain one circular chromosome smaller than 1,000 kb in size (Oshima et al., 2013). In this study, we found that '*Ca*. P. aurantifolia' NCHU2014 has the smallest circular chromosome with 635,584 bp among the phytoplasmas with complete genome sequences available, and '*Ca*. P. mali' AT has an even smaller linear chromosome with 601,943 bp (**Table 2**). In the genome of '*Ca*. P. aurantifolia' NCHU2014, only one PMU and one transposase gene (*tra5*) were annotated (**Figures 2**, **6**). This may indicate that events of insertion and recombination occurred rarely and, in part, resulted in a small chromosomal size of '*Ca*. P. aurantifolia' NCHU2014. Consistently, only two putative PMUs and one *tra5* gene were annotated in the small genome of '*Ca*. P. mali' AT (Kube et al., 2008). In contrast, multiple putative PMUs and *tra5* genes were annotated in the genomes of '*Ca*. P. asteris' AYWB and OY-M, which have larger chromosomes (Bai et al., 2006; Arashida et al., 2008; Oshima et al., 2013).

A total of 27 genes encoding putative secreted proteins were identified in the genome of '*Ca*. P. aurantifolia' NCHU2014. Among them, 13 were identified as homologs of those present in '*Ca*. P. asteris' AYWB and formed clusters near PMU, except $SAP11_{EpWB}$ (**Figures 2**, **7** and **Supplementary Table 2**). These results are consistent with previous studies regarding that effectors associated with PMUs may be transferred horizontally between different phytoplasmas (Chung et al., 2013; Cho et al., 2019). The SAP homologs with potential biochemical activities of destabilizing plant transcription factors were also identified in '*Ca*. P. aurantifolia' NCHU2014 (**Figure 7**). However, the sequence identities of those putative effectors between '*Ca*. P. aurantifolia' NCHU2014 (16SrII) and '*Ca*. P. asteris' AYWB (16SrI) are generally low. Among them, $SAP11_{EpWB}$ shares 36.7% sequence identity with $SAP11_{AYWB}$, and $SAP54_{EpWB}$ shares 60.4% sequence identity with $SAP54_{AYWB}$ (**Figure 8A** and **Supplementary Figure 3**). Nevertheless, $SAP11_{EpWB}$ was characterized with high activity in destabilizing AtTCP12 and AtTCP18, the integrators of branching signals (**Figure 8B**). The protein sequence of $SAP54_{EpWB}$ is identical with PHYL1 of '*Ca*. P. aurantifolia' NTU2011 (**Supplementary Figure 3**), which is demonstrated to induce the degradation of the floral meristem identity protein APETALA1 and floral organ identity proteins SEPALLATA1/2/3/4 in a proteasome-dependent

manner (Iwabuchi et al., 2020). These morphological changes of host plants, including increases of branching and young leaves, are expected to improve the fitness of phytoplasmas and their insect vectors, which may facilitate the spread of phytoplasma diseases (Tomkins et al., 2018; Dermastia, 2019).

Increasing evidence suggests that climate change, particularly global warming, is considered to play a role in facilitating the spread of phytoplasma diseases through the population dynamics of insect vectors (Krishnareddy, 2013). Recently, we demonstrated that multiple plant species, including soybean (*Glycine max* L.), mungbean (*Vigna radiata* L.), snake gourd (*Trichosanthes cucumerina* L.), threeflower tickclover (*Desmodium triflorum*), lilac tasselflower (*Emilia sonchifolia*), and *Ixeris Chinensis* could be attacked by 16SrII-V subgroup phytoplasmas in Taiwan (Chen et al., 2021; Chien et al., 2021a,b,c; Wang et al., 2021; Weng et al., 2021). Thus, a better understanding of the 16SrII-V subgroup of phytoplasmas is required to develop effective strategies to combat phytoplasma diseases, which represent an emerging threat to agriculture in Taiwan. In this study, a high-accuracy and complete phytoplasma genome was obtained based on the enrichment of uncultivated phytoplasma cells through immunoprecipitation to accelerate the understanding of pathogenicity of '*Ca*. P. aurantifolia' NCHU2014. Although up to 36% of the reads obtained with the antibody-enrichment procedure were mapped to the phytoplasma genome, reads originating from the plant host still account for a large portion of the total reads. This result likely stems from interactions of phytoplasmas with plant cells, which results in the co-purification of plant cells during the procedure. For further improvement, selecting infected plant samples with high phytoplasma titers may help. However, this is challenging because the number of phytoplasma cells can vary among individual plants propagated from the same infected plant (Christensen et al., 2004). Alternatively, further purification of the antibody used for immunoprecipitation may also improve the enrichment results. However, the polyclonal antibody that specifically recognizes '*Ca*. P. aurantifolia' NCHU2014 might not be suitable against a wide range of phytoplasmas because Imp proteins are highly variable and show low similarities between phytoplasmas even in the same species (Morton et al., 2003; Kakizawa et al., 2006b). Thus, generation of antibodies against individual Imps are necessary to enrich a wide range of phytoplasmas by immunoprecipitation. Moreover, those antibodies could be very useful for phytoplasma examinations during field surveys as Imps are highly abundant proteins in the cell membrane of phytoplasmas.

DATA AVAILABILITY STATEMENT

The complete genome of '*Ca*. P. aurantifolia' strain NCHU2014 has been deposited under the accession numbers CP040925 (chromosome) and CP040926 (plasmid). This genome sequencing project and the associated raw reads were deposited in the NCBI under BioProject PRJNA294131.

AUTHOR CONTRIBUTIONS

CT, Y-YC, C-JW, C-WW, and J-YY performed the experiments. Y-CL, J-RL, LC, and C-HK analyzed the data. CT, Y-CL, C-WW, Y-CC, and J-YY prepared the figures and **Supplementary Material**. J-YY and C-HK designed the experiments, acquired the funding, wrote the manuscript, and supervised the project. All authors contributed to the article and approved the submitted version.

FUNDING

This work was financially supported in part by grants-in-aid from the Ministry of Science and Technology (110-2628-B-005-002) and the Advanced Plant Biotechnology Center from the Featured Areas Research Center Program within the framework of the Higher Education Sprout Project by the Ministry of Education (MOE) in Taiwan to J-YY. Additional funding was provided by Academia Sinica and the Ministry of Science and Technology (106-2311-B-001-028-MY3) of Taiwan to C-HK.

ACKNOWLEDGMENTS

We thank Shu-Ting Cho for technical assistance.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2021.766221/full#supplementary-material

Supplementary Figure 1 | Molecular examinations of '*Ca*. P. aurantifolia' NCHU2014-infected periwinkle (*Catharanthus roseus*) and '*Ca*. P. luffae' NCHU2019-infected loofah (*Luffa aegyptiaca*). Genomic DNA samples prepared from healthy (H) and symptomatic (S) plants were examined for the 16S rRNA gene by nested PCR with the universal primers P1/P7 followed by R16F2n/R16R2. A DNA fragment of expected size 1.2 kb was observed only in the symptomatic plants and none in the healthy plants. Arrow indicates the 1.2 kb DNA fragment of phytoplasma 16S rRNA gene.

Supplementary Figure 2 | The pairwise genome alignment of '*Ca*. P. aurantifolia' strains NCHU2014 (CP040925) and NTU2011 (NZ_AMWZ01000001.1-13.1). Matches on the same strand and the opposite strand are labeled in red and blue, respectively.

Supplementary Figure 3 | Sequence comparison of SAP54/PHYL1 homologs identified in '*Ca*. P. asteris' AYWB (AYWB_224), '*Ca*. P. ziziphi' Jwb-nky (CWO85_00800), '*Ca*. P. ziziphi' Jwb-nky (CWO85_03080), '*Ca*. P. aurantifolia' NCHU2014 (EpWB_v2c2400), '*Ca*. P. aurantifolia' NTU2011 (PnWB_v1c0190), and '*Ca*. P. asteris' OY-M (PAM_049). SAP54/PHYL1 homologs without the signal peptide were aligned by MEGA 7.0 using ClustalW. The sequence alignment was then edited by Jalview software in which identical residues are shaded in blue (upper panel). The color gradient indicates the level of sequence conservation at each position. The sequence identity and sequence similarity between SAP54/PHYL1 homologs are presented on the lower panel.

Supplementary Table 1 | Prediction of putative secreted proteins based on SignalP-4.1 and SignalP-5.0.

Supplementary Table 2 | List of genes that encode putative secreted proteins. The genes are organized by the homologous gene clusters (HGC) and identified by locus tags. Annotation available from the AYWB or OY-M genomes are provided in the "note" field.

REFERENCES

Andersen, M. T., Liefting, L. W., Havukkala, I., and Beever, R. E. (2013). Comparison of the complete genome sequence of two closely related isolates of '*Candidatus* Phytoplasma australiense' reveals genome plasticity. *BMC Genomics* 14:529. doi: 10.1186/1471-2164-14-529

Arashida, R., Kakizawa, S., Hoshi, A., Ishii, Y., Jung, H. Y., Kagiwada, S., et al. (2008). Heterogeneic dynamics of the structures of multiple gene clusters in two pathogenetically different lines originating from the same phytoplasma. *DNA Cell Biol.* 27, 209–217. doi: 10.1089/dna.2007.0654

Armenteros, J. J. A., Tsirigos, K. D., Sønderby, C. K., Petersen, T. N., Winther, O., Brunak, S., et al. (2019). SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat. Biotechnol.* 37, 420–423. doi: 10.1038/s41587-019-0036-z

Bai, X., Correa, V. R., Toruño, T. Y., Ammar, E.-D., Kamoun, S., and Hogenhout, S. A. (2009). AY-WB phytoplasma secretes a protein that targets plant cell nuclei. *Mol. Plant Microbe Interact.* 22, 18–30. doi: 10.1094/MPMI-22-1-0018

Bai, X., Zhang, J., Ewing, A., Miller, S. A., Jancso Radek, A., Shevchenko, D. V., et al. (2006). Living with genome instability: the adaptation of phytoplasmas to diverse environments of their insect and plant hosts. *J. Bacteriol.* 188, 3682–3696. doi: 10.1128/JB.188.10.3682-3696.2006

Benson, D. A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Ostell, J., Pruitt, K. D., et al. (2018). GenBank. *Nucleic Acids Res.* 46, D41–D47.

Boonrod, K., Munteanu, B., Jarausch, B., Jarausch, W., and Krczal, G. (2012). An immunodominant membrane protein (Imp) of '*Candidatus* Phytoplasma mali' binds to plant actin. *Mol. Plant Microbe Interact.* 25, 889–895. doi: 10.1094/MPMI-11-11-0303

Chang, S.-H., Cho, S.-T., Chen, C.-L., Yang, J.-Y., and Kuo, C.-H. (2015). Draft genome sequence of a 16SrII-A subgroup phytoplasma associated with purple coneflower (Echinacea purpurea) witches' broom disease in Taiwan. *Genome Announc.* 3:e01398-15. doi: 10.1128/genomeA.01398-15

Chang, S. H., Tan, C. M., Wu, C.-T., Lin, T.-H., Jiang, S.-Y., Liu, R.-C., et al. (2018). Alterations of plant architecture and phase transition by the phytoplasma virulence factor SAP11. *J. Exp. Bot.* 69, 5389–5401. doi: 10.1093/jxb/ery318

Chen, Y.-M., Chien, Y.-Y., Chen, Y.-K., Liao, P.-Q., Tan, C. M., Chiu, Y.-C., et al. (2021). Identification of 16SrII-V phytoplasma associated with mungbean phyllody disease in Taiwan. *Plant Dis.* [Epub ahead print]. doi: 10.1094/PDIS-12-20-2683-SC

Chien, Y.-Y., Tan, C. M., Kung, Y.-C., Lee, Y.-C., Chiu, Y.-C., and Yang, J.-Y. (2021a). *Ixeris* chinensis is a new host for peanut witches' broom phytoplasma, a 16SrII-V subgroup strain, in Taiwan. *Plant Dis.* 105:210. doi: 10.1094/PDIS-06-20-1302-PDN

Chien, Y.-Y., Tan, C. M., Kung, Y.-C., Lee, Y.-C., Chiu, Y.-C., and Yang, J.-Y. (2021b). Lilac tasselflower (*Emilia sonchifolia*) is a new host for peanut witches' broom phytoplasma, a 16SrII-V subgroup strain in Taiwan. *Plant Dis.* 105:211. doi: 10.1094/PDIS-06-20-1304-PDN

Chien, Y.-Y., Tan, C. M., Kung, Y.-C., Lee, Y.-C., Chiu, Y.-C., and Yang, J.-Y. (2021c). Threeflower tickclover (*Desmodium triflorum*) is a new host for peanut witches' broom phytoplasma, a 16SrII-V subgroup strain, in Taiwan. *Plant Dis.* 105:209. doi: 10.1094/PDIS-06-20-1303-PDN

Cho, S.-T., Lin, C.-P., and Kuo, C.-H. (2019). Genomic characterization of the periwinkle leaf yellowing (PLY) phytoplasmas in Taiwan. *Front. Microbiol.* 10:2194. doi: 10.3389/fmicb.2019.02194

Cho, S.-T., Zwolińska, A., Huang, W., Wouters, R. H. M., Mugford, S. T., Hogenhout, S. A., et al. (2020). Complete genome sequence of "*Candidatus* Phytoplasma asteris" RP166, a plant pathogen associated with rapeseed phyllody disease in Poland. *Microbiol. Resour. Announc.* 9:e00760-20. doi: 10.1128/MRA.00760-20

Christensen, N. M., Axelsen, K. B., Nicolaisen, M., and Schulz, A. (2005). Phytoplasmas and their interactions with hosts. *Trends Plant Sci.* 10, 526–535.

Christensen, N. M., Nicolaisen, M., Hansen, M., and Schultz, A. (2004). Distribution of phytoplasmas in infected plants as revealed by real time PCR and bioimaging. *Mol. Plant Microbe Interact.* 17, 1175–1184. doi: 10.1094/MPMI.2004.17.11.1175

Chung, W. C., Chen, L. L., Lo, W. S., Lin, C. P., and Kuo, C. H. (2013). Comparative analysis of the peanut witches'-broom phytoplasma genome reveals horizontal transfer of potential mobile units and effectors. *PLoS One* 8:e62770. doi: 10.1371/journal.pone.0062770

Dermastia, M. (2019). Plant hormones in phytoplasma infected plants. *Front. Plant Sci.* 10:477. doi: 10.3389/fpls.2019.00477

Galetto, L., Bosco, D., Balestrini, R., Genre, A., Fletcher, J., and Marzachì, C. (2011). The major antigenic membrane protein of "*Candidatus* Phytoplasma asteris" selectively interacts with ATP synthase and actin of leafhopper vectors. *PLoS One* 6:e22571. doi: 10.1371/journal.pone.0022571

Garcion, C., Béven, L., and Foissac, X. (2021). Comparison of current methods for signal peptide prediction in phytoplasmas. *Front. Microbiol.* 12:661524. doi: 10.3389/fmicb.2021.661524

Gurr, G. M., Johnson, A. C., Ash, G. J., Wilson, B. A. L., Ero, M. M., Pilotti, C. A., et al. (2016). Coconut lethal yellowing diseases: a phytoplasma threat to palms of global economic and social significance. *Front. Plant Sci.* 7:1521. doi: 10.3389/fpls.2016.01521

Hogenhout, S. A., Oshima, K., Ammar, E. L. D., Kakizawa, S., Kingdom, H. N., and Namba, S. (2008). Phytoplasmas: bacteria that manipulate plants and insects. *Mol. Plant Pathol.* 9, 403–423. doi: 10.1111/j.1364-3703.2008.00472.x

Hogenhout, S. A., and Seruga Music, M. (2009). "Phytoplasma genomics, from sequencing to comparative and functional genomics-what have we learnt?" in *Phytoplasmas: Genomes, Plant Hosts and Vectors*, eds P. G. Weintraub and P. Jones (Wallingford: CABI), 19–36. doi: 10.1079/9781845935306.0019

Hoshi, A., Oshima, K., Kakizawa, S., Ishii, Y., Ozeki, J., Hashimoto, M., et al. (2009). A unique virulence factor for proliferation and dwarfism in plants identified from a phytopathogenic bacterium. *Proc. Natl. Acad. Sci. U.S.A.* 106, 6416–6421. doi: 10.1073/pnas.0813038106

Huang, W., MacLean, A. M., Sugio, A., Maqbool, A., Busscher, M., Cho, S.-T., et al. (2021). Parasitic modulation of host development by ubiquitin-independent protein degradation. *Cell* 184, 5201–5214.e12. doi: 10.1016/j.cell.2021.08.029

Hyatt, D., Chen, G.-L., LoCascio, P. F., Land, M. L., Larimer, F. W., and Hauser, L. J. (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 11:119. doi: 10.1186/1471-2105-11-119

Iwabuchi, N., Kitazawa, Y., Maejima, K., Koinuma, H., Miyazaki, A., Matsumoto, O., et al. (2020). Functional variation in phyllogen, a phyllody-inducing phytoplasma effector family, attributable to a single amino acid polymorphism. *Mol. Plant Pathol.* 21, 1322–1336. doi: 10.1111/mpp.12981

Kakizawa, S., Oshima, K., Ishii, Y., Hoshi, A., Maejima, K., Jung, H.-Y., et al. (2009). Cloning of immunodominant membrane protein genes of phytoplasmas and their in planta expression. *FEMS Microbiol. Lett.* 293, 92–101. doi: 10.1111/j.1574-6968.2009.01509.x

Kakizawa, S., Oshima, K., and Namba, S. (2006b). Diversity and functional importance of phytoplasma membrane proteins. *Trends Microbiol.* 14, 254–256. doi: 10.1016/j.tim.2006.04.008

Kakizawa, S., Oshima, K., Jung, H.-Y., Suzuki, S., Nishigawa, H., Arashida, R., et al. (2006a). Positive selection acting on a surface membrane protein of the plant-pathogenic phytoplasmas. *J. Bacteriol.* 188, 3424–3428. doi: 10.1128/JB.188.9.3424-3428.2006

Kanehisa, M., Sato, Y., and Morishima, K. (2016). BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. *J. Mol. Biol.* 428, 726–731. doi: 10.1016/j.jmb.2015.11.006

Kellner, F., Kim, J., Clavijo, B. J., Hamilton, J. P., Childs, K. L., Vaillancourt, B., et al. (2015). Genome-guided investigation of plant natural product biosynthesis. *Plant J.* 82, 680–692. doi: 10.1111/tpj.12827

Konnerth, A., Krczal, G., and Boonrod, K. (2016). Immunodominant membrane proteins of phytoplasmas. *Microbiology* 162, 1267–1273. doi: 10.1099/mic.0.000331

Krishnareddy, M. (2013). "Impact of climate change on insect vectors and vector-borne plant viruses and phytoplasma," in *Climate-Resilient Horticulture: Adaptation and Mitigation Strategies*, eds H. Singh, N. Rao, and K. Shivashankar (New Delhi: Springer). doi: 10.1007/978-81-322-0974-4_23

Ku, C., Lo, W.-S., and Kuo, C.-H. (2013a). Horizontal transfer of potential mobile units in phytoplasmas. *Mob. Genet. Elements* 3:e26145. doi: 10.4161/mge.26145

Ku, C., Chung, W.-C., Chen, L.-L., and Kuo, C.-H. (2013b). The complete plastid genome sequence of Madagascar periwinkle *Catharanthus roseus* (L.) G. Don: plastid genome evolution, molecular marker identification, and phylogenetic implications in asterids. *PLoS One* 8:e68518. doi: 10.1371/journal.pone.0068518

Kube, M., Mitrovic, J., Duduk, B., Rabus, R., and Seemüller, E. (2012). Current view on phytoplasma genomes and encoded metabolism. *Sci. World J.* 2012:185942. doi: 10.1100/2012/185942

Kube, M., Schneider, B., Kuhl, H., Dandekar, T., Heitmann, K., Migdoll, A. M., et al. (2008). The linear chromosome of the plant-pathogenic mycoplasma '*Candidatus* Phytoplasma mali'. *BMC Genomics* 9:306. doi: 10.1186/1471-2164-9-306

Kumari, S., Nagendran, K., Rai, A. B., Singh, B., Rao, G. P., and Bertaccini, A. (2019). Global status of phytoplasma diseases in vegetable crops. *Front. Microbiol.* 10:1349. doi: 10.3389/fmicb.2019.01349

Lagesen, K., Hallin, P. F., Rødland, E. A., Staerfeldt, H. H., Rognes, T., and Ussery, D. (2007). RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res.* 9, 3100–3108. doi: 10.1093/nar/gkm160

Lee, I. M., Davis, R. E., and Gundersen-Rindal, D. E. (2000). Phytoplasma: phytopathogenic mollicutes. *Annu. Rev. Microbiol.* 54, 221–255.

Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 34, 3094–3100. doi: 10.1093/bioinformatics/bty191

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 25, 1754–1760. doi: 10.1093/bioinformatics/btp324

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078–2079. doi: 10.1093/bioinformatics/btp352

Li, L., Stoeckert, C. J., and Roos, D. S. (2003). OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res.* 13, 2178–2189. doi: 10.1101/gr.1224503

Lowe, T. M., and Eddy, S. R. (1997). tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* 25, 955–964. doi: 10.1093/nar/25.5.955

MacLean, A. M., Orlovskis, Z., Kowitwanich, K., Zdziarska, A. M., Angenent, G. C., Immink, R. G., et al. (2014). Phytoplasma effector SAP54 hijacks plant reproduction by degrading MADS-box proteins and promotes insect colonization in a RAD23-dependent manner. *PLoS Biol.* 12:e1001835. doi: 10.1371/journal.pbio.1001835

MacLean, A. M., Sugio, A., Makarova, O. V., Findlay, K. C., Grieve, V. M., Toth, R., et al. (2011). Phytoplasma effector SAP54 induces indeterminate leaf-like flower development in Arabidopsis plants. *Plant Physiol.* 157, 831–841. doi: 10.1104/pp.111.181586

Maejima, K., Iwai, R., Himeno, M., Komatsu, K., Kitazawa, Y., Fujita, N., et al. (2014). Recognition of floral homeotic MADS domain transcription factors by a phytoplasmal effector, phyllogen, induces phyllody. *Plant J.* 78, 541–554. doi: 10.1111/tpj.12495

Marcone, C. (2014). Molecular biology and pathogenicity of phytoplasmas. *Ann. Appl. Biol.* 165, 199–221. doi: 10.1111/aab.12151

Morton, A., Davies, D. L., Blomquist, C. L., and Barbara, D. J. (2003). Characterization of homologues of the apple proliferation immunodominant membrane protein gene from three related phytoplasmas. *Mol. Plant Pathol.* 4, 109–114. doi: 10.1046/j.1364-3703.2003.00155.x

Namba, S. (2019). Molecular and biological properties of phytoplasmas. *Proc. Jpn. Acad. Ser. B Phys. Biol. Sci.* 95, 401–418. doi: 10.2183/pjab.95.028

Orlovskis, Z., Canale, M. C., Haryono, M., Lopes, J. R. S., Kuo, C. H., and Hogenhout, S. A. (2017). A few sequence polymorphisms among isolates of Maize bushy stunt phytoplasma associate with organ proliferation symptoms of infected maize plants. *Ann. Bot.* 119, 869–884. doi: 10.1093/aob/mcw213

Oshima, K., Ishii, Y., Kakizawa, S., Sugawara, K., Neriya, Y., Himeno, M., et al. (2011). Dramatic transcriptional changes in an intracellular parasite enable host switching between plant and insect. *PLoS One* 6:e23242. doi: 10.1371/journal.pone.0023242

Oshima, K., Kakizawa, S., Nishigawa, H., Jung, H. Y., Wei, W., Suzuki, S., et al. (2004). Reductive evolution suggested from the complete genome sequence of a plant-pathogenic phytoplasma. *Nat. Genet.* 36, 27–29. doi: 10.1038/ng1277

Oshima, K., Maejima, K., and Namba, S. (2013). Genomic and evolutionary aspects of phytoplasmas. *Front. Microbiol.* 4:230. doi: 10.3389/fmicb.2013.00230

Pecher, P., Moro, G., Canale, M. C., Capdevielle, S., Singh, A., MacLean, A., et al. (2019). Phytoplasma SAP11 effector destabilization of TCP transcription factors differentially impact development and defence of Arabidopsis versus maize. *PLoS Pathog.* 15:e1008035. doi: 10.1371/journal.ppat.1008035

Petersen, T. N., Brunak, S., von Heijne, G., and Nielsen, H. (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat. Methods* 8, 785–786. doi: 10.1038/nmeth.1701

Pierro, R., Semeraro, T., Luvisi, A., Garg, H., Vergine, M., De Bellis, L., et al. (2019). The distribution of phytoplasmas in south and east asia: an emerging threat to grapevine cultivation. *Front. Plant Sci.* 10:1108. doi: 10.3389/fpls.2019.01108

Polano, C., and Firrao, G. (2018). An effective pipeline based on relative coverage for the genome assembly of phytoplasmas and other fastidious prokaryotes. *Curr. Genomics* 19, 491–498. doi: 10.2174/1389202919666180314114628

Ricker, N., Qian, H., and Fulthorpe, R. R. (2012). The limitations of draft assemblies for understanding prokaryotic adaptation and evolution. *Genomics* 100, 167–175. doi: 10.1016/j.ygeno.2012.06.009

Robinson, J. T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., et al. (2011). Integrative genomics viewer. *Nat. Biotechnol.* 29, 24–26. doi: 10.1038/nbt.1754

Seruga Music, M., Samarzija, I., Hogenhout, S. A., Haryono, M., Cho, S.-T., and Kuo, C.-H. (2019). The genome of '*Candidatus* Phytoplasma solani' strain SA-1 is highly dynamic and prone to adopting foreign sequences. *Syst. Appl. Microbiol.* 42, 117–127. doi: 10.1016/j.syapm.2018.10.008

Sugio, A., Kingdom, H. N., MacLean, A. M., Grieve, V. M., and Hogenhout, S. A. (2011a). Phytoplasma protein effector SAP11 enhances insect vector reproduction by manipulating plant development and defense hormone biosynthesis. *Proc. Natl. Acad. Sci. U.S.A.* 108, E1254–E1263. doi: 10.1073/pnas.1105664108

Sugio, A., MacLean, A. M., Kingdom, H. N., Grieve, V. M., Manimekalai, R., and Hogenhout, S. A. (2011b). Diverse targets of phytoplasma effectors: from plant development to defense against insects. *Annu. Rev. Phytopathol.* 49, 175–195. doi: 10.1146/annurev-phyto-072910-095323

Sugio, A., MacLean, A. M., and Hogenhout, S. A. (2014). The small phytoplasma virulence effector SAP11 contains distinct domains required for nuclear targeting and CIN-TCP binding and destabilization. *New Phytol.* 202, 838–848. doi: 10.1111/nph.12721

Suzuki, S., Oshima, K., Kakizawa, S., Arashida, R., Jung, H.-Y., Yamaji, Y., et al. (2006). Interaction between the membrane protein of a pathogen and insect microfilament complex determines insect-vector specificity. *Proc. Natl. Acad. Sci. U.S.A.* 103, 4252–4257. doi: 10.1073/pnas.0508668103

Tan, C. M., Li, C.-H., Tsao, N.-W., Su, L.-W., Lu, Y.-T., Chang, S. H., et al. (2016). Phytoplasma SAP11 alters 3-isobutyl-2-methoxypyrazine biosynthesis in Nicotiana benthamiana by suppressing NbOMT1. *J. Exp. Bot.* 67, 4415–4425. doi: 10.1093/jxb/erw225

Tomkins, M., Kliot, A., Marée, A. F., and Hogenhout, S. A. (2018). A multi-layered mechanistic modelling approach to understand how effector genes extend beyond phytoplasma to modulate plant hosts, insect vectors and the environment. *Curr. Opin. Plant Biol.* 44, 39–48. doi: 10.1016/j.pbi.2018.02.002

Toruno, T. Y., Music, M. S., Simi, S., Nicolaisen, M., and Hogenhout, S. A. (2010). Phytoplasma PMU1 exists as linear chromosomal and circular extrachromosomal elements and has enhanced expression in insect vectors compared with plant hosts. *Mol. Microbiol.* 77, 1406–1415. doi: 10.1111/j.1365-2958.2010.07296.x

Tran-Nguyen, L. T., Kube, M., Schneider, B., Reinhardt, R., and Gibb, K. S. (2008). Comparative genome analysis of "*Candidatus* Phytoplasma australiense" (subgroup tuf-Australia I; rp-A) and "Ca. Phytoplasma asteris". Strains OY-M and AY-WB. *J. Bacteriol.* 190, 3979–3991. doi: 10.1128/JB.01301-07

Wang, C.-J., Chien, Y.-Y., Liao, P.-Q., Chiu, Y.-C., Chen, Y.-K., and Yang, J.-Y. (2021). First report of 16SrII-V phytoplasma associated with green manure soybean (*Glycine max* L.) in Taiwan. *Plant Dis.* [Epub ahead of print]. doi: 10.1094/PDIS-12-20-2714-PDN

Wang, J., Song, L., Jiao, Q., Yang, S., Gao, R., Lu, X., et al. (2018). Comparative genome analysis of jujube witches'-broom Phytoplasma, an obligate pathogen that causes jujube witches'-broom disease. *BMC Genomics* 19:689. doi: 10.1186/s12864-018-5075-1

Wang, N., Yang, H., Yin, Z., Liu, W., Sun, L., and Wu, Y. (2018). Phytoplasma effector SWP1 induces witches' broom symptom by destabilizing the TCP transcription factor BRANCHED1. *Mol. Plant Pathol.* 19, 2623–2634. doi: 10.1111/mpp.12733

Wei, W., Davis, R. E., Jomantiene, R., and Zhao, Y. (2008). Ancient, recurrent phage attacks and recombination shaped dynamic sequence-variable mosaics at the root of phytoplasma genome evolution. *Proc. Natl. Acad. Sci. U.S.A.* 105, 11827–11832. doi: 10.1073/pnas.0805237105

Weng, Y.-Y., Liou, W.-C., Chien, Y.-Y., Liao, P.-Q., Wang, C.-J., Chiu, Y.-C., et al. (2021). First report of 16SrII-V peanut witches' broom phytoplasma in snake gourd (*Trichosanthes cucumerina* L.) in Taiwan. *Plant Dis.* [Epub ahead of print]. doi: 10.1094/PDIS-12-20-2666-PDN

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 20 December 2021
doi: 10.3389/fmicb.2021.791756

Evaluating the Genetic Capacity of *Mycoplasmas* for Coenzyme A Biosynthesis in a Search for New Anti-mycoplasma Targets

*Tertius Alwyn Ras, Erick Strauss and Annelise Botes**

Department of Biochemistry, Stellenbosch University, Stellenbosch, South Africa

Mycoplasmas are responsible for a wide range of disease states in both humans and animals, in which their parasitic lifestyle has allowed them to reduce their genome sizes and curtail their biosynthetic capabilities. The subsequent dependence on their host offers a unique opportunity to explore pathways for obtaining and producing cofactors – such as coenzyme A (CoA) – as possible targets for the development of new anti-mycoplasma agents. CoA plays an essential role in energy and fatty acid metabolism and is required for membrane synthesis. However, our current lack of knowledge of the relevance and importance of the CoA biosynthesis pathway in mycoplasmas, and whether it could be bypassed within their pathogenic context, prevents further exploration of the potential of this pathway. In the universal, canonical CoA biosynthesis pathway, five enzymes are responsible for the production of CoA. Given the inconsistent presence of the genes that code for these enzymes across *Mycoplasma* genomes, this study set out to establish the genetic capacity of mycoplasmas to synthesize their own CoA *de novo*. Existing functional annotations and sequence, family, motif, and domain analysis of protein products were used to determine the existence of relevant genes in *Mycoplasma* genomes. We found that most *Mycoplasma* species do have the genetic capacity to synthesize CoA, but there was a differentiated prevalence of these genes across species. Phylogenetic analysis indicated that the phylogenetic position of a species could not be used to predict its enzyme-encoding gene combinations. Despite this, the final enzyme in the biosynthesis pathway – dephospho-coenzyme A kinase (DPCK) – was found to be the most common among the studied species, suggesting that it has the most potential as a target in the search for new broad-spectrum anti-mycoplasma agents.

Keywords: *Mycoplasma*, drug development, coenzyme A biosynthesis, biosynthetic variability, host dependence, dephospho-coenzyme A kinase, pantothenate kinase

OPEN ACCESS

Edited by:
Chih-Horng Kuo, Academia Sinica, Taiwan

Reviewed by:
Maria Suarez-Diez, Wageningen University and Research, Netherlands
Emilie Dordet Frisoni, Institut National de recherche pour l'agriculture, l'alimentation et l'environnement (INRAE), France

****Correspondence:***
Annelise Botes
annelise@sun.ac.za

Specialty section:
This article was submitted to Microbial Symbioses, a section of the journal Frontiers in Microbiology

Received: *08 October 2021*
Accepted: *22 November 2021*
Published: *20 December 2021*

Citation:
Ras TA, Strauss E and Botes A (2021) Evaluating the Genetic Capacity of Mycoplasmas for Coenzyme A Biosynthesis in a Search for New Anti-mycoplasma Targets. Front. Microbiol. 12:791756. doi: 10.3389/fmicb.2021.791756

INTRODUCTION

Mycoplasmas are responsible for a wide range of diseases in humans and animals, impacting health and economic activity – especially in the agricultural sector. They are cell wall-less bacteria of the class Mollicutes that are among the smallest known organisms capable of self-replication (Razin et al., 1998), with reduced genome sizes that have resulted in severely

curtailed biosynthetic capabilities (Rivera and Cedillo, 2015). Consequently, many mycoplasmas lead a parasitic lifestyle, in which they are dependent on their host for essential nutrients. This symbiotic relationship presents both a challenge and an opportunity for the development of new treatments for *Mycoplasma* infections. The challenge arises from the close relationship between pathogen and host, which makes it difficult to specifically target the *Mycoplasma*. In addition to this, antimicrobial resistance against commonly used cell wall synthesis inhibitors is a serious concern (Meyer and Van Rossum, 2014; Chernova et al., 2016) However, the mycoplasma's dependence on certain host-derived growth factors offers a unique opportunity for the discovery of inhibitors of a process or pathway that is central to this interaction, and that would be detrimental to the survival of the organism.

In this context, the mechanism by which mycoplasmas obtain the central metabolic cofactor coenzyme A (CoA) presents itself as an especially relevant case study. First, the enzymes involved in CoA biosynthesis have long been regarded as possible targets for antimicrobial development given their essential role in central energy and fatty acid metabolism, and the prediction that the bacterial enzymes could selectively be inhibited based on their divergence from those of their eukaryotic hosts (Spry et al., 2008; Moolman et al., 2014; Balish and Distelhorst, 2016). Second, CoA is also required for membrane biosynthesis and therefore growth of *Mycoplasma* (Pollack et al., 1997). Third, *in vitro* studies have shown that some of the genes encoding the enzymes responsible for CoA biosynthesis in mycoplasmas are dispensable. The extreme example is the organism with the "minimal bacterial genome" discovered following the genome reduction of *Mycoplasma mycoides*, which did not have any of the enzymes required for producing its own CoA (Yus et al., 2009; Hutchison et al., 2016). While *de novo* CoA biosynthesis may therefore also be a worthwhile target for the development of new anti-mycoplasma agents, our current lack of knowledge of the relevance and importance of this process in mycoplasmas, and whether it could be bypassed within their pathogenic context, prevents further exploration of the potential of this pathway.

In the universal, canonical CoA biosynthesis pathway five enzymes are responsible for the conversion of the precursor substrate pantothenic acid (Pan, or vitamin B_5) into CoA (**Figure 1A**). The first enzyme is pantothenate kinase (PanK), followed by CoaBC – a bifunctional protein in bacteria that has both phosphopantothenoylcysteine synthetase (PPCS) and phosphopantothenoylcysteine decarboxylase (PPCDC) activities. This is followed by phosphopantetheine adenylyltransferase (PPAT), with the last enzyme in the pathway being dephospho-coenzyme A kinase (DPCK; Strauss, 2010). An important possible deviation from the canonical pathway entails the use of the CoA degradation product pantetheine (PantSH) as an alternative substrate by PanK in a CoA salvage mechanism that allows for the bypassing of the PPCS and PPCDC activities of the CoaBC protein (**Figure 1A**). However, this is only possible for a certain subset of PanK enzymes, which occur as three distinct types: type I ($PanK_I$) and type III PanK ($PanK_{III}$) enzymes (encoded for by the *coaA* and *coaX* genes, respectively) that are found in prokaryotes, and type II PanKs ($PanK_{II}$) found in eukaryotes and selected *Staphylococcus* and *Bacillus* species. In particular, $PanK_{III}$ enzymes are not able to use PantSH as substrate and organisms with this type of PanK can therefore not make use of the CoA salvage pathway and bypass CoaBC. In addition, there has also been some evidence that certain cells are able to take up more advanced CoA biosynthetic intermediates, suggesting the possible existence of mechanisms whereby CoA could be obtained through the action of only PPAT and DPCK, or even DPCK alone (Sibon and Strauss, 2016). A schematic summary of all these mechanisms is presented in **Figure 1B**.

With this as background, we set out to establish the genetic capacity of *Mycoplasma* to synthesize their own CoA *de novo* using Pan as substrate, or through some truncated variations of the canonical pathway. Such an analysis is particularly warranted due to the inconsistent presence of the genes that code for the CoA biosynthesis enzymes across *Mycoplasma* genomes (Daugherty et al., 2002; Spry et al., 2008; Yus et al., 2009; De Jonge et al., 2013). While this could be a reflection of the parasitic lifestyle of these organisms, it could also be a result of insufficient annotation, since *Mycoplasma* genomes typically contain a large number of hypothetical protein-coding genes. Distinguishing between these possibilities is crucial if CoA biosynthesis is to be considered to hold any realistic potential for anti-mycoplasma drug development.

In this study, we determine the existence of genes encoding the CoA biosynthesis pathway enzymes among *Mycoplasma* genomes using existing functional annotations as well as sequence, family, motif, and domain analysis of protein products. We discovered an unexpectedly differentiated prevalence of these genes across *Mycoplasma* species, with a few specific combinations emerging. We therefore used phylogenetic analysis to determine the evolutionary relationship of the *Mycoplasma* species and the functional relationship of their CoA biosynthetic enzymes using 16S rRNA and amino acid data, respectively. This allowed us to uncover the enzyme targets that would hold the most potential for the development of new broad-spectrum or species-specific anti-mycoplasma agents.

MATERIALS AND METHODS

Selection of *Mycoplasma* Genomes Used in This Study

The annotated genomes of 62 *Mycoplasma* species were investigated in this study with some genomes being complete and others only available at contiguous sequence (contig) level (**Supplementary Table 1**). The 16S rRNA phylogeny of Wium et al. (2015) was used as starting point to select 28 *Mycoplasma* species that also had CoA biosynthesis pathway information available on the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway and/or SEED viewer subsystems databases. The species were selected to represent the typical 16S rRNA phylogenetic groupings (Weisburg et al., 1989; Johansson et al., 1998; Johansson and Petterson, 2002) as well as a range of

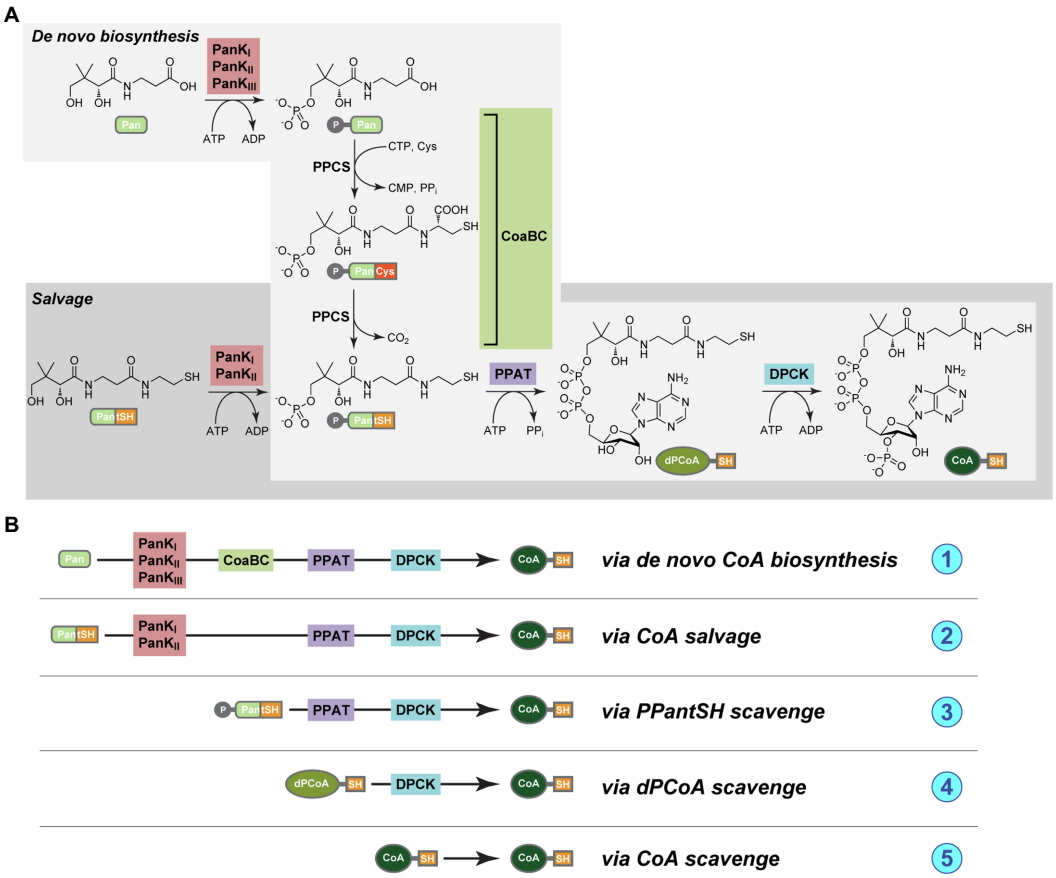


FIGURE 1 | (A) The pathways for *de novo* coenzyme A biosynthesis (light gray box) and coenzyme A (CoA) salvage (dark gray box). The compounds are pantothenic acid (Pan, or vitamin B_5), 4′-phosphopantothenic acid (P-Pan), pantetheine (PantSH), 4′-phosphopantetheine (PPantSH), dephospho-coenzyme A (dPCoA), CoA, and the enzymes pantothenate kinase (PanK, in red box), CoaBC [a bifunctional protein with both phosphopantothenoylcysteine synthetase (PPCS) and phosphopantothenoylcysteine decarboxylase (PPCDC) activities, in green box], phosphopantetheine adenylyltransferase (PPAT, in purple box), and finally dephospho-coenzyme A kinase (DPCK, blue box; Strauss, 2010). **(B)** A schematic representation of the various mechanisms whereby CoA could potentially be obtained, and of the enzyme combinations that would be required in each case. (1) *Via de novo* biosynthesis from Pan; (2) *via* salvage from PantSH; (3) *via* scavenge of PPantSH from the host or environment; (4) *via* scavenge of dPCoA from the host or environment; and (5) *via* scavenge of CoA from the host or environment.

hosts. To further increase the variety, additional species with pathway information were selected from the KEGG, SEED, and National Centre for Biotechnology Information (NCBI) databases. Due to our interest in ostrich-infecting *Mycoplasma*, the genome of *Mycoplasma* sp. Ms02 was also included in this study.

Given the exploratory nature of this study only one strain of a selected species was used. All the amino acid sequences were downloaded from the NCBI database and accession numbers are given in **Supplementary Table 2**.

Determining the Prevalence of Enzyme-Encoding Genes Using Existing Functional Annotations

Thirty-seven of the selected *Mycoplasma* species contained annotated gene information for one or more enzymes of the CoA biosynthesis pathway in the KEGG and/or SEED databases (Overbeek et al., 2005). The results of both databases were combined and compared to that on the NCBI database from which the amino acid sequences of the annotated genes were obtained. For species selected from NCBI that had no corresponding KEGG/SEED information available, the enzyme encoding genes were identified within their genomes using blastp searches with relevant amino acid sequences of annotated enzymes, identified *via* the KEGG or SEED databases, as query.

All the associated amino acid sequences were downloaded from the NCBI protein database for further protein a phylogenetic analysis (Pruitt et al., 2012). The arrangement of enzyme-encoding genes relative to one another within a genome was also determined in species with a complete genome or where genes were situated on the same contig within a draft genome.

Identification of Gene Homologs

For all the genomes in which one or more of the four CoA biosynthesis pathway enzyme-encoding genes were not annotated, gene homologs were identified using a combination of blastp and psi-blast searches (Altschul and Koonin, 1998). In such

cases, the amino acid sequences of annotated gene products were used as query. Identified sequences were deemed homologs if they exhibited an amino acid sequence identity of at least 30% over at least 70% of the query sequence. Identity was further confirmed using predicted protein family, domain, and motif information as well as the genomic location relative to other CoA enzyme-encoding genes.

Family, Motif, and Domain Analysis of Protein Products

Analysis of annotated and newly identified amino acid sequences of all the pathway enzymes for all the species were performed to confirm existing annotations as well as the identity of homologs as CoA biosynthetic enzymes. The workflow followed is schematically represented in **Figure 2**. NCBI Conserved Domain Database (CDD) v3.16 was used to predict the family relation and domains of protein products encoded by annotated CoA biosynthesis pathway enzyme-encoding genes as well as identified gene homologs. For additional verification, proteins were also analyzed using InterPro v65.0, which allows for several different member databases to be searched simultaneously (Blum et al., 2021).

Conserved motifs within the protein sequences of each enzyme were determined using the Multiple Expectation maximisation for Motif Elicitation (MEME) algorithm from the MEME suite v4.12.0. (Bailey et al., 2015). The search parameters were set to identify motifs of between six and 50 amino acids in length, with a maximum of four motifs per sequence. The predicted motifs were compared to known functional motifs of the corresponding enzymes.

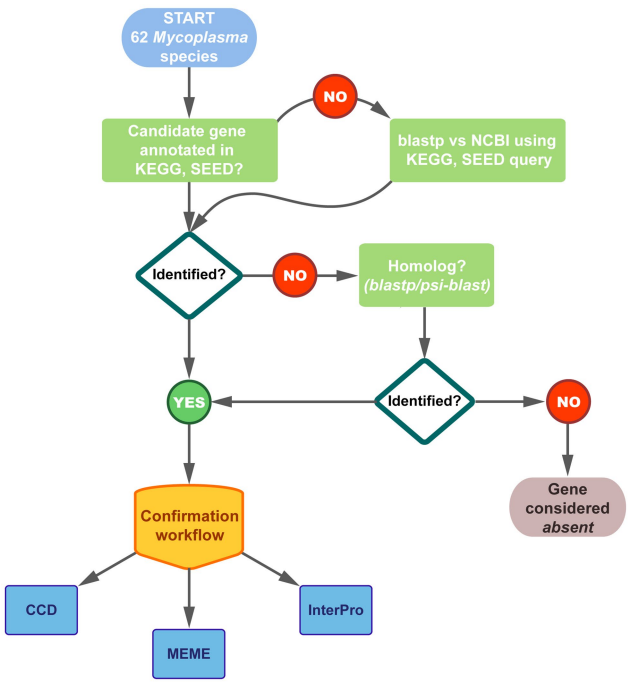


FIGURE 2 | Schematic representation of the workflow followed to identify the genes encoding CoA biosynthesis enzymes in the 62 studied *Mycoplasma* genomes. Genomes were selected from Kyoto Encyclopedia of Genes and Genomes (KEGG), SEED, and National Centre for Biotechnology Information (NCBI) databases. Thirty-seven of the species had annotations for one/more of the pathway genes in KEGG and/or SEED. Results were combined and compared to that on NCBI. For species on NCBI with no corresponding KEGG/SEED information, enzymes were identified through blastp searches. Fifty-four of the species had at least one of the CoA biosynthetic enzyme-encoding genes annotated and the identity of all these genes were confirmed using the indicated databases. Where no genes were annotated, homologs were searched using blastp or psi-blast applying a minimum threshold of 30% amino acid identity over minimum 70% of the query sequence. The confirmation workflow considered family, domain, and motif analyses using the indicated databases; see Materials and Methods for details.

Phylogenetic Analyses

For analysis of evolutionary relationships of *Mycoplasma* species, the 16S rRNA gene sequences were obtained from NCBI nucleotide database; the relevant accession numbers are listed in **Supplementary Table 2**. Included in this analysis were representative species of bacterial sister genera from which mycoplasmas have been shown to have evolved (Woese, 1987). These species, for which the 16S rRNA sequences were obtained from Genbank, are: *Clostridium innocuum* (KR364751.1), *Lactobacillus fermentum* (FJ462686.1), and *Streptococcus pneumoniae* (NR_115239.1). *Bacillus coahuilensis* (EF014447.1) was used as outgroup.

For analysis of the functional relationships of enzymes, their amino acid sequences were obtained from the NCBI protein database; the relevant accession numbers are listed in **Supplementary Table 2**. Amino acid sequences from the same representative species as before were included where available, and *B. coahuilensis* was again chosen as outgroup (**Supplementary Table 3**). No $PanK_{III}$ protein was included for *L. fermentum*, as it only possesses a $PanK_{I}$ enzyme. Similarly, no bifunctional CoaBC protein was included for *S. pneumoniae*, as it possesses separate PPCS and PPCDC proteins (Gerdes et al., 2002).

Using Clustal Omega (Sievers and Higgins, 2018) available at https://www.ebi.ac.uk/services, multiple sequence alignments (MSAs) were generated for 16S rRNA genes as well as the protein sequences of individual CoA biosynthesis enzymes. MSAs were manually adjusted in BioEdit v7.0.5.2 (Hall, 1999) to ensure optimal alignment, and where required, the Geneious v11.0.2 (Kearse et al., 2012) program due to its enhanced alignment functionalities. Amino acid sequences were used for the alignments since the nucleotide sequences of the respective enzyme-encoding genes are very heterogeneous and could not be aligned with confidence.

Maximum likelihood phylogenies were constructed using RAxML-HPC2 on XSEDE v8.2.10 (Stamatakis, 2014) *via* the CIPRES Science Gateway v3.3 web portal (Miller et al., 2010). Evaluation of clade support was performed with bootstrap analysis, which was set to automatically stop when the majority rule criterion is reached according to program recommendations. Bootstrap values ≥75% were considered resolved and well supported, values of ≤75% but ≥50% were considered resolved but moderately supported, and values ≤50% were considered poorly supported. Only values of ≥50% are indicated on the phylogenetic trees.

RESULTS

Determining the Prevalence of CoA Biosynthetic Enzyme-Encoding Genes Using Existing Functional Annotations

As a result of inconsistent annotation, a combination of database information from KEGG and SEED as well as blastp searches were used to find genes that had functional annotations corresponding to any of the enzymes involved in the CoA biosynthesis pathway (**Figure 2**). Of the 62 *Mycoplasma* species evaluated, 54 had at least one of the CoA biosynthetic enzyme-encoding genes annotated within their genome (**Table 1**). For species with pathway information on KEGG and SEED databases, it was found that the information between these databases did not always agree and in some instances also did not match genome information available on NCBI.

For those species with PanK-encoding genes, the protein products were indicated as "type III pantothenate kinase." The common gene annotation used in prokaryotes for $PanK_{III}$-encoding genes is *coaX*, but this annotation was not used in any of the included *Mycoplasma* genomes. Instead, genes were only referred to by their locus tag. There are, however, *Mycoplasma* genomes (not included in this study) in which the *coaX* annotation is used. Although, the PanK EC number (2.7.1.33) is correctly indicated in such instances, the protein product is incorrectly given as a transcriptional regulator due to initial incorrect annotation of this protein (Brand and Strauss, 2005).

Phosphopantothenoylcysteine synthetase and PPCDC activities were associated with a protein encoded by a single gene which in most genomes were annotated as *coaBC*, with the protein product indicated as a "bifunctional phosphopantothenoylcysteine decarboxylase/phosphopantothenate-cysteine ligase (CoaBC)." Within genomes used in this study, *Mycoplasma iowae* was the only exception and contained two different genes encoding for PPCS and PPCDC separately. However, these genes were not annotated as *coaB* and *coaC*, but only contained a locus tag with the protein product of PPCS indicated as a hypothetical protein. Separate genes are also found in all other strains of *M. iowae*, indicating that this separation is not due to a sequencing error in the genome of the strain chosen for this study. Although separate PPCS- and PPCDC-encoding genes were not found in any of the other species investigated in this study, this phenomenon is not unique to *M. iowae*. In the *Mycoplasma fermentans*, NCTC10117 genome (not used in this study) separate PPCS- and PPCDC-encoding genes can also be found and as is the case with all *M. iowae* strains, these genes were consistently separated by four nucleotides. Besides a locus tag as annotation, the separate PPCDC-encoding gene is often incorrectly annotated as *coaBC* with the protein product indicated as "DNA/pantothenate metabolism flavoprotein" (in reference to the gene first being discovered as affecting DNA synthesis; Spitzer and Weiss, 1985). The adjacent PPCS would then simply have a locus tag with the protein product indicated as "phosphopantothenate-cysteine ligase" (a synonym for phosphopantothenoylcysteine synthetase), or as "hypothetical protein."

The PPAT-encoding genes were annotated in some species as *coaD*, but for most species the gene only contained a locus tag with the protein product indicated as "phosphopantetheine adenylyltransferase."

The DPCK-encoding gene was the most commonly found in all the species, but with a few exceptions these genes were not annotated as *coaE* and only contained a locus tag. The protein products were, however, indicated as "dephospho-CoA kinase." Ten species only contained a DPCK-encoding gene in their genomes. Four of these coded for a bifunctional haloacid dehalogenase (HAD)-like/dephospho-CoA kinase (HAD-DPCK) protein. The DPCK-encoding part of this gene was found at the 3′ side of the open reading frame translating into a protein with a DPCK domain at the C-terminal end. *Mycoplasma conjunctivae* was the only species that had a HAD-DPCK-encoding gene in combination with another CoA biosynthesis enzyme-encoding gene, this being for the PPAT enzyme.

In species with both $PanK_{III}$- and CoaBC-encoding genes, the open reading frames always overlapped with 12–24 nucleotides, with the CoaBC-encoding gene found upstream of the $PanK_{III}$-encoding one. This is not a common feature in prokaryotes, as PanK and CoaBC-encoding genes are most commonly either adjacent (e.g., *Clostridium innocuum*) or >100,000 bp apart (e.g., *Lactobacillus* and *Bacillus* sp.). In *Mycoplasma* species with a $PanK_{III}$, but separate PPCS- and PPCDC-encoding genes, there was no consistent distance between the $PanK_{III}$- and the PPCS/PPCDC-encoding genes.

In species containing all four CoA biosynthesis enzyme-encoding genes the distances and orientation of the $PanK_{III}$- and CoaBC- or PPCS/PPCDC-encoding genes, relative to that for PPAT and DPCK, were not consistent. This was also true for species containing the combination of $PanK_{III}$-, PPAT-, and DPCK-encoding genes, with the exception being species in the Spiroplasma group. These displayed a distance in the range of 122–224 kb between the PPAT- and DPCK-encoding gene locations.

In species that contained only PPAT- and DPCK-encoding genes there was no consistent distance or arrangement between these genes. Again, species in the Spiroplasma group were the exception with distances in the range of 122–224 kb between the genes encoding for PPAT and DPCK, respectively. Among the species evaluated only members of the Spiroplasma and Hominis groups had PPAT- and DPCK-encoding gene combinations.

Identification of CoA Biosynthesis Enzyme-Encoding Gene Homologs

At the time this investigation was initiated (2018), no CoA biosynthesis pathway enzyme-encoding genes or gene homologs could be identified in eight of the 62 *Mycoplasma* genomes evaluated (**Table 1**; **Figure 2**). In genomes with no annotated $PanK_{III}$-encoding gene, a homolog was identified in *Mycoplasma gallinaceum* with a protein product of 248 aa in length. This falls within the typical range of annotated $PanK_{III}$ amino acid sequences of 220–275 aa.

In genomes with no annotated CoaBC-encoding gene, homologs were identified in *M. gallinaceum* and *Mycoplasma testudines*. The relative orientation and overlap between the

TABLE 1 | Summary of genes encoding for enzymes of the coenzyme A biosynthesis pathway found within the genomes of evaluated *Mycoplasma* species.

Phylogenetic group[a]	Mycoplasma species	Representative host	CoA biosynthetic pathway enzyme			
			PanK	CoaBC	PPAT	DPCK
Hominis group	*M. anatis**	Ducks	1	2	3	4
	*M. arginini**	Mammals	1	2	3	4
	*M. columborale**	Pigeons	1	2	3	4
	*M. cricetuli**	Hamsters	1	2	3	4[h]
	M. mobile^	Tench	1	2	3[p]	4
	*M. sturni**	Songbirds	1	2	3	4[h]
	M. synoviae^	Galliforms	1	2	3[p]	4[h]
	M. gallinaceum^	Galliforms	1[h]	2[h]	3	4[d]
	*M. alligatoris**	Alligators	1		3	4[h]
	*M. buteonis**	Raptors	1		3	4[h]
	M. crocodyli^	Crocodiles	1		3	4
	*M. molare**	Dogs	1		3	4
	M. pulmonis^	Mice	1		3	4
	M. agalactiae^	Small ruminants			3	4
	M. bovigenitalium^	Cattle			3	4
	M. bovis^	Cattle			3	4
	M. californicum^	Cattle			3	4
	*M. canis**	Dogs			3	4[h]
	*M. collis**	Rodents			3	4
	*M. columbinum**	Pigeons			3	4
	*M. felifaucium**	Pumas			3	4
	*M. felis**	Cats			3	4
	M. fermentans^	Humans			3	4
	*M. gallinarum**	Galliforms			3	4
	*M. iners**	Galliforms			3	4
	*M. leonicaptivi**	Lions			3	4
Hominis group	*M. lipofaciens**	Galliforms			3	4
	*M. opalescens**	Dogs			3	4
	*M. primatum**	Monkeys			3	4
	*M. simbae**	Lions			3	4
	M. conjunctivae^	Small ruminants			3	5
	M. bovoculi^	Cattle				5
	M. dispar^	Cattle				5
	M. flocculare^	Pigs				5[h]
	M. hyopneumoniae^	Pigs				5
	*M. ovipneumoniae**	Sheep				5
	M. hyorhinis^	Pigs				4
	M. sp. Ms02^	Ostrich				4
	M. arthritidis^	Rats				
	M. canadense^	Cattle				
	M. hominis^	Humans				
Spiroplasma group	*M. capricolum* subsp. *capricolum*^	Small ruminants			3	4
	M. capricolum subsp. *capripneumoniae*^	Goats			3	4
	M. leachii^	Cattle			3	4
	M. mycoides subsp. *capri* LC^	Small ruminants			3	4
	M. mycoides subsp. *mycoides* SC^	Cattle			3	4
	M. putrefaciens^	Small ruminants			3	4
	M. yeatsii^	Goats			3	4
Pneumoniae group	*M. iowae*^	Turkeys	1	2s[bh]	3	4
	*M. testudinis**	Tortoises	1	2[h]	3	4
	*M. alvi**	Cattle	1		3	4
	M. penetrans^	Humans	1		3	4
	*M. pirum**	Humans	1		3	4
	M. gallisepticum^	Galliforms				4

(Continued)

TABLE 1 | Continued

Phylogenetic group[a]	Mycoplasma species	Representative host	CoA biosynthetic pathway enzyme			
			PanK	CoaBC	PPAT	DPCK
Pneumoniae group	*M. genitalium*^	Humans				4
	*M. imitans**	Ducks, geese				4
	M. pneumoniae^	Humans				4
	M. haemocanis^[c]	Dogs				
	M. ovis^[c]	Sheep				
	M. parvum^[c]	Pigs				
	M. suis^[c]	Pigs				
	M. wenyonii^[c]	Cattle				

^*Complete genome.*
**Genome available as contigs.*
[a]*Based on 16S rRNA sequence data.*
[b]*Separate PPCS and PPCDC enzymes (no bifunctional protein) with PPCS annotated as "hypothetical protein."*
[c]*Belongs to the hemotropic cluster.*
[h]*Annotated as a "hypothetical protein."*
[p]*Annotated as a "putative protein."*
[d]*Disrupted open reading frame.*
1 = PanK type II; 2 = CoaBC; 3 = PPAT; 4 = DPCK; and 5 = HAD-DPCK. M. species = Mycoplasma species with no pathway enzymes identified.

PanK- and CoaBC-encoding gene homologs in *M. gallinaceum* was consistent with that found for annotated genes. However, *M. testudines* had a smaller overlap of only six nucleotides. The average length of the annotated CoaBC sequences was 360–410 aa. The two hypothetical CoaBC proteins are within this range with 385 aa and 380 aa for *M. gallinaceum* and *Mycoplasma testudinis*, respectively.

Furthermore, among genomes with no PPAT-encoding gene, a homolog could be identified in both *Mycoplasma synoviae* and *Mycoplasma mobile*. The typical length of annotated PPAT protein sequences ranges between 135 aa and 175 aa. Both putative PPAT proteins fall within this range with respective protein lengths of 145 aa and 148 aa for *M. mobile* and *M. synoviae*.

Seven species were found to contain a DPCK-encoding gene homolog, of which one was identified to encode for a HAD-DPCK (*Mycoplasma flocculare*). The average length of annotated DPCK sequences is 165–205 aa and the hypothetical DPCK proteins of *Mycoplasma canis* (189 aa), *Mycoplasma cricetuli* (191 aa), *Mycoplasma sturni* (187 aa), *Mycoplasma alligatoris* (185 aa), *Mycoplasma buteonis* (189 aa), and *M. synoviae* (168 aa) all fall within this range. The average length of annotated HAD-DPCK protein sequences is 445 aa, of which the HAD-domain is ~260 aa. The hypothetical HAD-DPCK protein of *M. flocculare* (447 aa) was found to be within this range.

Mycoplasma gallinaceum B2096 8B was the only species evaluated in this study that contained no DPCK-encoding gene despite having all three other genes. The KEGG database similarly indicated *M. gallinaceum* (strain not specified) does not contain this gene. A genetically close genome (*M. gallinaceum* NCTC10183) was, however, found to contain an annotated DPCK-encoding gene. The nucleotide sequence of this gene (coded by "LR214950.1:904112.904693") was therefore used in a blastn search to identify a possible homolog in B2096 8B. A hit with 96% identity and 100% query coverage was found in B2096 8B at the complement position of 322424.323005, which codes for a putative DPCK protein. This region currently contains a hypothetical disrupted protein indicating that the open reading frame was misidentified.

Family, Motif, and Domain Analysis of CoA Biosynthesis Protein Products

To confirm that genes annotated as hypothetical- or putative proteins were in fact homologs of CoA biosynthesis enzyme-encoding genes, their protein products were evaluated and compared to annotated genes in terms of protein family relationship, as well as domain and motifs contained within the proteins (**Figure 2**).

$PanK_{III}$-Encoding Genes

Except for *Mycoplasma anatis*, the CDD search results (**Supplementary Table 4**) confirmed the protein product of all the annotated $PanK_{III}$-encoding genes and identified homologs to be of the $PanK_{III}$ type. CDD predicted all of these proteins to belong to the ASKHA superfamily (cl17037), which is the expected assignment for $PanK_{III}$. The InterPro searches (**Supplementary Table 5**) supported the CDD results as it could predict that all of the proteins, including that of *M. anatis*, belong to Interpro family IPR004619 ("Type III pantothenate kinase"), along with the biological process gene ontology (GO) term, GO:0004594 (pantothenate kinase activity).

Similarities among proteins were further evaluated by identifying conserved motifs using the MEME algorithm. The advantage of using this algorithm is that it does not rely on existing database information for motif discovery, but this does mean that identified motifs may not correspond to a specific function. The four motifs identified in the PanK proteins (**Supplementary Table 6**; **Supplementary Figure 1**) all overlapped with $PanK_{III}$-specific domains predicted by CDD and InterPro. Based on active site information from UniProtKB for the $PanK_{III}$ of *Mycoplasma crocodyli* MP145, MEME Motif 1 contains a nucleotide-binding region within which the PHOSPHATE1-motif, attributed to the ASKHA superfamily, can be found (Yang et al., 2006). The MEME Motif 2 contains a conserved

amino acid (D) that acts as a proton acceptor and is part of the PAN-motif (*h*G*h*DR, *h* = hydrophobic residue) that is unique to $PanK_{III}$. Amino acids within this motif are involved in binding its substrate, Pan. MEME Motif 2 also contains a conserved amino acid (T) that is involved in ATP binding which forms part of the ASKHA-associated PHOSPHATE2-motif. The $PanK_{III}$-specific INTERFACE-motif is situated in the MEME Motif 3, but the sequence does not exactly match the indicated motif pattern of GGxIxPG (x = any residue). The PAN- and INTERFACE-motifs are said to distinguish $PanK_{III}$s from other ASKHA family proteins.

No MEME Motif 2 or 4 could be identified for the *M. anatis* and *M. buteonis* $PanK_{III}$ enzymes, while the $PanK_{III}$s of several species did not contain a Motif 4. Since MEME identifies ungapped motifs any insertions or deletions in the amino acid sequence will influence the identification of a motif within a specified region. Whether or not a motif was present had no relation to how many of the other pathway enzyme-encoding genes were present in the genome of the species concerned. This may indicate that although CoA biosynthesis-associated genes code for enzymes that function in the same pathway, they are not collectively under the same selection pressure.

CoaBC-Encoding Genes

Conserved Domain Database search results (**Supplementary Table 7**) confirmed the protein product of all annotated CoaBC-encoding genes and identified homologs to be "bi-functional phosphopantothenoylcysteine decarboxylase/phosphopantothenate-cysteine ligase" proteins. The same results were obtained for the separate PPCS and PPCDC proteins of *M. iowae*. CDD predicted all bifunctional proteins to belong to the DNA/pantothenate metabolism flavoprotein (DFP) superfamily (cl27193), as well as the flavoprotein superfamily (cl19190).

The InterPro searches (**Supplementary Table 8**) identified the InterPro family IPR005252 signature ("Coenzyme A biosynthesis bifunctional protein, CoaBC") in most of the CoaBC proteins, including the hypothetical protein of *M. gallinaceum*. InterPro also predicted the biological process GO term GO:0015937 ("CoA biosynthetic process") and molecular function GO terms, GO:0004632 ("PPCS activity") and GO:0004633 ("PPCDC activity"). Although no InterPro family IPR005252 signature was identified in the CoaBC of *M. mobile* or the hypothetical CoaBC of *M. testudinis*, the InterPro domain signatures, IPR003382 ("Flavoprotein") and IPR007085 (DFP, C-terminal), were identified and is also present in all of the other protein sequences. These respective InterPro domain signatures were also identified in PPCDC (IPR003382) and PPCS (IPR007085) of *M. iowae*. Except for *M. gallinaceum*, *M. testudinis*, and *M. mobile*, a N-terminal signaling region was predicted for CoaBC proteins, which implies that they are either membrane-bound or situated inside the *Mycoplasma* membrane. Similarly, a "Prokaryotic membrane lipoprotein lipid attachment site" was identified at the N-terminus in the PPCDC of *M. iowae*.

The regions of the MEME identified motifs (**Supplementary Table 9**) corresponded with the identified signature sequences predicted by CDD and InterPro. Two of the motifs (Motif 1 and Motif 2) were located in the PPCDC domain of the bifunctional protein as well as the separately encoded PPCDC protein of *M. iowae*. Motif 3 and Motif 4, on the other hand, were found in the PPCS domain of the bifunctional protein, but Motif 4 was absent from the separately encoded PPCS protein of *M. iowae*. Moreover, the protein sequences of *M. mobile* and *M. testudinis* were the only two CoaBC proteins which also had no Motif 4 in their PPCS domain (**Supplementary Figure 2**). Based on information from *C. innocuum*, a possible active site (proton donor) was contained within Motif 2 while all other possible nucleotide binding sites (CTP) were outside of the identified motifs. Changing the MEME parameters to a maximum of six motifs resulted in an additional nucleotide binding site falling within the identified motifs.

PPAT-Encoding Genes

In agreement with previous studies (Izard, 2002), the CDD results confirmed all the annotated and putative protein products of the PPAT-encoding genes as "phosphopantetheine adenylyltransferase" (**Supplementary Table 10**). Additionally, CDD predicted all the proteins to belong to the nucleotidyltransferase superfamily (cl00015). Similar results were obtained from the InterPro searches (**Supplementary Table 11**), where the family signature, IPR001980 (Phosphopantetheine adenylyltransferase), was identified in all the protein sequences. Furthermore, the InterPro predicted the GO term for the biological process and molecular function as GO:0015937 (CoA biosynthetic process) and GO:0004595 (PPAT activity), respectively. The regions used by CDD and Interpro for predictions in *M. mobile* were noticeably shorter (66% of the total sequence length) which is probably why it is annotated as a putative protein sequence.

Four motifs were identified using MEME (**Supplementary Table 12**) which overlapped with signature sequences identified by CDD and InterPro. There were segments of these predicted motifs that agreed with the literature-reported motifs (Izard and Geerlof, 1999; Izard, 2002) including prominent residues involved in substrate binding (**Supplementary Figure 3**). Yet again, the protein sequence of *M. mobile* was the outlier, in which only Motif 1 and 2 could be identified. The absence of Motif 3 and 4 correlates with the shorter recognition region of CDD and InterPro. Also, no Motif 3 could be identified within the protein sequences of *Mycoplasma alvi*, *Mycoplasma penetrans*, *Mycoplasma pirum*, *M. iowae*, and *M. testudinis*. Motif 3 specifically contains a portion of the active site where substrate binding takes place *via* an amide bond.

DPCK-Encoding Genes

The identity of all the DPCK protein sequences could be confirmed using CDD searches (**Supplementary Table 13**). In agreement with previous studies (O'Toole et al., 2003), all of the sequences, including the DPCK domain of the HAD-DPCK, were predicted to belong to both the nucleoside/nucleotide kinase (NK)-(cl17190) and CoaE (cl28605) superfamily. The

HAD domain of the HAD-DPCK proteins, including the hypothetical protein of *M. flocculare*, was predicted to belong to the HAD-like (cl21460) and Hydrolase 3 (cl26787) superfamily. The HAD proteins also contain a HAD-like domain predicted to belong to the Cof subfamily, which falls within the Class-IIB subfamily of the HAD-like superfamily. Proteins assigned as Cof proteins are generally cofactor phosphatases (in reference to the activity of the Cof protein of *Escherichia coli*, which acts on intermediates in thiamine biosynthesis) and other HAD family phosphatases (Kuznetsova et al., 2006).

The InterPro family signature, IPR001977 ("Dephospho-CoA kinase"), as well as the homologous superfamily signature, IPR027417 ("P-loop containing nucleoside triphosphate hydrolase") was identified in all of the annotated protein sequences (**Supplementary Table 14**). The regions matching these two signatures overlapped in all sequences. The corresponding GO terms predicted were: GO:0015937 ("CoA biosynthetic process"), referring to the biological process; along with GO:0004140 ("DPCK activity") and GO:0005524 ("ATP binding"), referring to the molecular function. For the eight hypothetical DPCK proteins, only the homologous superfamily signature (IPR027417) could be identified, but given that the region matching to this signature overlaps with that of the family signature (IPR001977) found in the annotated proteins, these sequences were viewed to be related. With regards to the HAD-like domain of the homologous HAD-DPCK protein of *M. flocculare*, InterPro could only identify the family signature, IPR006379 ("HAD-superfamily hydrolase, subfamily IIB").

The four MEME predicted motifs (**Supplementary Table 15**) in the DPCK protein overlapped with the regions identified by the CDD and InterPro searches. Except for *M. mobile*, all species from the Hominis group (**Figure 3**; **Table 1**) contained all four motifs in their DPCK protein. This included the DPCK domain of all HAD-DPCKs. The only exception was *M. mobile*, which, similar to species from the Pneumoniae and Spiroplasma groups, contained no Motif 4. In the case of *M. mobile*, this is due to a single amino acid deletion in the region of Motif 4. Based on information from *M. pneumoniae*, only Motif 1 is associated with a known active site in the form of a nucleotide (ATP) binding site that overlapped with a P-loop/Walker A motif (Walker et al., 1982; **Supplementary Figure 4**).

Correlating Evolutionary Relationship With CoA Biosynthesis Through Phylogenetic Analyses

A 16S rRNA gene phylogeny was generated to determine if the evolutionary relationships of *Mycoplasma* species reflect which of the CoA biosynthesis pathway enzyme-encoding genes are found in their genomes. Three well-resolved and supported groups were retrieved in this phylogeny, i.e., the Spiroplasma, Pneumoniae, and Hominis groups (**Figure 3**). Each group had identifiable and well-supported subclades; however, there was not always consistency between the combination of CoA synthesis pathway genes present in a species and its grouping within a specific clade. An exception to this was the Spiroplasma group which produced a single clade, in which all members only had PPAT and DPCK-encoding genes. Also, species with no CoA biosynthesis enzyme-encoding genes, and those with only a HAD-DPCK-encoding gene, grouped in a single clade in the Pneumoniae- and Hominis group respectively. In addition to this, among the species evaluated, the CoaBC enzyme appeared more frequently among Hominis group members.

Although the lack of one or more enzyme-encoding genes within species with incomplete genomes might influence deductions that can be made, this inconsistent distribution also holds for those with apparent complete genome information. However, when comparing taxa within the Pneumoniae or Hominis clades, evolutionary changes over time (branch lengths) were not reflected by a reduction in the number of CoA synthesis genes.

To determine if the grouping of species relative to the number of CoA synthesis genes in a genome can better be explained at a functional level, a MSA of the respective proteins was used for phylogenetic analysis. Since *M. iowae* contained separate genes encoding for PPCS and PPCDC, two separate alignments were created for the CoaBC enzyme: one without the *M. iowae* sequences, and one including the *M. iowae* PPCS and PPCDC proteins represented as a concatenated sequence.

The phylogeny based on the $PanK_{III}$ (**Figure 4**) revealed well-supported clades. The distribution of species within clades did not always agree with that of the 16S phylogeny, but similar to the 16S phylogeny there was no correlation between the species in a clade and the number of CoA synthesis enzymes present in their genomes. Sequences with a missing MEME Motif 4 (*M. mobile*, *Mycoplasma molare*, *Mycoplasma pulmonis*, and *M. testudinis*) were all located within the same clade, but the corresponding species did not have the same combination of CoA synthesis genes. *Mycoplasma buteonis* and *M. anatis* both have Motif 2 and 4 absent but are found in different clades due to sequence differences outside of the identified motifs.

The distribution of species in the CoaBC-based phylogeny also did not follow that of the 16S phylogeny (**Figure 5**). The concatenated sequence of the separate *M. iowae* PPCS and PPCDC protein sequences was retrieved in a basal position and therefore did not influence the tree topology. Except for *M. mobile* and *M. testudinis* that had a missing MEME Motif 4 and therefore grouped in a single clade, there was no apparent reason, beyond sequence similarities, for the distribution of species within this phylogeny.

In the PPAT phylogeny, a distinctive and well-supported clade for the Spiroplasma group could be retrieved similar to that in the 16S phylogeny (**Figure 6**). For the rest of the phylogeny, Pneumoniae- and Hominis groupings could be observed for most species, but the relationships did not always agree with that of the 16S phylogeny. The distribution of PPAT proteins was also not related to the presence of a PanK and/or CoaBC-encoding genes in the relevant species genome. However, with the exception of *Mycoplasma colis*, the PPAT proteins of species with only PPAT- and DPCK-encoding genes in their genome are always grouped in separate clades. This was also only for members of the Hominis and Spiroplasma

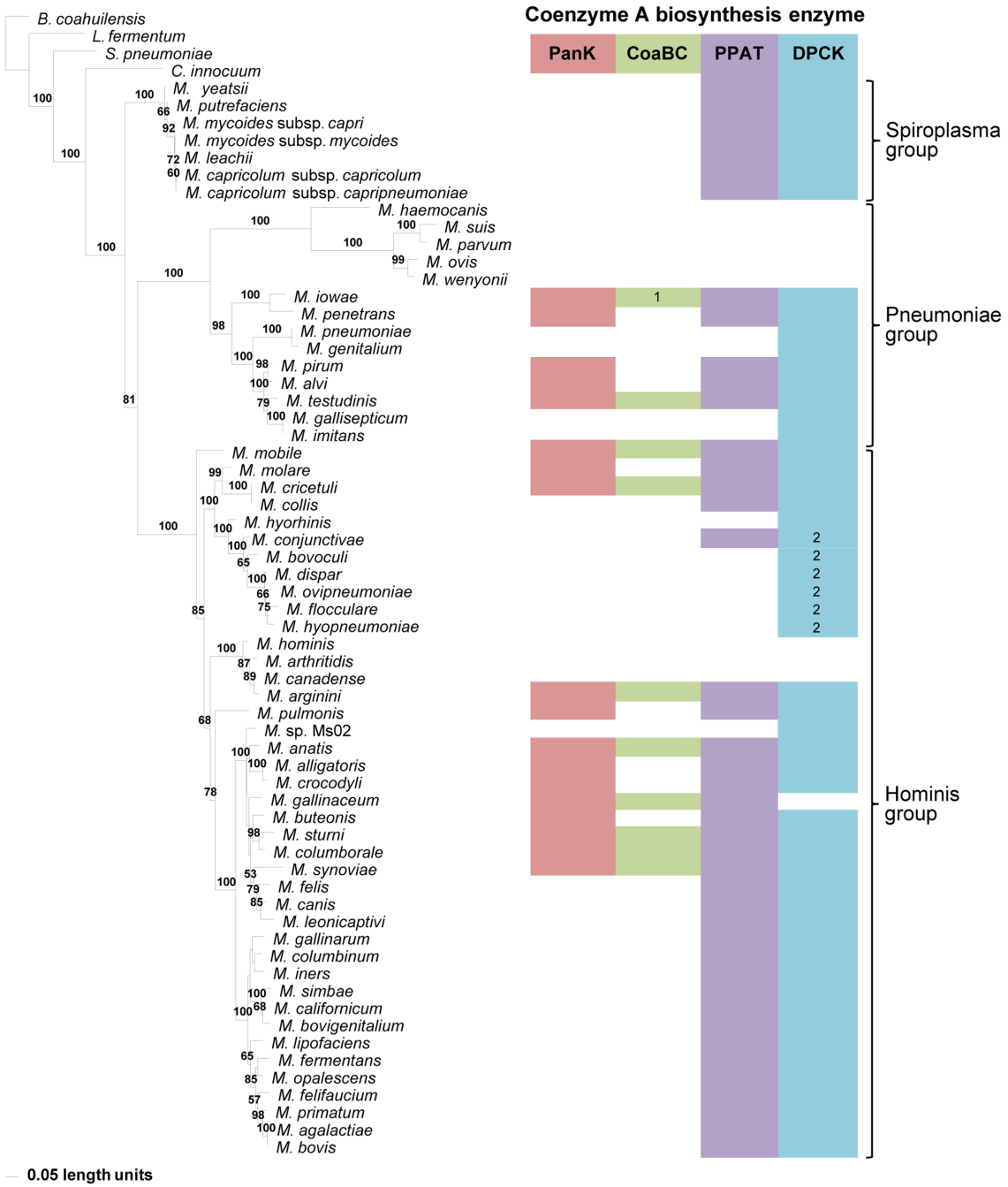


FIGURE 3 | The maximum likelihood phylogeny based on 16S rRNA gene sequences. Bootstrap values (≥50%) are indicated above lines. The identified coenzyme A biosynthetic pathway enzymes of each *Mycoplasma* species are indicated with a color chart: PanK (red), CoaBC (green), PPAT (purple), and DPCK (blue). Notations: 1 (green) – separate PPCS and PPCDC enzymes; 2 (blue) – HAD-DPCK protein.

groups since no species from the Pneumoniae group had a strictly PPAT-DPCK enzyme combination in their genome. Furthermore, no MEME Motif 3 could be identified in the PPAT sequences of *M. alvi*, *M. iowae*, *M. mobile*, *M. pirum*, *M. penetrans*, and *M. testudinis*, yet these species were grouped in two different clades indicating that phylogenetic distributions were not mainly influenced by motif sequences.

Compared to the other enzymes, the DPCK phylogeny was able to retrieve the same Spiroplasma, Pneumoniae, and Hominis groupings as in the 16S phylogeny (**Figure 7**). Also, except for *Mycoplasma arginini* and *M. cricetuli*, the DPCK proteins of all other species were grouped in the same well-supported clades as in the 16S phylogeny. This included species with a HAD domain as part of the DPCK protein. *Mycoplasma arginini*

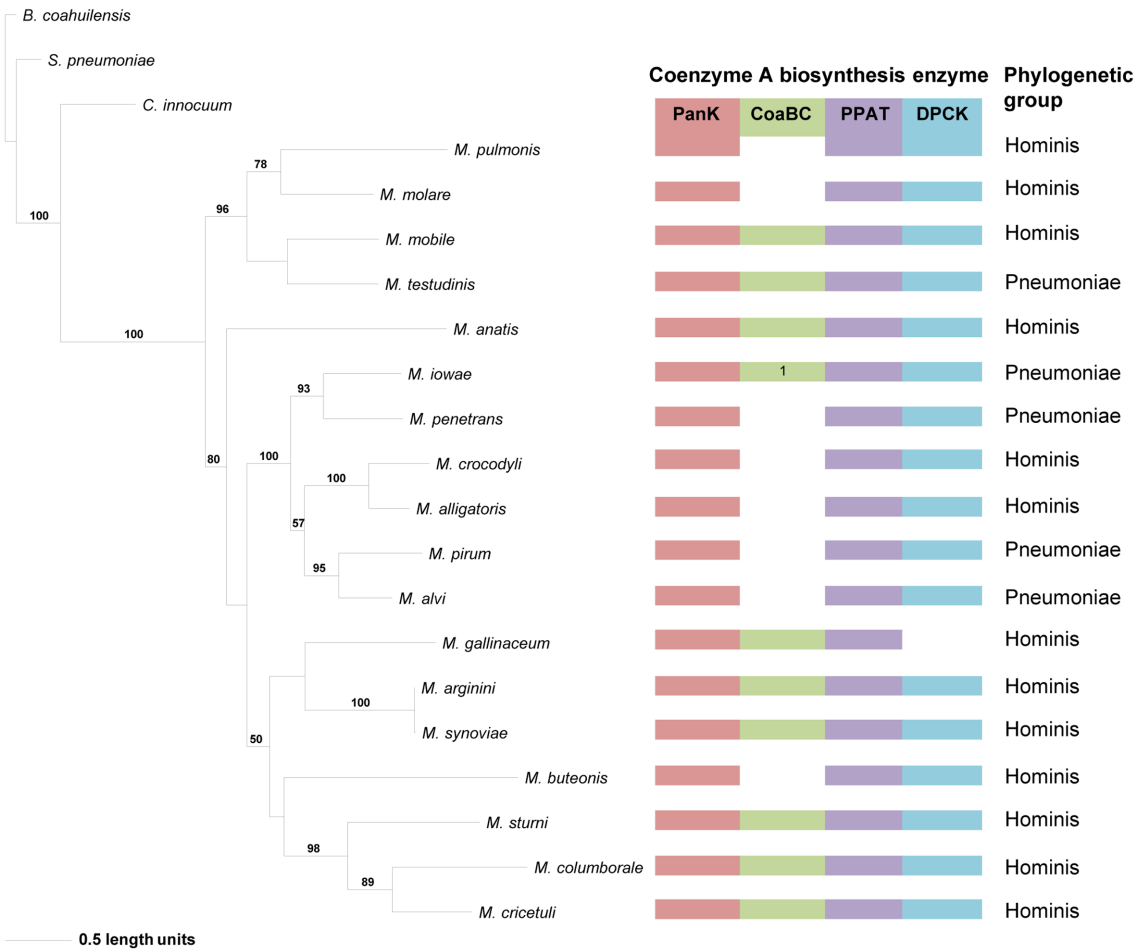


FIGURE 4 | The maximum likelihood phylogeny based on the PanK type III amino acid sequences. Bootstrap values (≥50%) are indicated above lines. Grouping of the species according to the 16S rRNA sequence-based phylogeny, indicated on the right. The identified coenzyme A biosynthetic pathway enzymes of each *Mycoplasma* species are indicated with a color chart: PanK (red), CoaBC (green), PPAT (purple), and DPCK (blue). Notation: 1 (green) –separate PPCS and PPCDC enzymes.

grouped with species that have no CoA synthesis genes in the 16S phylogeny, which were therefore not present in the DPCK phylogeny. The reason for the different grouping of *M. cricetuli* is not clear. All the species in which the MEME-predicted Motif 4 of DPCK enzymes was absent were from either the Spiroplasma or Pneumoniae group; these are then also found in a single clade in the respective groups.

DISCUSSION

Our study of CoA biosynthesis in *Mycoplasma* found that their genomes contain genes encoding the CoA pathway enzymes (**Figure 1A**), but all four genes were not present in every species and different combinations were found across genomes. Among the evaluated genomes available on NCBI, the typical gene annotations of *coaA*/*coaX*, *coaBC* (or *coaB*/*coaC*), *coaD*, and *coaE* were seldom found and therefore were of no use in confirming the presence of these genes within *Mycoplasma* genomes. Information from the KEGG and SEED databases was also limited since not all species – or strains of a species – are represented within these databases.

A description of protein products was, however, more consistent and this allowed the identification of associated genes which were then mostly annotated with a locus tag number. Where no protein products were indicated, enzyme homologs could be identified within several genomes using a combination of NCBI blastp or psi-blast searches using known protein sequences from a genetically close species as query. To further confirm, the identity of the gene its protein product was evaluated in terms of protein family, domain, and motif information using the web-based CDD, InterPro, and MEME Suite recourses.

Of the 62 *Mycoplasma* species evaluated in this study, 10 contained all four enzyme-encoding genes of the CoA synthesis pathway in their genomes. These 10 species were only found in the Hominis and Pneumoniae groups in the 16S phylogeny. Except for *M. iowae*, which contained separate PPCS- and PPCDC-encoding genes, the rest all had a bi-functional CoaBC-encoding gene which is more typical of prokaryotes. Although all strains of *M. iowae* contain separate genes, differences among

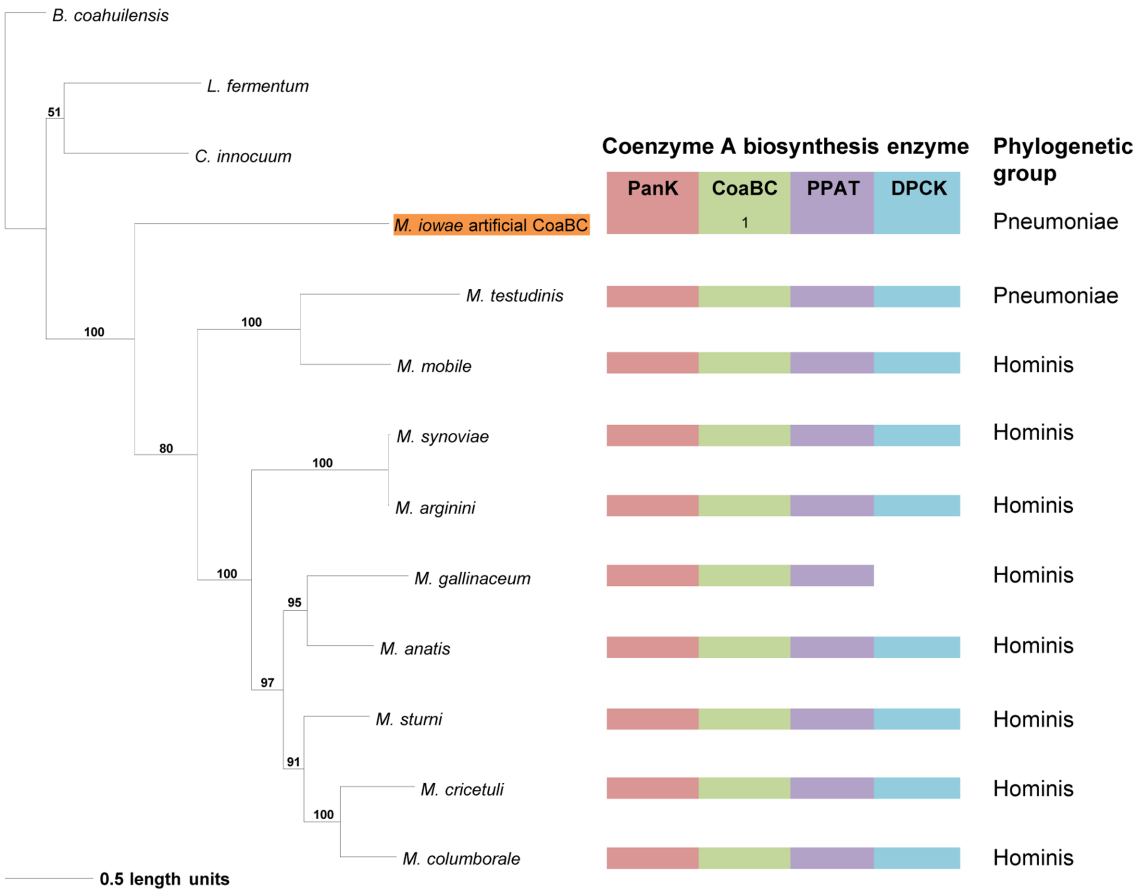


FIGURE 5 | The maximum likelihood phylogeny based on the CoaBC amino acid sequences, including *Mycoplasma iowae*. Bootstrap values (≥50%) are indicated above lines. Grouping of the species according to the 16S rRNA sequence-based phylogeny, indicated on the right. The identified coenzyme A biosynthetic pathway enzymes of each *Mycoplasma* species are indicated with a color chart: PanK (red), CoaBC (green), PPAT (purple), and DPCK (blue). Notation: 1 (green) –separate PPCS and PPCDC enzymes. The orange highlighted species is the concatenated PPCDC and PPCS protein sequence.

strains were found in *M. fermentans*. The open reading frame of all CoaBC-encoding genes always overlapped with that of PanK, which could imply that the genes are transcribed as a single operon and are therefore under the same selective pressure. However, where the PPCS- and PPCDC encoding genes are separated, the PanK-encoding gene is found thousands of base pairs upstream, but at no fixed distance and transcription as a single operon can therefore be excluded.

Except for *M. sturni*, the cellular localization of the PanK enzymes could not be predicted using InterPro, but prokaryotic PanKs are typically intracellular proteins. However, the PanK of *M. sturni* was predicted to be a transmembrane protein (**Supplementary Table 5**) with the extracellular domain containing the PAN- and INTERFACE-motifs that are specifically involved in interaction with the $PanK_{III}$ substrate, Pan (Yang et al., 2006). It could therefore bind extracellular Pan and convert it to 4′-phosphopantothenic acid (P-Pan) for further processing by CoaBC, which was also predicted to be an extracellular, membrane-bound protein. However, if the PanK enzymes in other *Mycoplasma* species are in fact situated in the cytoplasm, then the product of PanK, P-Pan, will have to be exported before it can be used as a substrate by CoaBC. Unassisted export of this substrate seems highly unlikely since it is very polar (consisting of a phosphate group and a carboxylic acid moiety) and would thus require a dedicated transporter to cross the membrane. Alternatively, P-Pan might be obtained from the host if readily available in the environment. If this is possible, then the PanK enzyme might not even be required for CoA synthesis in these organisms, but have an entirely different role. How and where the subsequent processing of the CoaBC product, 4′-phosphopantetheine (PPantSH), will take place is not clear since the cellular localization of the *Mycoplasma* PPAT could not be predicted. Prokaryotic PPAT enzymes are typically located in the cytoplasm and if this is also the case for *Mycoplasma* then PPantSH would have to be imported after being released from the CoaBC enzyme.

In eight of the evaluated species, a CoaBC-encoding gene was absent with only PanK-, PPAT-, and DPCK-encoding genes present. In the 16S phylogeny these species were only found in the Pneumoniae and Hominis groups, but none in the Spiroplasma group. This combination of enzymes would seem to indicate that CoA biosynthesis occurs *via* the salvage pathway (**Figure 1B**).

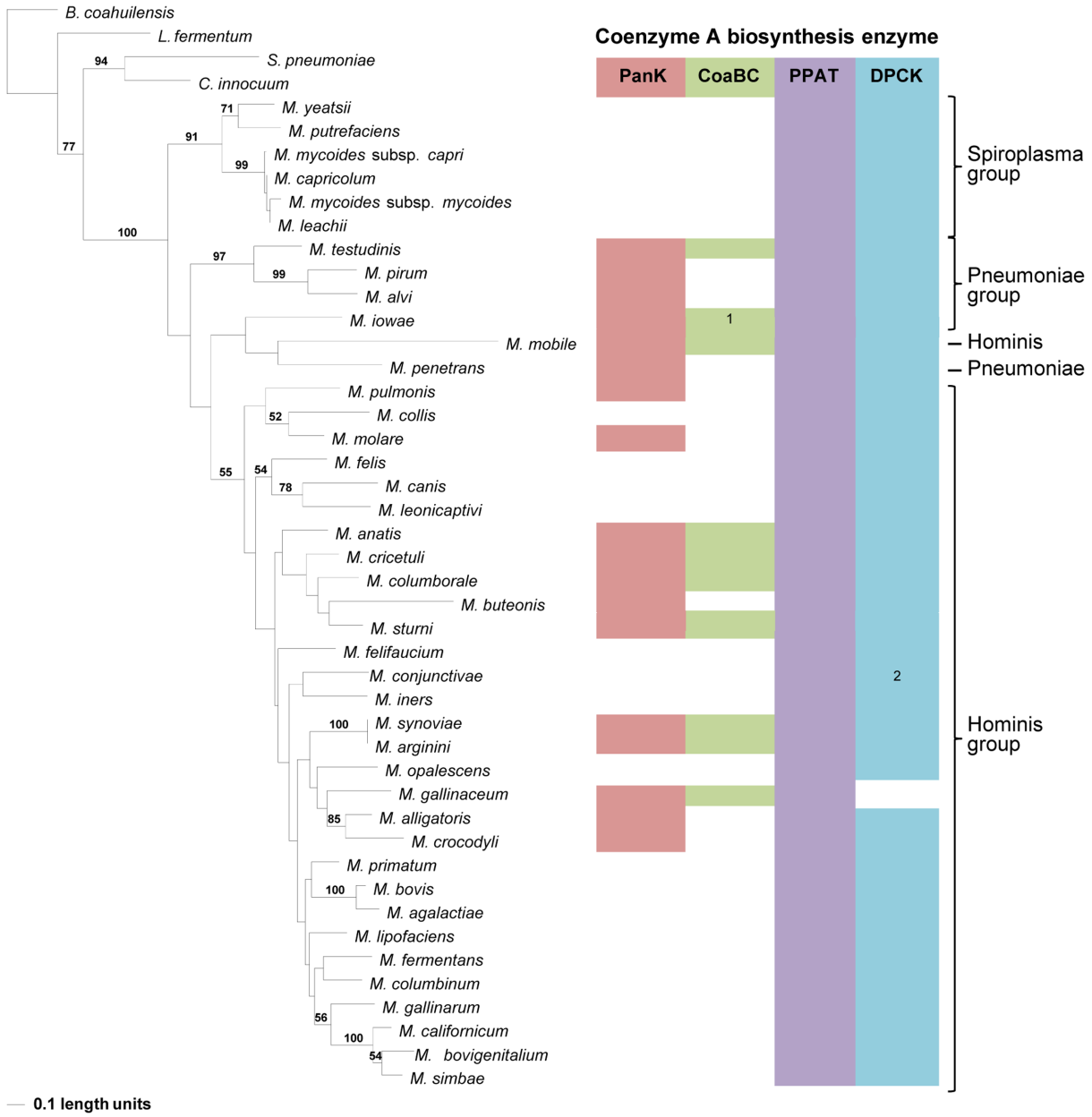


FIGURE 6 | The maximum likelihood phylogeny based on the PPAT amino acid sequences. Bootstrap values (≥50%) are indicated above lines. Suggestive groupings, which resemble that of the 16S rRNA sequence-based phylogeny, are indicated by the brackets (right), with *Mycoplasma mobile* and *Mycoplasma penetans* as exceptions. The identified coenzyme A biosynthetic pathway enzymes of each *Mycoplasma* species are indicated with a color chart: PanK (red), CoaBC (green), PPAT (purple), and DPCK (blue). Notations: 1 (green) – separate PPCS and PPCDC enzymes; 2 (blue) – HAD-DPCK protein.

However, *Mycoplasma* PanKs are all predicted to be $PanK_{III}$ type enzymes, none of which are known to accept PantSH – the substrate of the salvage pathway – as a substrate (Strauss, 2010). When comparing the PAN-motif (*h*G*h*DR) in *Mycoplasma* $PanK_{III}$ enzymes to that described for $PanK_{III}$s from other bacteria, the arginine residue is replaced by leucine or isoleucine in *Mycoplasma* (**Figure 8**). This arginine residue plays a prominent role in the active site of the PanK enzyme, since it is directly involved in an interaction with the substrate, Pan. Therefore, the substitution of this charged residue with an uncharged hydrophobic residue will most likely have major consequences with regards to the binding of Pan. The INTERFACE-motif in *Mycoplasma* similarly deviates from the general residue sequence. Since these two motifs together form a substrate-binding site in $PanK_{III}$s, the substitution of residues in both motifs may allow the enzymes to use PantSH as a substrate so that CoA biosynthesis could take place *via* the salvage pathway. It is also possible that the substitutions in these motifs have no impact on PanK substrate specificity, and that like other $PanK_{III}$s these enzymes are also not able to phosphorylate

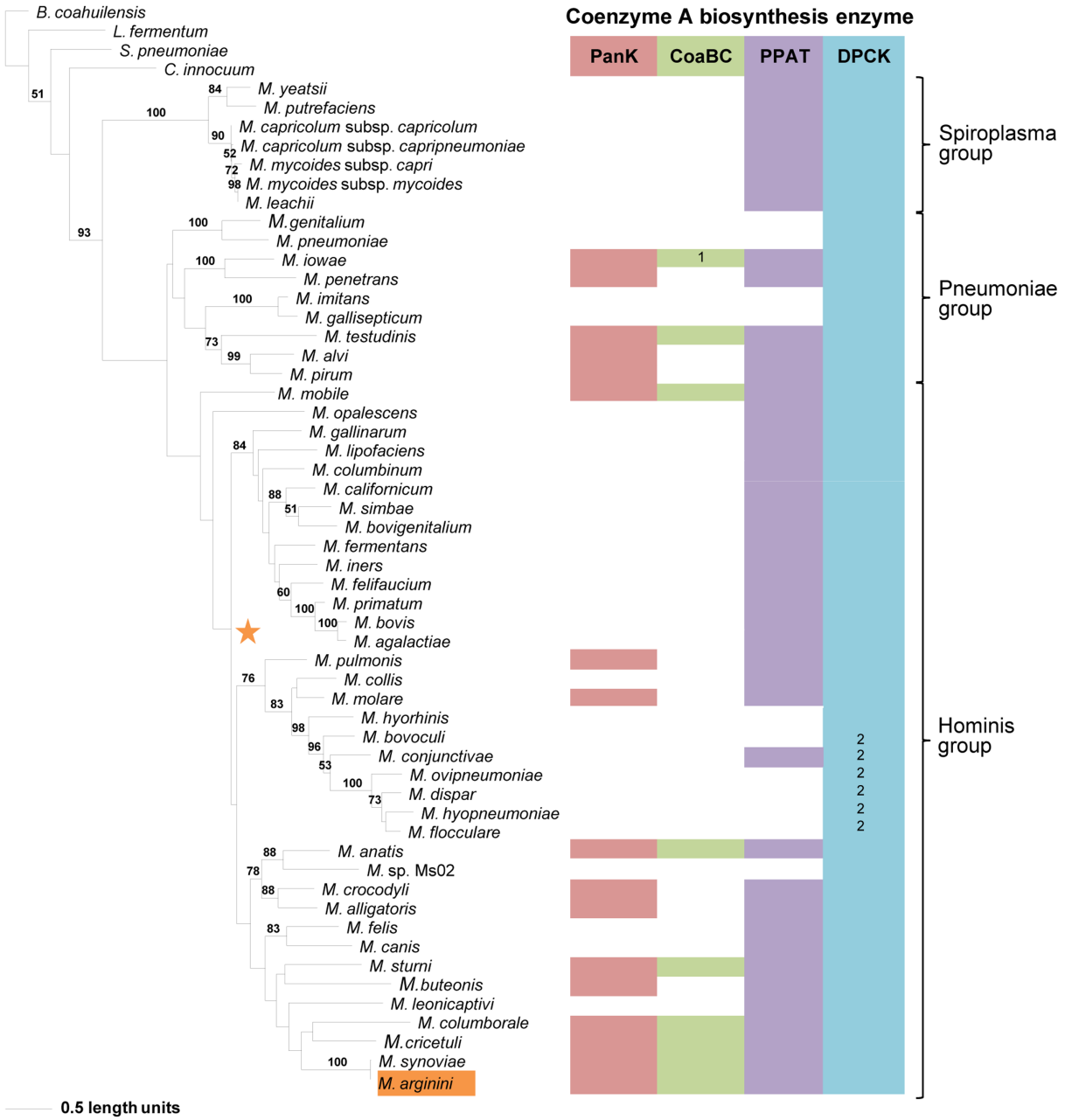


FIGURE 7 | The maximum likelihood phylogeny based on the DPCK amino acid sequences. Bootstrap values (≥50%) are indicated above lines. Suggestive groupings of the three prominent clades, which resemble that of the 16S rRNA sequence-based phylogeny, are indicated by the brackets (right). The identified coenzyme A biosynthetic pathway enzymes of each *Mycoplasma* species are indicated with a color chart: PanK (red), CoaBC (green), PPAT (purple), and DPCK (blue). Notations: 1 (green) –separate PPCS and PPCDC enzymes; 2 (blue) – HAD-DPCK protein. The orange highlighted species is the sub-clade species variation from the 16S rRNA sequence-based phylogeny. The orange star represents the location of *Mycoplasma arginini* in the 16S rRNA sequence-based phylogeny.

PantSH. They might thus have a role unrelated to CoA synthesis as suggested previously, with these species only relying on the PPAT and DPCK proteins to obtain CoA. Unfortunately, our attempts to date to experimentally evaluate the substrate specificity of the *Mycoplasma* PanKs were unsuccessful due to difficulties in obtaining soluble protein for use in activity assays.

The majority of *Mycoplasma* species evaluated in this study only contained genes encoding for PPAT and DPCK. These species were all situated in either the Hominis or Spiroplasma groups, but none in the Pneumoniae group. If these genes do produce functional proteins, PPAT would have to obtain its PPantSH substrate in the absence of PanK. Recently, it has become evident that certain cells, including some micro-organisms, might be able to obtain PPantSH from their surroundings to fulfil their CoA needs (Sibon and Strauss, 2016). Also, in *Drosophila melanogaster* Schneider 2 cells, PPantSH is able to cross the

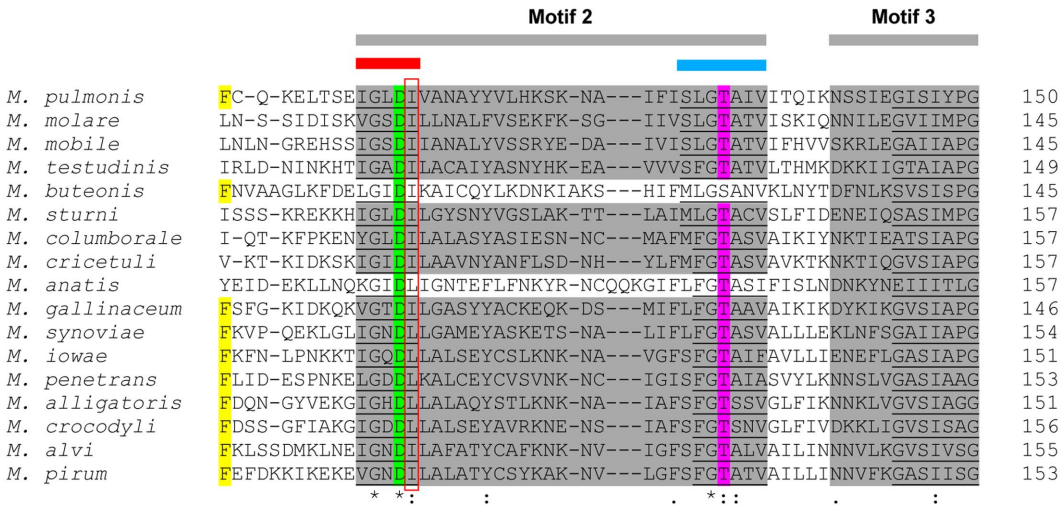


FIGURE 8 | A section of the multiple sequence alignment (MSA) of the PanK$_{III}$ amino acid sequences. The Multiple Expectation maximisation for Motif Elicitation (MEME) motifs are highlighted in gray. The PAN-motif (red bar) and INTERFACE-motif (blue bar) are both situated in MEME-predicted Motif 2. Insertions are indicated by dashes (-). The *Mycoplasma* PAN-motif mutation site (R substituted by L or I) is indicated by the red box. Yellow, green, and magenta highlighted residues represent conserved amino acids that are involved in substrate interactions.

membrane without the help of a transporter (Srinivasan et al., 2015). How *Mycoplasma* would obtain PPantSH remains an open question. However, all indications are that PPantSH is readily available in the environment. In addition to being an intermediate product of the *de novo* CoA biosynthetic pathway (required by the host), PPantSH can also be produced *via* the degradation of CoA and the acyl carrier proteins involved in lipid biosynthesis (Jackowski and Rock, 1984b). Furthermore, *E. coli* was found to excrete PPantSH, but also never re-imported it (Jackowski and Rock, 1984a). Thus, PPantSH supplied by the host or surrounding microbial communities would allow *Mycoplasma* to indeed salvage PPantSH from the environment for conversion to CoA by the action of PPAT and DPCK (**Figure 1B**).

The most common enzyme-encoding gene that could be identified among the investigated mycoplasmas was that for DPCK. With some exceptions, this gene was present in all of the mycoplasmas. Eleven of the species contained only a DPCK-enzyme encoding gene and these species were distributed over the Hominis and Pneumoniae groups. Five of these were HAD-DPCK-encoding genes and these were only found among the Hominis group members. *Mycoplasma conjunctivae* was the only species to contain a HAD-DPCK-encoding gene in combination with another CoA-synthesis pathway gene (PPAT).

The fact that only the final enzyme in the CoA synthesis pathway could be identified for some species suggests that the DPCK enzymes scavenge their substrate – dephospho-coenzyme A (dPCoA) – either from the host or the surrounding environment (**Figure 1B**). The DPCK enzymes of *Mycoplasma gallisepticum* and *Mycoplasma imitans* were both predicted to contain signal peptides and be located extracellularly, which would allow easy access to dPCoA from the host or environment. The cellular localization of the rest could not be predicted, but given that DPCK enzymes are typically found intracellularly, this would require the *Mycoplasma* to have a (as of yet unidentified) dPCoA transport system. The presence of only DPCK and their concomitant requirement for the uptake of dPCoA has also been predicted for various *Rickettsia* and *Chlamydia* spp. (Strauss, 2010; Driscoll et al., 2017). Furthermore, the characterization of a mitochondrial transporter of dPCoA in *D. melanogaster* suggests that a prokaryotic version of such a transporter might exist (Vozza et al., 2017). The obvious dependency of mycoplasmas on their host for many metabolites and intermediates – due to their restricted metabolic biosynthetic potential – highlights the necessity for a vast set of transporters. However, while many *Mycoplasma* transporters are vaguely annotated based on similarities to established transporter classes or membrane segments, little is known regarding their actual substrate preferences or specificities (Großhennig et al., 2013).

The role of the HAD domain in the bifunctional HAD-DPCK enzyme is not clear. As indicated before, the HAD protein is predicted to belong to the Cof subfamily of proteins which consists of phosphatases. These types of phosphatases are promiscuous and seem to be associated with numerous vitamin and cofactor metabolic processes (Lawhorn et al., 2004; Kuznetsova et al., 2006). While the phosphatase activity might suggest a role in regulation (by removing phosphates required for biosynthesis) this has not yet been experimentally tested.

Only eight *Mycoplasma* species with no CoA biosynthesis enzyme-encoding genes could be identified. Five of these form part of the hemotropic *Mycoplasma* cluster (**Table 1**; Brown, 2010), which is a division within the Pneumoniae group while the rest were from the Hominis group. The mycoplasmas in the hemotropic cluster are also referred to as hemoplasmas and are blood-borne pathogens that infect the erythrocyte (Messick, 2004). The absence of genes in the eight species was confirmed by a comparative genomics study that showed them to not contain any enzymes involved in CoA metabolism, the pentose-phosphate pathway or pyruvate dehydrogenase complex (Guimaraes et al., 2014). In the case of the hemoplasma, the nutrient-rich

environment of the blood provides the ideal conditions for additional degenerative evolution, and thereby metabolic reduction. However, it is interesting to note that the blood-borne *Plasmodium* parasite that causes malaria also infects erythrocytes, and yet CoA biosynthesis is the only vitamin-requiring pathway that is essential in this organism (Spry et al., 2008). Clearly, the need for CoA depends on the organism and its metabolic requirements, and for many pathogens, this remains largely unexplored.

The phylogeny produced using 16S rRNA gene sequences reflected the evolutionary relationships of the investigated mycoplasmas as reported in the literature (Weisburg et al., 1989; Johansson et al., 1998; Maniloff, 2002; Brown, 2010; Wium et al., 2015). Except for the Spiroplasma group, where all species only contained genes encoding for PPAT and DPCK, the evolutionary relationships of species were not found to correlate with the combination of CoA synthesis genes contained within their genomes.

Apart from the DPCK phylogeny, the distribution of species in the phylogeny of individual enzymes did not resemble the evolutionary relationships found in the 16S rRNA phylogeny. Moreover, the functional relationship of one enzyme in the CoA synthesis pathway also did not determine which of the other enzymes would be present in a species. However, members of the Spiroplasma group did produce distinctive clades in both the PPAT and DPCK phylogenies. With a few exceptions, this was also the case for members of the Hominis group with only PPAT and DPCK proteins.

From our results, it can be concluded that most *Mycoplasma* species do have the genetic capacity to synthesize CoA, but there is an unpredictability according to the universally accepted 16S rRNA phylogeny among species in terms of which enzyme-encoding gene combinations can be expected, and therefore what the required substrate for CoA synthesis would be in each case. This variation may be entirely dependent on the nutritional conditions at the site of infection. Their parasitic lifestyle on the one hand, and their nutritionally exacting nature on the other, typically causes *Mycoplasma* to exhibit a rather strict host, organ, and tissue specificity (Razin, 2006). They are therefore reliant on the nutritional milieu provided by the host or the microbial community at the site of infection, which would in turn influence their biosynthetic capacity or requirements. We note, for example, that those species with the most complete pathways mainly have non-mammalian hosts. It is possible that in these hosts the ability to obtain CoA *via* truncated pathways is limited due to low levels of circulating CoA biosynthesis intermediates. This observation suggests that instead of correlating with evolutionary relationships, the combination of CoA synthesis genes is more likely to correlate with the mycoplasmas' immediate nutritional environment. However, testing such a proposal will only be possible once more information becomes available on the levels of circulating metabolites in the various hosts.

Despite the variation in the CoA synthesis capacity of *Mycoplasma*, there is the potential for the CoA biosynthetic pathway as a drug/vaccine target. A comprehensive investigation into the CoA synthesis pathway of the targeted *Mycoplasma* will, however, have to be performed in order to identify the genetic variation present within the genome of the specific species (or within strains of the same species, as these might also show differences). The most consistently present enzyme-encoding gene was that of DPCK, suggesting that the development of a broad-spectrum anti-mycoplasmal agent acting on CoA biosynthesis should target this enzyme. The potential of DPCK as a selective target is further underscored by the fact that the main DPCK activity of the human host is found on a bifunctional PPAT-DPCK protein called CoA synthase (COASY), indicating that selective targeting may be possible. On the other hand, for those species in which no CoA biosynthesis enzyme-encoding genes could be identified, the likely requirement for a specific CoA transporter also offers an opportunity for the development of new therapies that would target this protein and/or its function. Overall, our findings suggest that CoA biosynthesis in *Mycoplasma* holds many significant opportunities for further study and investigation, specifically in the search for new treatments of *Mycoplasma* infections.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

AB and ES contributed to conception and design of the study. TR performed all experimental work and analysis and wrote first draft of the manuscript. AB supervised the project with the assistance of ES and revised the manuscript for submission with support from ES. All authors contributed to the article and approved the submitted version.

FUNDING

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors. Postgraduate student funding was received from the National Research Foundation and the Faculty of Science, Stellenbosch University.

ACKNOWLEDGMENTS

We thank Dirk Bellstedt for his valuable comments on the interpretation of the phylogenies. We also acknowledge Sandisiwe Matyesini for her efforts in expressing mycoplasma PanK for substrate analysis.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2021.791756/full#supplementary-material

REFERENCES

Altschul, S. F., and Koonin, E. V. (1998). Iterated profile searches with PSI-BLAST--a tool for discovery in protein databases. *Trends Biochem. Sci.* 23, 444–447. doi: 10.1016/s0968-0004(98)01298-5

Bailey, T. L., Johnson, J., Grant, C. E., and Noble, W. S. (2015). The MEME suite. *Nucleic Acids Res.* 43, W39–W49. doi: 10.1093/nar/gkv416

Balish, M. F., and Distelhorst, S. L. (2016). Potential molecular targets for narrow-spectrum agents to combat *Mycoplasma pneumoniae* infection and disease. *Front. Microbiol.* 7:205. doi: 10.3389/fmicb.2016.00205

Blum, M., Chang, H. Y., Chuguransky, S., Grego, T., Kandasaamy, S., Mitchell, A., et al. (2021). The InterPro protein families and domains database: 20 years on. *Nucleic Acids Res.* 49, D344–D354. doi: 10.1093/nar/gkaa977

Brand, L. A., and Strauss, E. (2005). Characterization of a new pantothenate kinase isoform from helicobacter pylori. *J. Biol. Chem.* 280, 20185–20188. doi: 10.1074/jbc.C500044200

Brown, D. R. (2010). "Phylum XVI. Tenericutes Murray 1984a, 356VP (Effective publication: Murray 1984b, 33.)," in *Bergey's Manual of Systematic Bacteriology: Volume 4*. eds. W. B. Whitman and A. C. Parte (New York, USA: Springer), 567–723.

Chernova, O. A., Medvedeva, E. S., Mouzykantov, A. A., Baranova, N. B., and Chernov, V. M. (2016). Mycoplasmas and their antibiotic resistance: the problems and prospects in controlling infections. *Acta Nat.* 8, 24–34. doi: 10.32607/20758251-2016-8-2-24-34

Daugherty, M., Polanuyer, B., Farrell, M., Scholle, M., Lykidis, A., De Crécy-Lagard, V., et al. (2002). Complete reconstitution of the human coenzyme A biosynthetic pathway via comparative genomics. *J. Biol. Chem.* 277, 21431–21439. doi: 10.1074/jbc.M201708200

De Jonge, B. L. M., Walkup, G. K., Lahiri, S. D., Huynh, H., Neckermann, G., Utley, L., et al. (2013). Discovery of inhibitors of 4′-phosphopantetheine adenylyltransferase (PPAT) to validate PPAT as a target for antibacterial therapy. *Antimicrob. Agents Chemother.* 57, 6005–6015. doi: 10.1128/AAC.01661-13

Driscoll, T. P., Verhoeve, V. I., Guillotte, M. L., Lehman, S. S., Rennoll, S. A., Beier-Sexton, M., et al. (2017). Wholly rickettsia! Reconstructed metabolic profile of the quintessential bacterial parasite of eukaryotic cells. *MBio* 8, e00859–e00917. doi: 10.1128/mBio.00859-17

Gerdes, S. Y., Scholle, M. D., D'Souza, M., Bernal, A., Baev, M. V., Farrell, M., et al. (2002). From genetic footprinting to antimicrobial drug targets: examples in cofactor biosynthetic pathways. *J. Bacteriol.* 184, 4555–4572. doi: 10.1128/JB.184.16.4555-4572.2002

Großhennig, S., Schmidl, S. R., Schmeisky, G., Busse, J., and Stülke, J. (2013). Implication of glycerol and phospholipid transporters in *Mycoplasma pneumoniae* growth and virulence. *Infect. Immun.* 81, 896–904. doi: 10.1128/IAI.01212-12

Guimaraes, A. M. S., Santos, A. P., Do Nascimento, N. C., Timenetsky, J., and Messick, J. B. (2014). Comparative genomics and phylogenomics of hemotrophic mycoplasmas. *PLoS One* 9:e91445. doi: 10.1371/journal.pone.0091445

Hall, T. A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for windows 95/98/NT. *Nucleic Acids Symp. Ser.* 41, 95–98.

Hutchison, C. A., Chuang, R.-Y., Noskov, V. N., Assad-Garcia, N., Deerinck, T. J., Ellisman, M. H., et al. (2016). Design and synthesis of a minimal bacterial genome. *Science* 351:aad6253. doi: 10.1126/science.aad6253

Izard, T. (2002). The crystal structures of phosphopantetheine adenylyltransferase with bound substrates reveal the enzyme's catalytic mechanism. *J. Mol. Biol.* 315, 487–495. doi: 10.1006/jmbi.2001.5272

Izard, T., and Geerlof, A. (1999). The crystal structure of a novel bacterial adenylyltransferase reveals half of sites reactivity. *EMBO J.* 18, 2021–2030. doi: 10.1093/emboj/18.8.2021

Jackowski, S., and Rock, C. O. (1984a). Metabolism of 4′-phosphopantetheine in *Escherichia coli*. *J. Bacteriol.* 158, 115–120. doi: 10.1128/jb.158.1.115-120.1984

Jackowski, S., and Rock, C. O. (1984b). Turnover of the 4′-phosphopantetheine prosthetic group of acyl carrier protein. *J. Biol. Chem.* 259, 1891–1895. doi: 10.1016/s0021-9258(17)43491-0

Johansson, K.-E., Heldtander, M. U. K., and Petterson, B. (1998). "Characterization of mycoplasmas by PCR and sequence analysis with universal 16S rDNA primers," in *Methods in Molecular Biology: Mycoplasma Protocols*. eds. R. J. Miles and R. A. J. Nicholas (Totowa, USA: Humana Press), 145–165.

Johansson, K.-E., and Petterson, B. (2002). "Taxonomy of mollicutes," in *Molecular Biology and Pathogenicity of Mycoplasmas*. eds. S. Razin and R. Herrmann (New York: Springer Science & Business Media), 1–29.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., et al. (2012). Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28, 1647–1649. doi: 10.1093/bioinformatics/bts199

Kuznetsova, E., Proudfoot, M., Gonzalez, C. F., Brown, G., Omelchenko, M. V., Borozan, I., et al. (2006). Genome-wide analysis of substrate specificities of the Escherichia coli haloacid dehalogenase-like phosphatase family. *J. Biol. Chem.* 281, 36149–36161. doi: 10.1074/jbc.M605449200

Lawhorn, B. G., Gerdes, S. Y., and Begley, T. P. (2004). A genetic screen for the identification of thiamin metabolic genes. *J. Biol. Chem.* 279, 43555–43559. doi: 10.1074/jbc.M404284200

Maniloff, J. (2002). "Phylogeny and evolution," in *Molecular Biology and Pathogenicity of Mycoplasmas*. eds. S. Razin and R. Herrmann (New York: Springer Science & Business Media), 31–43.

Messick, J. B. (2004). Hemotrophic mycoplasmas (hemoplasmas): a review and new insights into pathogenic potential. *Vet. Clin. Pathol.* 33, 2–13. doi: 10.1111/j.1939-165X.2004.tb00342.x

Meyer, P. M., and Van Rossum, A. M. C. (2014). *Mycoplasma pneumoniae* in children: carriage, pathogenesis, and antibiotic resistance. *Curr. Opin. Infect. Dis.* 27, 220–227. doi: 10.1097/QCO.0000000000000063

Miller, M., Pfeiffer, W., and Schwartz, T. (2010). "Creating the CIPRES science gateway for inference of large phylogenetic trees," In *Proceedings of the Gateway Computing Environments Workshop (GCE)*; November 14, 2010; New Orleans, LA, 1–8.

Moolman, W., de Villiers, M., and Strauss, E. (2014). Recent advances in targeting coenzyme A biosynthesis and utilization for antimicrobial drug development. *Biochem. Soc. Trans.* 42, 1080–1086. doi: 10.1042/BST20140131

O'Toole, N., Barbosa Joao, A. R. G., Li, Y., Hung, L.-W., Matte, A., and Cygler, M. (2003). Crystal structure of a trimeric form of dephosphocoenzyme A kinase from *Escherichia coli*. *Protein Sci.* 12, 327–336. doi: 10.1110/ps.0227803

Overbeek, R., Begley, T., Butler, R. M., Choudhuri, J. V., Chuang, H. Y., Cohoon, M., et al. (2005). The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. *Nucleic Acids Res.* 33, 5691–5702. doi: 10.1093/nar/gki866

Pollack, J. D., Williams, M. V., and McElhaney, R. N. (1997). The comparative metabolism of the mollicutes (Mycoplasmas): the utility for taxonomic classification and the relationship of putative gene annotation and phylogeny to enzymatic function in the smallest free-living cells. *Crit. Rev. Microbiol.* 23, 269–354. doi: 10.3109/10408419709115140

Pruitt, K. D., Tatusova, T., Brown, G. R., and Maglott, D. R. (2012). NCBI reference sequences (RefSeq): current status, new features and genome annotation policy. *Nucleic Acids Res.* 40, D130–D135. doi: 10.1093/nar/gkr1079

Razin, S. (2006). "The genus *Mycoplasma* and related genera (class Mollicutes)," in *The Prokaryotes*. eds. M. Dworkin, S. Falkow, E. Rosenberg, K. Schleifer and E. Stackebrandt (New York: Springer-Verlag).

Razin, S., Yogev, D., and Naot, Y. (1998). Molecular biology and pathogenicity of mycoplasmas. *Microbiol. Mol. Biol. Rev.* 62, 1094–1156. doi: 10.1128/MMBR.62.4.1094-1156.1998

Rivera, A., and Cedillo, L. (2015). Mycoplasmas and nucleases. *J. Pure Appl. Microbiol.* 9, 2115–2125.

Sibon, O., and Strauss, E. (2016). Coenzyme A: to make it or uptake it? *Nat. Rev. Mol. Cell Biol.* 17, 605–606. doi: 10.1038/nrm.2016.110

Sievers, F., and Higgins, D. G. (2018). Clustal omega for making accurate alignments of many protein sequences. *Protein Sci.* 27, 135–145. doi: 10.1002/pro.3290

Spitzer, E. D., and Weiss, B. (1985). dfp gene of *Escherichia coli* K-12, a locus affecting DNA synthesis, codes for a flavoprotein. *J. Bacteriol.* 164, 994–1003. doi: 10.1128/jb.164.3.994-1003.1985

Spry, C., Kirk, K., and Saliba, K. J. (2008). Coenzyme A biosynthesis: an antimicrobial drug target. *FEMS Microbiol. Rev.* 32, 56–106. doi: 10.1111/j.1574-6976.2007.00093.x

Srinivasan, B., Baratashvili, M., van der Zwaag, M., Kanon, B., Colombelli, C., Lambrechts, R. A., et al. (2015). Extracellular 4′-phosphopantetheine is a source for intracellular coenzyme A synthesis. *Nat. Chem. Biol.* 11, 784–792. doi: 10.1038/nchembio.1906

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312–1313. doi: 10.1093/bioinformatics/btu033

Strauss, E. (2010). "7.11—coenzyme A biosynthesis and enzymology," in *Comprehensive Natural Products II*. eds. H.-W. Liu and L. Mander (Oxford: Elsevier), 351–410.

Vozza, A., De Leonardis, F., Paradies, E., De Grassi, A., Pierri, C. L., Parisi, G., et al. (2017). Biochemical characterization of a new mitochondrial transporter of dephosphocoenzyme A in *Drosophila melanogaster*. *Biochim. Biophys. Acta Bioenerg.* 1858, 137–146. doi: 10.1016/J.BBABIO.2016.11.006

Walker, J. E., Saraste, M., Runswick, M. J., and Gay, N. J. (1982). Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. *EMBO J.* 1, 945–951. doi: 10.1002/j.1460-2075.1982.tb01276.x

Weisburg, W. G., Tully, J. G., Rose, D. L., Petzel, J. P., Oyaizu, H., Yang, D., et al. (1989). A phylogenetic analysis of the mycoplasmas: basis for their classification. *J. Bacteriol.* 171, 6455–6467. doi: 10.1128/jb.171.12.6455-6467.1989

Wium, M., Botes, A., and Bellstedt, D. U. (2015). The identification of oppA gene homologues as part of the oligopeptide transport system in mycoplasmas. *Gene* 558, 31–40. doi: 10.1016/j.gene.2014.12.036

Woese, C. R. (1987). Bacterial evolution. *Microbiol. Rev.* 51, 221–271. doi: 10.1128/mmbr.51.2.221-271.1987

Yang, K., Eyobo, Y., Brand, L. A., Martynowski, D., Tomchick, D., Strauss, E., et al. (2006). Crystal structure of a type III pantothenate kinase: insight into the mechanism of an essential coenzyme A biosynthetic enzyme universally distributed in bacteria. *J. Bacteriol.* 188, 5532–5540. doi: 10.1128/JB.00469-06

Yus, E., Maier, T., Michalodimitrakis, K., Van Noort, V., Yamada, T., Chen, W. H., et al. (2009). Impact of genome reduction on bacterial metabolism and its regulation. *Science* 326, 1263–1268. doi: 10.1126/science.1177263

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 28 February 2022
doi: 10.3389/fmicb.2022.773608

Comparative Genome Analysis of '*Candidatus* Phytoplasma luffae' Reveals the Influential Roles of Potential Mobile Units in Phytoplasma Evolution

Ching-Ting Huang[1], Shu-Ting Cho[1], Yu-Chen Lin[1], Choon-Meng Tan[2], Yi-Ching Chiu[2,3], Jun-Yi Yang[2,3,4] and Chih-Horng Kuo[1,3,5,6]**

[1] Institute of Plant and Microbial Biology, Academia Sinica, Taipei, Taiwan, [2] Institute of Biochemistry, National Chung Hsing University, Taichung, Taiwan, [3] Ph.D. Program in Microbial Genomics, National Chung Hsing University and Academia Sinica, Taichung, Taiwan, [4] Advanced Plant Biotechnology Center, National Chung Hsing University, Taichung, Taiwan, [5] Molecular and Biological Agricultural Sciences Program, Taiwan International Graduate Program, National Chung-Hsing University and Academia Sinica, Taipei, Taiwan, [6] Biotechnology Center, National Chung Hsing University, Taichung, Taiwan

OPEN ACCESS

Edited by:
Robert Czajkowski,
University of Gdańsk, Poland

Reviewed by:
Edel Pérez-López,
Laval University, Canada
Christophe Garcion,
INRA Centre Bordeaux-Aquitaine, France

***Correspondence:**
Jun-Yi Yang
jyang@nchu.edu.tw
Chih-Horng Kuo
chk@gate.sinica.edu.tw

Specialty section:
This article was submitted to Microbial Symbioses, a section of the journal Frontiers in Microbiology

Received: 10 September 2021
Accepted: 07 February 2022
Published: 28 February 2022

Citation:
Huang C-T, Cho S-T, Lin Y-C, Tan C-M, Chiu Y-C, Yang J-Y and Kuo C-H (2022) Comparative Genome Analysis of 'Candidatus Phytoplasma luffae' Reveals the Influential Roles of Potential Mobile Units in Phytoplasma Evolution. Front. Microbiol. 13:773608. doi: 10.3389/fmicb.2022.773608

Phytoplasmas are insect-transmitted plant pathogens that cause substantial losses in agriculture. In addition to economic impact, phytoplasmas induce distinct disease symptoms in infected plants, thus attracting attention for research on molecular plant-microbe interactions and plant developmental processes. Due to the difficulty of establishing an axenic culture of these bacteria, culture-independent genome characterization is a crucial tool for phytoplasma research. However, phytoplasma genomes have strong nucleotide composition biases and are repetitive, which make it challenging to produce complete assemblies. In this study, we utilized Illumina and Oxford Nanopore sequencing technologies to obtain the complete genome sequence of '*Candidatus* Phytoplasma luffae' strain NCHU2019 that is associated with witches' broom disease of loofah (*Luffa aegyptiaca*) in Taiwan. The fully assembled circular chromosome is 769 kb in size and is the first representative genome sequence of group 16SrVIII phytoplasmas. Comparative analysis with other phytoplasmas revealed that NCHU2019 has a remarkably repetitive genome, possessing a pair of 75 kb repeats and at least 13 potential mobile units (PMUs) that account for ∼25% of its chromosome. This level of genome repetitiveness is exceptional for bacteria, particularly among obligate pathogens with reduced genomes. Our genus-level analysis of PMUs demonstrated that these phytoplasma-specific mobile genetic elements can be classified into three major types that differ in gene organization and phylogenetic distribution. Notably, PMU abundance explains nearly 80% of the variance in phytoplasma genome sizes, a finding that provides a quantitative estimate for the importance of PMUs in phytoplasma genome variability. Finally, our investigation found that in addition to horizontal gene transfer, PMUs also contribute to intra-genomic duplications of effector

genes, which may provide redundancy for subfunctionalization or neofunctionalization. Taken together, this work improves the taxon sampling for phytoplasma genome research and provides novel information regarding the roles of mobile genetic elements in phytoplasma evolution.

Keywords: plant pathogen, phytoplasma, genomics, molecular evolution, mobile genetic element, effector

INTRODUCTION

Phytoplasmas are uncultivated bacteria associated with plant diseases in several hundred species (Lee et al., 2000; Hogenhout et al., 2008; Bertaccini et al., 2014; Namba, 2019). In infected plants, phytoplasma cells are restricted to phloem tissues and can secrete effector proteins that cause developmental abnormalities of the hosts (Sugio et al., 2011b). Typical symptoms of phytoplasma infections include stunting, dwarfism, virescence (i.e., greening of flowers), phyllody (i.e., abnormal development of floral parts into leaf-like tissues), and witches' broom (i.e., proliferation of stems and leaves), which result in substantial agricultural losses.

For classification of these uncultivated bacteria, a system based on restriction fragment length polymorphism (RFLP) analysis of their 16S rRNA genes was developed in the 1990s (Lee et al., 1993, 1998; Gundersen and Lee, 1996) and at least 33 16S rRNA gene RFLP (16Sr) groups have been described (Zhao et al., 2009; Zhao and Davis, 2016). Later, a provisional genus-level taxon '*Candidatus* Phytoplasma' was proposed to accommodate these bacteria (The IRPCM Phytoplasma/Spiroplasma Working Team - Phytoplasma taxonomy group, 2004) and at least 41 '*Ca*. P.' species have been described or proposed (Bertaccini and Lee, 2018). Based on analysis of 16S rRNA and other conserved genes, phytoplasmas are divided into three major phylogenetic clusters (Hogenhout and Seruga Music, 2009; Chung et al., 2013; Seruga Music et al., 2019). Early genomics studies were mainly conducted for clusters I (i.e., '*Ca*. P. asteris' of group 16SrI and '*Ca*. P. australiense' of 16SrXII) (Oshima et al., 2004; Bai et al., 2006; Tran-Nguyen et al., 2008) and II (i.e., '*Ca*. P. mali' of 16SrX) (Kube et al., 2008). In comparison, cluster III harbors the highest level of diversity, yet has received limited attention for comparative genomics studies (Chung et al., 2013; Wang et al., 2018a).

To improve our understanding of phytoplasma genome diversity, we conducted whole genome sequencing of a '*Ca*. P. luffae' strain collected in Taiwan. The species-level taxon '*Ca*. P luffae' belongs to group 16SrVIII in cluster III and is associated with witches' broom disease of loofah (*Luffa aegyptiaca*) (Davis et al., 2017). The availability of a complete genome sequence from this taxon provides a complete view of its gene content, which can facilitate the study of its pathogenesis mechanisms and other aspects of its biology. More importantly, with the increased availability of genome sequences from diverse phytoplasmas (**Table 1**), we performed genus-level comparative analysis to obtain a more comprehensive picture of their genomic diversity. This improves upon previous works that are limited to comparisons of closely related taxa or have sparse sampling (Bai et al., 2006; Tran-Nguyen et al., 2008; Kube et al., 2012; Saccardo et al., 2012; Andersen et al., 2013; Chung et al., 2013; Orlovskis et al., 2017; Wang et al., 2018a; Cho et al., 2019, 2020a; Seruga Music et al., 2019; Kirdat et al., 2021; Zhao et al., 2021). Furthermore, our focused analysis of the potential mobile units (PMUs) (Bai et al., 2006) revealed the influential roles of these mobile genetic elements in the evolution of phytoplasma genome organization and effector gene content.

MATERIALS AND METHODS

Biological Materials

The strain NCHU2019 was collected from a naturally infected loofah plant found in Dacheng Township (Changhua County, Taiwan; 23.861860 N, 120.291322 E) on July 4, 2019 (**Figure 1A**). After initial collection, the bacterium was transferred to lab-grown loofah plants (cultivar A-Jun, Known-You Seed Co., Kaohsiung, Taiwan) through grafting and maintained in a plant growth facility in National Chung Hsing University (Taichung, Taiwan) (**Figure 1B**). To confirm the presence and identity of this phytoplasma strain, a partial sequence of the rRNA operon was PCR amplified using the phytoplasma-specific primer set P1/P7 for Sanger sequencing as described (Chung et al., 2013).

Genome Sequencing, Assembly, and Annotation

The procedures for genome sequencing and analysis were based on those described in our previous work on phytoplasma genomes (Chung et al., 2013; Cho et al., 2019, 2020a,b). All kits were used according to the manufacturer's protocols and all bioinformatics tools were used with the default settings unless stated otherwise.

For whole genome shotgun sequencing, leaves from one artificially infected plant exhibiting typical symptoms (i.e., small leaves and witches' broom) (**Figure 1B**) were collected for total genomic DNA extraction using the Wizard Genomic DNA Purification Kit (A1120; Promega, Madison, WI, United States). For Illumina sequencing, the DNA sample was processed using the KAPA Library Preparation Kit (KK8234) and the Invitrogen SizeSelect Gels (G6610-02) to obtain ∼550 bp fragments, followed by MiSeq 2 × 300 bp paired-end sequencing (v3 chemistry). For Oxford Nanopore Technologies (ONT) sequencing, the library was prepared using the ONT Ligation Kit (SQK-LSK109) without shearing or size selection, followed by MinION sequencing (R9.4.1 chemistry) and Guppy v3.3.3 base-calling.

The *de novo* genome assembly involved two stages. In the first stage, only the Illumina reads were used for running Velvet v1.2.10 (Zerbino and Birney, 2008) with k-mer length set to

TABLE 1 | List of the genome assemblies analyzed. For each strain, information regarding the 16S rRNA gene (16Sr) group, genome accession number, assembly status, genome size, coding sequence (CDS) count, potential mobile unit (PMU) gene count, and combined length of PMU genes are provided. The PMU gene information is based on homologs of eight core genes (i.e., *tra5*, *dnaB*, *dnaG*, *tmk*, *hflB*, *himA*, *ssb*, and *rpoD*), including full-length genes and putative pseudogenes. A complete list of these genes is provided in **Supplementary Table 1**. Other genes located in proximity, such as those encoding hypothetical proteins or putative secreted proteins, are excluded. *Acholeplasma laidlawii* is included as an outgroup.

Strain	16Sr group	Accession	Assembly	Genome size (bp)	CDS count	PMU gene count	PMU gene length (bp)
'*Ca.* P. asteris' AYWB	I	GCF_000012225.1	Complete	706,569	514	112	49,188
'*Ca.* P. asteris' De Villa	I	GCF_004214875.1	Complete	600,116	521	21	11,958
'*Ca.* P. asteris' DY2014	I	GCA_005093185.1	8 contigs	824,596	775	89	48,761
'*Ca.* P. asteris' LD1	I	GCF_001866375.1	8 contigs	599,264	513	29	15,441
'*Ca.* P. asteris' M3	I	GCF_001712875.1	Complete	576,118	490	27	14,541
'*Ca.* P. asteris' OY-M	I	GCF_000009845.1	Complete	853,092	708	137	72,399
'*Ca.* P. tritici' WBD	I	GCF_000495255.1	6 contigs	611,462	471	45	19,549
'*Ca.* P. aurantifolia' NCHU2014	II	GCA_001307505.2	Complete	635,584	471	23	16,378
'*Ca.* P. pruni' CX	III	GCF_001277135.1	46 contigs	598,511	434	31	22,474
'*Ca.* P. ziziphi' Jwb-nky	V	GCF_003640545.1	Complete	750,803	641	69	49,824
'*Ca.* P. luffae' NCHU2019	VIII	GCA_018024475.1	Complete	769,143	725	117	84,321
'*Ca.* P. mali' AT	X	GCF_000026205.1	Complete	601,943	495	20	15,084
'*Ca.* P. oryzae' Mbita	XI	GCF_001578535.1	28 contigs	533,195	432	17	14,259
'*Ca.* P. sacchari' SCGS	XI	GCF_009268105.1	28 contigs	502,197	402	15	13,620
'*Ca.* P. australiense' NZSb11	XII	GCF_000397185.1	Complete	959,779	828	131	78,642
'*Ca.* P. australiense' PAa	XII	GCF_000069925.1	Complete	879,959	699	108	55,565
'*Ca.* P. solani' SA-1	XII	GCA_003698095.1	19 contigs	821,322	709	66	45,234
'*Ca.* P. cynodontis' LW01	XIV	GCF_009268075.1	23 contigs	483,935	425	13	11,856
'*Ca.* P. pini' MDPP	XXI	GCF_007821455.1	16 contigs	474,136	392	15	11,685
Acholeplasma laidlawii PG-8A	N/A	GCF_000018785.1	Complete	1,496,992	1,378	10	8,811

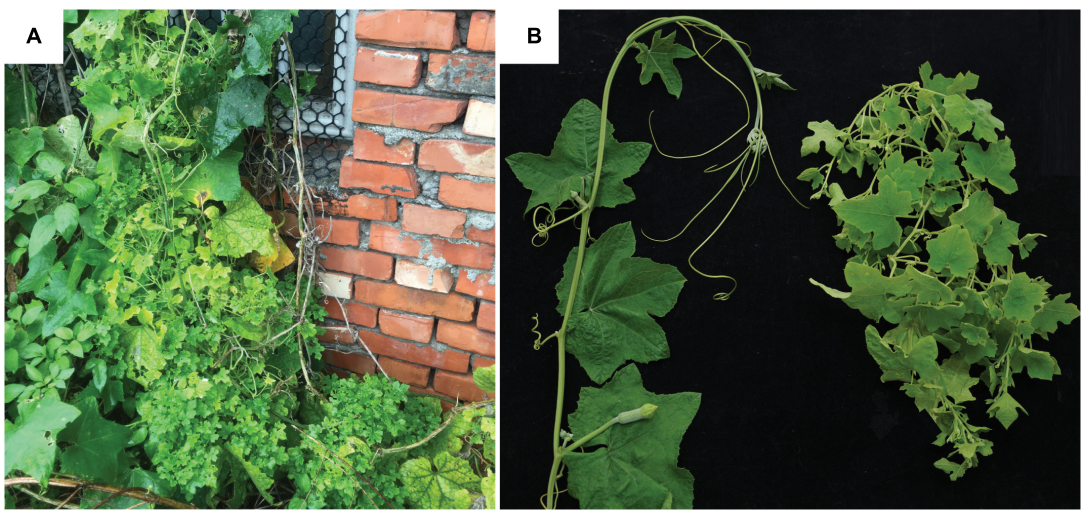


FIGURE 1 | Infection symptoms. **(A)** The loofah plant exhibiting phytoplasma infection symptoms that was collected in Changhua, Taiwan. **(B)** Loofah plants grown in the lab. Left, healthy control; right, grafted with the infected plant shown in panel **(A)**.

151 and minimum contig length set to 2,000 bp. To identify putative phytoplasma contigs, all contigs were queried against the NCBI non-redundant protein database (Benson et al., 2018) using BLASTX v2.10.0 (Camacho et al., 2009). Those having a hit with an *e*-value of $<1e$-15 to phytoplasma proteins were selected for manual inspection. False positive contigs derived from plant chloroplast and mitochondrial genomes were identified based on the difference in sequencing depth and verified based on BLASTN v2.10.0 (Camacho et al., 2009) searches against the NCBI standard nucleotide collection database (Benson et al., 2018).

To start the second stage of assembly, the putative phytoplasma contigs identified from the first stage were used as the reference for extracting phytoplasma reads from the Illumina data set using BWA v0.7.17 (Li and Durbin, 2009) with an alignment score cutoff of 30 and from the ONT data set using Minimap2 v2.15 (Li, 2018) with an alignment score

cutoff of 1,000. The extracted reads were processed together for *de novo* hybrid assembly using Unicycler v0.4.9b. The contigs were validated as belonging to the phytoplasma genome by BLAST searches according to the process utilized in the first stage and used as the starting point for an iterative process of assembly improvement. In each iteration, all Illumina and ONT raw reads were mapped to the draft assembly as described. The mapping results were programmatically checked using the "mpileup" function of SAMTOOLS v1.9 (Li et al., 2009) and manually inspected using IGV v2.5.0 (Robinson et al., 2011) to identify possible assembly errors. Regions with raw reads mapping results exhibiting abnormalities were cut and re-arranged manually based on the continuity of ONT long reads, then validated using the mapping results in the next iteration. During the early iterations, the mapping results of ONT reads were used to provide scaffolding information and validate the overall organization of the circular chromosome, particularly the junctions between repetitive and unique regions. Additionally, the reads mapped to contig ends were visually inspected using IGV for manual selection of representative reads that can extend contigs and fill gaps. During the later iterations, the mapping results of Illumina reads were used to validate bp-scale indels and possible sequencing errors introduced by the ONT reads. The process was repeated until the complete genome assembly was obtained and all regions are supported by the raw reads mapping results. Additionally, the "depth" function of SAMTOOLS is used to calculate the sequencing coverage.

To provide a genome size estimate based on k-mer distribution, all Illumina reads mapped to the finalized assembly with an alignment score above 200 were extracted. Based on these reads, occurrences of k-mers in the size range between 17 and 63 were calculated by using jellyfish v2.2.8 (Marçais and Kingsford, 2011). The genome size was estimated by dividing the total k-mer count with the peak depth as suggested previously (Lu et al., 2016).

The procedure of gene prediction was performed using RNAmmer v1.2 (Lagesen et al., 2007), tRNAscan-SE v1.3.1 (Lowe and Eddy, 1997), and Prodigal v2.6.3 (Hyatt et al., 2010). The annotation was based on the homologous gene clusters present in other phytoplasma genomes (**Table 1**) as identified by BLASTP v 2.10.0 (Camacho et al., 2009) and OrthoMCL v1.3 (Li et al., 2003), followed by manual curation based on information obtained from GenBank (Benson et al., 2018), KEGG (Kanehisa et al., 2010), and COG (Tatusov et al., 2003) databases. Additionally, putative secreted proteins were predicted using SignalP v5.0 (Armenteros et al., 2019) based on the Gram-positive bacteria model. Those candidates with transmembrane domains identified by TMHMM v2.0 (Krogh et al., 2001) were removed and the remaining ones were required to have a signal peptide length in the range of 21–52 amino acids. For visualization, the Circos v0.69-6 (Krzywinski et al., 2009) was used to draw the genome map.

Comparative Analysis

For comparative analysis with other representative phytoplasma genomes (**Table 1**), homologous gene clusters were identified using OrthoMCL (Li et al., 2003). Multiple sequence alignments of homologous genes were prepared using MUSCLE v3.8.31 (Edgar, 2004) and visualized using JalView v2.11 (Waterhouse et al., 2009). Maximum likelihood phylogenies were inferred using PhyML v3.3 (Guindon and Gascuel, 2003) with the default LG substitution model and visualized using FigTree v1.4.4. PHYLIP v3.697 (Felsenstein, 1989) was used for bootstrap analysis. Classification of the 16S RFLP group was performed using *i*PhyClassifier (Zhao et al., 2009).

For whole-genome comparison, FastANI v1.1 (Jain et al., 2018) was used to calculate the proportion of genomic segments mapped and the average nucleotide identity (ANI). For pairwise genome alignments, Mummer v3.23 (Kurtz et al., 2004) was used with the options "–maxmatch –mincluster 30" and the results were visualized using genoPlotR v0.8.9 (Guy et al., 2010).

The PMU analysis was based on the eight core genes (i.e., *tra5*, *dnaB*, *dnaG*, *tmk*, *hflB*, *himA*, *ssb*, and *rpoD*) defined previously (Bai et al., 2006). To ensure uniform annotation and to include possible pseudogenes, all genome assemblies were examined by using representative PMU genes as queries to run TBLASTN searches. Hits that have a high-scoring segment pair (HSP) accounting for at least 30% of the query length and at least 50% sequence similarity within HSP were identified (**Table 1** and **Supplementary Table 1**). For defining a putative PMU region, at least four intact core genes are required to be in proximity and have the same orientation. This criterion was chosen to identify the less degenerated PMU regions for the purpose of establishing a classification scheme. The limitation is that the more fragmented PMU regions would be excluded from the analysis. Exceptions were made for two putative PMU regions that do not meet this criterion, namely, AYWB_4 and AYWB_5, because '*Ca*. P. asteris' AYWB is often used as a reference for comparative analysis of PMUs. All statistical tests were performed in the R statistical environment (R Core Team, 2019); correlation coefficients were calculated using the "cor.test" function, linear regression was performed using the "lm" function and visualized using the "plot" and "abline" functions.

RESULTS AND DISCUSSION

Genomic Characterization of NCHU2019

The shotgun sequencing generated ∼4.7 Gb of Illumina raw reads and ∼3.5 Gb of ONT raw reads. In the first stage of *de novo* assembly with only the Illumina reads, a draft assembly with 49 contigs totaling 450,754 bp was obtained (longest contig: 31,507 bp; shortest contig: 2,001 bp; N50: 12,506 bp). Based on the mapping results to this first draft assembly, we extracted 0.9% (136,116 out of 15,655,220) of all Illumina raw reads and 1.7% (9,416 out of 539,354) of all ONT raw reads. These reads correspond to a sequencing depth of 71- and 110-fold, respectively. In the second stage of *de novo* hybrid assembly using these extracted reads from both sequencing platforms, a second draft assembly with 10 contigs totaling 668,430 bp was obtained. These 10 contigs include two long ones that are 552,199 and 103,217 bp in size, as well as eight short ones that are shorter than 3,000 bp. Examination of the mapping results to this second draft

assembly found 186,979 Illumina reads with 73-fold coverage and 13,242 ONT reads with 95-fold coverage.

During the iterative assembly improvement process, a pair of duplicated chromosomal segments that are each ∼75 kb in size and multiple smaller repeats were identified. These regions were manually corrected to generate the finalized assembly. All junctions involving repetitive regions were verified based on visual inspection of the ONT long read mapping results. For the finalized assembly, we obtained a circular contig that corresponds to the phytoplasma chromosome; no plasmid was found. This circular contig is 769,143 bp in size with 23.3% G + C content (**Figure 2**). For the final round of raw reads mapping examination, 149,857 Illumina reads and 13,649 ONT reads were mapped to this circular contig, corresponding to 51- and 80-fold sequencing depth, respectively. The decrease of sequencing depth compared to the draft assemblies is likely explained by the repetitive regions being resolved. As a support for this inference, the sequencing depth exhibits a nearly uniform distribution across the entire finalized circular contig, with repetitive regions having similar depth as non-repetitive regions (**Supplementary Figure 1**).

For genome size estimation based on k-mer analysis, we utilized $k = 21$ as a representative data set for visualization (**Supplementary Figure 2A**). We found that while the peak depth is 51, the frequency distribution is nearly flat in the range between ∼40 and 55. Based on this distribution, the genome size is estimated to be in the range between 660 and 842 kb (**Supplementary Figure 2B**). The assembled genome size of 769 kb is near the middle point (i.e., 751 kb) of these estimates.

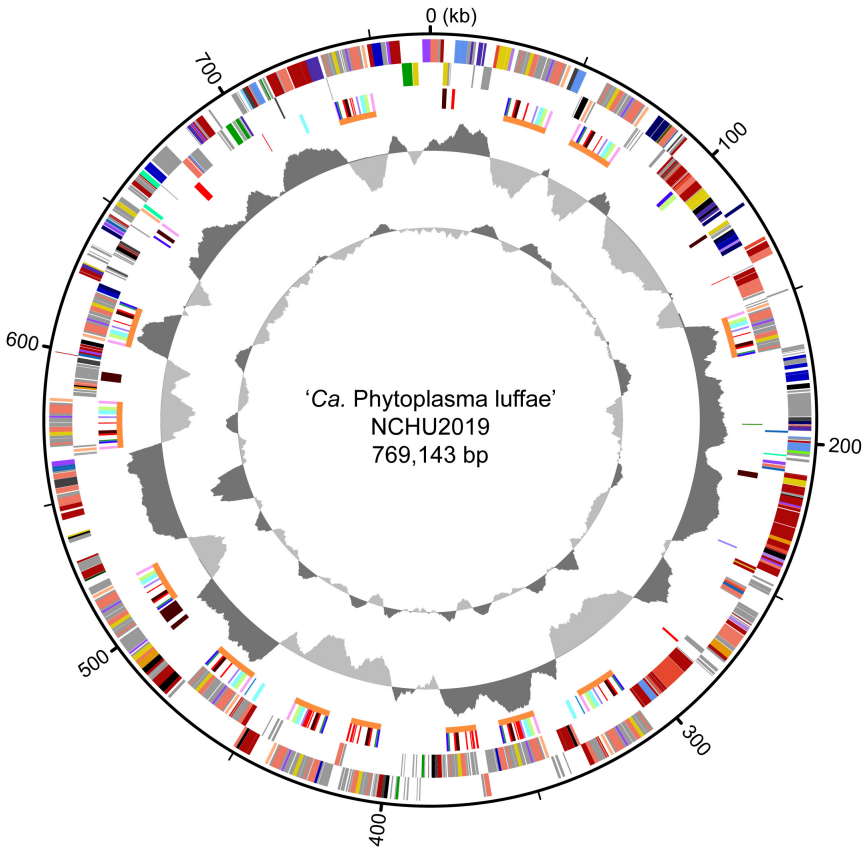
The annotation of this phytoplasma genome contains two complete sets of 16S-23S-5S rRNA gene clusters, 31 tRNA genes, 725 coding sequences (CDSs), and 13 pseudogenes. Both copies of the 16S rRNA gene are 100% identical to the reference sequence of '*Ca*. P. luffae' LfWB (GenBank accession AF248956). Among the CDSs, 317 (44%) lacked any COG functional category assignments. Among the CDSs that were assigned to specific functional categories, those assigned to information storage and processing (e.g., replication, transcription, and translation) represent the largest group that account for 32% of all CDSs. In comparison, those assigned to cell process and metabolism account for 9 and 13%, respectively. These observations are consistent with findings from characterization of other phytoplasmas and more distantly related Mollicutes (e.g., *Spiroplasma*, *Entomoplasma*, and *Mycoplasma*) (Chen et al., 2012; Kube et al., 2012; Lo et al., 2018). The observation that a large fraction of genes lack specific functional annotation may be attributed to the elevated evolutionary rates of Mollicutes and their distant phylogenetic relationships from model organisms (Wu and Eisen, 2008; Wu et al., 2009; McCutcheon and Moran, 2012). The observation that genes related to information storage and processing genes are relatively abundant compared to those involved in metabolism is common among symbiotic bacteria with small genomes (McCutcheon and Moran, 2012; Lo et al., 2016).

Other than the low G + C content and reduced gene content, phytoplasma genomes are generally known to be repetitive, partly due to the presence of PMUs (Sugio and Hogenhout, 2012). Interestingly, the genome of this strain is far more repetitive than other phytoplasmas that have been characterized. On average, PMU core genes account for ∼4.7% (Std. Dev. = 2.7%) of the genome size among those 19 representative phytoplasma strains analyzed (**Table 1** and **Supplementary Table 1**). For strain NCHU2019, there are 117 PMU core genes organized into at least 13 distinct PMU regions (**Figure 2** and **Supplementary Table 2**) and account for 11% of the chromosome length. Additionally, a pair of chromosomal segments, each ∼75 kb in size, were found to be duplicated (positions: 315,975-391,140 and 391,446-466,612; sequence identity: 75,160/75,167 = 99.99%). The same set of 79 CDSs arranged in the same order were found in each of the repeats. Out of these CDSs, 27 correspond to hypothetical proteins. Two notable genes are *ruvA*/*B*, which are involved DNA recombination and repair. The duplication of these genes made NCHU2019 the only phytoplasma strain with two copies of *ruvA*/*B* reported to date. Together, these two duplicated segments and the 13 PMU regions account for 25% of the chromosome length. Explanation for the high genome repetitiveness of strain NCHU2019 compared to other phytoplasmas is unclear.

Comparisons With Other Phytoplasmas

Based on the established classification scheme and a previous study of 16S rRNA gene phylogeny (Davis et al., 2017), '*Ca*. P. luffae' belongs to group 16SrVIII and is most closely related to '*Ca*. P. malaysianum' (group 16SrXXXII; GenBank accession EU371934) (Nejat et al., 2013). These two species-level taxa have 95.9% sequence identity (i.e., 1,463/1,526 aligned nucleotides) in their 16S rRNA genes. However, no genome sequence is available for '*Ca*. P. malaysianum' for comparative analysis.

Other than '*Ca*. P. malaysianum,' the next closest relatives of '*Ca*. P. luffae' include phytoplasmas belonging to groups 16SrV ('*Ca*. P. ziziphi'), 16SrVI ('*Ca*. P. sudamericanum' and '*Ca*. P. trifolii'), and 16SrVII ('*Ca*. P. fraxini'). Among these, one complete genome sequence (GenBank accession CP025121) is available for '*Ca*. P. ziziphi' strain Jwb-nky (Wang et al., 2018a), which represents the most closely related lineage for comparative analysis (**Figure 3A**). Comparison based on the 16S rRNA gene sequences indicated that '*Ca*. P. luffae' NCHU2019 and '*Ca*. P. ziziphi' Jwb-nky have 94.9% sequence identity (i.e., 1,447/1,524 aligned nucleotides). For whole-genome comparison, only 49% of the chromosomal segments can be mapped between these two strains and these segments have <80% ANI. Pairwise genome alignment indicated that the most of the conserved regions correspond to PMUs (**Figure 4**), which further supports that the sequence divergence between these two genomes is too high for nucleotide-level comparisons. The lack of chromosome-level synteny conservation was commonly reported in previous comparisons between complete genome sequences of phytoplasmas (Bai et al., 2006; Tran-Nguyen et al., 2008; Andersen et al., 2013; Orlovskis et al., 2017; Wang et al., 2018a), even for '*Ca*. P. asteris' strains sharing >99.9% 16S rRNA gene sequence identity and >98.1% genome-wide ANI (Cho et al., 2020a). These observations may be attributed to the strong nucleotide composition biases, the high mutation accumulation rates, and the influence of PMUs (Cho et al., 2020a).



INFORMATION
- [J] Translation
- [K] Transcription
- [L] Replication

POORLY CHARACTERIZED
- [R] General prediction only
- [S] Function unknown
- [?] No COG assignment

CELL PROCESS
- [D] Cell cycle
- [V] Defense
- [T] Signal transduction
- [M] Cell wall, membrane, & envelope
- [N] Motility
- [W] Extracellular structure
- [U] Trafficking, secretion, & transport
- [O] Posttranslational modification
- [X] Mobilome

METABOLISM
- [C] Energy
- [G] Carbohydrate
- [E] Amino acid
- [F] Nucleotide
- [H] Coenzyme
- [I] Lipid
- [P] Inorganic ion
- [Q] Secondary metabolites

FIGURE 2 | Genome map of '*Candidatus* Phytoplasma luffae' NCHU2019. Rings from outside in: (1) Scale marks (kb). (2 and 3) Coding sequences on the forward and reverse strand, respectively. Color-coded by functional categories. (4) Genes associated with potential mobile units (PMUs), color-coded by annotation (see **Figure 5**). Gene clusters that represent individual PMUs are labeled by orange lines. Genes encoding putative secreted proteins are in red. (5) GC skew (positive: dark gray; negative: light gray). (6) GC content (above average: dark gray; below average; light gray). One high GC peak located in the 531–544 kb region corresponds to two adjacent rRNA operons.

Considering the high level of nucleotide sequence divergence between '*Ca.* P. luffae' NCHU2019 and '*Ca.* P. ziziphi' Jwb-nky, we also examined gene content based on protein sequence comparisons. In the pairwise comparison, 344 homologous gene clusters are shared between these two strains, while 175 and 241 are specific to NCHU2019 and Jwb-nky, respectively. These counts of shared and strain-specific genes are comparable to those found in previous studies of phytoplasma genome comparisons at between-species level (Chung et al., 2013; Tan et al., 2021). Although relatively large numbers of strain-specific genes were identified, >80% of these genes lack functional annotation, making the inference of their roles difficult. As such, the gene content regarding metabolic capacity and transporters of '*Ca.* P. luffae' NCHU2019 is expected to be highly similar to that of '*Ca.* P. ziziphi' Jwb-nky, which was described in detail previously (Wang et al., 2018a).

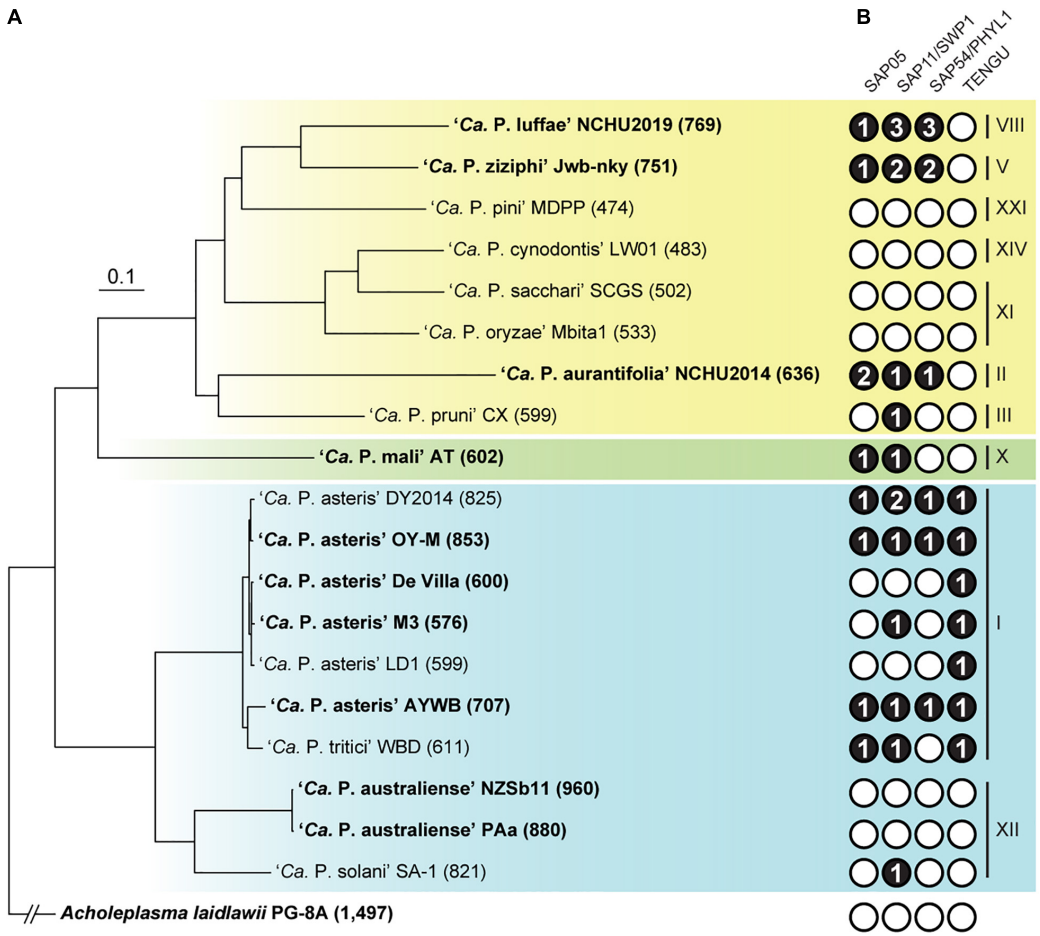


FIGURE 3 | Evolutionary relationships among representative phytoplasmas with genome sequences available. The 16S rRNA gene (16Sr) group assignments are labeled on the right. The three major phylogenetic clusters are indicated by colored backgrounds (I: blue; II; green; III: yellow). *Acholeplasma laidlawii* is included as the outgroup. **(A)** Maximum likelihood phylogeny inferred based on a concatenated alignment of 80 conserved single-copy genes (27,742 aligned amino acid sites). All internal nodes received >80% bootstrap support based on 1,000 re-sampling. Strains with complete genome assemblies available are highlighted in bold. The number in parentheses following the strain name indicates the genome size (in kb). **(B)** Distribution of known effector genes. Gene presence and absence are indicated by filled (with copy numbers) and empty circles, respectively.

For comparisons with those more divergent phytoplasmas with genomic information available (**Table 1**), only 134 homologous gene clusters are conserved among the 19 representatives analyzed. This estimate of phytoplasma core genome is much lower than the ∼200 genes consistently reported in previous studies (Chen et al., 2012; Chung et al., 2013; Tan et al., 2021), an observation that is likely explained by the inclusion of several draft assemblies in this study. Detailed description regarding the functions of these ∼200 core genes were reported previously (Chen et al., 2012; Kube et al., 2012; Chung et al., 2013).

Detailed Characterization of Potential Mobile Units

To better understand the roles of PMUs in phytoplasma genome evolution, we conducted detailed characterization of these mobile genetic elements. Among the 13 PMUs identified in '*Ca.* P. luffae' NCHU2019 (**Figure 5A**), 11 are considered as complete ones and range from 14 to 18 kb in size. The remaining two (i.e., #6 and #7) are both 12 kb in size and appear to be truncated based on the lack of multiple PMU core genes (i.e., *tmk*, *dnaB*, *dnaG*, and *tra5*). These 13 PMUs are variable in the gene content in between *rad50* and *tmk*, while the sequences of shared core genes are highly conserved. For example, the *tmk* and *dnaB* homologs among these PMUs have identical sequences (**Figure 6**). It is unclear if the sequence conservation of these core genes is explained by lack of mutation accumulation, purifying selection, or assembly artifacts. Regarding chromosomal locations, these PMUs are interspersed across the entire chromosome and there is no obvious pattern of clustering (**Figures 2**, **4**). Two sets of PMUs (i.e., #4–6 and #7–9) are associated with the two 75-kb repeat regions. Notably, one set of junctions for these large repeats (chromosomal positions 315,975 and 466,612) are located inside PMU #4 (positions 308,032.323,983) and PMU #9 (positions 458,604.475,923). This finding suggests

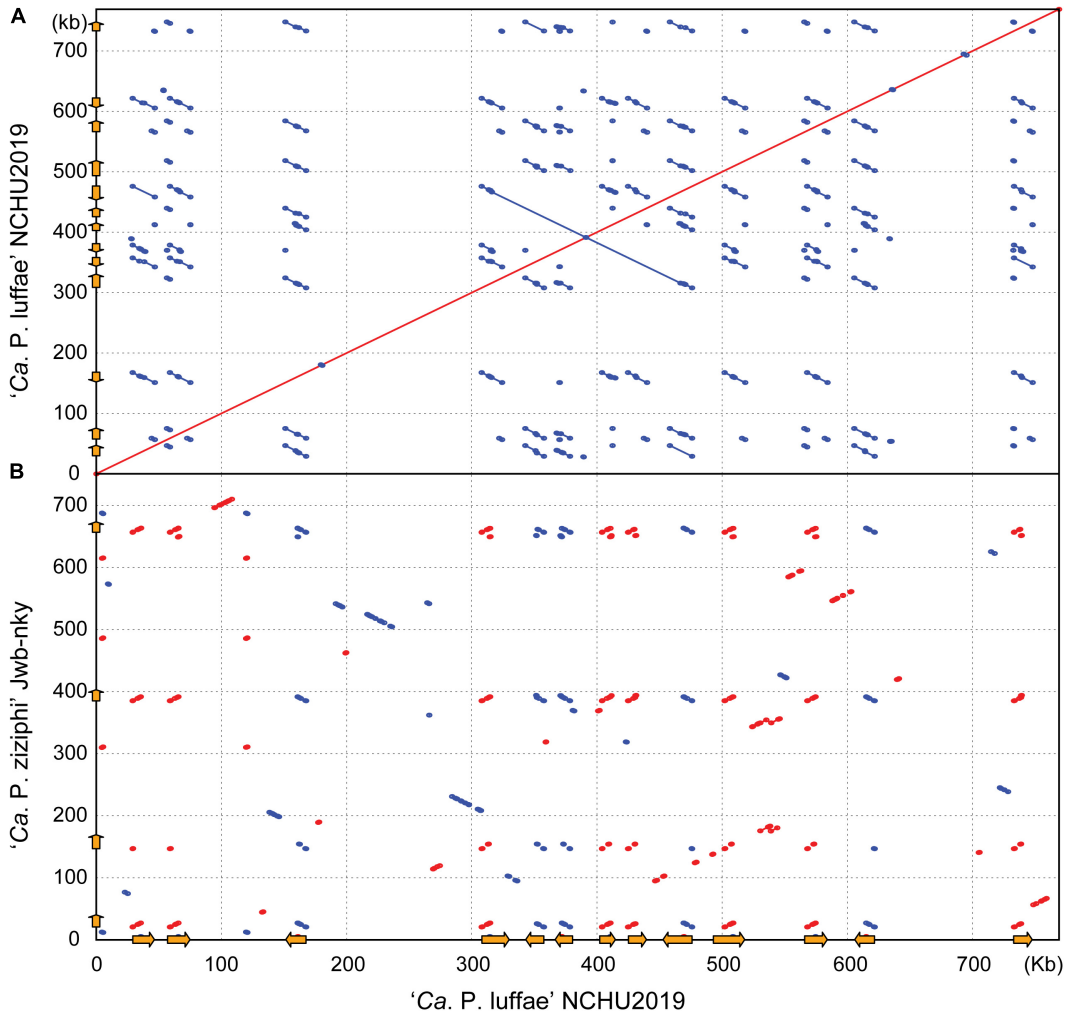


FIGURE 4 | Pairwise genome alignments. The genome of '*Ca*. P. luffae' NCHU2019 was used as the reference for comparison with **(A)** itself and **(B)** '*Ca*. P. ziziphi' Jwb-nky. Matches on the same strand and the opposite strand are indicated in red and blue, respectively. Potential mobile units (PMUs) are illustrated as orange arrows on both axes.

that homologous recombination facilitated by PMUs may have promoted the segmental duplication of this chromosome, which increased the PMU copy number.

Based on initial characterization of the PMUs in '*Ca*. P. asteris' AYWB genome, eight core genes were identified (Bai et al., 2006). As more genome sequences become available from diverse phytoplasmas, we were able to include representatives of 13 '*Ca*. P.' species from all three phylogenetic clusters (**Table 1**) for comparison. Compared to other phytoplasmas, '*Ca*. P. luffae' NCHU2019 likely has the highest number of intact PMUs and is distinctive in that all of its PMUs are similar (**Figure 5A** and **Supplementary Table 2**). Among the 11 complete PMUs in this genome, the eight PMU core genes and *rad50* are all organized in the same order. The minor variation in gene organization is mainly in between *rad50* and *tmk*, where genes encode different putative secreted proteins and hypothetical proteins are found. For comparison, in the closely related '*Ca*. P. ziziphi' Jwb-nky of cluster III (i.e., the yellow clade in **Figure 3A**), the four PMUs exhibit much higher levels of diversity in gene organization between *hflB* and *tmk* (**Figure 5A**). For the distantly related '*Ca*. P. asteris' AYWB and '*Ca*. P. australiense' PAa of cluster I (i.e., the blue clade in **Figure 3A**), high levels of intra-genomic PMU diversity are observed (**Figure 5B** and **Supplementary Table 2**).

Based on the presence/absence pattern and order of eight PMU core genes defined previously (Bai et al., 2006), we found that the PMUs in these phytoplasmas can be classified into three major types (**Figure 5B**). The strains omitted in the visualization all have close relatives belonging to the same 16Sr group and share similar PMUs when there are identifiable ones (**Supplementary Table 2**). Among the three major types, type A PMUs, in which *tmk* is upstream of *dnaB*, are the first reported ones (Bai et al., 2006) and also the most common ones that are found in phylogenetic clusters I (e.g., '*Ca*. P. asteris,' '*Ca*. P. australiense,' and '*Ca*. P. solani') and III (e.g., '*Ca*. P. luffae,' '*Ca*. P. ziziphi,' and '*Ca*. P. aurantifolia'). Type B PMUs are shorter, have *tmk* downstream of *dnaB*, and are found only in '*Ca*. P. australiense'

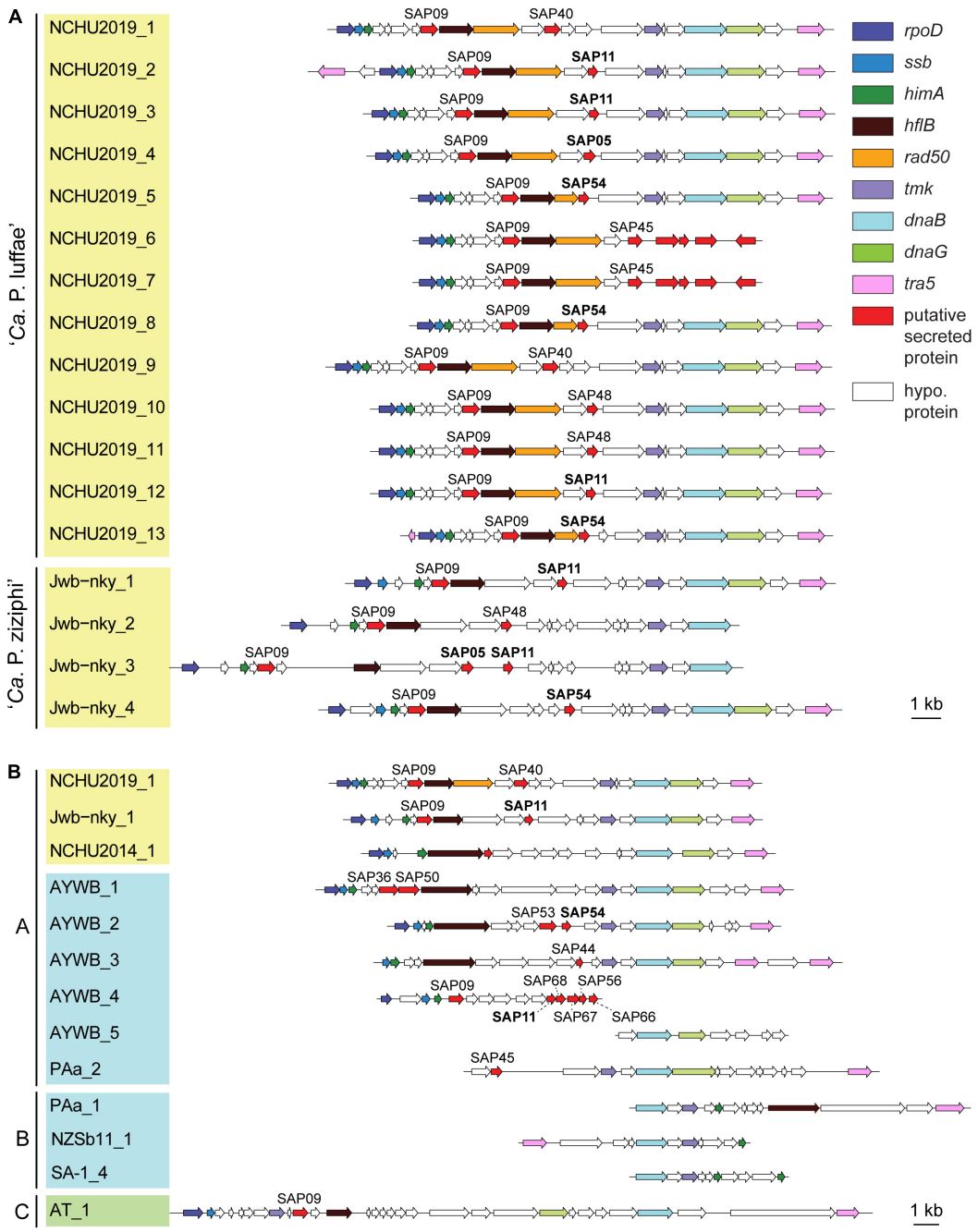


FIGURE 5 | Gene organization of potential mobile units (PMUs). Each individual PMU is labeled by the phytoplasma strain name and a numerical identifier. Background colors for the PMU identifiers indicate the three phylogenetic clusters of phytoplasmas (I: blue; II; green; III: yellow). Genes are drawn to scale and color-coded according to annotation. Homologs of putative effectors identified in the '*Ca*. P. asteris' AYWB genomes are labeled, those with experimental evidence (i.e., SAP05, SAP11, and SAP54) are highlighted in bold. Detailed information regarding the coding sequences in these regions is provided in **Supplementary Table 2**. **(A)** All PMUs in '*Ca*. P. luffae' NCHU2019 and '*Ca*. P. ziziphi' Jwb-nky. Grouped by genomes. **(B)** Representative PMUs from selected phytoplasmas. Grouped by PMU types.

(Tran-Nguyen et al., 2008; Andersen et al., 2013) and '*Ca*. P. solani' (Seruga Music et al., 2019) that belong to cluster I. Type C is the rarest type, with only one representative found in '*Ca*. P. mali' that belongs to cluster II (Kube et al., 2008; Chung et al., 2013), and has *hflB* and *dnaG* located in between *tmk* and *dnaB*. In addition to gene organization, molecular phylogenies of *dnaB* and *tmk* also revealed sequence divergence among homologs from different PMU types (**Figure 6**). These patterns provide further support for our PMU classification scheme. It is interesting to note that regardless of PMU copy numbers, most

of the phytoplasmas with genome sequences available harbor only one single type of PMUs. Two exceptions are '*Ca*. P. australiense' PAa and '*Ca*. P. solani' SA-1, both harboring PMUs belong to types A and B. Future improvements in sampling more diverse phytoplasma genomes, particularly cluster II lineages, are necessary to provide a more comprehensive understanding of PMU diversity. Another point worth mentioning is that several of the genome sequences examined lack any identifiable PMU region according to our criterion, while still harbor multiple PMU genes (e.g., '*Ca*. P. asteris' strains De Villa and M3). It is likely that the PMU regions in these genomes were degenerated through mutation accumulation.

Potential Mobile Unit and Phytoplasma Genome Size Variation

One notable observation regarding phytoplasma genomes is the extensive size variation at both between- and within-species levels (**Table 1**). Phylogenetic relatedness, as quantified by core genome sequence divergence, does not appear to provide a reliable predictor for genome sizes (**Figure 3A**). Previous within-species comparisons suggested that PMU abundance is an important factor for genome size variation (Bai et al., 2006; Andersen et al., 2013). To further test if this pattern holds true for genus-level analysis, we performed regression analysis to examine the correlation between the combined length of PMU core genes and genome size. The main reason of utilizing PMU core genes, rather than putative PMU regions, is because that several genomes (e.g., '*Ca*. P. asteris' OY-M) harbor putative PMU regions with highly variable degrees of fragmentation and the exact boundaries of those regions are difficult to be defined unambiguously. Furthermore, even for '*Ca*. P. luffae' NCHU2019, in which the 13 PMU regions are well defined, there are quite a few orphan PMU genes located outside of those 13 regions and scatter across the entire chromosome (**Figure 2**). As such, analysis based on the PMU core genes provides a more objective and accurate quantification for the relative abundance of PMUs. Strikingly, when all 19 representative phytoplasma genome assemblies were examined together, the combined length of PMU core genes explains 79% of the variance in genome sizes (r = 0.89, p = 3.7e–07) (**Figure 7**). Due to the concern that draft assemblies cannot provide accurate information regarding these two metrics, we also performed regression analysis with only the 10 complete assemblies and obtained similar results (r = 0.87, p = 9.7e–04).

The evolution of bacterial genome sizes is a topic that received much research attention and previous studies have identified multiple relevant factors, such as composition of gene content, prevalence of mobile genetic elements, effective population sizes, level of mutational biases toward deletion, and ecological niches (Mira et al., 2001; Konstantinidis and Tiedje, 2004; Ochman and Davalos, 2006; Kuo and Ochman, 2009; Kuo et al., 2009; Novichkov et al., 2009; McCutcheon and Moran, 2012; Lo et al., 2016; Sabater-Muñoz et al., 2017; Weinert and Welch, 2017). Due to the complexity of bacterial genome size evolution and the multitude of factors involved in the process, the finding that one single factor (i.e., PMU abundance) having such a strong correlation with genome size among phytoplasmas is surprising. Notably, as genome reduction appears to be a common and recurring theme of symbiont evolution, the roles of PMU in genome expansion of some phytoplasmas require further investigation. Because PMUs are known to be associated with effector genes and likely can transfer horizontally between closely- (Cho et al., 2019; Seruga Music et al., 2019) or distantly-related phytoplasmas (Chung et al., 2013; Ku et al., 2013), the involvement of PMUs in phytoplasma effector gene content evolution is particularly important. However, it is worth noting that even though horizontal transfers of PMUs may provide novel combinations of effector gene content, PMU evolution may have been dominated by vertical inheritance or at least transfers between closely related lineages based on the observation that in most cases related lineages have similar PMUs.

Effector Genes

An important feature of phytoplasmas is their ability to modulate host plant development through small secreted proteins known as effectors (Sugio et al., 2011b; Sugio and Hogenhout, 2012). Four notable effectors have been experimentally characterized, including SAP05 (Gamboa et al., 2019; Huang and Hogenhout, 2019; Huang et al., 2021), SAP11/SWP1 (Bai et al., 2009; Sugio et al., 2011a; Lu et al., 2014; Chang et al., 2018; Wang et al., 2018c,b), SAP54/PHYL1 (MacLean et al., 2011; Maejima et al., 2014; Orlovskis and Hogenhout, 2016), and TENGU (Hoshi et al., 2009; Sugawara et al., 2013; Minato et al., 2014). The expanded genome sequence availability allowed us to investigate the phylogenetic distribution of homologous effector genes among diverse phytoplasmas. The highly variable distribution patterns suggest that the effector gene content may have rapid turnover during the diversification of phytoplasmas (**Figure 3**). For example, at within-species level comparison among '*Ca*. P. asteris' strains, the patterns of presence/absence and gene copy number are variable for three of these effectors (**Figure 3B**) although the level of core genome sequence divergence is very low as indicated by the short branch lengths (**Figure 3A**). This finding is consistent with our previous in-depth characterization of 16SrI phytoplasmas (Cho et al., 2020a). At genus-level, TENGU appears to be restricted to and conserved among 16SrI lineages in cluster I (i.e., '*Ca*. P. asteris' and '*Ca*. P. tritici'), while the other three effectors are variable. Based on these patterns, it is likely that TENGU originated in the most recent common ancestor (MRCA) of 16SrI phytoplasmas. However, for the other three effectors, it is unclear if the MRCA of all extant phytoplasmas harbor these genes or not. If yes, then multiple independent gene losses are required to explain the distribution of these genes among extant phytoplasmas. Alternatively, multiple independent origins, likely mediated by PMU-mediated horizontal gene transfer (Chung et al., 2013; Cho et al., 2019; Seruga Music et al., 2019), are required to explain the gene distribution pattern. Another interesting observation is that several phytoplasmas do not possess any of these four effector genes. It is likely that these phytoplasmas harbor novel effector genes that are yet to be characterized and further investigation is required to obtain a more complete picture of phytoplasma effector diversity.

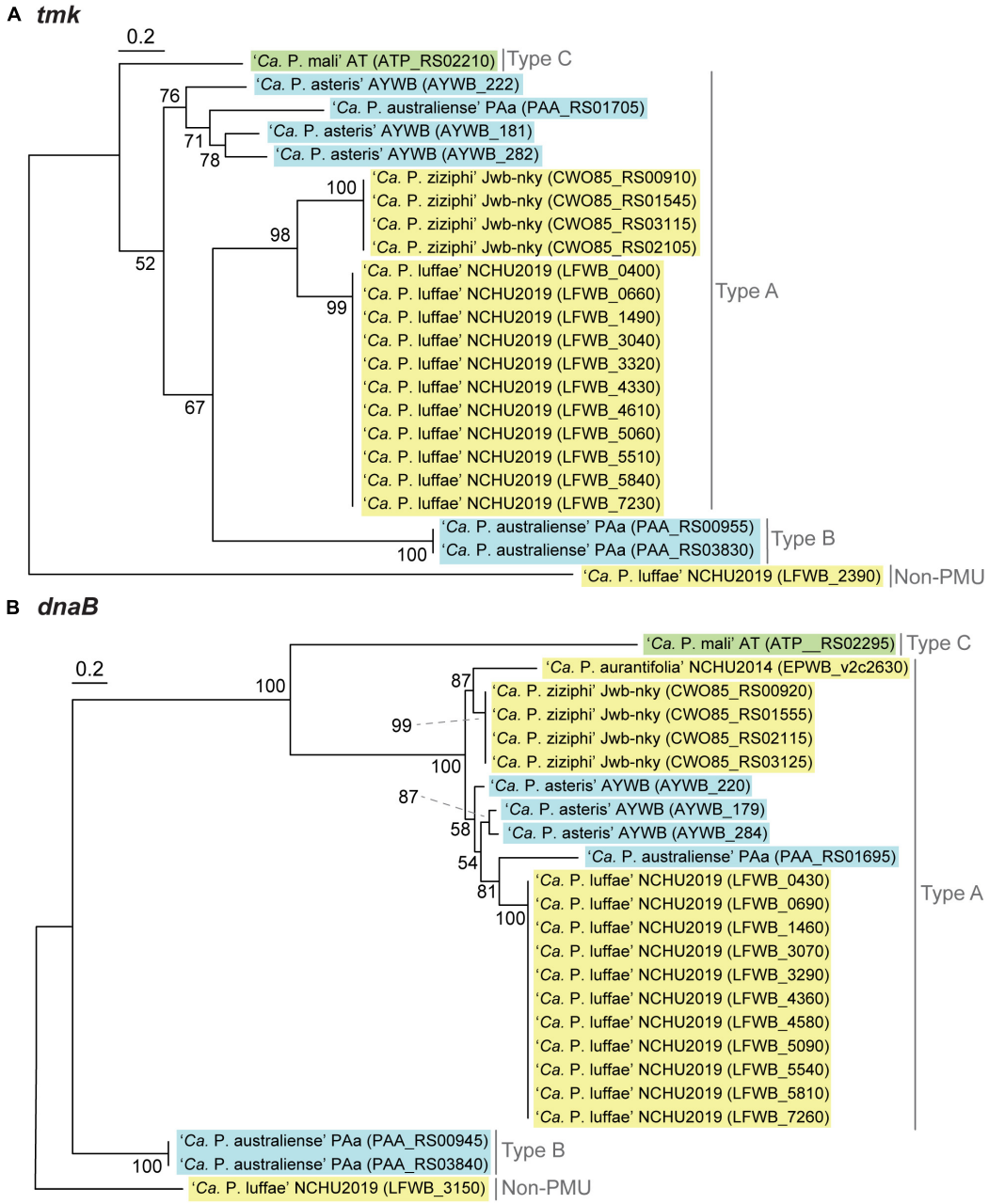


FIGURE 6 | Maximum likelihood phylogenies of PMU core genes **(A)** *tmk* (228 aligned amino acid sites). **(B)** *dnaB* (531 aligned amino acid sites). In both panels, a non-PMU homolog is included as the outgroup. Numbers next to internal branches indicate the bootstrap support levels based on 1,000 re-sampling. Background colors for the gene identifiers indicate the three phylogenetic clusters of phytoplasmas (I: blue; II; green; III: yellow).

For closer inspections of these four effectors, we performed multiple sequence alignments to examine the protein sequence divergence among homologs (**Supplementary Figure 3**). Consistent with the expectation derived from their phylogenetic distribution patterns, the three effector genes found among diverse phytoplasmas have higher levels of sequence divergence compared to phylogenetically restricted TENGU homologs.

For copy number variation, '*Ca.* P. luffae' NCHU2019 stands out as having the highest copy numbers for SAP11 and SAP54 (**Figure 3B**). These homologous genes are all located within PMUs (**Figure 5**) and have nearly identical sequences (**Supplementary Figure 3**), which suggest that recent intra-genomic PMU duplications are responsible for expansions in effector gene copy numbers. Similar patterns were observed for the SAP11 and SAP54 homologs in '*Ca.* P. ziziphi' Jwb-nky. Experimental characterization of the two '*Ca.* P. ziziphi' SAP11 homologs demonstrated that both can stimulate lateral bud outgrowth for witches' broom symptoms when expressed

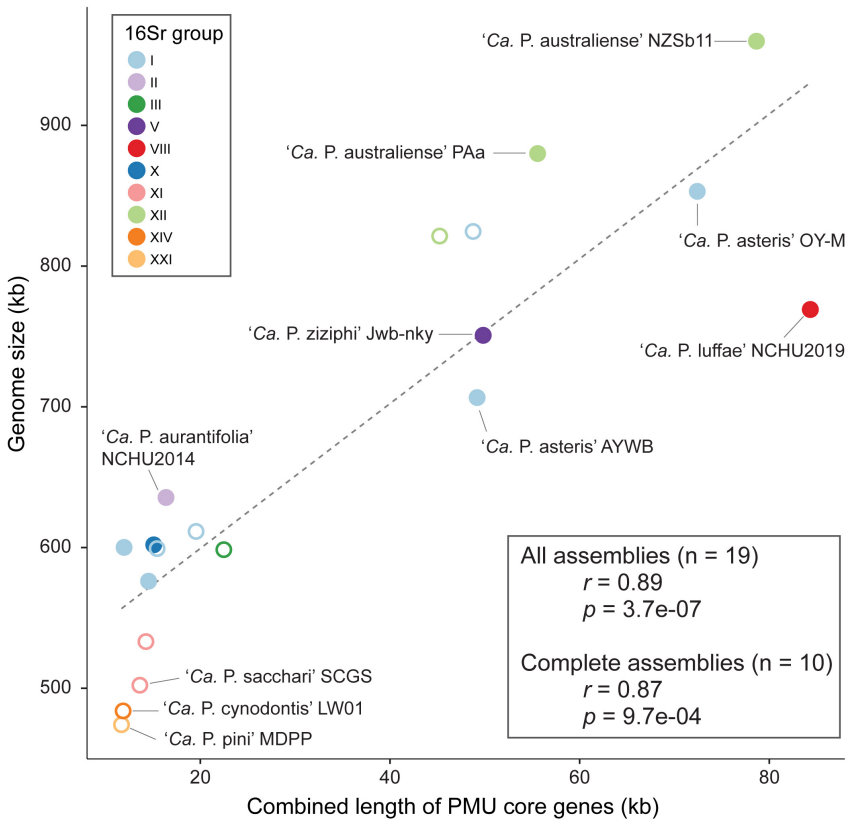


FIGURE 7 | Correlation between the combined length of PMU core genes and genome size. Strains with complete and draft genome assemblies are indicated by filled and open circles, respectively. The linear regression line was based on all available assemblies.

in *Nicotiana benthamiana* (Zhou et al., 2021). Intriguingly, in phytoplasma-infected jujube plants, these two SAP11 homologs have different expression patterns among tissue types. These findings suggest that such gene duplication events may lead to subfunctionalization or even neofunctionalization, thus promoting the genetic diversity of effectors.

In addition to these four effectors that have been well-characterized, there are likely additional effectors among diverse phytoplasmas. For example, two putative effectors were reported recently for '*Ca.* P. mali' (Mittelberger et al., 2019; Strohmayer et al., 2019). The functions and evolution of these and other novel effectors warrant further studies.

CONCLUSION

In this work, we determined the complete genome sequence of an uncultivated '*Ca.* P. luffae' strain associated with the witches' broom disease of loofah. This assembly provides the first representative genome sequence for the 16SrVIII group of phytoplasmas and improves the taxon sampling of these diverse plant-pathogenic bacteria. For comparative genomics analysis conducted at genus level, we provided a global view of the PMUs (i.e., phytoplasma-specific mobile genetic elements) and identified three major PMU types that differ in gene organization and phylogenetic distribution. Importantly, statistical analysis revealed that PMU abundance explains nearly 80% of the variance in phytoplasma genome sizes, providing a quantitative estimate on the importance of these elements. Finally, our investigation of effector genes highlighted the genetic diversity associated with phytoplasma virulence and supported the roles of PMUs in shaping such diversity.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://www.ncbi.nlm.nih.gov/, SRR11921288; https://www.ncbi.nlm.nih.gov/, SRR11921289; and https://www.ncbi.nlm.nih.gov/genbank/, CP054393.

AUTHOR CONTRIBUTIONS

J-YY and C-HK conceptualized the study, acquired funding, and supervised the project. J-YY, Y-CC, and C-MT provided the biological materials. C-MT and S-TC coordinated the sequencing and conducted the initial genome assembly. S-TC completed the

assembly. Y-CL and C-TH validated the assembly. S-TC, C-TH, and Y-CL conducted the comparative analysis and prepared the figures. C-TH and C-HK wrote the draft. All authors approved the submitted version.

FUNDING

Work in the J-YY lab was supported by grants-in-aid from the Ministry of Science and Technology (110-2628-B-005-002) and the Advanced Plant Biotechnology Center from the Featured Areas Research Center Program within the framework of the Higher Education Sprout Project by the Ministry of Education in Taiwan. Work in the C-HK lab was supported by Academia Sinica and the Ministry of Science and Technology (106-2311-B-001-028-MY3) of Taiwan. The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

ACKNOWLEDGMENTS

We thank Ai-Ping Chen and Shu-Jen Chou for technical assistance. The Illumina sequencing library preparation service was provided by the Genomic Technology Core Facility (Institute of Plant and Microbial Biology, Academia Sinica, Taipei, Taiwan). The Illumina paired-end sequencing service was provided by the Genomics Core Facility (Institute of Molecular Biology, Academia Sinica, Taipei, Taiwan). We thank the reviewers for their helpful comments that improved this work.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2022.773608/full#supplementary-material

Supplementary Figure 1 | Coverage pattern of raw read mapping results to the genome assembly. Rings from outside in: (1) Scale marks (kb). (2) Sequencing depth calculated from Illumina (dark blue; outward) and ONT (light blue; inward) raw reads mapped to the assembly using a sliding window of 200-bp. A filtering step was applied to include only the reads with high alignment scores. For Illumina reads, the cutoff was set at 200, which identified 120,804 reads corresponding to an average sequencing depth of 47-fold. For ONT reads, the cutoff was set at 1,000, which identified 10,562 reads corresponding to an average sequencing depth of 79-fold. (3) Potential mobile unit (PMU) regions (orange). (4) Two duplicated chromosomal segments (green; positions: 316–391 and 391–467 kb).

Supplementary Figure 2 | Genome size estimation based on k-mer analysis. **(A)** A representative frequency distribution of k-mer depth (k = 21). The peak depth is 51. **(B)** Correspondence between genome size estimates and k-mer depths.

Supplementary Figure 3 | Protein sequence alignments of four characterized phytoplasma effectors. Individual effectors are identified using locus tags. Shaded background colors indicate the levels of sequence conservation. **(A)** SAP05. **(B)** SAP11/SWP1. **(C)** SAP54/PHYL1. **(D)** TENGU.

Supplementary Table 1 | List of potential mobile unit (PMU) core genes in the genome sequences analyzed.

Supplementary Table 2 | List of potential mobile unit (PMU) regions among the genome sequences analyzed. One strain, '*Ca*. P. asteris' DY2014, was omitted because the PMU regions in this genome were reported to have complex structures and described in detail previously (Cho et al., 2019).

Supplementary Data 1 | Concatenated multiple sequence alignment of conserved single copy-genes used for inferring the molecular phylogeny shown in **Figure 3**. Simple text file in PHYLIP format.

REFERENCES

Andersen, M. T., Liefting, L. W., Havukkala, I., and Beever, R. E. (2013). Comparison of the complete genome sequence of two closely related isolates of '*Candidatus* Phytoplasma australiense' reveals genome plasticity. *BMC Genomics* 14:529. doi: 10.1186/1471-2164-14-529

Armenteros, J. J. A., Tsirigos, K. D., Sønderby, C. K., Petersen, T. N., Winther, O., Brunak, S., et al. (2019). SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat. Biotechnol.* 37, 420–423. doi: 10.1038/s41587-019-0036-z

Bai, X., Correa, V. R., Toruño, T. Y., Ammar, E.-D., Kamoun, S., and Hogenhout, S. A. (2009). AY-WB phytoplasma secretes a protein that targets plant cell nuclei. *Mol. Plant. Microbe Interact.* 22, 18–30. doi: 10.1094/MPMI-22-1-0018

Bai, X., Zhang, J., Ewing, A., Miller, S. A., Jancso Radek, A., Shevchenko, D. V., et al. (2006). Living with genome instability: the adaptation of phytoplasmas to diverse environments of their insect and plant hosts. *J. Bacteriol.* 188, 3682–3696. doi: 10.1128/JB.188.10.3682-3696.2006

Benson, D. A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Ostell, J., Pruitt, K. D., et al. (2018). GenBank. *Nucleic Acids Res.* 46, D41–D47.

Bertaccini, A., and Lee, I.-M. (2018). "Phytoplasmas: An update," in *Phytoplasmas: Plant Pathogenic Bacteria - I: Characterisation and Epidemiology of Phytoplasma - Associated Diseases*, eds G. P. Rao, A. Bertaccini, N. Fiore, and L. W. Liefting (Singapore: Springer).

Bertaccini, A., Duduk, B., Paltrinieri, S., and Contaldo, N. (2014). Phytoplasmas and phytoplasma diseases: a severe threat to agriculture. *Am. J. Plant Sci.* 5, 1763–1788. doi: 10.4236/ajps.2014.512191

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., et al. (2009). BLAST+: architecture and applications. *BMC Bioinformatics* 10:421. doi: 10.1186/1471-2105-10-421

Chang, S. H., Tan, C. M., Wu, C.-T., Lin, T.-H., Jiang, S.-Y., Liu, R.-C., et al. (2018). Alterations of plant architecture and phase transition by the phytoplasma virulence factor SAP11. *J. Exp. Bot.* 69, 5389–5401. doi: 10.1093/jxb/ery318

Chen, L.-L., Chung, W.-C., Lin, C.-P., and Kuo, C.-H. (2012). Comparative analysis of gene content evolution in phytoplasmas and mycoplasmas. *PLoS One* 7:e34407. doi: 10.1371/journal.pone.0034407

Cho, S.-T., Kung, H.-J., Huang, W., Hogenhout, S. A., and Kuo, C.-H. (2020a). Species boundaries and molecular markers for the classification of 16SrI phytoplasmas inferred by genome analysis. *Front. Microbiol.* 11:1531. doi: 10.3389/fmicb.2020.01531

Cho, S.-T., Lin, C.-P., and Kuo, C.-H. (2019). Genomic characterization of the periwinkle leaf yellowing (PLY) phytoplasmas in Taiwan. *Front. Microbiol.* 10:2194. doi: 10.3389/fmicb.2019.02194

Cho, S.-T., Zwolińska, A., Huang, W., Wouters, R. H. M., Mugford, S. T., Hogenhout, S. A., et al. (2020b). Complete genome sequence of '*Candidatus* Phytoplasma asteris' RP166, a plant pathogen associated with rapeseed phyllody disease in Poland. *Microbiol. Resour. Announc.* 9:e760-20. doi: 10.1128/MRA.00760-20

Chung, W.-C., Chen, L.-L., Lo, W.-S., Lin, C.-P., and Kuo, C.-H. (2013). Comparative analysis of the peanut witches'-broom phytoplasma genome reveals horizontal transfer of potential mobile units and effectors. *PLoS One* 8:e62770. doi: 10.1371/journal.pone.0062770

Davis, R. E., Zhao, Y., Wei, W., Dally, E. L., and Lee, I.-M. (2017). '*Candidatus* Phytoplasma luffae', a novel taxon associated with witches' broom disease of loofah, Luffa aegyptica Mill. *Int. J. Syst. Evol. Microbiol.* 67, 3127–3133. doi: 10.1099/ijsem.0.001980

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792–1797. doi: 10.1093/nar/gkh340

Felsenstein, J. (1989). PHYLIP - phylogeny inference package (version 3.2). *Cladistics* 5, 164–166.

Gamboa, C., Cui, W., Quiroga, N., Fernández, C., Fiore, N., and Zamorano, A. (2019). Identification of 16SrIII-J phytoplasma effectors using a viral vector. *Phytopathogenic Mollicutes* 9:229. doi: 10.5958/2249-4677.2019.00115.4

Guindon, S., and Gascuel, O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* 52, 696–704. doi: 10.1080/10635150390235520

Gundersen, D. E., and Lee, I.-M. (1996). Ultrasensitive detection of phytoplasmas by nested-PCR assays using two universal primer pairs. *Phytopathol. Mediterr.* 35, 144–151.

Guy, L., Roat Kultima, J., and Andersson, S. G. E. (2010). genoPlotR: comparative gene and genome visualization in R. *Bioinformatics* 26, 2334–2335. doi: 10.1093/bioinformatics/btq413

Hogenhout, S. A., and Seruga Music, M. (2009). "Phytoplasma genomics, from sequencing to comparative and functional genomics - what have we learnt?," in *Phytoplasmas: Genomes, Plant Hosts and Vectors*, eds P. G. Weintraun and P. Jones (Wallingford: CABI), 19–36. doi: 10.1079/9781845935306.0019

Hogenhout, S. A., Oshima, K., Ammar, E.-D., Kakizawa, S., Kingdom, H. N., and Namba, S. (2008). Phytoplasmas: bacteria that manipulate plants and insects. *Mol. Plant Pathol.* 9, 403–423. doi: 10.1111/j.1364-3703.2008.00472.x

Hoshi, A., Oshima, K., Kakizawa, S., Ishii, Y., Ozeki, J., Hashimoto, M., et al. (2009). A unique virulence factor for proliferation and dwarfism in plants identified from a phytopathogenic bacterium. *Proc. Natl. Acad. Sci.* 106, 6416–6421. doi: 10.1073/pnas.0813038106

Huang, W., and Hogenhout, S. A. (2019). "Phytoplasma effectors have converged onto degrading plant transcription factors with fundamental roles in plant development and defense to insects," in *Paper presentation at the IS-MPMI XVIII Congress Abstracts* (Glasgow: IS-MPMI).

Huang, W., MacLean, A. M., Sugio, A., Maqbool, A., Busscher, M., Cho, S.-T., et al. (2021). Parasitic modulation of host development by ubiquitin-independent protein degradation. *Cell* 184, 5201–5214. doi: 10.1016/j.cell.2021.08.029

Hyatt, D., Chen, G.-L., LoCascio, P., Land, M., Larimer, F., and Hauser, L. (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 11:119. doi: 10.1186/1471-2105-11-119

Jain, C., Rodriguez-R, L. M., Phillippy, A. M., Konstantinidis, K. T., and Aluru, S. (2018). High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat. Commun.* 9:5114.

Kanehisa, M., Goto, S., Furumichi, M., Tanabe, M., and Hirakawa, M. (2010). KEGG for representation and analysis of molecular networks involving diseases and drugs. *Nucleic Acids Res.* 38, D355–D360. doi: 10.1093/nar/gkp896

Kirdat, K., Tiwarekar, B., Thorat, V., Sathe, S., Shouche, Y., and Yadav, A. (2021). '*Candidatus* Phytoplasma sacchari', a novel taxon - associated with Sugarcane Grassy Shoot (SCGS) disease. *Int. J. Syst. Evol. Microbiol.* 71:4591. doi: 10.1099/ijsem.0.004591

Konstantinidis, K. T., and Tiedje, J. M. (2004). Trends between gene content and genome size in prokaryotic species with larger genomes. *Proc. Natl. Acad. Sci.* 101, 3160–3165. doi: 10.1073/pnas.0308653100

Krogh, A., Larsson, B., von Heijne, G., and Sonnhammer, E. L. L. (2001). Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.* 305, 567–580. doi: 10.1006/jmbi.2000.4315

Krzywinski, M., Schein, J., Birol, İ, Connors, J., Gascoyne, R., Horsman, D., et al. (2009). Circos: an information aesthetic for comparative genomics. *Genome Res.* 19, 1639–1645. doi: 10.1101/gr.092759.109

Ku, C., Lo, W.-S., and Kuo, C.-H. (2013). Horizontal transfer of potential mobile units in phytoplasmas. *Mob. Genet. Elem.* 3:e26145. doi: 10.4161/mge.26145

Kube, M., Mitrovic, J., Duduk, B., Rabus, R., and Seemüller, E. (2012). Current view on phytoplasma genomes and encoded metabolism. *Sci. World J.* 2012, 1–25. doi: 10.1100/2012/185942

Kube, M., Schneider, B., Kuhl, H., Dandekar, T., Heitmann, K., Migdoll, A., et al. (2008). The linear chromosome of the plant-pathogenic mycoplasma '*Candidatus* Phytoplasma' mali. *BMC Genomics* 9:306. doi: 10.1186/1471-2164-9-306

Kuo, C.-H., and Ochman, H. (2009). Deletional bias across the three domains of life. *Genome Biol. Evol.* 1, 145–152. doi: 10.1093/gbe/evp016

Kuo, C.-H., Moran, N. A., and Ochman, H. (2009). The consequences of genetic drift for bacterial genome complexity. *Genome Res.* 19, 1450–1454. doi: 10.1101/gr.091785.109

Kurtz, S., Phillippy, A., Delcher, A. L., Smoot, M., Shumway, M., Antonescu, C., et al. (2004). Versatile and open software for comparing large genomes. *Genome Biol.* 5:R12. doi: 10.1186/gb-2004-5-2-r12

Lagesen, K., Hallin, P., Rodland, E. A., Staerfeldt, H.-H., Rognes, T., and Ussery, D. W. (2007). RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res.* 35, 3100–3108. doi: 10.1093/nar/gkm160

Lee, I.-M., Davis, R. E., and Gundersen-Rindal, D. E. (2000). Phytoplasma: phytopathogenic mollicutes. *Annu. Rev. Microbiol.* 54, 221–255.

Lee, I.-M., Gundersen-Rindal, D. E., Davis, R. E., and Bartoszyk, I. M. (1998). Revised classification scheme of phytoplasmas based on RFLP analyses of 16S rRNA and ribosomal protein gene sequences. *Int. J. Syst. Evol. Microbiol.* 48, 1153–1169. doi: 10.1099/00207713-48-4-1153

Lee, I.-M., Hammond, R., Davis, R., and Gundersen, D. (1993). Universal amplification and analysis of pathogen 16S rDNA for classification and identification of mycoplasmalike organisms. *Phytopathology* 83, 834–842. doi: 10.1094/phyto-83-834

Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 34, 3094–3100. doi: 10.1093/bioinformatics/bty191

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with burrows–wheeler transform. *Bioinformatics* 25, 1754–1760. doi: 10.1093/bioinformatics/btp324

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* 25, 2078–2079. doi: 10.1093/bioinformatics/btp352

Li, L., Stoeckert, C. J., and Roos, D. S. (2003). OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res.* 13, 2178–2189. doi: 10.1101/gr.1224503

Lo, W.-S., Gasparich, G. E., and Kuo, C.-H. (2018). Convergent evolution among ruminant-pathogenic *Mycoplasma* involved extensive gene content changes. *Genome Biol. Evol.* 10, 2130–2139. doi: 10.1093/gbe/evy172

Lo, W.-S., Huang, Y.-Y., and Kuo, C.-H. (2016). Winding paths to simplicity: genome evolution in facultative insect symbionts. *FEMS Microbiol. Rev.* 40, 855–874. doi: 10.1093/femsre/fuw028

Lowe, T., and Eddy, S. (1997). tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* 25, 955–964. doi: 10.1093/nar/25.5.955

Lu, M., An, H., and Li, L. (2016). Genome survey sequencing for the characterization of the genetic background of *Rosa roxburghii* Tratt and leaf ascorbate metabolism genes. *PLoS One* 11:e0147530. doi: 10.1371/journal.pone.0147530

Lu, Y.-T., Li, M.-Y., Cheng, K.-T., Tan, C. M., Su, L.-W., Lin, W.-Y., et al. (2014). Transgenic plants that express the phytoplasma effector SAP11 show altered phosphate starvation and defense responses. *Plant Physiol.* 164, 1456–1469. doi: 10.1104/pp.113.229740

MacLean, A. M., Sugio, A., Makarova, O. V., Findlay, K. C., Grieve, V. M., Tóth, R., et al. (2011). Phytoplasma effector SAP54 induces indeterminate leaf-like flower development in *Arabidopsis* plants. *Plant Physiol.* 157, 831–841. doi: 10.1104/pp.111.181586

Maejima, K., Iwai, R., Himeno, M., Komatsu, K., Kitazawa, Y., Fujita, N., et al. (2014). Recognition of floral homeotic MADS domain transcription factors by a phytoplasmal effector, phyllogen, induces phyllody. *Plant J.* 78, 541–554. doi: 10.1111/tpj.12495

Marçais, G., and Kingsford, C. (2011). A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* 27, 764–770. doi: 10.1093/bioinformatics/btr011

McCutcheon, J. P., and Moran, N. A. (2012). Extreme genome reduction in symbiotic bacteria. *Nat. Rev. Microbiol.* 10, 13–26. doi: 10.1038/nrmicro2670

Minato, N., Himeno, M., Hoshi, A., Maejima, K., Komatsu, K., Takebayashi, Y., et al. (2014). The phytoplasmal virulence factor TENGU causes plant sterility by downregulating of the jasmonic acid and auxin pathways. *Sci. Rep.* 4:7399. doi: 10.1038/srep07399

Mira, A., Ochman, H., and Moran, N. A. (2001). Deletional bias and the evolution of bacterial genomes. *Trends Genet.* 17, 589–596. doi: 10.1016/s0168-9525(01)02447-7

Mittelberger, C., Stellmach, H., Hause, B., Kerschbamer, C., Schlink, K., Letschka, T., et al. (2019). A novel effector protein of apple proliferation phytoplasma

disrupts cell integrity of *Nicotiana* spp. protoplasts. *Int. J. Mol. Sci.* 20:4613. doi: 10.3390/ijms20184613

Namba, S. (2019). Molecular and biological properties of phytoplasmas. *Proc. Jpn. Acad. Ser. B* 95, 401–418. doi: 10.2183/pjab.95.028

Nejat, N., Vadamalai, G., Davis, R. E., Harrison, N. A., Sijam, K., Dickinson, M., et al. (2013). '*Candidatus* Phytoplasma malaysianum', a novel taxon associated with virescence and phyllody of Madagascar periwinkle (*Catharanthus roseus*). *Int. J. Syst. Evol. Microbiol.* 63, 540–548. doi: 10.1099/ijs.0.041467-0

Novichkov, P. S., Wolf, Y. I., Dubchak, I., and Koonin, E. V. (2009). Trends in prokaryotic evolution revealed by comparison of closely related bacterial and archaeal genomes. *J. Bacteriol.* 191, 65–73. doi: 10.1128/JB.01237-08

Ochman, H., and Davalos, L. M. (2006). The nature and dynamics of bacterial genomes. *Science* 311, 1730–1733. doi: 10.1126/science.1119966

Orlovskis, Z., and Hogenhout, S. A. (2016). A bacterial parasite effector mediates insect vector attraction in host plants independently of developmental changes. *Front. Plant Sci.* 7:885. doi: 10.3389/fpls.2016.00885

Orlovskis, Z., Canale, M. C., Haryono, M., Lopes, J. R. S., Kuo, C.-H., and Hogenhout, S. A. (2017). A few sequence polymorphisms among isolates of maize bushy stunt phytoplasma associate with organ proliferation symptoms of infected maize plants. *Ann. Bot.* 119, 869–884. doi: 10.1093/aob/mcw213

Oshima, K., Kakizawa, S., Nishigawa, H., Jung, H.-Y., Wei, W., Suzuki, S., et al. (2004). Reductive evolution suggested from the complete genome sequence of a plant-pathogenic phytoplasma. *Nat. Genet.* 36, 27–29. doi: 10.1038/ng1277

R Core Team (2019). *R: a Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing

Robinson, J. T., Thorvaldsdottir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., et al. (2011). Integrative genomics viewer. *Nat. Biotechnol.* 29, 24–26. doi: 10.1038/nbt.1754

Sabater-Muñoz, B., Toft, C., Alvarez-Ponce, D., and Fares, M. A. (2017). Chance and necessity in the genome evolution of endosymbiotic bacteria of insects. *ISME J.* 11, 1291–1304. doi: 10.1038/ismej.2017.18

Saccardo, F., Martini, M., Palmano, S., Ermacora, P., Scortichini, M., Loi, N., et al. (2012). Genome drafts of four phytoplasma strains of the ribosomal group 16SrIII. *Microbiology* 158, 2805–2814. doi: 10.1099/mic.0.061432-0

Seruga Music, M., Samarzija, I., Hogenhout, S. A., Haryono, M., Cho, S.-T., and Kuo, C.-H. (2019). The genome of '*Candidatus* Phytoplasma solani' strain SA-1 is highly dynamic and prone to adopting foreign sequences. *Syst. Appl. Microbiol.* 42, 117–127. doi: 10.1016/j.syapm.2018.10.008

Strohmayer, A., Moser, M., Si-Ammour, A., Krczal, G., and Boonrod, K. (2019). '*Candidatus* Phytoplasma mali' genome encodes a protein that functions as an E3 ubiquitin ligase and could inhibit plant basal defense. *Mol. Plant. Microbe Interact.* 32, 1487–1495. doi: 10.1094/MPMI-04-19-0107-R

Sugawara, K., Honma, Y., Komatsu, K., Himeno, M., Oshima, K., and Namba, S. (2013). The alteration of plant morphology by small peptides released from the proteolytic processing of the bacterial peptide TENGU. *Plant Physiol.* 162, 2005–2014. doi: 10.1104/pp.113.218586

Sugio, A., and Hogenhout, S. A. (2012). The genome biology of phytoplasma: modulators of plants and insects. *Curr. Opin. Microbiol.* 15, 247–254. doi: 10.1016/j.mib.2012.04.002

Sugio, A., Kingdom, H. N., MacLean, A. M., Grieve, V. M., and Hogenhout, S. A. (2011a). Phytoplasma protein effector SAP11 enhances insect vector reproduction by manipulating plant development and defense hormone biosynthesis. *Proc. Natl. Acad. Sci.* 108, E1254–E1263. doi: 10.1073/pnas.1105664108

Sugio, A., MacLean, A. M., Kingdom, H. N., Grieve, V. M., Manimekalai, R., and Hogenhout, S. A. (2011b). Diverse targets of phytoplasma effectors: from plant development to defense against insects. *Annu. Rev. Phytopathol.* 49, 175–195. doi: 10.1146/annurev-phyto-072910-095323

Tan, C. M., Lin, Y.-C., Li, J.-R., Chien, Y.-Y., Wang, C.-J., Chou, L., et al. (2021). Accelerating complete phytoplasma genome assembly by immunoprecipitation-based enrichment and MinION-based DNA sequencing for comparative analyses. *Front. Microbiol.* 12:766221. doi: 10.3389/fmicb.2021.766221

Tatusov, R., Fedorova, N., Jackson, J., Jacobs, A., Kiryutin, B., Koonin, E., et al. (2003). The COG database: an updated version includes eukaryotes. *BMC Bioinformatics* 4:41. doi: 10.1186/1471-2105-4-41

The IRPCM Phytoplasma/Spiroplasma Working Team - Phytoplasma taxonomy group (2004). '*Candidatus* Phytoplasma', a taxon for the wall-less, non-helical prokaryotes that colonize plant phloem and insects. *Int. J. Syst. Evol. Microbiol.* 54, 1243–1255. doi: 10.1099/ijs.0.02854-0

Tran-Nguyen, L. T. T., Kube, M., Schneider, B., Reinhardt, R., and Gibb, K. S. (2008). Comparative genome analysis of '*Candidatus* Phytoplasma australiense' (subgroup *tuf*-Australia I; *rp*-A) and '*Ca*. Phytoplasma asteris' strains OY-M and AY-WB. *J. Bacteriol.* 190, 3979–3991. doi: 10.1128/JB.01301-07

Wang, J., Song, L., Jiao, Q., Yang, S., Gao, R., Lu, X., et al. (2018a). Comparative genome analysis of jujube witches'-broom phytoplasma, an obligate pathogen that causes jujube witches'-broom disease. *BMC Genomics* 19:689. doi: 10.1186/s12864-018-5075-1

Wang, N., Li, Y., Chen, W., Yang, H. Z., Zhang, P. H., and Wu, Y. F. (2018b). Identification of wheat blue dwarf phytoplasma effectors targeting plant proliferation and defence responses. *Plant Pathol.* 67, 603–609.

Wang, N., Yang, H., Yin, Z., Liu, W., Sun, L., and Wu, Y. (2018c). Phytoplasma effector SWP1 induces witches' broom symptom by destabilizing the TCP transcription factor BRANCHED1. *Mol. Plant Pathol.* 19, 2623–2634. doi: 10.1111/mpp.12733

Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M., and Barton, G. J. (2009). Jalview version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* 25, 1189–1191. doi: 10.1093/bioinformatics/btp033

Weinert, L. A., and Welch, J. J. (2017). Why might bacterial pathogens have small genomes? *Trends Ecol. Evol.* 32, 936–947. doi: 10.1016/j.tree.2017.09.006

Wu, D., Hugenholtz, P., Mavromatis, K., Pukall, R., Dalin, E., Ivanova, N. N., et al. (2009). A phylogeny-driven genomic encyclopaedia of Bacteria and Archaea. *Nature* 462, 1056–1060. doi: 10.1038/nature08656

Wu, M., and Eisen, J. (2008). A simple, fast, and accurate method of phylogenomic inference. *Genome Biol.* 9:R151. doi: 10.1186/gb-2008-9-10-r151

Zerbino, D. R., and Birney, E. (2008). Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res.* 18, 821–829. doi: 10.1101/gr.074492.107

Zhao, Y., and Davis, R. E. (2016). Criteria for phytoplasma 16Sr group/subgroup delineation and the need of a platform for proper registration of new groups and subgroups. *Int. J. Syst. Evol. Microbiol.* 66, 2121–2123. doi: 10.1099/ijsem.0.000999

Zhao, Y., Wei, W., Davis, R. E., Lee, I.-M., and Bottner-Parker, K. D. (2021). The agent associated with blue dwarf disease in wheat represents a new phytoplasma taxon, '*Candidatus* Phytoplasma tritici'. *Int. J. Syst. Evol. Microbiol.* 71:004604. doi: 10.1099/ijsem.0.004604

Zhao, Y., Wei, W., Lee, I.-M., Shao, J., Suo, X., and Davis, R. E. (2009). Construction of an interactive online phytoplasma classification tool, *i*PhyClassifier, and its application in analysis of the peach X-disease phytoplasma group (16SrIII). *Int. J. Syst. Evol. Microbiol.* 59, 2582–2593. doi: 10.1099/ijs.0.010249-0

Zhou, J., Ma, F., Yao, Y., Deng, M., Chen, M., and Zhang, S. (2021). JWB phytoplasma effectors SJP1 and SJP2 induce lateral bud outgrowth by repressing the ZjBRC1-controlled auxin efflux channel. *Plant Cell Environ.* 44, 3257–3272. doi: 10.1111/pce.14141

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Microbiology

TYPE Original Research
PUBLISHED 24 October 2022
DOI 10.3389/fmicb.2022.1044771

OPEN ACCESS

EDITED BY
Allison Hansen,
University of California,
United States

REVIEWED BY
Rosario Gil,
University of Valencia,
Spain
Amanda May Vivian Brown,
Texas Tech University,
United States

*CORRESPONDENCE
Shigeyuki Kakizawa
s.kakizawa@aist.go.jp
Takema Fukatsu
t-fukatsu@aist.go.jp

†These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to Microbial Symbioses, a section of the journal Frontiers in Microbiology

RECEIVED 15 September 2022
ACCEPTED 07 October 2022
PUBLISHED 24 October 2022

CITATION
Kakizawa S, Hosokawa T, Oguchi K, Miyakoshi K and Fukatsu T (2022) *Spiroplasma* as facultative bacterial symbionts of stinkbugs. *Front. Microbiol.* 13:1044771. doi: 10.3389/fmicb.2022.1044771

Spiroplasma as facultative bacterial symbionts of stinkbugs

Shigeyuki Kakizawa[1†*], Takahiro Hosokawa[2†], Kohei Oguchi[1,3], Kaori Miyakoshi[1] and Takema Fukatsu[1,4,5*]

[1]Bioproduction Research Institute, National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba, Japan, [2]Department of Biology, Faculty of Science, Kyushu University, Fukuoka, Japan, [3]Misaki Marine Biological Station (MMBS), School of Science, The University of Tokyo, Miura, Japan, [4]Department of Biological Sciences, Graduate School of Science, The University of Tokyo, Tokyo, Japan, [5]Graduate School of Life and Environmental Sciences, University of Tsukuba, Tsukuba, Japan

Many insects are associated with facultative symbiotic bacteria, and their infection prevalence provides an important clue to understand the biological impact of such microbial associates. Here we surveyed diverse stinkbugs representing 13 families, 69 genera, 97 species and 468 individuals for *Spiroplasma* infection. Diagnostic PCR detection revealed that 4 families (30.8%), 7 genera (10.1%), 11 species (11.3%) and 21 individuals (4.5%) were *Spiroplasma* positive. All the 21 stinkbug samples with *Spiroplasma* infection were subjected to PCR amplification and sequencing of *Spiroplasma*'s 16S rRNA gene. Molecular phylogenetic analysis uncovered that the stinkbug-associated *Spiroplasma* symbionts were placed in three distinct clades in the Spiroplasmataceae, highlighting multiple evolutionary origins of the stinkbug-*Spiroplasma* associations. The *Spiroplasma* phylogeny did not reflect the host stinkbug phylogeny, indicating the absence of host-symbiont co-speciation. On the other hand, the *Spiroplasma* symbionts associated with the same stinkbug family tended to be related to each other, suggesting the possibility of certain levels of host-symbiont specificity and/or ecological symbiont sharing. Amplicon sequencing analysis targeting bacterial 16S rRNA gene, FISH visualization of the symbiotic bacteria, and rearing experiments of the host stinkbugs uncovered that the *Spiroplasma* symbionts are generally much less abundant in comparison with the primary gut symbiotic bacteria, localized to various tissues and organs at relatively low densities, and vertically transmitted to the offspring. On the basis of these results, we conclude that the *Spiroplasma* symbionts are, in general, facultative bacterial associates of low infection prevalence that are not essential but rather commensalistic for the host stinkbugs, like the *Spiroplasma* symbionts of fruit flies and aphids, although their impact on the host phenotypes should be evaluated in future studies.

KEYWORDS

***Spiroplasma*, symbiosis, bacteria, stinkbug, Hemiptera, Pentatomidae**

Introduction

Diverse insects are generally in symbiotic association with microorganisms (Buchner, 1965; Bourtzis and Miller, 2003). Some symbionts are obligatory microbial partners essential for their hosts *via* helping digestion, supplementing essential nutrients, or undertaking other important biological roles (Moran et al., 2008; Douglas, 2009; Brune,

2014). Other symbionts are facultative microbial associates affecting their hosts either positively *via* conferring context-dependent benefits such as defense against natural enemies, resistance against parasites and pathogens, tolerance against abiotic stressors, etc. (Flórez et al., 2015; Van Arnam et al., 2018; Lemoine et al., 2020), neutrally with no apparent phenotypic consequences, or negatively *via* reducing host growth, survival and/or fecundity in a variety of ways (Oliver et al., 2010; Feldhaar, 2011).

Members of the alphaproteobacterial genus *Wolbachia* are known as the most widespread facultative symbiotic bacteria of insects and other arthropods (Werren et al., 2008; Kaur et al., 2021). While *Wolbachia* symbionts are famous for their capabilities of manipulating host reproduction by inducing cytoplasmic incompatibility, parthenogenesis, feminization or male-killing (Werren et al., 2008; Kaur et al., 2021), which generally affect the host fitness negatively, some *Wolbachia* strains were reported to entail positive fitness consequences *via* conferring resistance to pathogen infections (Hedges et al., 2008; Teixeira et al., 2008), supplying essential nutrients (Hosokawa et al., 2010; Nikoh et al., 2014; Moriyama et al., 2015), or slightly enhancing growth and/or fecundity often in a context-dependent manner (Zug and Hammerstein, 2015). In addition, the diverse insects may also host such facultative symbiotic bacteria as *Rickettsia*, *Cardinium*, *Sodalis*, *Arsenophonus*, *Spiroplasma*, etc. (Zchori-Fein and Perlman, 2004; Perlman et al., 2006; Duron et al., 2008; Nováková et al., 2009; Hosokawa et al., 2015). In comparison with the extensively compiled data on the infection prevalence and diversity of *Wolbachia* in natural host species and populations (ex. Zug and Hammerstein, 2012; Gerth et al., 2014), it has been less explored and still to be established how these facultative microbial associates other than *Wolbachia* are prevalent in the natural host diversity (Duron et al., 2008).

Members of the bacterial genus *Spiroplasma*, belonging to the class Mollicutes of the phylum Tenericutes (or Mycoplasmatota as recently proposed; see Oren and Garrity, 2021), are, typically, helical in shape, actively motile, and lacking outer cell wall (Gasparich et al., 2020). Initially, they were described as plant disease agents *S. kunkeli* for corn stunt disease (Davis et al., 1972) and *S. citri* for citrus stubborn disease (Saglio et al., 1973), then identified as male-killing sex ratio distorters of fruit flies (Poulson and Sakaguchi, 1961) later named *S. poulsonii* (Williamson et al., 1999), and thus far recognized as widely associated with diverse insects and other terrestrial arthropods (Clark, 1982), and more recently, also found from aquatic invertebrates including crustaceans, sea cucumbers and jellyfish (Nunan et al., 2005; Wang et al., 2005, 2011; He et al., 2017; Viver et al., 2017). Of the diverse *Spiroplasma* species and strains, some were reported as either highly or weakly pathogenic to plants, insects and crustaceans (Daniels, 1983; Fukatsu et al., 2001; Wang et al., 2005; Simon et al., 2011), some were shown to provide their hosts with defensive benefit against parasitic wasps, nematodes and fungi (Jaenike et al., 2010; Xie et al., 2010; Łukasik et al., 2013; Xie et al., 2014), some were noted for causing prominent male-killing phenotypes in flies, lady beetles, butterflies, moths, planthoppers and aphids (Jiggins et al., 2000; Hurst et al., 2003; Haselkorn, 2010; Simon et al., 2011; Sanada-Morimura et al., 2013), but the remaining majority are recognized solely by PCR detection and/or sequencing of bacterial gene fragments, with little biological information being available (Anbutsu and Fukatsu, 2011; Ballinger and Perlman, 2019).

In this context, stinkbugs (Hemiptera: Pentatomoidea) represent a notable insect group in that their obligatory and facultative symbiotic bacteria have been well surveyed at species and population levels (Kikuchi and Fukatsu, 2003; Matsuura et al., 2012; Hosokawa et al., 2015, 2016). The majority of plant-sucking stinkbugs develop a midgut symbiotic organ consisting of an assemblage of numerous sac- or tube-like crypts, whose inner cavities harbor a dense population of specific symbiotic bacteria (Buchner, 1965; Salem et al., 2015; Takeshita and Kikuchi, 2017). These gut symbionts are generally indispensable for normal growth, survival and/or reproduction of the host stinkbugs, being mutualistic microbial associates whose biological roles are presumed as provisioning of essential amino acids, vitamins and other metabolites to their hosts (Nikoh et al., 2011; Kaiwa et al., 2014; Salem et al., 2014). In addition to the primary gut symbiotic bacteria, these stinkbugs are also infected with such facultative symbiotic bacteria as *Wolbachia*, *Sodalis*, *Rickettsia*, *Spiroplasma, Lariskella*, and others. Of these, extensive infection surveys have been conducted for *Wolbachia*, *Sodalis* and *Lariskella* (Kikuchi and Fukatsu, 2003; Matsuura et al., 2012; Hosokawa et al., 2015), whereas *Spiroplasma* and *Rickettsia* have been detected in a few stinkbug species (Kikuchi et al., 2009; Caspi-Fluger et al., 2014; Matsuura et al., 2014; Amiri et al., 2020; Dally et al., 2020).

In this study, we surveyed diverse stinkbugs representing 13 families, 69 genera, 97 species and 468 individuals for *Spiroplasma* infection, thereby elucidating the infection frequency and the diversity of *Spiroplasma* associated with the stinkbugs. The *Spiroplasma*-positive stinkbug samples were subjected to molecular phylogenetic analysis and amplicon sequencing analysis targeting bacterial 16S rRNA gene, which uncovered sporadic and facultative nature of the stinkbug-*Spiroplasma* associations in general. In a stinkbug species that exhibited frequent *Spiroplasma* infection in natural populations, we investigated infection prevalence, vertical transmission and *in vivo* localization of *Spiroplasma* in detail.

Materials and methods

Insect materials

The stinkbug samples examined in this study are listed (Supplementary Tables S1, S2). Laboratory strains of the pea aphid *Acyrthosiphon pisum* (*Spiroplasma* negative) and the fruit fly *Drosophila melanogaster* (*Spiroplasma* positive and negative) were used as negative and positive control samples for *Spiroplasma* detection (Supplementary Table S3). The stinkbugs were collected in the field, of which some were preserved in acetone or 99% ethanol, some were preserved in ultracold freezers at –80°C, and

others were freshly dissected in phosphate buffered saline (PBS) under a dissection microscope using fine forceps and razor blades. The white-spotted stinkbug *Eysarcoris ventralis* was maintained on sunflower seeds and water containing 0.05% ascorbic acid at 25°C under a long-day condition of 16 h light and 8 h dark.

DNA sample preparation

After surface sterilization with 70% ethanol, the stinkbug samples were individually subjected to DNA extraction using QIAamp DNA mini kit (Qiagen). For relatively large insects (larger than 5 mm in size), the dissected alimentary tract was subjected to DNA preparation, whereas for relatively small insects (about 5 mm or smaller in size), the whole body was subjected to DNA preparation. For each sample, the dissected alimentary tract or the whole body was homogenized and digested in a lysis buffer containing proteinase K, purified using a spin column, and recovered with an elution buffer. For the initial proteinase K digestion, 200 µl of lysis buffer was used for most samples. For some samples large in size, 1,000 µl of lysis buffer was used for digestion and an aliquot (100–200 µl depending on the tissue size) was applied to the spin column for purification. The DNA samples were recovered with 200 µl of elution buffer, which were used as PCR templates without further dilution.

Diagnostic PCR screening of *Spiroplasma*

Diagnostic PCR specifically targeting 16S rRNA gene of *Spiroplasma* was performed with the primers 16SA1 (5′-AGA GTT TGA TCM TGG CTC AG-3′; Fukatsu and Nikoh, 1998) and SpR5 (5′- CTG CAG CAC CGA ACT TAG TC-3′; Bastian and Foster, 2001) under the temperature profile of 98°C for 1 min followed by 30 cycles of 98°C for 10 s, 55°C for 15 s and 68°C for 45 s.

Sequencing of 16S rRNA gene of *Spiroplasma*

For sequencing of 1.5 kb region of *Spiroplasma* 16S rRNA gene, in addition to the 0.8 kb region amplified by the primers 16SA1 and SpR5, a partially overlapping 0.9 kb region amplified by the primers SpF (5′- GCG CAG ACG GTT TAA CAA G-3′; Alexeeva et al., 2006) and 16SB1 (5'-TAC GGY TAC CTT GTT ACG ACT T-3′; Fukatsu and Nikoh, 1998) was directly sequenced as described previously (Kakizawa and Kamagata, 2014). The nucleotide sequences determined in this study were deposited in the DNA Data Bank of Japan[1] under accession numbers LC685076- LC685096 (Supplementary Table S1).

1 http://www.ddbj.nig.ac.jp/index-e.html

Molecular phylogenetic analysis

Multiple alignments were generated by the program Clustal Omega (Sievers et al., 2011). Ambiguously aligned nucleotide sites were manually removed. Phylogenetic analyses were conducted by maximum likelihood and neighbor-joining methods. The nucleotide substitution models were selected using MEGA version X (Kumar et al., 2018). Maximum likelihood trees and neighbor-joining trees were constructed by using MEGA version X. Bootstrap values were obtained with 1,000 replications.

Amplicon sequencing analysis

The stinkbugs and other insect samples subjected to amplicon sequencing of bacterial 16S rRNA gene are listed (Supplementary Tables S3, S4). The V3-V4 region of bacterial 16S rRNA gene was amplified by the 2-step tailed PCR method as follows. The 1st PCR was performed with the primers 1st-341f_MIX (5′-ACA CTC TTT CCC TAC ACG ACG CTC TTC CGA TCT -NNN NN-C CTA CGG GNG GCW GCA G-3′; Muyzer et al., 1993) and 1st-805r_MIX (5′-GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC T-NN NNN- GAC TAC HVG GGT ATC TAA TCC-3′; Herlemann et al., 2011) under the temperature profile of 94°C for 2 min followed by 30 cycles of 94°C for 30 s, 50°C for 30 s and 72°C for 30 s, followed by 72°C for 5 min. PCR was performed with either ExTaqHS polymerase (TaKaRa) or TksGflex polymerase (TaKaRa). PCR products were purified with AMPure XP magnetic beads (BECKMAN COULTER). The 2nd PCR was performed with the primers 2ndF (5′-AAT GAT ACG GCG ACC ACC GAG ATC TAC AC -Index2- ACA CTC TTT CCC TAC ACG ACG C-3′) and 2ndR (5′-CAA GCA GAA GAC GGC ATA CGA GAT -Index1- GTG ACT GGA GTT CAG ACG TGT G-3′) using ExTaqHS polymerase under the temperature profile of 94°C for 2 min followed by 10 cycles of 94°C for 30 s, 60°C for 30 s and 72°C for 30 s, followed by 72°C for 5 min. Index1 and Index2 sequences were according to the Nextera XT Index Kit (Illumina). The 2nd PCR was for adding sequence indexes and tags for Illumina sequencing, and the cycles were minimized to reduce PCR-derived biases. The PCR products were purified with AMPure XP. Quantity of the library was analyzed by Synergy H1 with QuantiFluor dsDNA System (BioTek), and quality of the library was analyzed by Fragment Analyzer with dsDNA 915 Reagent Kit (Advanced Analytical Technologies). Finally, the library was sequenced on an Illumina MiSeq system with MiSeq Reagent Kit v3 (Illumina) that generated 2 × 300 bp paired end reads. The sequence data were analyzed as follows. First, the reads matching to the primer sequences were extracted by fastx_barcode_splitter tool of FASTX-Toolkit (ver. 0.0.14). The primer sequences were removed from all the reads by fastx_trimer of FASTX-Toolkit. Low quality reads were removed using sickle (ver. 1.33) based on the quality score 20, and reads shorter than 130 bp were also removed. All paired reads were combined using FLASH (ver.1.2.11) script

TABLE 1 Summary of *Spiroplasma* detection from diverse stinkbugs representing 4 superfamilies, 13 families, 69 genera, 97 species and 468 individuals.

Superfamily	Family	Genus	Species	Individual
		Spiroplasma-positive taxa/taxa examined		
Coreoidea	Alydidae	1/1	1/1	1/1
	Coreidae	0/6	0/6	0/10
Lygaeoidea	Lygaeidae	0/1	0/1	0/1
	Rhyparochromidae	0/1	0/1	0/2
Pyrrhocoroidea	Pyrrhocoridae	0/1	0/1	0/1
Pentatomoidea	Acanthosomatidae	1/4	4/10	6/25
	Cydnidae	2/3	2/5	5/31
	Dinidoridae	0/1	0/1	0/17
	Parastrachiidae	0/1	0/1	0/6
	Pentatomidae	3/38	4/53	9/297
	Plataspidae	0/2	0/4	0/6
	Scutelleridae	0/8	0/10	0/64
	Urostylididae	0/2	0/3	0/7
Total		7/69	11/97	21/468

For detail, see Supplementary Table S1.

based on >10-bp overlapping criteria. Subsequently, chimeric sequences were removed using the dada2 plugin of QIIME2 (ver.2021.4). In QIIME2, the chimeric-filtered sequences were clustered into operational taxonomic units (OTUs) using EzBioCloud 16S database.[2] The OTUs were tabulated on each taxonomic level from phylum to genus and their relative abundances were calculated using a workflow script in QIIME2. All the above procedures were performed by Bioengineering Lab. Co., Ltd. (Kanagawa, Japan).

Fluorescence *in situ* hybridization

Whole-mount fluorescence *in situ* hybridization (FISH) targeting bacterial 16S rRNA of *Spiroplasma* symbiont was performed as described previously (Matsuura et al., 2014). The dissected insect tissues were fixed in 4% paraformaldehyde solution in PBS for 3 h and then thoroughly washed three times in PBST (0.2% Tween20 in PBS). The fixed insect tissues were washed twice in a hybridization buffer (20 mM Tris–HCl, 0.9 M NaCl, 0.01% SDS, 30% formamide). For detection of *Spiroplasma* symbiont, the probe Spr403 (5′- AlexaFluor 555 - TAC TTA CTG TTC TTC CCT TAC A-3′; Matsuura et al., 2014) was used. Samples were incubated overnight at room temperature in the hybridization buffer containing 50 nM of the probe. After washed twice in PBST, nuclear DNA were stained with 4.5 µM of 4′,6-diamidino-2-phenylindole (DAPI; Thermo Fisher Scientific) for 1 h at room temperature, and then washed with PBT again (Koga et al., 2012). Samples were washed twice in PBST and mounted in 50% glycerol solution in PBS and observe a laser scanning confocal microscope (LSM 700, Carl Zeiss, Germany).

2 https://www.ezbiocloud.net

Results

Screening of *Spiroplasma* infection among diverse stinkbugs

Our large collection of stinkbug samples, consisting of 13 families, 69 genera, 97 species and 468 individuals (Supplementary Table S1), were subjected to diagnostic PCR detection of *Spiroplasma* infection. Of these, 4 families (30.8%), 7 genera (10.1%), 11 species (11.3%) and 21 individuals (4.5%) were diagnosed as *Spiroplasma* positive (Table 1). The following species contained *Spiroplasma* positive samples: *Riptortus pedestris* (1/1, Alydidae); *Acanthosoma denticaudum* (1/1, Acanthosomatidae); *Acanthosoma forficula* (1/1, Acanthosomatidae); *Acanthosoma labiduroides* (1/1, Acanthosomatidae); *Acanthosoma spinicolle* (3/3, Acanthosomatidae); *Adomerus triguttulus* (3/11, Cydnidae); *Macroscytus japonensis* (2/7, Cydnidae); *Eurydema dominulus* (1/10, Pentatomidae); *Eurydema rugosa* (1/17, Pentatomidae); *Eysarcoris ventralis* (6/7, Pentatomidae); and *Gonopsis affinis* (1/4, Pentatomidae) (Supplementary Table S1). These results suggested that, although not conclusive due to limited sample sizes, the *Spiroplasma* symbionts are, in general, facultative microbial associates of the stinkbugs.

Phylogenetic placement of *Spiroplasma* strains detected from diverse stinkbugs

All the 21 stinkbug samples positive of *Spiroplasma* infection were subjected to PCR amplification and sequencing of 1.5 kb region of 16S rRNA gene of *Spiroplasma*. The 16S rRNA gene sequences were subjected to molecular phylogenetic analysis together with *Spiroplasma* and allied bacterial sequences retrieved from GenBank and other DNA databases. The *Spiroplasma* symbionts detected from the stinkbugs were placed in either of the clades "*S. citri*," "*S. poulsoni*" or "*S. minum*" (*sensu* Barré et al., 2004; Paredes et al., 2015; Ballinger and Perlman, 2019) in the family Spiroplasmataceae, the class Mollicutes, and the phylum Tenericutes (or Mycoplasmatota) (Figure 1). Note that *Spiroplasma* symbionts previously reported from stinkbugs also fell in these clades: the clade "*S. poulsoni*" for the coffee stinkbug *Antestiopsis thunbergii* (Pentatomidae) (Matsuura et al., 2014) and the clade "*S. mirum*" for the corn stinkbug *Eurygaster integriceps* (Scutelleridae) (Amiri et al., 2020). Globally, the phylogenetic relationship of the *Spiroplasma* symbionts did not reflect the phylogenetic relationship of the host stinkbugs. On the other hand, the *Spiroplasma* symbionts associated with the same stinkbug family tended to be related to each other: all the symbionts of cydnid stinkbugs were placed in the "*S. minum*" clade; almost all the symbionts of pentatomid stinkbugs were

placed in the "*S. citri*" clade; and the symbionts of acanthosomatid stinkbugs were placed either in the "*S. poulsonii*" clade or the "*S. citri*" clade; (Figure 1). Furthermore, different individuals of the same stinkbug species tended to be associated with the same *Spiroplasma* symbiont as in *A. triguttulus*, *M. japonensis* and *E. ventralis*, although *A. spinicolle* represented an exceptional case (Figure 1).

Microbiome of stinkbugs with and without *Spiroplasma* infection

From 19 stinkbug samples, of which 8 and 11 were *Spiroplasma* positive and negative, respectively, the alimentary tracts were dissected and subjected to amplicon sequencing analysis of the V3-V4 region of bacterial 16S rRNA gene, with aphids and fruit flies as negative and positive control samples. *Spiroplasma* reads were specifically detected from the 8 *Spiroplasma* positive stinkbug samples (Figure 2; Supplementary Table S3), confirming the results of diagnostic PCR and molecular phylogenetic analysis (Supplementary Table S1; Figure 1). Reflecting the fact that stinkbugs generally harbor specific gut symbionts in their midgut symbiotic organ (Salem et al., 2015; Hosokawa et al., 2016; Takeshita and Kikuchi, 2017), the *Spiroplasma* reads generally occupied only a very small fraction of the total reads, ranging from 0.07 to 4.23%, with an exceptional case of 23.02% in *R. pedestris* (Figure 2; Supplementary Table S3). Phylum-, order- and family-level assignments of the reads verified that the majority of the reads were certainly derived from the gut symbionts of the stinkbugs: the Betaproteobacteria (= *Burkholderia*) dominant in *R. pedestris* (Coreoidea: Alydidae) (*cf.* Kikuchi et al., 2011); the Firmicutes (or Bacillota) (= *Clostridium*, *Lactococcus*) and the Actinobacteria (or Actinomycetota) (*Coriobacterium*, *Gordonibacter*) dominant in *Pyrrhocoris sinuaticollis* (Pyrrhocoroidea: Pyrrhocoridae) (*cf.* Sudakaran et al., 2012); and the Gammaproteobacteria (= Enterobacteriales) dominant in the other stinkbugs (Pentatomoidea: Acanthosomatidae, Pentatomidae, Scutelleridae and Urostylididae) (*cf.* Kikuchi et al., 2009; Kaiwa et al., 2014; Hosokawa et al., 2016, 2019) (Supplementary Tables S5–S7).

Infection prevalence, vertical transmission, and *in vivo* localization of *Spiroplasma* in *Eysarcoris ventralis*

Among the 11 *Spiroplasma* positive stinkbug species, we focused on the white-spotted stinkbug *E. ventralis* (Figure 3A), because this species was maintainable on sunflower seeds in the laboratory at least for some period and exhibited a high *Spiroplasma* infection rate (6/7 = 85.7%) in our initial screening (Supplementary Table S1). We additionally collected 25 individuals of *E. ventralis* from 4 localities in Japan (Supplementary Table S2), and the insects were subjected to a series of experiments. Diagnostic PCR of dissected tissues and organs of *E. ventralis* (Figures 3B,C) identified a high *Spiroplasma* infection rate (22/25 = 88.0%; Supplementary Table S2). In some samples, *Spiroplasma* infection was not detected from the dissected midgut symbiotic organ but identified from the rest of the insect body, suggesting that *Spiroplasma* is mainly distributed among other tissues and organs than the midgut symbiotic organ (Supplementary Table S2). An adult female (= Ibaraki1; Supplementary Table S2) laid an egg mass in the laboratory, the hatchlings were reared on sunflower seeds, three newborn nymphs and three adults were subjected to diagnostic PCR, and all the insects were diagnosed as *Spiroplasma* positive (Supplementary Table S2). These results strongly suggested that *Spiroplasma* is passed to the next generation vertically in *E. ventralis*. We dissected three individuals of *E. ventralis* (= Okinawa2, Okinawa3 and Ibaraki15), and the dissected tissue samples were subjected to amplicon sequencing analysis of bacterial 16S rRNA gene. While the proteobacterial gut symbiont was predominant in the dissected midgut samples, *Spiroplasma* was dominant in all the other tissue samples (Figure 3D; Supplementary Tables S8–S10), suggesting that *Spiroplasma* infection is found throughout the body parts of *E. ventralis*.

Finally, we additionally collected five adult insects of *E. ventralis*, and their dissected tissues were subjected to FISH visualization of *Spiroplasma* cells (Supplementary Table S2). FISH signals were detected from all the insects, but density of the signals varied among individuals. In the fat body, intracellular signals were consistently detected, although distribution and density of the signals were quite variable (Figures 3E,F). In the midgut symbiotic organ, intracellular signals were scarcely observed, whereas sporadic signals outside the crypts were occasionally found, which might represent *Spiroplasma* cells in the hemolymph (Figures 3G,H).

Discussion

In this study, we surveyed diverse stinkbugs representing 13 families, 69 genera, 97 species and 468 individuals for *Spiroplasma* symbionts. Thus far, *Spiroplasma* infections have been detected from diverse insects and other arthropods, plants, and some marine invertebrates (Anbutsu and Fukatsu, 2011; Ballinger and Perlman, 2019; Gasparich et al., 2020), but large-scale data of infection prevalence in natural host populations have been reported only in a few cases. To our knowledge, 10/136 (7.4%) species and 76/2052 (3.7%) individuals of diverse arthropods (Duron et al., 2008), 6/19 (31.6%) species and 284/2907 (9.8%) individuals of fruit flies (Watts et al., 2009), and 1/21 (11.3%) species and 24/566 (4.2%) individuals of aphids (Romanov et al., 2020) represent such reports. In comparison with these previous studies, our results that 11/97 (11.3%) species and 21/468 (4.2%) individuals of stinkbugs were *Spiroplasma* positive (Table 1; Supplementary Table S1) seem to present similar and relatively low infection rates, supporting the notion that the *Spiroplasma*

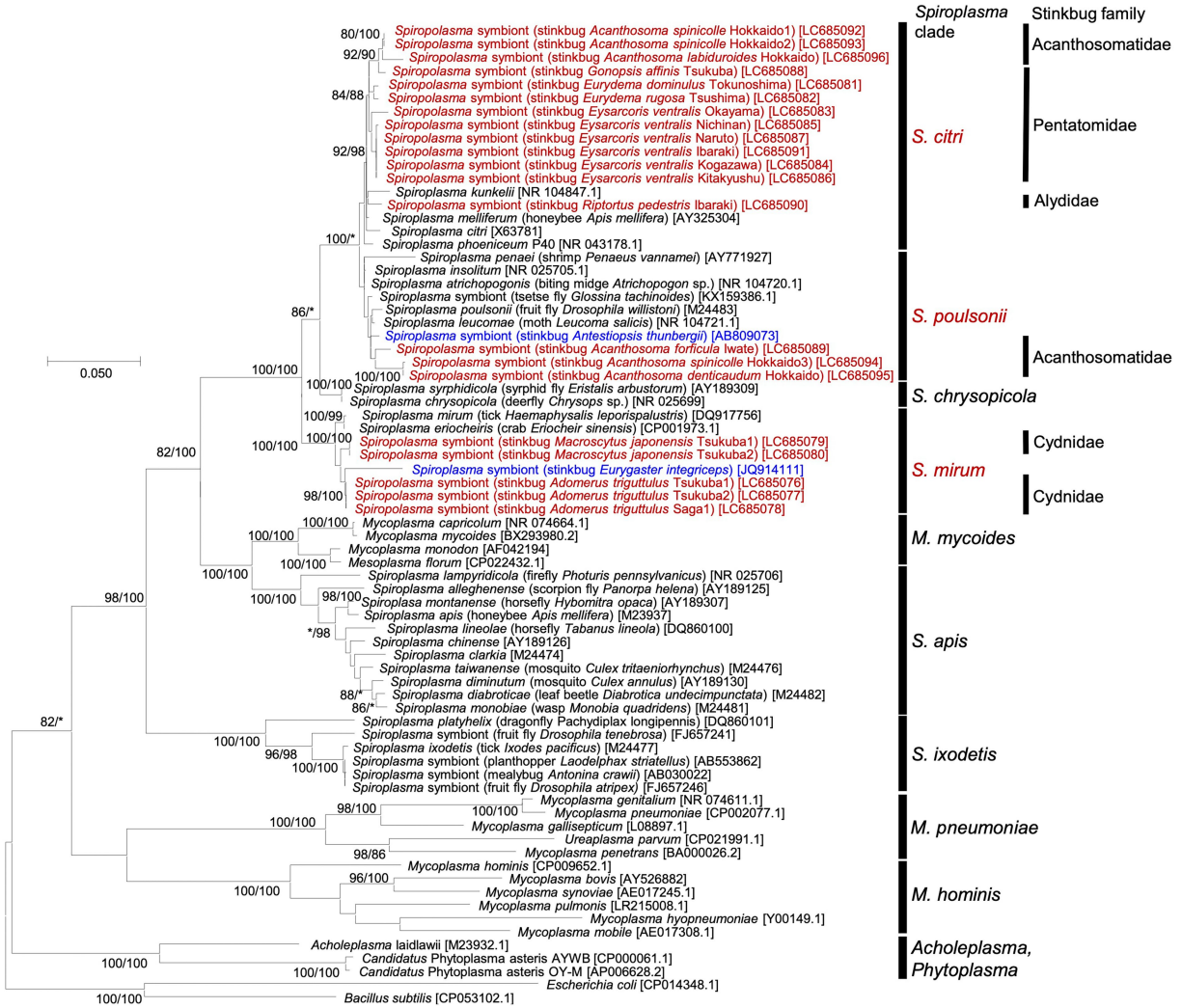


FIGURE 1
Phylogenic relationship of *Spiroplasma* symbionts detected from diverse stinkbugs based on 16S rRNA gene sequences. A maximum-likelihood phylogeny inferred from 1,632 aligned nucleotide sites is shown. Bootstrap support probabilities were obtained from 1,000 resamplings, of which the values of 80% or higher are shown at the nodes in the order of maximum likelihood/neighbor joining. Asterisks indicate support values lower than 80%. For each bacterial sequence, host-related information in parentheses and accession number in brackets are shown. The *Spiroplasma* sequences detected from stinkbugs in this study are highlighted in red, the *Spiroplasma* sequences detected from stinkbugs in previous studies (Matsuura et al., 2014; Amiri et al., 2020) are highlighted in blue, and the other sequences of *Spiroplasma* and allies retrieved from the databases are shown in black. The major *Spiroplasma* clades are depicted on the right side according to previously published papers (Paredes et al., 2015; Ballinger and Perlman, 2019) and the MolliGen database (Barré et al., 2004). On the right side, host stinkbug families for the *Spiroplasma* positive samples are also displayed.

symbionts are facultative microbial associates of insects and other arthropods in general.

Here it should be noted that the low detection rates of *Spiroplasma* in field-collected stinkbugs may entail underestimation for the following reasons. (i) Limited sample size: Since a limited number of samples, often a single specimen, were examined for each host species, *Spiroplasma* infections at low levels in natural host populations may be frequently overlooked. (ii) Fluctuating symbiont density: Facultative symbionts generally exhibit relatively low infection densities, the infection densities may drastically vary depending on environmental conditions, and thus samples with very low infection density may be diagnosed as *Spiroplasma* negative. (iii) PCR efficiency, specificity and sensitivity: Primer and amplicon sizes may affect PCR efficiency, specificity and sensitivity. We note that the amplicon size of diagnostic PCR, 0.8 kb, adopted in this study is relatively large, which may be less efficient for *Spiroplasma* detection in comparison with 0.2–0.5 kb amplicons that are commonly used for diagnostic PCR. In contrast to the possibility of false negatives, all the *Spiroplasma* infections detected from 21 stinkbug species in this study are true positives on the ground that they were confirmed by gene sequencing.

Amplicon sequencing of bacterial 16S rRNA gene revealed that the *Spiroplasma* symbionts are quantitatively minor in

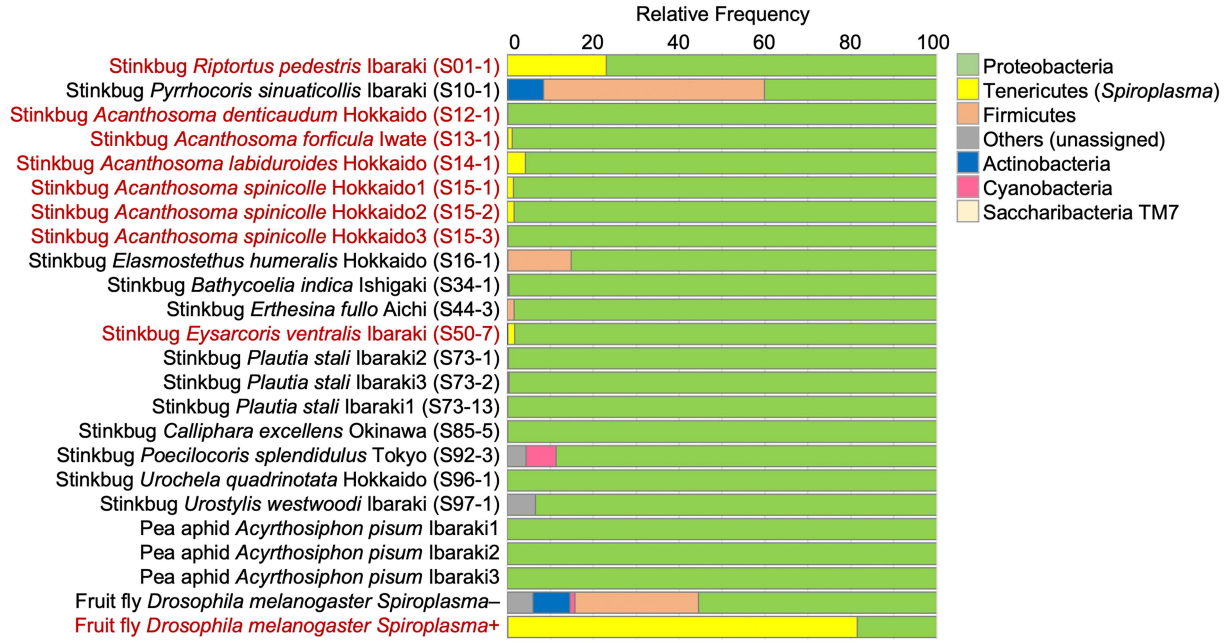


FIGURE 2
The relative abundance of the top seven bacterial phyla detected from stinkbugs and other insects by 16S rRNA gene amplicon sequencing. The phylum Tenericutes (or Mycoplasmatota) corresponds to *Spiroplasma*. The insect names in red indicate *Spiroplasma* positive ones. Sample numbers are shown in parentheses (Supplementary Table S1).

comparison with the primary gut symbionts (Figure 2), confirming the notion that the *Spiroplasma* symbionts are facultative microbial associates of relatively low infection densities in the stinkbugs. On the other hand, in dissected tissues and body parts containing no midgut symbiotic organ, the *Spiroplasma* symbionts were detected as a major bacterial component (Figure 3D). The apparent *Spiroplasma* predominance in the dissected tissues is likely due to the absence of other bacteria rather than high *Spiroplasma* density in these tissues. In combination with the FISH observations (Figures 3E–H), these results suggest that the *Spiroplasma* symbionts are distributed in various tissues and organs at relatively low densities, at least in *E. ventralis*, like many other facultative symbiotic bacteria of insects and other arthropods (Oliver et al., 2010; Feldhaar, 2011).

Molecular phylogenetic analysis uncovered that the stinkbug-associated *Spiroplasma* symbionts are placed in at least three distinct clades in the Spiroplasmataceae (Figure 1), highlighting multiple evolutionary origins of the stinkbug-*Spiroplasma* associations. Obviously, the *Spiroplasma* phylogeny does not reflect the host stinkbug phylogeny (Figure 1), indicating the absence of host-symbiont co-speciation. On the other hand, the *Spiroplasma* symbionts associated with the same stinkbug family tend to be phylogenetically related to each other (ex. "*S. mirum*" with Cydnidae, "*S. poulsonii*" with Acanthosomatidae; see Figure 1). These patterns can be accounted for by the following ecological, physiological and/or evolutionary processes. (i) Physiological host specificity: Closely-related host insects may provide similar intra-host conditions for symbiotic bacteria, which may promote the preferential association of specific symbiont genotypes to specific host taxa. (ii) Ecological symbiont sharing mediated by horizontal symbiont transfers *via* common host plants: Closely-related host insects tend to use the same group of food plants, and plant-mediated symbiont transfers may cause the preferential association of specific symbiont genotypes to specific host taxa. It should be noted that plant-pathogenic *Spiroplasma* species are generally vectored by plant-sucking hemipteran insects (Gasparich et al., 2020) and some facultative symbionts of whiteflies and leafhoppers, such as *Rickettsia*, *Cardinium* and *Wolbachia*, have been reported to be horizontally transmitted *via* plants (Caspi-Fluger et al., 2014; Chrostek et al., 2017). (iii) Ecological symbiont sharing mediated by horizontal symbiont transfers *via* common parasites: Closely-related host insects tend to be exploited by common parasites, and parasite-mediated symbiont transfers may result in the preferential association of specific symbiont genotypes to specific host taxa. Previous studies reported that parasitoid wasps and blood-sucking mites may mediate such horizontal symbiont transmission (Jaenike et al., 2007; Gehrer and Vorburger, 2012). In fact, the evolutionary dynamics of the stinkbug-*Spiroplasma* associations seems likely to have been shaped by combination of these processes, which should be pursued in more depth in the future.

The large-scale infection prevalence data of *Spiroplasma* in stinkbugs are of particular interest in that such data have been also collected for other facultative symbionts, *Wolbachia* and *Sodalis*, in diverse stinkbugs (Kikuchi and Fukatsu, 2003; Hosokawa et al., 2015), which provide an opportunity to examine how different

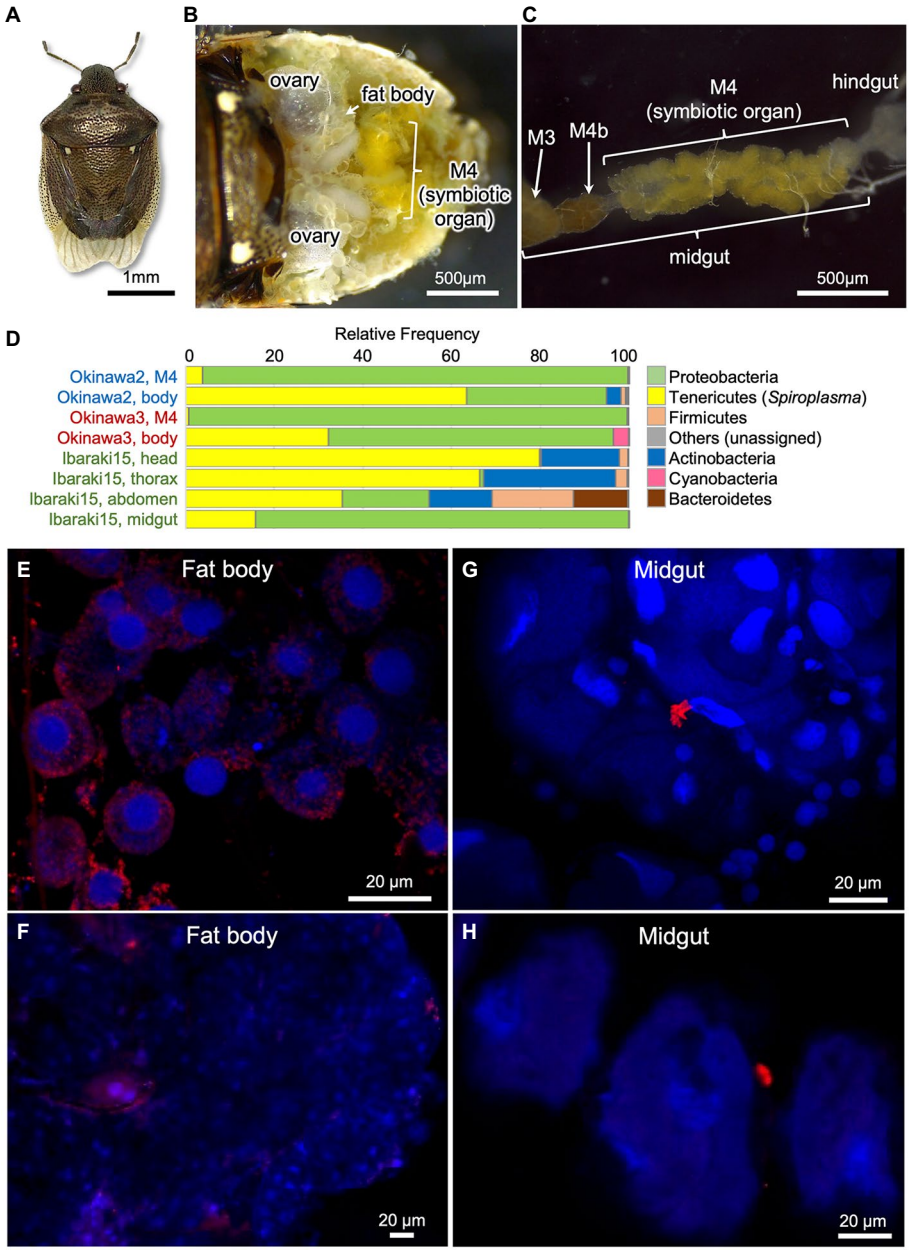


FIGURE 3
Spiroplasma symbiont associated with the white-spotted stinkbug *Eysarcoris ventralis*. **(A)** An adult insect. **(B)** Dissected abdomen. **(C)** Isolated alimentary tract. Abbreviations: M3, midgut third section; M4, midgut fourth section as the symbiotic organ; M4b, bulb-like section connected to M4 (see Oishi et al., 2019; Moriyama et al., 2022). **(D)** The relative abundance of the top seven bacterial phyla detected from three individuals of *E. ventralis* by 16S rRNA gene amplicon sequencing. The phylum Tenericutes (or Mycoplasmatota) corresponds to *Spiroplasma*. The sample names in the same color indicate those derived from the same insect individual. **(E–H)** FISH localization of *Spiroplasma* symbiont in fat body **(E,F)** and midgut symbiotic organ **(G,H)** of *E. ventralis*. **(E,F)** represent images derived from different individuals, so do **(G,H)**. Red signals represent *Spiroplasma* cells visualized by FISH whereas blue signals show insect nuclear DNA visualized by DAPI staining.

facultative symbiotic bacteria exhibit their specific infection patterns in the same host insect group. The infection frequency patterns are largely similar across the distinct symbiotic bacteria, *Wolbachia*, *Sodalis* and *Spiroplasma*: most stinkbug species were symbiont negative, some species exhibited intermediate infection frequencies, and several species showed 100% symbiont infection (Supplementary Figure S1). The predominance of the symbiont-free stinkbug species strongly suggests that the facultative symbiotic bacteria are generally not essential for the host stinkbugs. The intermediate infection frequencies may be realized by symbiont-mediated cytoplasmic incompatibility as known for many *Wolbachia* strains (Werren et al., 2008; Kaur et al., 2021), by symbiont-mediated defense against parasites and/or pathogens as known for some *Wolbachia* and *Spiroplasma* strains (Flórez et al.,

2015; Ballinger and Perlman, 2019), or by symbiont-mediated fitness facilitation as known for a variety of facultative symbiotic bacteria (Oliver et al., 2010; Zug and Hammerstein, 2015). The fixed symbiont infections in a limited number of host species may have been maintained by symbiont-induced cytoplasmic incompatibility as known for diverse *Wolbachia* strains (Werren et al., 2008; Kaur et al., 2021), or by fitness improvement *via* nutritional supplementation as known for *Wolbachia* of bedbugs (Hosokawa et al., 2010; Nikoh et al., 2014; Moriyama et al., 2015) and *Sodalis pierantonius* in grain weevils (Oakeson et al., 2014; Vigneron et al., 2014). What mechanisms underpin the 100% infections with *Wolbachia* and *Sodalis* in several stinkbug species (Supplementary Figures S1A,B) is currently unknown and to be examined in future studies. Notably, no stinkbug species exhibited 100% infection with *Spiroplasma* in this study (Supplementary Figure S1C). To our knowledge, no *Spiroplasma* symbionts have been reported to exhibit 100% infection rates in natural arthropod populations (Duron et al., 2008; Watts et al., 2009; Anbutsu and Fukatsu, 2011; Romanov et al., 2020). These observations suggest the possibility that, although speculative, *Spiroplasma* might be inherently unlikely to evolve essential mutualism with arthropod hosts. This may be relevant to the fact that the genomes of *Spiroplasma* and allied bacteria, constituting the class Mollicutes, are highly reduced ancestrally, being devoid of most of the metabolic genes needed for establishment of nutritional mutualism *via* synthesis of amino acids and vitamins (Lo et al., 2016). On the other hand, some *Spiroplasma* genomes contain a number of laterally transferred genes (Lo et al., 2015), evidencing existence of a potential route for acquisition of such metabolic genes. Since our survey of *Spiroplasma* diversity has been quite limited in comparison with the enormous arthropod diversity in nature, it is conceivable that 100% infection cases might be discovered in wider surveys in the future.

In conclusion, we present the distribution and the diversity of the stinkbug-*Spiroplasma* symbiotic associations, thereby laying the foundation for future studies on the stinkbug-*Spiroplasma* symbiosis. The next step will be the establishment of an experimentally tractable stinkbug-*Spiroplasma* model system that is maintainable in the laboratory. In this study, we tried to establish a laboratory rearing system for *E. ventralis* that exhibited frequent *Spiroplasma* infections in natural populations, but it was not successful: fed with sunflower seeds and water supplemented with ascorbic acid, adult insects of *E. ventralis* survived for over a month, but produced few eggs and finally died out. Only a female managed to produce an egg mass, and emergence of *Spiroplasma*-infected offspring from the eggs strongly suggested vertical transmission of the *Spiroplasma* symbiont, but more data are needed to verify this idea. In this context, we regard the bean bug *Riptortus pedestris* as a promising model system, on the grounds that *R. pedestris* is easily and stably maintainable in the laboratory and widely used for studies on the gut symbiotic bacteria of the genus *Burkholderia* (Takeshita and Kikuchi, 2017; Kaltenpoth and Flórez, 2020). In this study, we were able to examine only a field-collected sample of *R. pedestris*, in which the *Spiroplasma* infection density was remarkably high (Figure 2). Thus far, we have inspected several laboratory strains of *R. pedestris*, but unfortunately, they were all free of *Spiroplasma* infection. However, we expect that a wide field survey in the next season will lead to the establishment of *Spiroplasma*-infected laboratory strains of *R. pedestris*, which will enable us to experimentally investigate important biological aspects of the stinkbug-*Spiroplasma* symbiotic association such as fitness effects, reproductive phenotypes, vertical transmission route and efficiency, etc.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found at: https://www.ddbj.nig.ac.jp/, LC685076-LC685096; https://www.ddbj.nig.ac.jp/, DRR358028-DRR358051; https://www.ddbj.nig.ac.jp/, DRR359516-DRR359523.

Author contributions

SK, TH, and TF: conceived the study. TH, KO, and TF: collected the stinkbug samples. SK, TH, and KM: performed diagnostic PCR screening and 16S rRNA gene sequencing. SK and TH: conducted molecular phylogenetic analysis. SK and KM: performed amplicon sequencing analysis. KO: conducted FISH. TF and SK: wrote the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This study was supported by the JST ERATO grant number JPMJER1902 to SK and TF, and the JSPS KAKENHI grant number JP17H06388 to TF. KO was supported by the JSPS research fellowships for young scientists (21J01321 to KO).

Acknowledgments

We thank Atsushi Kikuchi, Akio Tanigawa, Bin Hirota, Chihiro Himuro, Eiichi Hara, Emi Kobayashi, Genta Okude, Gen Sakurai, Hiroyuki Hirayama, Hiromi Mukai, Hirokazu Toju, Katsura Ito, Koichi Inadomi, Kiichi Shimizu, Kyosuke Okuda, Keiko Ohno-Shiromoto, Keiichi Takahashi, Mitsuo Baba, Mantaro Hironaka, Masaaki Kimura, Minoru Moriyama, Mitsutaka Sakakibara, Mikio Takai, Mariko Taguchi, Narumi Baba, Nahomi Kaiwa, Nobuo Tsurusaki, Ryutaro Suzuki, Shuhei Kada, Shin-ich Kudo, Satoshi Maehara, Sumio Tojo, Tetsuhiro Kawagoe, Teruyuki Niimi, Tomohito Noda, Tomonari Nozaki, Taku Tsukada, Yoshiko Ayabe, Yuki G. Baba, Yoshitomo Kikuchi, Yutaka Osada, and Yu Matsuura for insect samples, and Minoru Moriyama and Ryuichi Koga for technical advice on molecular experiments.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary materials

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2022.1044771/full#supplementary-material

References

Alexeeva, I., Elliott, E. J., Rollins, S., Gasparich, G. E., Lazar, J., and Rohwer, R. G. (2006). Absence of *Spiroplasma* or other bacterial 16S rRNA genes in brain tissue of hamsters with scrapie. *J. Clin. Microbiol.* 44, 91–97. doi: 10.1128/jcm.44.1.91-97.2006

Amiri, A., Bandani, A. R., and Kafil, M. (2020). Gut compartments and ovary bacterial symbionts of the Sunn pest. *J. Asia Pacif. Entomol.* 23, 723–730. doi: 10.1016/j.aspen.2020.06.002

Anbutsu, H., and Fukatsu, T. (2011). *Spiroplasma* as a model insect endosymbiont. *Environ. Microbiol. Rep.* 3, 144–153. doi: 10.1111/j.1758-2229.2010.00240.x

Ballinger, M. J., and Perlman, S. J. (2019). The defensive *Spiroplasma*. *Curr. Opin. Insect Sci.* 32, 36–41. doi: 10.1016/j.cois.2018.10.004

Barré, A., de Daruvar, A., and Blanchard, A. (2004). MolliGen, a database dedicated to the comparative genomics of Mollicutes. *Nucleic Acids Res.* 32, 307D–3310D. doi: 10.1093/nar/gkh114

Bastian, F. O., and Foster, J. W. (2001). *Spiroplasma* sp. 16S rDNA in Creutzfeldt-Jakob disease and scrapie as shown by PCR and DNA sequence analysis. *J. Neuropathol. Exp. Neurol.* 60, 613–620. doi: 10.1093/jnen/60.6.613

Bourtzis, K., and Miller, T. A. (2003). *Insect Symbiosis. 1st Edn*. Boca Raton: CRC Press.

Brune, A. (2014). Symbiotic digestion of lignocellulose in termite guts. *Nat. Rev. Microbiol.* 12, 168–180. doi: 10.1038/nrmicro3182

Buchner, P. (1965). *Endosymbiosis of Animals with Plant Microorganisms*. New York, Interscience Publishers.

Caspi-Fluger, A., Inbar, M., Steinberg, S., Friedmann, Y., Freund, M., Mozes-Daube, N., et al. (2014). Characterization of the symbiont *Rickettsia* in the mirid bug *Nesidiocoris tenuis* (Reuter) (Heteroptera: Miridae). *Bull. Entomol. Res.* 104, 681–688. doi: 10.1017/S0007485314000492

Chrostek, E., Pelz-Stelinski, K., Hurst, G. D. D., and Hughes, G. L. (2017). Horizontal transmission of intracellular insect symbionts via plants. *Front. Microbiol.* 8:2237. doi: 10.3389/fmicb.2017.02237

Clark, T. B. (1982). Spiroplasmas: diversity of arthropod reservoirs and host-parasite relationships. *Science* 217, 57–59. doi: 10.1126/science.217.4554.57

Dally, M., Lalzar, M., Belausov, E., Gottlieb, Y., Coll, M., and Zchori-Fein, E. (2020). Cellular localization of two *Rickettsia* symbionts in the digestive system and within the ovaries of the mirid bug, *Macrolophous pygmaeus*. *Insects* 11:530. doi: 10.3390/insects11080530

Daniels, M. J. (1983). Mechanisms of *Spiroplasma* pathogenicity. *Annu. Rev. Phytopathol.* 21, 29–43. doi: 10.1146/annurev.py.21.090183.000333

Davis, R. E., Worley, J. F., Whitcomb, R. F., Ishijima, T., and Steere, R. L. (1972). Helical filaments produced by a mycoplasma-like organism associated with corn stunt disease. *Science* 176, 521–523. doi: 10.1126/science.176.4034.521

Douglas, A. E. (2009). The microbial dimension in insect nutritional ecology. *Funct. Ecol.* 23, 38–47. doi: 10.1111/j.1365-2435.2008.01442.x

Duron, O., Bouchon, D., Boutin, S., Bellamy, L., Zhou, L., Engelstädter, J., et al. (2008). The diversity of reproductive parasites among arthropods: *Wolbachia* do not walk alone. *BMC Biol.* 6:27. doi: 10.1186/1741-7007-6-27

Feldhaar, H. (2011). Bacterial symbionts as mediators of ecologically important traits of insect hosts. *Ecol. Entomol.* 36, 533–543. doi: 10.1111/j.1365-2311.2011.01318.x

Flórez, L. V., Biedermann, P. H. W., Engl, T., and Kaltenpoth, M. (2015). Defensive symbioses of animals with prokaryotic and eukaryotic microorganisms. *Nat. Prod. Rep.* 32, 904–936. doi: 10.1039/C5NP00010F

Fukatsu, T., and Nikoh, N. (1998). Two intracellular symbiotic bacteria from the mulberry psyllid *Anomoneura mori* (Insecta, Homoptera). *Appl. Environ. Microbiol.* 64, 3599–3606. doi: 10.1128/AEM.64.10.3599-3606.1998

Fukatsu, T., Tsuchida, T., Nikoh, N., and Koga, R. (2001). *Spiroplasma* symbiont of the pea aphid, *Acyrthosiphon pisum* (Insecta: Homoptera). *Appl. Environ. Microbiol.* 67, 1284–1291. doi: 10.1128/AEM.67.3.1284-1291.2001

Gasparich, G. E., Kuo, C.-H., and Foissac, X. (2020). Spiroplasma. *Bergey's Manual of Systematics of Archaea and Bacteria*. Wiley. doi: 10.1002/9781118960608.gbm01262.pub2a.

Gehrer, L., and Vorburger, C. (2012). Parasitoids as vectors of facultative bacterial endosymbionts in aphids. *Biol. Lett.* 8, 613–615. doi: 10.1098/rsbl.2012.0144

Gerth, M., Gansauge, M.-T., Weigert, A., and Bleidorn, C. (2014). Phylogenomic analyses uncover origin and spread of the *Wolbachia* pandemic. *Nat. Commun.* 5:5117. doi: 10.1038/ncomms6117

Haselkorn, T. S. (2010). The *Spiroplasma* heritable bacterial endosymbiont of *Drosophila*. *Fly (Austin)* 4, 80–87. doi: 10.4161/fly.4.1.10883

He, L.-S., Zhang, P.-W., Huang, J.-M., Zhu, F.-C., Danchin, A., and Wang, Y. (2017). The enigmatic genome of an obligate ancient *Spiroplasma* symbiont in a hadal holothurian. *Appl. Environ. Microbiol.* 84, e01965–e01917. doi: 10.1128/AEM.01965-17

Hedges, L. M., Brownlie, J. C., O'Neill, S. L., and Johnson, K. N. (2008). *Wolbachia* and virus protection in insects. *Science* 322:702. doi: 10.1126/science.1162418

Herlemann, D. P. R., Labrenz, M., Jürgens, K., Bertilsson, S., Waniek, J. J., and Andersson, A. F. (2011). Transitions in bacterial communities along the 2000 km salinity gradient of the Baltic Sea. *ISME J.* 5, 1571–1579. doi: 10.1038/ismej.2011.41

Hosokawa, T., Imanishi, M., Koga, R., and Fukatsu, T. (2019). Diversity and evolution of bacterial symbionts in the gut symbiotic organ of jewel stinkbugs (Hemiptera: Scutelleridae). *Appl. Entomol. Zool.* 54, 359–367. doi: 10.1007/s13355-019-00630-4

Hosokawa, T., Kaiwa, N., Matsuura, Y., Kikuchi, Y., and Fukatsu, T. (2015). Infection prevalence of *Sodalis* symbionts among stinkbugs. *Zool. Let.* 1:5. doi: 10.1186/s40851-014-0009-5

Hosokawa, T., Koga, R., Kikuchi, Y., Meng, X.-Y., and Fukatsu, T. (2010). *Wolbachia* as a bacteriocyte-associated nutritional mutualist. *Proc. Natl. Acad. Sci. U. S. A.* 107, 769–774. doi: 10.1073/pnas.0911476107

Hosokawa, T., Matsuura, Y., Kikuchi, Y., and Fukatsu, T. (2016). Recurrent evolution of gut symbiotic bacteria in pentatomid stinkbugs. *Zool. Let.* 2:24. doi: 10.1186/s40851-016-0061-4

Hurst, G. D., Anbutsu, H., Kutsukake, M., and Fukatsu, T. (2003). Hidden from the host: *Spiroplasma* bacteria infecting *Drosophila* do not cause an immune response, but are suppressed by ectopic immune activation. *Insect Mol. Biol.* 12, 93–97. doi: 10.1046/j.1365-2583.2003.00380.x

Jaenike, J., Polak, M., Fiskin, A., Helou, M., and Minhas, M. (2007). Interspecific transmission of endosymbiotic *Spiroplasma* by mites. *Biol. Lett.* 3, 23–25. doi: 10.1098/rsbl.2006.0577

Jaenike, J., Unckless, R., Cockburn, S. N., Boelio, L. M., and Perlman, S. J. (2010). Adaptation via symbiosis: recent spread of a *Drosophila* defensive symbiont. *Science* 329, 212–215. doi: 10.1126/science.1188235

Jiggins, F. M., Hurst, G. D. D., Jiggins, C. D., v d Schulenburg, J. H. G., and Majerus, M. E. N. (2000). The butterfly *Danaus chrysippus* is infected by a male-killing *Spiroplasma* bacterium. *Parasitology* 120, 439–446. doi: 10.1017/S0031182099005867

Kaiwa, N., Hosokawa, T., Nikoh, N., Tanahashi, M., Moriyama, M., Meng, X.-Y., et al. (2014). Symbiont-supplemented maternal investment underpinning host's ecological adaptation. *Curr. Biol.* 24, 2465–2470. doi: 10.1016/j.cub.2014.08.065

Kakizawa, S., and Kamagata, Y. (2014). A multiplex-PCR method for strain identification and detailed phylogenetic analysis of AY-group phytoplasmas. *Plant Dis.* 98, 299–305. doi: 10.1094/PDIS-03-13-0216-RE

Kaltenpoth, M., and Flórez, L. V. (2020). Versatile and dynamic symbioses between insects and *Burkholderia* bacteria. *Annu. Rev. Entomol.* 65, 145–170. doi: 10.1146/annurev-ento-011019-025025

Kaur, R., Shropshire, J. D., Cross, K. L., Leigh, B., Mansueto, A. J., Stewart, V., et al. (2021). Living in the endosymbiotic world of *Wolbachia*: a centennial review. *Cell Host Microbe* 29, 879–893. doi: 10.1016/j.chom.2021.03.006

Kikuchi, Y., and Fukatsu, T. (2003). Diversity of *Wolbachia* endosymbionts in heteropteran bugs. *Appl. Environ. Microbiol.* 69, 6082–6090. doi: 10.1128/aem.69.10.6082-6090.2003

Kikuchi, Y., Hosokawa, T., and Fukatsu, T. (2011). An ancient but promiscuous host–symbiont association between *Burkholderia* gut symbionts and their heteropteran hosts. *ISME J.* 5, 446–460. doi: 10.1038/ismej.2010.150

Kikuchi, Y., Hosokawa, T., Nikoh, N., Meng, X.-Y., Kamagata, Y., and Fukatsu, T. (2009). Host-symbiont co-speciation and reductive genome evolution in gut symbiotic bacteria of acanthosomatid stinkbugs. *BMC Biol.* 7:2. doi: 10.1186/1741-7007-7-2

Koga, R., Meng, X.-Y., Tsuchida, T., and Fukatsu, T. (2012). Cellular mechanism for selective vertical transmission of an obligate insect symbiont at the bacteriocyte–embryo interface. *Proc. Natl. Acad. Sci. U. S. A.* 109, E1230–E1237. doi: 10.1073/pnas.1119212109

Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. (2018). MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* 35, 1547–1549. doi: 10.1093/molbev/msy096

Lemoine, M. M., Engl, T., and Kaltenpoth, M. (2020). Microbial symbionts expanding or constraining abiotic niche space in insects. *Curr. Opin. Insect Sci.* 39, 14–20. doi: 10.1016/j.cois.2020.01.003

Lo, W.-S., Gasparich, G. E., and Kuo, C.-H. (2015). Found and lost: the fates of horizontally acquired genes in arthropod-symbiotic *Spiroplasma*. *Genome Biol. Evol.* 7, 2458–2472. doi: 10.1093/gbe/evv160

Lo, W.-S., Huang, Y.-Y., and Kuo, C.-H. (2016). Winding paths to simplicity: genome evolution in facultative insect symbionts. *FEMS Microbiol. Rev.* 40, 855–874. doi: 10.1093/femsre/fuw028

Łukasik, P., van Asch, M., Guo, H., Ferrari, J., Charles, J., and Godfray, H. (2013). Unrelated facultative endosymbionts protect aphids against a fungal pathogen. *Ecol. Lett.* 16, 214–218. doi: 10.1111/ele.12031

Matsuura, Y., Hosokawa, T., Serracin, M., Tulgetske Genet, M., Miller Thomas, A., and Fukatsu, T. (2014). Bacterial symbionts of a devastating coffee plant pest, the stinkbug *Antestiopsis thunbergii* (Hemiptera: Pentatomidae). *Appl. Environ. Microbiol.* 80, 3769–3775. doi: 10.1128/AEM.00554-14

Matsuura, Y., Kikuchi, Y., Meng Xian, Y., Koga, R., and Fukatsu, T. (2012). Novel clade of alphaproteobacterial endosymbionts associated with stinkbugs and other arthropods. *Appl. Environ. Microbiol.* 78, 4149–4156. doi: 10.1128/AEM.00673-12

Moran, N. A., McCutcheon, J. P., and Nakabachi, A. (2008). Genomics and evolution of heritable bacterial symbionts. *Annu. Rev. Genet.* 42, 165–190. doi: 10.1146/annurev.genet.41.110306.130119

Moriyama, M., Hayashi, T., and Fukatsu, T. (2022). A mucin protein predominantly expressed in the female-specific symbiotic organ of the stinkbug *Plautia stali*. *Sci. Rep.* 12:7782. doi: 10.1038/s41598-022-11895-1

Moriyama, M., Nikoh, N., Hosokawa, T., and Fukatsu, T. (2015). Riboflavin provisioning underlies *Wolbachia*'s fitness contribution to its insect host. *MBio* 6, e01732–e01715. doi: 10.1128/mBio.01732-15

Muyzer, G., de Waal, E. C., and Uitterlinden, A. G. (1993). Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16S rRNA. *Appl. Environ. Microbiol.* 59, 695–700. doi: 10.1128/aem.59.3.695-700.1993

Nikoh, N., Hosokawa, T., Moriyama, M., Oshima, K., Hattori, M., and Fukatsu, T. (2014). Evolutionary origin of insect–*Wolbachia* nutritional mutualism. *Proc. Natl. Acad. Sci. U. S. A.* 111, 10257–10262. doi: 10.1073/pnas.1409284111

Nikoh, N., Hosokawa, T., Oshima, K., Hattori, M., and Fukatsu, T. (2011). Reductive evolution of bacterial genome in insect gut environment. *Genome Biol. Evol.* 3, 702–714. doi: 10.1093/gbe/evr064

Nováková, E., Hypša, V., and Moran, N. A. (2009). *Arsenophonus*, an emerging clade of intracellular symbionts with a broad host distribution. *BMC Microbiol.* 9:143. doi: 10.1186/1471-2180-9-143

Nunan, L. M., Lightner, D. V., Oduori, M. A., and Gasparich, G. E. (2005). *Spiroplasma penaei* sp. nov., associated with mortalities in *Penaeus vannamei*, Pacific white shrimp. *Int. J. Syst. Evol. Microbiol.* 55, 2317–2322. doi: 10.1099/ijs.0.63555-0

Oakeson, K. F., Gil, R., Clayton, A. L., Dunn, D. M., von Niederhausern, A. C., Hamil, C., et al. (2014). Genome degeneration and adaptation in a nascent stage of symbiosis. *Genome Biol. Evol.* 6, 76–93. doi: 10.1093/gbe/evt210

Oishi, S., Moriyama, M., Koga, R., and Fukatsu, T. (2019). Morphogenesis and development of midgut symbiotic organ of the stinkbug *Plautia stali* (Hemiptera: Pentatomidae). *Zool. Let.* 5:16. doi: 10.1186/s40851-019-0134-2

Oliver, K. M., Degnan, P. H., Burke, G. R., and Moran, N. A. (2010). Facultative symbionts in aphids and the horizontal transfer of ecologically important traits. *Annu. Rev. Entomol.* 55, 247–266. doi: 10.1146/annurev-ento-112408-085305

Oren, A., and Garrity, G. G. (2021). Valid publication of the names of forty-two phyla of prokaryotes. *Int. J. Syst. Evol. Microbiol.* 71:005056. doi: 10.1099/ijsem.0.005056

Paredes, J. C., Herren, J. K., Schüpfer, F., Marin, R., Claverol, S., Kuo, C.-H., et al. (2015). Genome sequence of the *Drosophila melanogaster* male-killing *Spiroplasma* strain MSRO endosymbiont. *mBio* 6:e02437-14. doi: 10.1128/mBio.02437-14

Perlman, S. J., Hunter, M. S., and Zchori-Fein, E. (2006). The emerging diversity of *Rickettsia*. *Proc. R. Soc. B* 273, 2097–2106. doi: 10.1098/rspb.2006.3541

Poulson, D. F., and Sakaguchi, B. (1961). Nature of "sex-ratio" agent in *Drosophila*. *Science* 133, 1489–1490. doi: 10.1126/science.133.3463.1489

Romanov, D. A., Zakharov, I. A., and Shaikevich, E. V. (2020). *Wolbachia*, *Spiroplasma*, and *Rickettsia* symbiotic bacteria in aphids (Aphidoidea). *Vavilovskii Zhurnal Genet. Selektsii* 24, 673–682. doi: 10.18699/vj20.661

Saglio, P., Lhospital, M., Laflèche, D., Dupont, G., Bovè, J. M., Tully, J. G., et al. (1973). *Spiroplasma citri* gen. and sp. n.: a mycoplasma-like organism sssociated with "stubborn" disease of citrus. *Int. J. Syst. Evol. Microbiol.* 23, 191–204. doi: 10.1099/00207713-23-3-191

Salem, H., Bauer, E., Strauss, A. S., Vogel, H., Marz, M., and Kaltenpoth, M. (2014). Vitamin supplementation by gut symbionts ensures metabolic homeostasis in an insect host. *Proc. R. Soc. B* 281:20141838. doi: 10.1098/rspb.2014.1838

Salem, H., Florez, L., Gerardo, N., and Kaltenpoth, M. (2015). An out-of-body experience: the extracellular dimension for the transmission of mutualistic bacteria in insects. *Proc. R. Soc. B* 282:20142957. doi: 10.1098/rspb.2014.2957

Sanada-Morimura, S., Matsumura, M., and Noda, H. (2013). Male killing caused by a *Spiroplasma* symbiont in the small brown planthopper, *Laodelphax striatellus*. *J. Heredity* 104, 821–829. doi: 10.1093/jhered/est052

Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., et al. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal omega. *Mol. Syst. Biol.* 7:539. doi: 10.1038/msb.2011.75

Simon, J.-C., Boutin, S., Tsuchida, T., Koga, R., Le Gallic, J.-F., Frantz, A., et al. (2011). Facultative symbiont infections affect aphid reproduction. *PLoS One* 6:e21831. doi: 10.1371/journal.pone.0021831

Sudakaran, S., Salem, H., Kost, C., and Kaltenpoth, M. (2012). Geographical and ecological stability of the symbiotic mid-gut microbiota in European firebugs, *Pyrrhocoris apterus* (Hemiptera, Pyrrhocoridae). *Mol. Ecol.* 21, 6134–6151. doi: 10.1111/mec.12027

Takeshita, K., and Kikuchi, Y. (2017). *Riptortus pedestris* and *Burkholderia* symbiont: an ideal model system for insect–microbe symbiotic associations. *Res. Microbiol.* 168, 175–187. doi: 10.1016/j.resmic.2016.11.005

Teixeira, L., Ferreira, Á., and Ashburner, M. (2008). The bacterial symbiont *Wolbachia* induces resistance to RNA viral infections in *Drosophila melanogaster*. *PLoS Biol.* 6:e1000002:e2. doi: 10.1371/journal.pbio.1000002

Van Arnam, E. B., Currie, C. R., and Clardy, J. (2018). Defense contracts: molecular protection in insect-microbe symbioses. *Chem. Soc. Rev.* 47, 1638–1651. doi: 10.1039/C7CS00340D

Vigneron, A., Masson, F., Vallier, A., Balmand, S., Rey, M., Vincent-Monégat, C., et al. (2014). Insects recycle endosymbionts when the benefit is over. *Curr. Biol.* 24, 2267–2273. doi: 10.1016/j.cub.2014.07.065

Viver, T., Orellana, L. H., Hatt, J. K., Urdiain, M., Díaz, S., Richter, M., et al. (2017). The low diverse gastric microbiome of the jellyfish *Cotylorhiza tuberculata* is dominated by four novel taxa. *Environ. Microbiol.* 19, 3039–3058. doi: 10.1111/1462-2920.13763

Wang, W., Gu, W., Ding, Z., Ren, Y., Chen, J., and Hou, Y. (2005). A novel *Spiroplasma* pathogen causing systemic infection in the crayfish *Procambarus clarkii* (Crustacea: decapod), in China. *FEMS Microbiol. Lett.* 249, 131–137. doi: 10.1016/j.femsle.2005.06.005

Wang, W., Gu, W., Gasparich, G. E., Bi, K., Ou, J., Meng, Q., et al. (2011). *Spiroplasma eriocheiris* sp. nov., associated with mortality in the Chinese mitten crab, *Eriocheir sinensis*. *Int. J. Syst. Evol. Microbiol.* 61, 703–708. doi: 10.1099/ijs.0.020529-0

Watts, T., Haselkorn, T. S., Moran, N. A., and Markow, T. A. (2009). Variable incidence of *Spiroplasma* infections in natural populations of *Drosophila* species. *PLoS One* 4:e5703. doi: 10.1371/journal.pone.0005703

Werren, J. H., Baldo, L., and Clark, M. E. (2008). *Wolbachia*: master manipulators of invertebrate biology. *Nat. Rev. Microbiol.* 6, 741–751. doi: 10.1038/nrmicro1969

Williamson, D. L., Sakaguchi, B., Hackett, K. J., Whitcomb, R. F., Tully, J. G., Carle, P., et al. (1999). *Spiroplasma poulsonii* sp. nov., a new species associated with male-lethality in *Drosophila willistoni*, a neotropical species of fruit fly. *Int. J. Syst. Evol. Microbiol.* 49, 611–618. doi: 10.1099/00207713-49-2-611

Xie, J., Butler, S., Sanchez, G., and Mateos, M. (2014). Male killing *Spiroplasma* protects *Drosophila melanogaster* against two parasitoid wasps. *Heredity* 112, 399–408. doi: 10.1038/hdy.2013.118

Xie, J., Vilchez, I., and Mateos, M. (2010). *Spiroplasma* bacteria enhance survival of *Drosophila hydei* attacked by the parasitic wasp *Leptopilina heterotoma*. *PLoS One* 5:e12149. doi: 10.1371/journal.pone.0012149

Zchori-Fein, E., and Perlman, S. J. (2004). Distribution of the bacterial symbiont *Cardinium* in arthropods. *Mol. Ecol.* 13, 2009–2016. doi: 10.1111/j.1365-294X.2004.02203.x

Zug, R., and Hammerstein, P. (2012). Still a host of hosts for *Wolbachia*: analysis of recent data suggests that 40% of terrestrial arthropod species are infected. *PLoS One* 7:e38544. doi: 10.1371/journal.pone.0038544

Zug, R., and Hammerstein, P. (2015). Bad guys turned nice? A critical assessment of *Wolbachia* mutualisms in arthropod hosts. *Biol. Rev. Camb. Philos. Soc.* 90, 89–111. doi: 10.1111/brv.12098

frontiers | Frontiers in Microbiology

TYPE Original Research
PUBLISHED 02 November 2022
DOI 10.3389/fmicb.2022.996938

OPEN ACCESS

EDITED BY
Shigeyuki Kakizawa,
National Institute of Advanced Industrial Science and Technology (AIST), Japan

REVIEWED BY
Maxuel Andrade,
National Center for Research in Energy and Materials, Brazil
Frank Bastian,
Tulane Medical Center, United States

*CORRESPONDENCE
Wei Sun
swjscdc@aliyun.com
Wei Gu
skywei426@sina.com

SPECIALTY SECTION
This article was submitted to Microbial Symbioses, a section of the journal Frontiers in Microbiology

RECEIVED 18 July 2022
ACCEPTED 10 October 2022
PUBLISHED 02 November 2022

CITATION
Liu P, Li Y, Ye Y, Chen J, Li R, Zhang Q, Li Y, Wang W, Meng Q, Ou J, Yang Z, Sun W and Gu W (2022) The genome and antigen proteome analysis of *Spiroplasma mirum*. *Front. Microbiol.* 13:996938. doi: 10.3389/fmicb.2022.996938

The genome and antigen proteome analysis of *Spiroplasma mirum*

Peng Liu[1], Yuxin Li[1], Youyuan Ye[1], Jiaxin Chen[1], Rong Li[1], Qinyi Zhang[1], Yuan Li[1], Wen Wang[2], Qingguo Meng[2], Jingyu Ou[1], Zhujun Yang[1], Wei Sun[3]* and Wei Gu[2]*

[1]Hunan Provincial Key Laboratory for Special Pathogens Prevention and Control, Basic Medical School, Hengyang Medical School, Institute of Pathogenic Biology, University of South China, Hengyang, China, [2]Key Laboratory for Aquatic Crustacean Diseases, College of Marine Science and Engineering, Nanjing Normal University, Nanjing, China, [3]Jiangsu Provincial Center for Disease Control and Prevention, Nanjing, China

Spiroplasma mirum, small motile wall-less bacteria, was originally isolated from a rabbit tick and had the ability to infect newborn mice and caused cataracts. In this study, the whole genome and antigen proteins of *S. mirum* were comparative analyzed and investigated. Glycolysis, pentose phosphate pathway, arginine metabolism, nucleotide biosynthesis, and citrate fermentation were found in *S. mirum*, while trichloroacetic acid, fatty acids metabolism, phospholipid biosynthesis, terpenoid biosynthesis, lactose-specific PTS, and cofactors synthesis were completely absent. The Sec systems of *S. mirum* consist of SecA, SecE, SecDF, SecG, SecY, and YidC. Signal peptidase II was identified in *S. mirum*, but no signal peptidase I. The relative gene order in *S. mirum* is largely conserved. Genome analysis of available species in *Mollicutes* revealed that they shared only 84 proteins. *S. mirum* genome has 381 pseudogenes, accounting for 31.6% of total protein-coding genes. This is the evidence that spiroplasma genome is under an ongoing genome reduction. Immunoproteomics, a new scientific technique combining proteomics and immunological analytical methods, provided the direction of our research on *S. mirum*. We identified 49 proteins and 11 proteins (9 proteins in common) in *S. mirum* by anti-*S. mirum* serum and negative serum, respectively. Forty proteins in *S. mirum* were identified in relation to the virulence. All these proteins may play key roles in the pathogeny and can be used in the future for diagnoses and prevention.

KEYWORDS

genome, antigen proteome, *Spiroplasma mirum*, spiroplasma, pathogenesis

Introduction

Spiroplasmas, belonging to the class *Mollicutes* that includes *Mycoplasma*, *Phytoplasma,* and so on, featured by small genome sizes and a lack of peptidoglycan layer (Liu et al., 2018; Sasajima and Miyata, 2021). Spiroplasmas are geographically widespread as bacterial pathogens or symbionts in insects and plants (Liu P. et al., 2017) as well as important pathogens in commercially exploited aquatic crustaceans (Wang et al., 2004, 2005; Nunan et al., 2005). Spiroplasmas perform helical structure with well-defined cytoskeleton-like structure (Kürner et al., 2005; Trachtenberg, 2006) also show snake-like motility (Shaevitz et al., 2005) and unique infective characters including male-killing endosymbionts (Watts et al., 2009), which possibly involved in certain neurodegenerative diseases such as transmissible spongiform encephalopathies (Bastian et al., 2007). In recent years, this microorganism was studied as an attractive minimal model system (Trachtenberg, 2004), and its ultrastructural and macro-molecular characters have been well-described (Trachtenberg, 2004; Shaevitz et al., 2005; Trachtenberg et al., 2008).

Genome research was necessary and useful technique to investigate many species, it helps us to answer previously intractable biological questions (Sussman, 2015). Many bacterial whole genomes have been sequenced in the class of *Mollicutes* (including *Mycoplasma*, *Phytoplasma*, and *Spiroplasma*), the whole genome sequence of a few spiroplasma bacteria has yet to be published (Liu P. et al., 2017; Chen et al., 2019). Genomic research can provide us with more information on the characteristics of the morphology, structure, physiological metabolism, and genetic evolution of the bacteria (Chen et al., 2019). *Spiroplasma eriocheiris*, a pathogen of Chinese mitten crab *Eriocheir sinensis* and close to *Spiroplasma mirum* (Wang et al., 2011), performs chemotaxis without the conventional two-component system which was commonly found in bacterial chemotaxis. The cells are polarized by a tip structure and a dumbbell-shaped core in the tip connected by a fat ribbon, they are forming the internal structure of *S. eriocheiris*. The spiroplasma genomes reported so far do not have orthologs of other bacterial motility systems but have one tubulin homolog-FtsZ and five to seven MreB protein. MreB is related to actin, which is responsible for many eukaryotic motility systems (Takahashi et al., 2020). Sixteen proteins were identified as the components of the internal structure by mass spectrometry, including Fibril protein and four types of MreB proteins (Liu P. et al., 2017). MreB5, one of the five MreB paralogs, was proven the contributing to cell elongation and is essential for the transition from rod-to-helical shape in *Spiroplasma* (Harne et al., 2020). Cytoskeleton-like proteins (Fibril, Mrebs, and EF-Tu) were also phosphorylated suggesting that phosphorylation may play a crucial role in the formation of the cytoskeleton-like structure (Liu et al., 2020).

With the completion of a large number of biological genome sequences, while it is found that genome research alone does not explain the fundamental problem, because the ultimate function of various physiological functions in organisms is protein. Immunoproteomics as an emerging field of proteome analysis plays a key role in screening immunogenic proteins. Such a method has been used to successfully identify antigenic epitopes in several pathogens, such as *Mycoplasma bovis*, *Brucella abortus*, *Neisseria meningitidis,* and *S. eriocheiris* (Liu Y. et al., 2017).

Therefore, the study of proteomics has become the main research topic and direction in the post-genome era (Vandemoortele et al., 2016), and the most important issue for pathogenic microorganisms is to find the main immune proteins or antigen proteins (Aslam et al., 2017). That is to find the protein that plays a major role in the pathogenic process. The main method to study microbial immunoproteomics is to extract the whole protein or membrane protein of microorganism, electrophoresis, then hybridization with antiserum, so as to screen out antigen protein and then identify it by mass spectrometry or sequencing (Vandemoortele et al., 2016). Several mycoplasma immunoproteomics studies have been carried out using this method, such as *Mycoplasma genitalium*, *Mycoplasma pneumoniae*, *Mycoplasma hyopneumoniae*, *Mycoplasma synoviae,* and *Mycoplasma mycoides* (Khan et al., 2018).

Spiroplasma mirum originally isolated from a rabbit tick (*Haemaphysalis leporis-palustis* ATCC 29335 = CIP 103925) (Megraud et al., 1983). *S. mirum* and *S. eriocheiris are* close to each other in the phylogenetic tree and both can cause the cataract of neonatal rat (Hou et al., 2017). In this study, we mainly conducted comparative genomics studies on *S. mirum* and *S. eriocheiris* to find out the characteristics of the physiological metabolism, and genetic evolution of the spiroplasma, and to determine their different pathogenicity. We screened the antigenic proteins and provide a comprehensive view of the immunogenic proteins of *S. mirum* and also certainly provide valuable information for the identification of virulent proteins or diagnosis of pathogenic mechanisms.

Materials and methods

Bacterial strain

Spiroplasma mirum, Suckling Mouse Cataract Agent (SMCA) (Megraud et al., 1983), was originally isolated from rabbit ticks (*Haemaphysalis leporis-palustis*). We purchased

SMCA [CIP 103925] from ATCC and cultured and cloned it with the same method (Itoh et al., 1989).

Genome sequencing strategy

The complete genome sequences of *S. mirum* were firstly determined by high-throughput sequencing via pyrosequencing. A total of 170,589 sequences with an average length of 255 bp were obtained for the *S. mirum* genome, resulting in 37-fold genome coverage.

Assembly was performed using Newbler software of the 454-suite package, producing 19 contigs ranging from 0.5 to 490 kb for *S. mirum*.

Multiplex PCR was performed to determine the relationship of contigs, and closure of the gaps was performed by sequencing the PCR products. Phred, Phrap, and Consed software package[1] was used for the final assembly and edition. Regions with poor sequencing quality and homopolymers were resequenced. The final consensus accuracy was 99.9999% for the *S. mirum* genome.

Gene prediction and annotation

Putative protein-coding sequences (CDSs) were identified by Glimmer3 (Delcher et al., 1999) and ZCURVE 1.0 (Guo et al., 2003). Functional annotation of CDSs was performed through BLASTP searches against GenBank's non-redundant (nr) protein database, followed by manual inspection. Transfer RNA genes were predicted by tRNAScan-SE (v1.23) (Lowe and Eddy, 1997). The metabolic pathways were constructed using KEGG database (Juncker et al., 2003; Kanehisa et al., 2004). Lipoprotein was determined by LipoP 1.0 (Juncker et al., 2003) and searching PROSITE motif PS51257. Protein domain prediction and COG assignment were performed by RPS-BLAST using NCBI CDD library.

An all-to-all BLASTP was applied to determine the paralogs. Two sequences in a pair are paralogs if the remaining HSPs cover at least 80% of the shorter protein's length and if the identity is greater or equal to 50%.

Comparative genomics

Orthologous proteins between *S. eriocheiris* and *S. mirum* were identified by all-vs-all reciprocal-BLASTP search, with criteria of similarity >20% and coverage >80%. Orthologs of known Mollicutes genomes were obtained from the MBGD database (Uchiyama, 2003). Unique genes were verified by TBLASTN search using protein sequences of each strain against the other strain's genome sequence. Comparison of the gene order between *S. eriocheiris* and *S. mirum* was analyzed by Artemis Comparison Tool (ACT) (Carver et al., 2005). Protein synteny plot of *S. eriocheiris* and *S. mirum* was constructed by blast-score-ratio (Rasko et al., 2005).

Identification of antigen proteins

The spiroplasma cells were activated in R2 medium to exponential phase and collected by centrifugation. After washing and resuspension, the sample was sonicated on ice and centrifuged to remove the cellular debris. Rabbit antiserum against *S. mirum* was prepared by following the methods described by Liu Y. et al. (2017). The serum was conjugated to Protein A Beads at 4°C for 4 h with gentle shaking. Conjugated antibody was added to the solution protein of spiroplasma and incubated at 4°C for 4 h. After a wash, the proteins were eluted by Glycine from the beads.

Dried protein pellets were reconstituted in 100 mM ammonium bicarbonate and protein concentrations were determined by Bio-Rad protein assay. Proteins were digested with trypsin and inactivated by heating at 99°C for 5 min. The solution was cleared by centrifugation and vacuum dried to remove bicarbonate salts.

Peptides were continuously separated by strong cation column (SCX) followed by C18 columns (Dionex) before being subjected to MS/MS analysis in an LTQ-Orbitrap mass spectrometer (Thermo Electron, Bremen, Germany). The mass spectrum data were initially searched against the database constructed using the predicted proteins of *S. mirum* with the aid of the Sequest search engine.

Phylogenetic tree construction

We analyzed the genome of *S. mirum*, it revealed that *S. mirum* shares a total of 84 common proteins with *S. eriocheiris*, 14 Mycoplasmas, 4 Phytoplasmas, 3 Ureaplasmas, 1 Acholeplasma, and part of the genome sequence analysis of *Spiroplasma citri*. Concatenated protein sequences of 84 orthologous proteins of *mycoplasma* species were first aligned using MUSCLE (Edgar, 2004), then the conserved alignment blocks were extracted by the Gblocks program (Castresana, 2000). The phylogenetic tree was built using the maximum likelihood method implemented in PHYML (Guindon and Gascuel, 2003) using the following parameters: 100 replications for bootstrap analysis, "JTT" for substitution model, "estimated" for the proportion of invariable sites, "estimated" for gamma distribution parameters, "4" for the number of substitution categories, "yes" to optimize tree topology, and "BIONJ" for starting tree(s). Graphical representation and edition of the phylogenetic tree were performed with MEGA4.

1 http://www.phrap.org/phredphrapconsed.html

Nucleotide sequence accession number

The complete genome sequence and annotation of *S. mirum* were deposited in GenBank under accession numbers CP002082.

Results

Genome features

The genomes of *S. mirum* contain a single, circular chromosome with 1,132,591 bp in length. The base numbering start point was chosen at the first base of the *dnaA* gene, adjacent to the replication origin (*oriC*; **Figure 1**). The basic characteristics of the genomes of *S. mirum* and *S. eriocheiris* are shown in **Table 1**. *S. mirum* genome length is 1, 132, 591 bp, and G + C content is 29.41%. *S. mirum* and *S. eriocheiris are* close to each other in the phylogenetic tree (**Figure 2**), both species could cause cataracts in rats (Hou et al., 2017). The *S. eriocheiris* genome size is 1, 364,757 bp, with 29.79% G + C content (Liu P. et al., 2017). The specific genes of *S. mirum* genomes are listed in **Supplementary Table 1**. Generally, there is a complete base excision repair system in the *Spiroplasma* genomes. *S. eriocheiris* genome contains several genes involved in homologous recombination, including *recA*, *ruvA/B*, *recD*, *recO*, *recR*, and *recU*, while in *S. mirum* only the *recA* and *recD* genes remain complete (**Supplementary Table 2**), giving us a hint that *S. mirum* lost the capacity for recombinational repair.

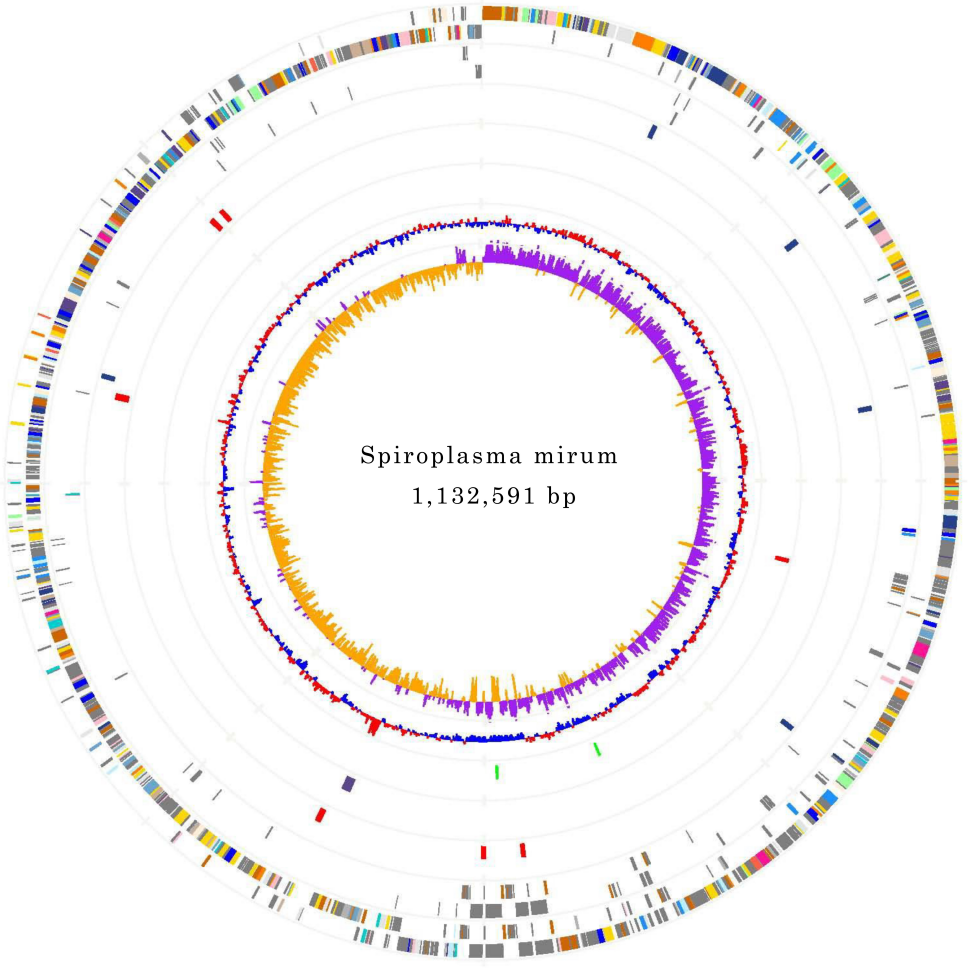


FIGURE 1
Chromosome Atlas of *Spiroplasma mirum*. Moving inside, each concentric circle represents a different genomic feature. The outermost circle shows predicted protein-coding sequences on both strands, colored by functional categories according to COG classification. The second circle displays specific genes of *S. mirum* genome compared with the other. The third circle illustrates tRNA genes on plus strand **(blue)** and minus strand **(red)**, and ribosomal RNA genes **(green)**. The fourth circle presents IS elements. The fifth circle shows mean centered G + C content (red-above mean, blue-below mean). The sixth circle **(innermost)** represents GC skew (G - C)/(G + C) calculated using a 1 kb window.

TABLE 1 General features of *Spiroplasma eriocheiris* and *Spiroplasma mirum* genome.

Spiroplasma	*Spiroplasma eriocheiris*	*Spiroplasma mirum*
Genome size (base pairs)	1,364,757 bp	1,132,591 bp
G + C content (%)	29.79	29.41
Protein coding genes		
With assigned functions	755	731
Conserved hypothetical	174	138
Hypothetical	313	338
Total	1,242	1,207
Average gene length	976 bp	745 bp
Coding density (%)	88.7%	79.2%
Pseudogenes	53	381
rRNA operons	1	1
tRNA genes	32	33
IS elements	0	3

We identified 22 transcription factors and one sigma factor in the *S. eriocheiris* genome, and 13 transcription factors and one sigma factor in the *S. mirum* genome (**Supplementary Table 3**).

The *S. mirum* contains 1, 207 CDSs, 1 ribosomal RNA operon, and 33 tRNA genes. The average length of CDSs is 745 bp, and CDSs and the totality of rRNA and tRNA genes accounted for 79.2 and 0.6% of the genome, respectively. The GC contents on the GC and the third codon were 30.58 and 20.93%, respectively. A total of 731 CDSs (60.6%) have clear biological functions, 138 CDSs (11.4%) are similar to conservative proteins with unknown function, 338 CDSs (28%) are hypothetical proteins, COG classification was possible for 735 CDSs (60.9%). We assume that the *S. eriocheiris* contains 1, 242 CDSs, with an average length of 976 bp, and the GC contents on the GC and the third codon were 30.66 and 20.65%, respectively. A total of 755 CDSs (60.8%) have been identified, which have clear biological functions, another 174 CDSs (14%) are similar to conservative proteins with unknown function, 313 CDSs (25.2%) are hypothetical proteins, and 769 CDSs (61.9%) can be classified by COG (**Figure 3**). The genome contains 1 ribosomal RNA operon and 32 tRNA genes. CDSs and stable RNA genes accounted for 88.7 and 0.5% of the genome, respectively. There are no insertion sequences (IS element) or phages in the *S. eriocheiris* genome, while there are three IS elements in the *S. mirum* genome, one of which may be a phage. The phage with a length of 16,008 bp includes 21 genes (From SMM_0576 to SMM_0597). Particularly, SMM_578, SMM_579, SMM_580, SMM_581, and SMM_583 were identified as putative adhesin p58, P12, P54, P123, and P18 of *S. citri*, respectively. Thus, the phage may be acquired from *S. citri*. The phage may have facilitated extensive genome rearrangements in *S. mirum* and contributed to horizontal gene transfers that led to species-specific adaptation to different eukaryotic hosts and acquisition of pathogenicity. In addition, the common ancestor of the *Spiroplasma*, *Entomoplasma*, and

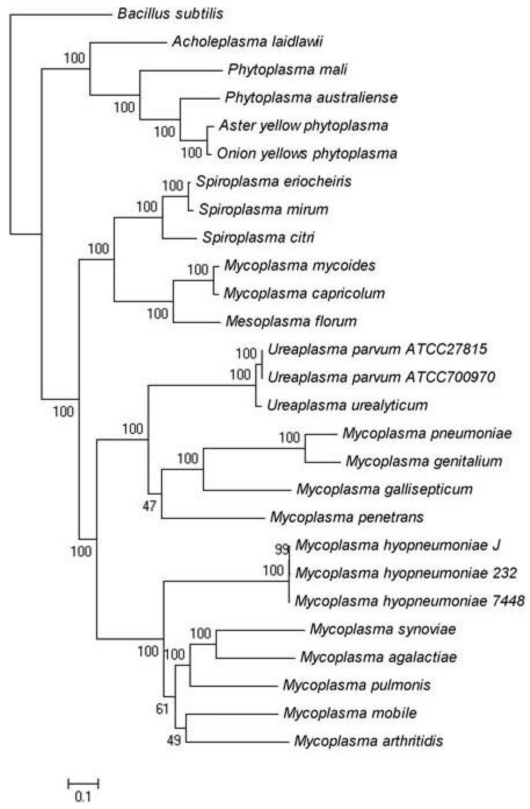


FIGURE 2
Maximum-likelihood tree of *Spiroplasma* and *Mycoplasma* species for which complete genomes are available.

Mycoplasma clade may have had a relatively large genome and flexible metabolic capacity; the extremely reduced genomes of *Mycoplasma* and *Spiroplasma* species are likely to be the result of independent gene losses (Lo et al., 2013). It is thought that genome reduction is aided by genetic isolation-bacteria that live in monocultures in special host organs, or inside host cells, have less access to other bacterial species from which they can obtain genes (Waterworth et al., 2020).

The genome of *S. mirum* is 232,166 bp smaller than *S. eriocheiris*. Genome reduction is a common phenomenon in numerous pathogens. The pattern of genome reduction observed in *S. atrichopogonis* has been found in several other arthropod-associated bacteria, such as *S. poulsonii* (Lo et al., 2015). Generally, the stable establishment of a bacterium in a host cell results in genome reduction (Campbell et al., 2015). The loss of functions required for living independently but not within a host gives rise to diminished genomes in various symbionts. Although the phenomenon of genome reduction can be explained by existing evolutionary models, the initiation of the process is not thoroughly understood. The key step preceding genome reduction in the symbiont was likely the horizontal acquisition of the putative lagriamide lga biosynthetic gene cluster. The parasitic bacterium does need to produce

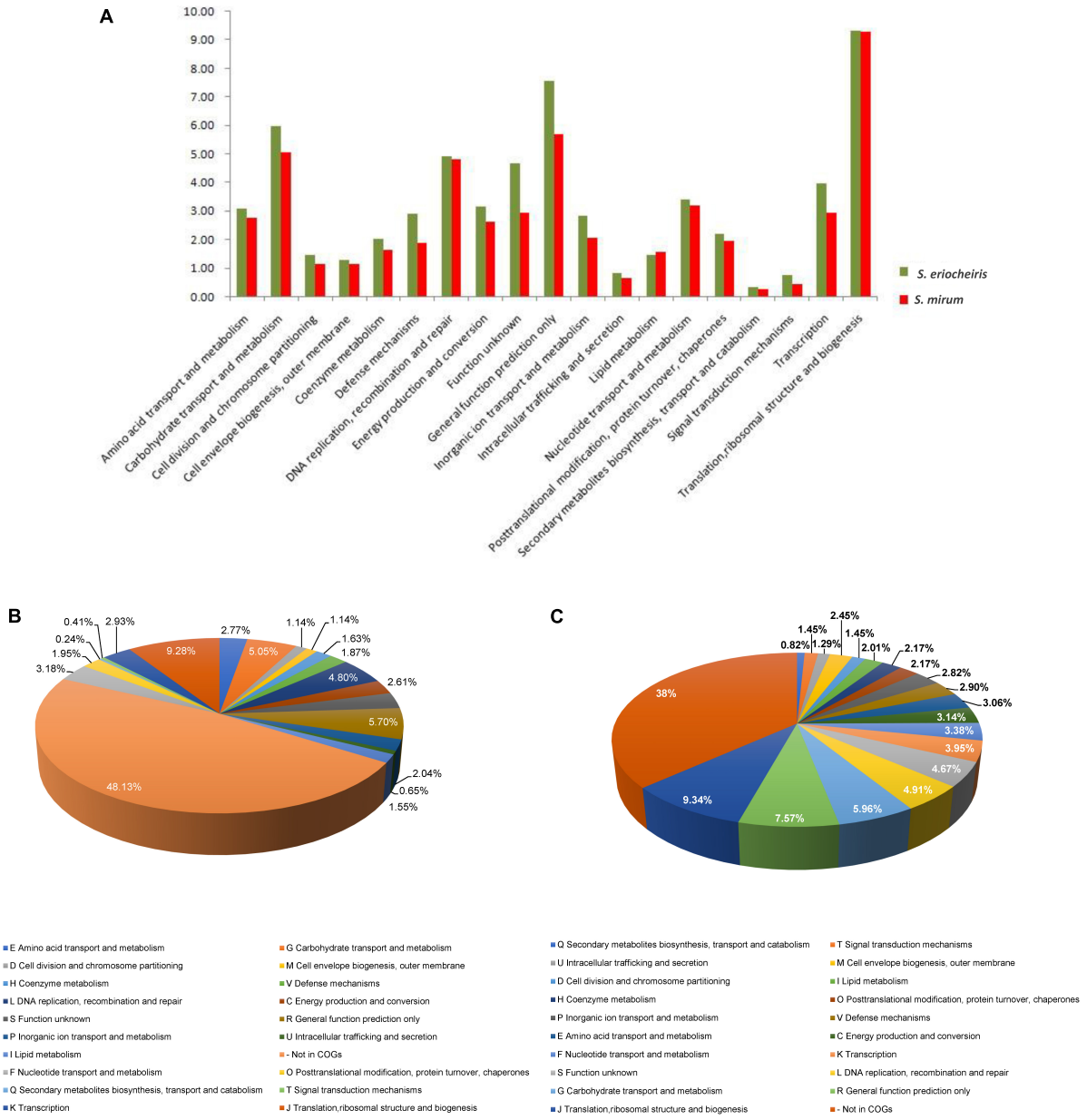


FIGURE 3
(A) Comparison of COG distribution of *Spiroplasma eriocheiris* and *Spiroplasma mirum*. **(B,C)** COG classification of *S. mirum* and *S. eriocheiris*, respectively.

nutrients that the host provides, leading to the loss of genes that it would need to live independently and to a consequent reduction in genome size (McCutcheon and Moran, 2011).

Basic metabolic

Basic feature of *S. mirum* metabolism elucidated by genomic analysis permitted the metabolic reconstruction of spiroplasma (**Figure 4**). Spiroplasmas are parasitic in specific hosts and their specific tissues, indicating that they have special nutritional requirements and parasitic history. Genomic sequences will provide evidence to reveal the complex nutritional requirement of spiroplasma. Both *S. eriocheiris* and *S. mirum* can utilize glucose in the Embden–Meyerhof–Parnas (EMP) pathway, The oxidative branch of the pentose phosphate pathway was completely absent, as well as the TCA cycle, and finally, pyruvate can be converted into lactic acid, formic acid, and acetic acid. *S. eriocheiris* can also produce acetaldehyde. However, the arginine dihydrolase pathway was found in both strains.

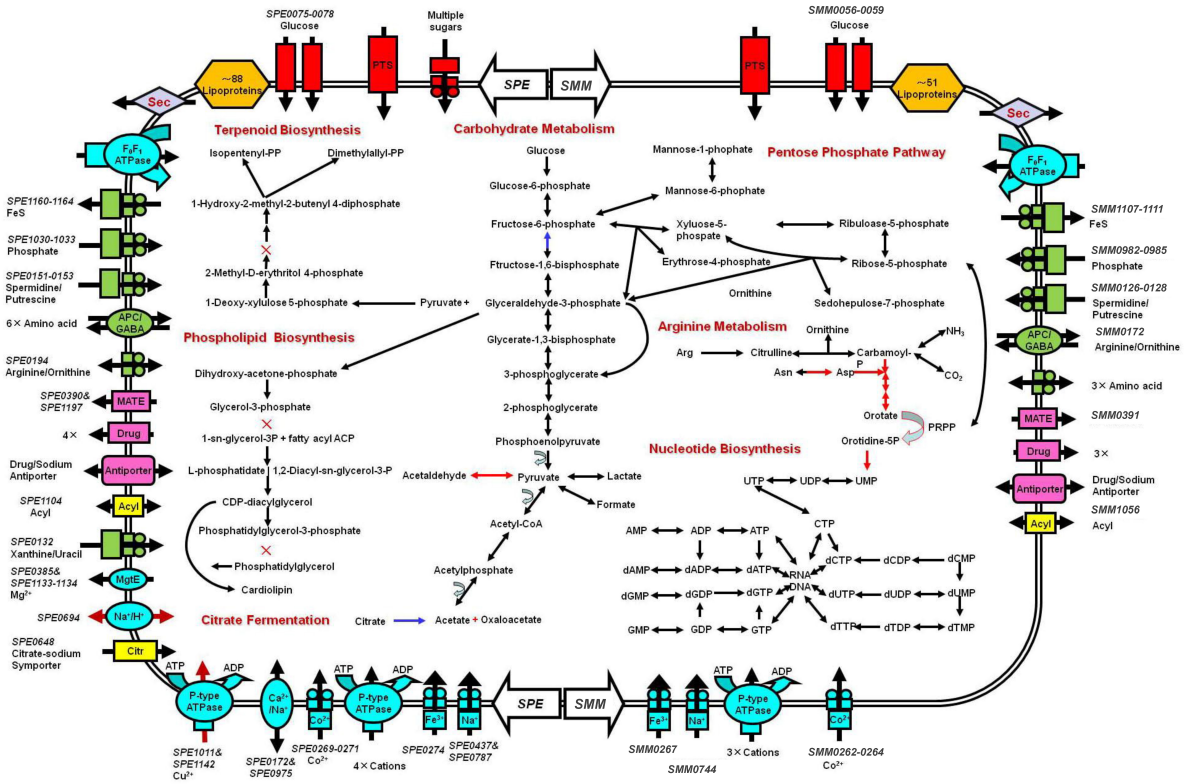


FIGURE 4
Schematic representation of metabolic pathways of *Spiroplasma eriocheiris* **(left side)** and *Spiroplasma mirum* **(right side)**. Shown are transporters and the main elements of metabolic pathways, deduced from the set of genes with predicted functions. Black arrows: pathways found in both *S. eriocheiris* and *S. mirum* genomes. Blue arrows: *S. eriocheiris* special in *Mollicutes*. Red arrows: pathways found in the *S. eriocheiris* genome but not in the *S. mirum* genome. × : absent enzyme in pathways. Transporters are colored according to their substrates: blue, cations; green, anions and amino acids; rad, carbohydrates; pink, multidrug, and metabolic product efflux. Arrows indicate the direction of substrate transport. In some cases, we cannot identify all the components of these transporters.

Compared with other flexor-membranous microorganisms, *S. eriocheiris* has fructose 1,6-diphosphate fructose in the glycolysis pathway, which can catalyze 1,6-diphosphate to form fructose 6-phosphate in gluconeogenesis. However, due to the absence of citric acid cycle, the gluconeogenesis pathway is still incomplete (**Figure 4**). Both genomes show a complete absence of pentose phosphate pathway oxidation pathways and incomplete non-oxidation pathways, this is similar to other mycoplasmas (Pollack et al., 2002; Westberg et al., 2004).

The tricarboxylic acid cycle is completely absent in the genome, so do mycoplasma. Arginine hydrolysis pathway exists in both spiroplasma genomes, resulting in ornithine and ATP, CO_2, and ammonia. Both purine and pyrimidine cannot be synthesized in *S. mirum*, but *S. eriocheiris* can synthesize pyrimidine through 1-diphosphate-5-phosphate ribose (PRPP). The pyrimidine is synthesized. Most amino acids cannot be synthesized in spiroplasma, oxaloacetic acid (produced by citric acid fermentation) can produce aspartic acid, pyruvate can produce serine, asparagine, and glycine can be converted from aspartic acid and serine, respectively. Consistent with their parasitic life characteristics, spiroplasma loses most of their synthetic abilities, such as fatty acids and various cofactors (except CoA). GDSL family lipases can catalyze the decomposition of lipids to obtain fatty acids in *S. eriocheiris* and *S. mirum*, and then fatty acids can be oxidized, but the enzymes that catalyze fatty acids β-oxidation are missing in the genome. terpenoids and phospholipids cannot be synthesized in spiroplasma because of the lack of some enzymes, which must be obtained from the host or medium to maintain their survival (**Figure 4**).

Antigen proteins

We used the IgG and antibodies of different antigens in serum can specifically bind to other antigens and Protein A, we use polyclonal antibodies to screen antigens in spiroplasma. In the experiment, *S. mirum* were tested, and negative serum control group was set up to remove non-specific immune binding. We identified 49 and 11 proteins of positive and negative serum, including 9 common proteins. Therefore, 40

antigenic proteins of *S. mirum* were screened, and the results are shown in **Table 2**. While 45 antigenic proteins of *S. eriocheiris* were identified (Liu Y. et al., 2017). There are six antigenic proteins of *S. mirum* are homologous proteins that can be used as antigen proteins of *S. eriocheiris*, including extension factor EF-G, pyruvate dehydrogenase (β subunit, E2 subunit, E3 subunit), thioredoxin reductase and DegV family proteins.

The oligoendopeptidase F (SMM0146), belonging to thimet oligopeptidase family member, was proved as a surface-exposed proteases of *M. hyopneumoniae* (Jarocki et al., 2019). EF-Tu (SMM_0065) and EF-G (SMM_0064) were identified as antigenic proteins. EF-Tu has evolved the ability to perform diverse functions on the extracellular surface of a wide variety of pathogenic bacteria. While moonlighting functions vary among microbial species, there is a common theme for roles in adherence and in immune regulation (Harvey et al., 2019). EF-G, which is the third most conserved trGTPase among all domains of life. The elongation phase of translation is an important regulatory node in health and disease (Xu et al., 2021). F0F1 ATP synthase (SMM_0099) has transmembrane proton transport function (Feniouk and Junge, 2005) and can be considered candidate antigens to minimize cross reaction in the diagnosis of brucellosis and useful sources for Brucella vaccine development (Ko et al., 2012). Only one lipoprotein (SMM_0101) was screened, lipoprotein is a surface-exposed molecule that is immunodominant and is used as the major antigen for serological diagnosis of *Mycoplasma penetrans* infection (Sasaki et al., 2002). RelA/SpoT family protein (SMM_0491) is considered as enzymes, which control bacterial physiology through synthesis and degradation of the nucleotide alarmone, and acts as toxins of toxin–antitoxin modules (Kurata et al., 2021). Furthermore, putative copper homeostasis protein CutC (SMM_0941) was found in the antigenic protein. Numerous exciting studies have revealed that copper plays an indispensable role in the microbial pathogen-host axis for entities ranging from pathogenic bacteria to deadly fungal species (Li et al., 2019). Fructose 1,6-bisphosphate aldolase is a ubiquitous cytosolic enzyme that catalyzes the fourth step of glycolysis. Apart from their conserved role in carbohydrate metabolism, aldolases have been reported to perform numerous non-enzymatic functions. Fructose-1,6-bisphosphate aldolase of *M. bovis* is a plasminogen-binding adhesion (Gao et al., 2018).

Discussion

Comparative genome analysis

We analyzed the genome of *S. mirum*, 2 spiroplasmas, 14 mycoplasmas, 4 phytoplasmas, 3 ureaplasmas, and 1 acholeplasma, which revealed that *S. mirum* share a total of 84 common proteins with them (**Supplementary Table 4**). *Bacillus subtilis* as an external group. *M. mycoides*, Mycoplasma *capricolum,* and *Mesoplasma florum* have a closer evolutionary relationship to spiroplasma than other *Mollicutes*, this is consistent with the analysis results of the phylogenetic tree constructed with 16S rRNA (**Figure 2**) (Stülke et al., 2009). We found 684 spiroplasma-specific proteins, including 187 *S. eriocheiris*-specific proteins, 182 *S. mirum*-specific proteins, and 315 common proteins exist in two spiroplasma (**Supplementary Table 5**). No restriction-modification (RM) system was revealed in *Spiroplasma*; thus, the barrier of horizontal gene transfer does not exist. Although the functions of most common proteins are unknown, several adhesion-like lipoprotein may be associated with spiroplasma-specific virulence or host-specific invasion.

From the comparative analysis of the *S. mirum* and *S. eriocheiris* genomes, it can be found that *S. eriocheiris* genome is 232 kb bigger than *S. mirum*, but those two genomes have a highly conserved genome structure and gene sequence (**Figure 5**). 273 CDSs of *S. eriocheiris* genomes did not find homologous sequences in *S. mirum*. These CDSs include two lactose PTS systems and some gene clusters. One of the gene clusters functions as citric acid fermentation, consisting of genes such as CitC, CitDEF, CitG, CitS, and CitX, the sodium citrate transporter (CitS) is responsible for the upregulation of citric acid in anaerobic conditions (van der Rest et al., 1992). Citric acid lyase (CitDEF) catalyzes the decomposition of citric acid into acetic acid and oxaloacetic acid. However, oxaloacetic acid cannot further become pyruvate, malic acid, or isocitrate, so citric acid cannot be used as a carbon source and energy source. It is worth to notice that this gene cluster has not been found in the other *Mollicutes*, which means that *S. eriocheiris* may have some special characteristics. Glucuronic acid can be synthesized in *S. eriocheiris*, but it cannot be synthesized in *S. mirum* because of the lack of D-mannose redox enzyme. Another gene cluster catalyzes the first four steps of uracil nucleoside synthesis, from carbamoyl phosphate to lactate 5-phosphate, which is also present in the *M. penetrans* (hf-2) genome (Sasaki et al., 2002). A total of 129 CDSs of *S. mirum* did not find homologous sequences in *S. eriocheiris* genome, and the function of these specific proteins is still unknown.

Phosphotransferase system

Phosphotransferase system (PTS system) in bacteria consists of two parts: Histidine-containing protein (HPr) that can phosphorylate all carbohydrates and enzyme I. Enzyme II can only catalyze specific substrates, which are composed of 3 functional subunits AII, IIB, IIC, and IID only exist in mannose. Therefore, enzyme II can be divided into four categories: glucose, lactose, mannitol, and mannose. There are 9 categories of PTS systems in the genome of *S. eriocheiris*, including 3 lactose PTS system and 3 fructose PTS system. While in the *S. mirum* genome, most of the genes of the PTS system

TABLE 2 Immunogenic proteins of *Spiroplasma mirum* identified by anti-*S. mirum* serum and negative serum.

SMM-positive	Gene annotation	SMM-negative
SMM0024	Hypothetical protein	
SMM0033	Single-strand DNA-binding protein	
SMM0064	Translation elongation factor G	
SMM0065	Translation elongation factor Tu	
SMM0066	LemA family protein	SMM0066
SMM0099	F_0F_1 ATP synthase subunit epsilon	
SMM0101	Putative lipoprotein	SMM0101
SMM0146	Oligoendopeptidase F	
SMM0156	Heat-inducible transcription repressor HrcA	SMM0156
SMM0159	Molecular chaperone DnaJ	
SMM0170	Arginine deiminase	
SMM0171	Ornithine carbamoyltransferase	SMM0171
SMM0172	Arginine-ornithine antiporter	SMM0172
SMM0233	Adenylosuccinate lyase	
SMM0239	Hypothetical protein	
SMM0299	50S ribosomal protein L17	
SMM0315	Asparagine-tRNA ligase	
SMM0324	Putative lipoate-protein ligase A	SMM0324
SMM0326	Pyruvate dehydrogenase E1 component beta subunit	
SMM0327	Pyruvate dehydrogenase E2 (dihydrolipoamide acetyltransferase) component	
SMM0328	Pyruvate dehydrogenase E3 (dihydrolipoamide dehydrogenase) component	
SMM0388	ABC-type transport system ATP-binding protein	
SMM0422	Endoribonuclease L-PSP	
SMM0437	Hypothetical protein	
	Hypothetical protein	SMM0444
SMM0491	RelA/SpoT family protein	
SMM0532	Truncated transmembrane protein	
SMM0535	Hypothetical protein	
SMM0564	Hypothetical protein	
SMM0585	Hypothetical protein	
SMM0598	hypothetical protein	
SMM0640	Conserved hypothetical protein	
SMM0643	Hypothetical protein	
SMM0693	6-phosphofructokinase	
SMM0696	30S ribosomal protein S2	SMM0696
SMM0834	N-terminal truncated NADH-dependent flavin oxidoreductase	SMM0834
SMM0915	Truncated aldose 1-epimerase	
SMM0928	N-terminal truncated F0F1 ATP synthase subunit C	
SMM0940	Putative phosphate acetyltransferase	
SMM0941	Putative copper homeostasis protein CutC	
SMM1003	Hypothetical protein	
SMM1014	Conserved hypothetical protein	
SMM1020	Putative thioredoxin reductase	
SMM1051	Putative tRNA-binding domain-containing protein	
SMM1054	Putative DegV family protein	
SMM1077	Hypothetical protein	
SMM1121	Lactose phosphotransferase system repressor	
SMM1158	Hypothetical protein	
SMM1170	Fructose-bisphosphate aldolase	
	Cell shape determining protein MreB5	SMM1188
SMM1189	Putative deoxyribonuclease	SMM1189

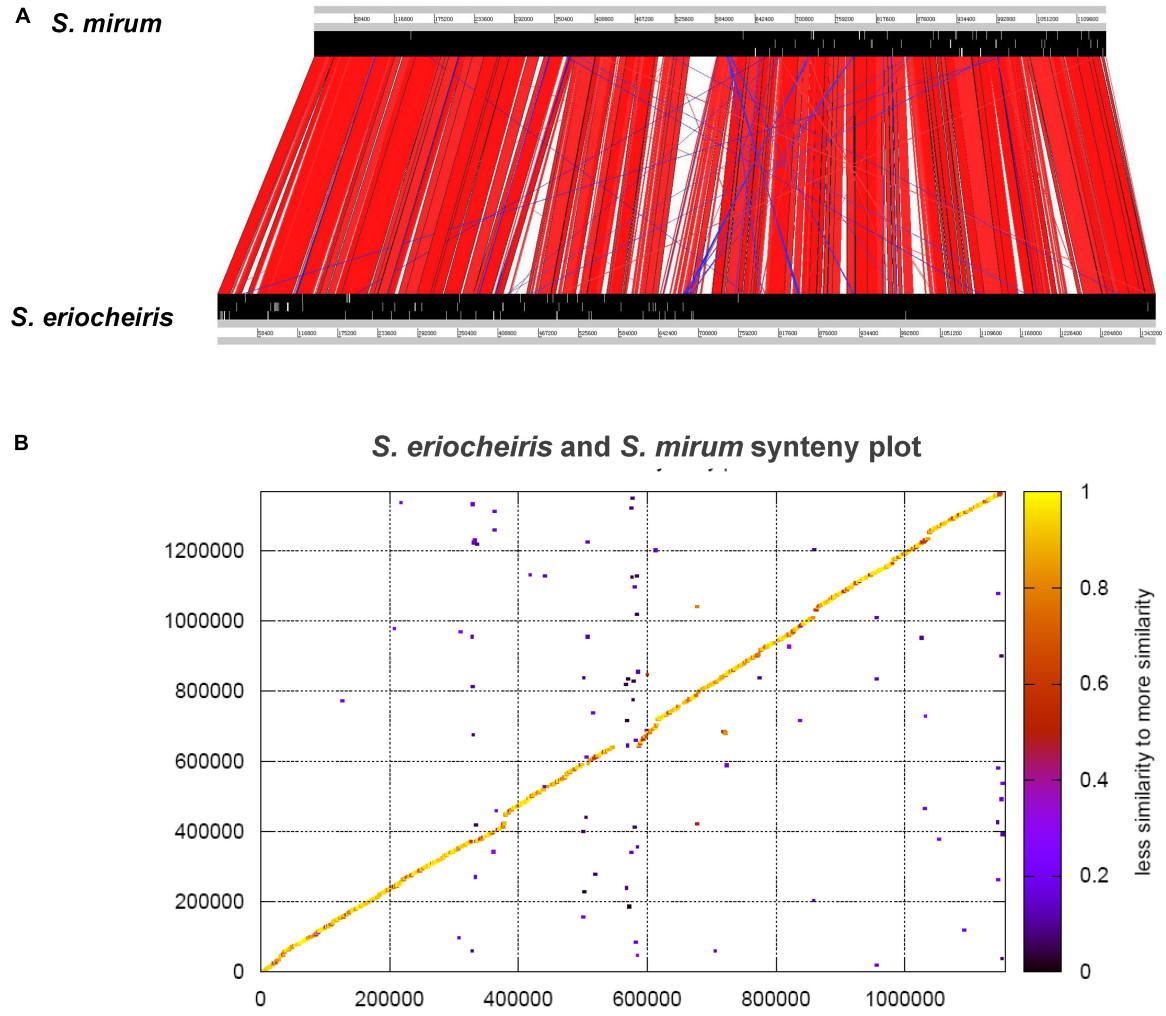


FIGURE 5
(A) Comparison of the gene order between *Spiroplasma eriocheiris* and *Spiroplasma mirum*. **(B)** Protein synteny plot of *S. eriocheiris* and *S. mirum*.

have been lacking, and only 9 of the 23 genes may have functions except for one glucose and one phage fructose PTS system (**Supplementary Table 6**). Both fructose and glucose PTS systems are encoded by single genes in both genomes.

ATP-binding cassette transport system

Because spiroplasma lacks many metabolic and substance synthesis pathways, most of the nutrients they need are obtained from host or artificial medium. The need for a lot of exotic nutrients means that spiroplasma must have a lot of transfer systems. But from the genome, we find that the number of genes in the ATP-binding cassette transport system (ABC transport system) of these two spiroplasma and their proportion to the total number of genes is much less than those of other bacteria, and many are broken genes or pseudogenes. ABC transporters in the genomes of *S. eriocheiris* and *S. mirum* are shown in **Supplementary Table 7**. Five proteins in *S. eriocheiris* (SPE_0075, SPE_1062, SPE_1069, SPE_1074, and SPE_1075) and one protein in *S. mirum* (SMM_0056) are similar to the solute-binding protein of an ABC transporter in *S. citri*. It has been reported that the ability of *S. citri* to be transmitted by *Circulifer haematoceps* is clearly lost by disruption of a gene encoding a putative solute-binding protein of an ABC transporter and restored by the addition of this gene (Boutareaud et al., 2004). The ATP copper ion transporter in *S. eriocheiris* may be related to the specific infection of mitten crab, because the concentration of copper ion in hemocytes of Chinese mitten crab is particularly high. Moreover *S. eriocheiris* there are more sodium ion-related transporters than *S. mirum*, which is consistent with *S. eriocheiris* physiological characteristics such as high salinity

tolerance, because Chinese mitten crab lives in seawater (high salinity) during spawning period.

Secretion system

The secretory system plays a very important role in the excretion of toxins and pathogenicity in bacteria. The spiroplasma genome only has Sec secretory pathway, which consists of SecA, SecE, SecDF, SecG, SecY, and YidC (**Supplementary Table 8**). These proteins also constitute the simplest Sec secretion pathway in spiroplasma, although the Sec secretion pathway of spiroplasma is incomplete when compared to the secretion pathway in *Escherichia coli*, these proteins are consistent with the minimal required set of secretion and translocation proteins. The genome analysis indicates that the composition of Sec secretion pathway of spiroplasma is very similar to that of mycoplasma, but the important feature is the lack of SecB proteins in spiroplasma. It is speculated that there are other molecular chaperones to perform SecB functions, like *Bacillus subtilis*. The other special thing is that most of the *Mollicutes* have double-functional SecDF proteins. It is located on the membrane and pumps secretory proteins out of the membrane by hydrolysis ATP with SecA and SecY support. Because the spiroplasma have very special motility, the contribution of Sec system in spiroplasma motility is not as great as mycoplasma.

Signal peptidase is divided into signal peptidase I and signal peptidase II, which are responsible for disconnecting signal peptides at the N terminal of protein, but we only found signal peptidase II in the spiroplasma genome, the same as other *Mollicutes* members. Signal peptidase is considered to play an important role in the pathogenesis of *Mollicutes*, because it plays a crucial role in the formation of adhesion proteins in *M. hyopneumoniae* and *M. pneumoniae*; 37 and 15 extracellular proteins were found in *S. eriocheiris* and *S. mirum*, respectively. These proteins may be secreted by the Sec system, but the function of most proteins is unknown. Besides the hypothetical proteins and transporters, endonuclease I and GDSL family lipases have been found in *S. eriocheiris* genome and these two proteins may play a role in host invading. Verotoxin, a known functional extramembrane protein, which is an important toxin in bacteria, was found in *S. mirum*.

Outer membrane proteins and virulence proteins

There are the *S. mirum* contains 243 membrane proteins, 51 lipoproteins, and 15 extra-membrane proteins, while 282 membrane proteins, 88 lipoproteins, and 37 extra-membrane proteins in the genome of *S. eriocheiris*. Lipoproteins are usually located on the surface of the cell membrane of Gram-positive bacteria, but because spiroplasma evolved from clostridium, and has no cell wall. Lipoproteins located in the outer membrane of spiroplasma are considered to be involved in the colonization of the hosts, ensuring vertical transmission (Paredes et al., 2015). While only one lipoprotein (SMM0101) was identified by antigen proteomics. This result may imply that not all of the lipoprotein is immunogenic. Spiralin is the most studied protein in spiroplasma, and it is also a unique protein in spiroplasma. There are two Spiralin protein (SPE0615 and SPE1172) in the genome of *S. eriocheiris*, while only 87 bp fragments at the C terminal of one Spiralin protein exist in the *S. mirum*. The similarity of Spiralin protein between *S. mirum, S. citri*, *Spiroplasma kunkelii*, *Spiroplasma melliferum* and *Spiroplasma phoenicerum* is low, which may be caused by the difference of host and parasitic environment, and the Spiralin protein has polymorphism.

The Pre-lipoprotein acetyltransferase is a membrane lipoprotein, which can catalyze the first step of lipoprotein synthesis, transfer the acetyl group to the N terminal of the protein, and connect with cysteine to form lipoprotein. Prolipoprotein diacylglyceryl transferase (Lgt) and lipoprotein signal peptidase II (Lsp II) are unique enzymes (Hutchings et al., 2009) and associated with lipoprotein metabolism in prokaryotes. Alterations in the expression of pre-lipoprotein acetyltransferase and signal peptidase can cause changes in virulence, transport, and signaling systems. There are two Lgt and one Lsp II encoding genes in both genomes of *S. eriocheiris* and *S. mirum* (**Supplementary Table 9**).

According to the genome analysis, there are two kinds of adhesion proteins (SARPs) (SPE0025 and SPE0529) in *S. eriocheiris* (**Supplementary Table 9**). Adhesion proteins play a very important role in the interaction between spiroplasma and insect cells (Yu et al., 2000). The transposons in *S. mirum* contain five adhesion-like proteins (SMM_0578-0581 and SMM_0583; **Supplementary Table 9**), these proteins are homologous to P58, P123, P54, and P18 in *S. citri*, and these proteins are specific in spiroplasma (Bai and Hogenhout, 2002). The ability of *S. citri* to infect leafhoppers depends on a protein, identified as a substrate-binding protein for ABC transporters (Boutareaud et al., 2004). There are 5 Substrate-binding proteins (SPE0075, SPE 1062, SPE 1069, SPE 1074, and SPE 1075) and 1 ABC transporters homologous (SMM0056) present in *S. eriocheiris* and *S. mirum*. Those proteins may promote the invasion process.

Endo-beta-N-acetylglucosaminidase (SPE_0847 and SMM_0801) was shown to be a pathogenicity determinant of mammal pathogens and depressed immune response and interfered with host defense (Valisena et al., 1991). SPE_0813 and SMM_0767 encoded a truncated toxin protein, which is lethal to mice and toxic to Vero cells (Amimoto et al., 2007). SPE_0973 encodes a truncated toxin A, which is *S. eriocheiris* specific and is homologous with the N-terminal of toxin A of *Clostridium*

difficile. SMM_0541 encodes an extracellular verocytotoxin 1, which is *S. mirum* specific and shows homology with the shiga toxin 1A subunit (**Supplementary Table 9**). The presence of domains similar to proteins of the incomplete sec secretion system in pathogenic bacteria suggests that *S. mirum* possesses a related translocation system. The Virulence substances may be delivered by the sec secretion system.

FeoB, a high-affinity ferrous iron uptake membrane protein, is present in many bacteria except for Mollicute genomes and contributes considerably to bacterial virulence (Cartron et al., 2006). We identified a FeoB-encoding gene in spiroplasma genomes (SPE_0274 and SMM_0267), which might enable spiroplasmas to compete for iron from the host and contribute to their pathogenicity.

Two copper-transporting P-type ATPases (SPE_1011, SPE_1142) of *S. eriocheiris* might be associated with its pathogenicity and host specificity. It is known that *S. eriocheiris* is transmitted in blood of crabs (Wang et al., 2004), which consists of a large amount of copper cation. NADH peroxidase (SMM_0037) and putative thiol peroxidase (SMM_1137 and SMM_0497) were found in the genome of *S. mirum,* these enzymes may protect *S. mirum* from oxidative. The presence of domains similar to proteins of the sec secretion system in pathogenic bacteria suggests that spiroplasma possesses a related translocation system.

Genetic degradation

Spiroplasma mirum have 383 pseudogenes, which covered 11.7% genome length, 31.6% gene numbers (**Supplementary Table 10**). This indicates that the genome of *S. mirum* is degenerating. In contrast, only 53 pseudogenes were revealed in the *S. eriocheiris* genome, and 23 pseudogenes can be compensated for by other collateral homologous genes. We found 282 genes in the genome of *S. eriocheiris* constitute 90 collateral homologous gene families, while 188 genes in *S. mirum* can form 56 collateral homologous gene families (**Supplementary Table 11**). The complete collateral homologous genes of 96 pseudogenes in *S. mirum* were found. This means that other pseudogenes are not necessary for *S. mirum*. The main function of pseudogenes is related to carbohydrate transport and metabolism in *S. mirum*. The encoding density of *S. mirum* genome is 79.2%, which is much lower than that of other bacteria. A large amount of non-coding DNA may be caused by genetic degradation. The GC content of the non-coding regions in *S. eriocheiris* and *S. mirum* was 22.9 and 24.9%, respectively, which was lower than the coding region. Pseudogenes may be caused by previously important genes that are not strictly evolved and gradually accumulated mutations (Lawrence et al., 2001). Usually, the high deletion rate of chromosomes will lead to the formation of pseudogenes, and the intracellular bacterium will lead to an increase in deletion rate and mutation rate due to the constant environment. However, the limited function of mismatch repair system and the deletion of homologous recombinant gene may be the main reasons for the increase in pseudogenes. The pathogen has acquired genes from other bacteria despite going through genome reduction, suggesting that isolation has not yet played a major role in this case of genome reduction, with horizontal gene gains still offering a potential route for adaptation (Waterworth et al., 2020).

Horizontal transfer genes

Because there is no restriction modification system in the spiroplasma genome, the hindrance of horizontal transfer gene does not exist. 51 *S. eriocheiris* proteins and 15 *S. mirum* proteins are homologous to non-mollicutes bacteria (**Supplementary Table 12**). These genes may obtain from other bacteria genes after the evolution of spiroplasma. These horizontal transfer genes can be divided into 6 gene clusters in *S. eriocheiris* genome, including 3 PTS systems, 1 citric acid fermentation and 1 glucuronic acid fermentation. The acquisition of some horizontal transfer genes represents some special properties of spiroplasma, such as fructose 1,6-bisphosphate enzyme mentioned in metabolism. In addition, we also found some virulence factors in the horizontal transfer system. Endo-Beta-N-acetylglucosidase has been found in both spiroplasma genome, which is an important pathogenic factor of animal pathogen and can reduce host immune function (Valisena et al., 1991). Both *S. eriocheiris* and *S. mirum* contain this protein consistent with their characteristics as animal pathogen.

One disconnected toxin protein was found in *S. eriocheiris* (SPE_0319) and *S. mirum* (SMM_0306) genomes and which is homologous to the N terminal of the cytotoxic TpeL protein of *Clostridium perfringens*. TpeL protein, a novel virulence factor, plays an important role in mouse lethal processes (Amimoto et al., 2007). Because the N terminal sequence is conserved in many clostridium toxins, we speculate that TpeL protein may be an important virulence factor of spiroplasma. A disconnected toxin A protein is also present in the genome of *S. eriocheiris*, which is homologous to the N terminal of *Clostridium difficile* toxin A protein. As discussed in Sec system, the presence of Verotoxin I (SMM_0541) in the genome of *S. mirum*, which is homologous to the virulence factors of *Shigella dysenteriae* and Shigatoxigenic *E. coli*, named shigella toxin IA subunit. Ferritin FeoB is a very important virulence factor in many bacteria. SPE0274 and SMM0267 can encode complete FeoB, which may be helpful for spiroplasma to despoil ions in host and produce virulence. These horizontal transfer genes do not exist in other *Mollicutes* members but are important virulence factors in some *Clostridium* and *bacilli*. This reveals that the pathogenic characteristics of spiroplasma are similar to intestinal pathogens such as *Clostridium* or *bacillus*.

The *S. eriocheiris* and *S. mirum* genomics research reveals the commonness and difference in spiroplasma in aquatic environment and terrestrial environment. In an evolutionary sense, we can continue to study the evolutionary origin and classification of cell wall-less microbes in the *Mollicutes*, as well as their differences and evolutionary relationships with other bacteria. Further genomic analysis of two spiroplasma may identify factors associated with their special living environment (parasitic history) and pathogenicity due to aquatic and terrestrial habitat differences.

Antigen proteomics

Antigen proteomics is a new research field of pathogen in recent years. At present, there is still less study of spiroplasma antigen proteomics, and the study of mycoplasma antigen proteomics has just begun. The current research on mycoplasma has *M. genitalium*, *M. pneumoniae*, *M. hyopneumoniae*, and *M. synoviae*. *S. mirum* belongs to serological group V, while *S. eriocheiris* belongs to a new *Spiroplasma* group XLIII (Wang et al., 2011). The two species have very different hosts and environments. So, it is not surprising that these two strains have very different cell surface proteins and antigen-related genes. We used immunoprecipitation to screen the antigenic proteins of *S. mirum*. Negative serum was used as a control to determine the main antigen protein. We've screened out 40 antigen proteins of *S. mirum* were screened out, six of them are homologous proteins that can be used as antigen proteins of *S. eriocheiris*, including extension factor EF-G, pyruvate dehydrogenase (β subunit, E2 subunit, and E3 subunit), thioredoxin reductase, and DegV family proteins. Elongation factor thermo unstable (EF-Tu) and pyruvate dehydrogenase E1-β subunit are also *M. hyopneumoniae* antigen proteins (Pinto et al., 2007). The EF-Tu is also the antigen protein of *M. synoviae* (Bercic et al., 2008). Mmm SC PG1 of *M. mycoides* shared with extraordinary EF-Tu, extension factors EF-G, pyruvate dehydrogenase, fatty acids protein ligase A and fructose diphosphate aldolase as antigen proteins (Jores et al., 2009), *M. mycoides* subsp. mycoides strain B237 shared pyruvate dehydrogenase as antigen protein (Naseem et al., 2010). In addition, translation elongation factor G is likely to be a drug target for the malarial parasite. Therefore, the EF-G could be evaluated as potential proteins of diagnostic markers and target proteins of research on pathogenicity (Liu Y. et al., 2017).

Previously, we have screened out 45 antigen proteins of *S. eriocheiris* (Liu Y. et al., 2017), many of which are homologous to proteins found in mycoplasma antigen proteomics, such as pyruvate dehydrogenase E1-β subunit is also *M. pneumoniae* antigen protein (Pinto et al., 2007); enolases, ATP synthetases β subunits, pyruvate kinases, and DanK can act as antigen proteins of *M. synoviae* (Bercic et al., 2008). *Mycoplasma*, which has the closest evolutionary relationship with spiroplasma, has been studied in immunoproteomics. Jores et al. screened the immunogenic protein of *M. mycoides* subsp. mycoides small clonal strain MmmSC PG1T and used experiments. The serum prepared in the laboratory screened out 24 immunogenic proteins. 11 of these 24 antigenic proteins were homologous to the antigenic proteins of *S. eriocheiris* sinensis. They are elongation factor-G, leucyl peptidase, Pyruvate dehydrogenase (α subunit, β subunit, E2 subunit, and E3 subunit), acetate kinase, glyceraldehyde 3-phosphate dehydrogenase, triose phosphate isomerase, phosphoglycerate mutase, and ATP synthase β subunit and others (Jores et al., 2009). Naseem et al. screened 22 immunogenic proteins in the small clonal strain B237 of *M. mycoides* subsp. mycoides, including DNA-guided RNA polymerase proteins such as β subunit and acyl carrier protein phosphodiesterase, are also antigen proteins of *S. eriocheiris* (Naseem et al., 2010). The other antigenic proteins specific to the microbes of the class lamina are mostly lipoproteins or membrane proteins with variable morphology and structure. A putative copper homeostasis protein CutC (SMM0941) was identified in the antigenic protein. The bacterial copper efflux system plays a predominant role in regulating pathogen fitness during infection. Analyses of copper homeostasis in bacteria and fungi extensively demonstrate that copper is utilized by the host immune system as an anti-microbial agent. The expression of copper efflux and detoxification from microbial pathogens is induced to counteract the host's copper bombardment, which in turn disrupts these machineries, resulting in the attenuation of microbial survival in host tissue (Li et al., 2019).

Conclusion

Now we have entered the post-genome era because there are no technical obstacles to genome sequencing, so proteomics (including immunoproteomics and antigenic histology) has also entered a period of rapid development. Regarding pathogenic microorganisms, people are most concerned about which proteins of microorganisms play a major role in the pathogenic process, so it is of great significance for us to carry out this antigen-screening experiment. However, we only preliminarily screened the antigen proteins of spiroplasma, and the specific functions of each protein need to be further studied.

Spiroplasma was attractive for research due to their very small, helical structure and unusual mode of motility by means of a contractile cytoskeleton-like structure, functioning as a linear motor (Trachtenberg, 2006; Trachtenberg et al., 2008). Some structural, dynamic, and proteomic approaches have elucidated the basic organizing principles of their minimal, yet functional, cytoskeleton-like architecture (Liu P. et al., 2017). The full genome information of spiroplasma should provide further significant opportunities to understand the pathogenesis and mechanism of this unique cytoskeleton-like architecture. The comparative genomic analysis of *S. eriocheiris* and *S. mirum*

reveals some remarkable differences in genes related to their special characteristics associated with their living environments. The complete genome sequence of *S. mirum* from rabbit tick reveals that Glycolysis, pentose phosphate pathway, arginine metabolism, nucleotide biosynthesis, and citrate fermentation were found in *S. mirum*. Trichloroacetic acid, fatty acids metabolism, phospholipid biosynthesis, terpenoid biosynthesis, and cofactors synthesis were completely absent. But lactose-specific PTS was not identified in *S. mirum*. A total of 273 *S. eriocheiris* CDSs had no homologous in *S. mirum* genome, including lactose-specific PTS and citrate fermentation gene cluster. *S. mirum* genome had 381 pseudogenes, accounting for 31.6% of total protein-coding genes. This is the evidence that spiroplasma genome is under an ongoing genome reduction. Interestingly, *S. mirum* was infective for suckling mice and produced cataracts. But we could not find genes associated with this character up to now. However, further analyses and experiments based on the genomic data should be helpful to probe the mechanisms of *S. mirum* infections in vertebrate hosts.

We identified the antigen proteins and antigen membrane proteins of *S. mirum*. 40 proteins in *S. mirum* are identified in relation to virulence. The results reported in this study will elucidate the immune relationship between the host and the pathogen *S. mirum*, as well as benefit analytical techniques for identifying bioindicators and more precise diagnosis of the pathology. Currently, there is no explicit understanding of the exact role of these potential immunogenic proteins; future experiments should reveal the immunogenicity of these immunogenic proteins, and the interaction between these proteins and host proteins can be identified and proved.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary material**.

Author contributions

PL, YXL, YY, JC, WS, WG (Antigen proteome analysis parts), RL, QZ, YAL, WW, QM, JO, and ZY (Genome part) contributed to the conception and design. PL, YXL (Antigen proteome analysis parts), and YY (Genome part) contributed to the collection and assembly of data. PL, YXL, and YY wrote the manuscript. PL, WS, and WG approved the final manuscript. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the Natural Science Foundation of Hunan Province, China (2019JJ50493); Scientific Research Foundation of Hunan Provincial Education Department, China (19B489); Research Foundation of University of South China (190XQD015); and Hunan Provincial College Students' Innovation and Entrepreneurship Training Program (S202210555220 and 202210555076).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2022.996938/full#supplementary-material

References

Amimoto, K., Noro, T., Oishi, E., and Shimizu, M. (2007). A novel toxin homologous to large clostridial cytotoxins found in culture supernatant of *Clostridium perfringens* type C. *Microbiology* 153(Pt 4), 1198–1206. doi: 10.1099/mic.0.2006/002287-0

Aslam, B., Basit, M., Nisar, M. A., Khurshid, M., and Rasool, M. H. (2017). Proteomics: Technologies and their applications. *J. Chromatogr. Sci.* 55, 182–196.

Bai, X., and Hogenhout, S. A. (2002). A genome sequence survey of the mollicute corn stunt spiroplasma *Spiroplasma kunkelii*. *FEMS Microbiol. Lett.* 210, 7–17. doi: 10.1111/j.1574-6968.2002.tb11153.x

Bastian, F. O., Sanders, D. E., Forbes, W. A., Hagius, S. D., Walker, J. V., Henk, W. G., et al. (2007). Spiroplasma spp. from transmissible spongiform encephalopathy brains or ticks induce spongiform encephalopathy in ruminants. *J. Med. Microbiol.* 56(Pt 9), 1235–1242.

Bercic, R. L., Slavec, B., Lavric, M., Narat, M., Bidovec, A., Dovc, P., et al. (2008). Identification of major immunogenic proteins of *Mycoplasma synoviae* isolates. *Vet. Microbiol.* 127, 147–154. doi: 10.1016/j.vetmic.2007.07.020

Boutareaud, A., Danet, J. L., Garnier, M., and Saillard, C. (2004). Disruption of a gene predicted to encode a solute binding protein of an ABC transporter reduces transmission of *Spiroplasma citri* by the leafhopper *Circulifer haematoceps*. *Appl. Environ. Microbiol.* 70, 3960–3967. doi: 10.1128/AEM.70.7.3960-3967.2004

Campbell, M. A., Van Leuven, J. T., Meister, R. C., Carey, K. M., Simon, C., and McCutcheon, J. P. (2015). Genome expansion via lineage splitting and genome reduction in the cicada endosymbiont Hodgkinia. *Proc. Natl. Acad. Sci. U.S.A.* 112, 10192–10199. doi: 10.1073/pnas.1421386112

Cartron, M., Maddocks, S., Gillingham, P., Craven, C., and Andrews, S. (2006). Feo - transport of ferrous iron into bacteria. *Biometals* 19, 143–157.

Carver, T. J., Rutherford, K. M., Berriman, M., Rajandream, M. A., Barrell, B. G., and Parkhill, J. (2005). ACT: The artemis comparison tool. *Bioinformatics* 21, 3422–3423.

Castresana, J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Biol. Evol.* 17, 540–552.

Chen, S., Hao, H., Yan, X., Liu, Y., and Chu, Y. (2019). Genome-wide analysis of *Mycoplasma dispar* provides insights into putative virulence factors and phylogenetic relationships. *G3* 9, 317–325. doi: 10.1534/g3.118.200941

Delcher, A. L., Harmon, D., Kasif, S., White, O., and Salzberg, S. L. (1999). Improved microbial gene identification with GLIMMER. *Nucleic Acids Res.* 27, 4636–4641.

Edgar, R. (2004). MUSCLE: A multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* 5:113. doi: 10.1186/1471-2105-5-113

Feniouk, B. A., and Junge, W. (2005). Regulation of the F0F1-ATP synthase: The conformation of subunit epsilon might be determined by directionality of subunit gamma rotation. *FEBS Lett.* 579, 5114–5118. doi: 10.1016/j.febslet.2005.08.030

Gao, X., Bao, S., Xing, X., Fu, X., Zhang, Y., Xue, H., et al. (2018). Fructose-1,6-bisphosphate aldolase of *Mycoplasma bovis* is a plasminogen-binding adhesin. *Microb. Pathog.* 124, 230–237. doi: 10.1016/j.micpath.2018.08.032

Guindon, S., and Gascuel, O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* 52, 696–704.

Guo, F. B., Ou, H. Y., and Zhang, C. T. (2003). ZCURVE: A new system for recognizing protein-coding genes in bacterial and archaeal genomes. *Nucleic Acids Res.* 31, 1780–1789. doi: 10.1093/nar/gkg254

Harne, S., Duret, S., Pande, V., Bapat, M., Béven, L., and Gayathri, P. (2020). MreB5 is a determinant of rod-to-helical transition in the cell-wall-less bacterium spiroplasma. *Curr. Biol.* 30, 4753–4762.e7. doi: 10.1016/j.cub.2020.08.093

Harvey, K. L., Jarocki, V. M. I, Charles, G., and Djordjevic, S. P. (2019). The diverse functional roles of elongation factor Tu (EF-Tu) in microbial pathogenesis. *Front. Microbiol.* 10:2351. doi: 10.3389/fmicb.2019.02351

Hou, L., Gu, W., Zhu, H., Yao, W., Wang, W., and Meng, Q. (2017). *Spiroplasma eriocheiris* induces mouse 3T6-Swiss albino cell apoptosis that associated with the infection mechanism. *Mol. Immunol.* 91, 75–85. doi: 10.1016/j.molimm.2017.08.002

Hutchings, M. I., Palmer, T., Harrington, D. J., and Sutcliffe, I. C. (2009). Lipoprotein biogenesis in Gram-positive bacteria: Knowing when to hold 'em, knowing when to fold 'em. *Trends Microbiol.* 17, 13–21. doi: 10.1016/j.tim.2008.10.001

Itoh, K. I, Pan, J., and Koshimizu, K. (1989). A proposed life cycle model of *Spiroplasma mirum* based on scanning electron microscopical observations of growth in liquid culture. *Microbiol. Immunol.* 33, 821–832. doi: 10.1111/j.1348-0421.1989.tb00968.x

Jarocki, V. M., Raymond, B. B. A., Tacchi, J. L., Padula, M. P., and Djordjevic, S. P. (2019). *Mycoplasma hyopneumoniae* surface-associated proteases cleave bradykinin, substance P, neurokinin A and neuropeptide Y. *Sci. Rep.* 9:14585. doi: 10.1038/s41598-019-51116-w

Jores, J., Meens, J., Buettner, F. F., Linz, B., Naessens, J., and Gerlach, G. F. (2009). Analysis of the immunoproteome of *Mycoplasma mycoides* subsp. *mycoides* small colony type reveals immunogenic homologues to other known virulence traits in related *Mycoplasma* species. *Vet. Immunol. Immunopathol.* 131, 238–245. doi: 10.1016/j.vetimm.2009.04.016

Juncker, A. S., Willenbrock, H., Von Heijne, G., Brunak, S., Nielsen, H., and Krogh, A. (2003). Prediction of lipoprotein signal peptides in Gram-negative bacteria. *Protein Sci.* 12, 1652–1662.

Kanehisa, M., Goto, S., Kawashima, S., Okuno, Y., and Hattori, M. (2004). The KEGG resource for deciphering the genome. *Nucleic Acids Res.* 32, D277–D280.

Khan, F. A., Zhao, G., Guo, Y., Faisal, M., Chao, J., Chen, X., et al. (2018). Proteomics identification and characterization of MbovP730 as a potential DIVA antigen of *Mycoplasma bovis*. *Oncotarget* 9, 28322–28336. doi: 10.18632/oncotarget.22265

Ko, K. Y., Kim, J. W., Her, M., Kang, S. I., Jung, S. C., Cho, D. H., et al. (2012). Immunogenic proteins of *Brucella abortus* to minimize cross reactions in brucellosis diagnosis. *Vet. Microbiol.* 156, 374–380. doi: 10.1016/j.vetmic.2011.11.011

Kurata, T., Brodiazhenko, T., Alves Oliveira, S. R., Roghanian, M., Sakaguchi, Y., Turnbull, K. J., et al. (2021). RelA-SpoT Homolog toxins pyrophosphorylate the CCA end of tRNA to inhibit protein synthesis. *Mol. Cell* 81, 3160–3170.e9. doi: 10.1016/j.molcel.2021.06.005

Kürner, J., Frangakis, A. S., and Baumeister, W. (2005). Cryo-electron tomography reveals the cytoskeletal structure of *Spiroplasma melliferum*. *Science* 307, 436–438. doi: 10.1126/science.1104031

Lawrence, J. G., Hendrix, R. W., and Casjens, S. (2001). Where are the pseudogenes in bacterial genomes? *Trends Microbiol.* 9, 535–540.

Li, C., Li, Y., and Ding, C. (2019). The role of copper homeostasis at the host-pathogen axis: From bacteria to fungi. *Int. J. Mol. Sci.* 20:175.

Liu, P., Du, J., Zhang, J., Wang, J., Gu, W., Wang, W., et al. (2018). The structural and proteomic analysis of *Spiroplasma eriocheiris* in response to colchicine. *Sci. Rep.* 8:8577. doi: 10.1038/s41598-018-26614-y

Liu, P., Hou, L., Liu, M., Xu, X., Gao, Q., Deng, J., et al. (2020). Phosphoproteomic analysis of *Spiroplasma eriocheiris* and crosstalk with acetylome reveals the role of post-translational modifications in metabolism. *Curr. Proteomics* 17, 392–403.

Liu, P., Zheng, H., Meng, Q., Terahara, N., Gu, W., Wang, S., et al. (2017). Chemotaxis without conventional two-component system, based on cell polarity and aerobic conditions in helicity-switching swimming of *Spiroplasma eriocheiris*. *Front. Microbiol.* 8:58. doi: 10.3389/fmicb.2017.00058

Liu, Y., Xu, Y., Li, S., Xu, X., Gao, Q., Yuan, M., et al. (2017). Identification of proteome, antigen protein and antigen membrane protein from *Spiroplasma eriocheiris*. *Lett. Appl. Microbiol.* 65, 395–402. doi: 10.1111/lam.12784

Lo, W. S., Chen, L. L., Chung, W. C., Gasparich, G. E., and Kuo, C. H. (2013). Comparative genome analysis of *Spiroplasma melliferum* IPMB4A, a honeybee-associated bacterium. *BMC Genomics* 14:22. doi: 10.1186/1471-2164-14-22

Lo, W. S., Gasparich, G. E., and Kuo, C. H. (2015). Found and lost: The fates of horizontally acquired genes in arthropod-symbiotic spiroplasma. *Genome Biol. Evol.* 7, 2458–2472. doi: 10.1093/gbe/evv160

Lowe, T. M., and Eddy, S. R. (1997). tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* 25, 955–964.

McCutcheon, J. P., and Moran, N. A. (2011). Extreme genome reduction in symbiotic bacteria. *Nat. Rev. Microbiol.* 10, 13–26.

Megraud, F., Gamon, L. B., and McGarrity, G. J. (1983). Characterization of *Spiroplasma mirum* (suckling mouse cataract agent) in a rabbit lens cell culture. *Infect. Immun.* 42, 1168–1175. doi: 10.1128/iai.42.3.1168-1175.1983

Naseem, S., Meens, J., Jores, J., Heller, M., Dübel, S., Hust, M., et al. (2010). Phage display-based identification and potential diagnostic application of novel antigens from *Mycoplasma mycoides* subsp. *mycoides* small colony type. *Vet. Microbiol.* 142, 285–292. doi: 10.1016/j.vetmic.2009.09.071

Nunan, L. M., Lightner, D. V., Oduori, M. A., and Gasparich, G. E. (2005). *Spiroplasma penaei* sp. nov., associated with mortalities in *Penaeus vannamei*, Pacific white shrimp. *Int. J. Syst. Evol. Microbiol.* 55(Pt 6), 2317–2322. doi: 10.1099/ijs.0.63555-0

Paredes, J. C., Herren, J. K., Schüpfer, F., Marin, R., Claverol, S., Kuo, C. H., et al. (2015). Genome sequence of the *Drosophila melanogaster* male-killing *Spiroplasma* strain MSRO endosymbiont. *mBio* 6:e02437-14. doi: 10.1128/mBio.02437-14

Pinto, P. M., Chemale, G., de Castro, L. A., Costa, A. P., Kich, J. D., Vainstein, M. H., et al. (2007). Proteomic survey of the pathogenic *Mycoplasma hyopneumoniae* strain 7448 and identification of novel post-translationally modified and antigenic proteins. *Vet. Microbiol.* 121, 83–93. doi: 10.1016/j.vetmic.2006.11.018

Pollack, J. D., Myers, M. A., Dandekar, T., and Herrmann, R. (2002). Suspected utility of enzymes with multiple activities in the small genome *Mycoplasma* species: The replacement of the missing household nucleoside diphosphate kinase gene and activity by glycolytic kinases. *Omics* 6, 247–258. doi: 10.1089/15362310260256909

Rasko, D. A., Myers, G. S., and Ravel, J. (2005). Visualization of comparative genomic analyses by BLAST score ratio. *BMC Bioinformatics* 6:2. doi: 10.1186/1471-2105-6-2

Sasajima, Y., and Miyata, M. (2021). Prospects for the mechanism of spiroplasma swimming. *Front. Microbiol.* 12:706426. doi: 10.3389/fmicb.2021.706426

Sasaki, Y., Ishikawa, J., Yamashita, A., Oshima, K., Kenri, T., Furuya, K., et al. (2002). The complete genomic sequence of *Mycoplasma penetrans*, an intracellular bacterial pathogen in humans. *Nucleic Acids Res.* 30, 5293–5300. doi: 10.1093/nar/gkf667

Shaevitz, J., Lee, J., and Fletcher, D. (2005). *Spiroplasma* swim by a processive change in body helicity. *Cell* 122, 941–945. doi: 10.1016/j.cell.2005.07.004

Stülke, J., Eilers, H., and Schmidl, S. R. (2009). "Mycoplasma and spiroplasma," in *Encyclopedia of microbiology*, 3rd Edn, ed. M. Schaechter (Oxford: Elsevier), 208–219.

Sussman, H. E. (2015). 20 years of genome research. Preface. *Genome Res.* 25:xv. doi: 10.1101/gr.199026.115

Takahashi, D., Fujiwara, I., and Miyata, M. (2020). Phylogenetic origin and sequence features of MreB from the wall-less swimming bacteria Spiroplasma. *Biochem. Biophys. Res. Commun.* 533, 638–644. doi: 10.1016/j.bbrc.2020.09.060

Trachtenberg, S. (2004). Shaping and moving a spiroplasma. *J. Mol. Microbiol. Biotechnol.* 7, 78–87.

Trachtenberg, S. (2006). The cytoskeleton of spiroplasma: A complex linear motor. *J. Mol. Microbiol. Biotechnol.* 11, 265–283. doi: 10.1159/000094060

Trachtenberg, S., Dorward, L. M., Speransky, V. V., Jaffe, H., Andrews, S. B., and Leapman, R. D. (2008). Structure of the cytoskeleton of *Spiroplasma melliferum* BC3 and its interactions with the cell membrane. *J. Mol. Biol.* 378, 778–789. doi: 10.1016/j.jmb.2008.02.020

Uchiyama, I. (2003). MBGD: Microbial genome database for comparative analysis. *Nucleic Acids Res.* 31, 58–62.

Valisena, S., Varaldo, P., and Satta, G. (1991). Staphylococcal endo-beta-N-acetylglucosaminidase inhibits response of human lymphocytes to mitogens and interferes with production of antibodies in mice. *J. Clin. Invest.* 87, 1969–1976. doi: 10.1172/JCI115224

van der Rest, M. E., Siewe, R. M., Abee, T., Schwarz, E., Oesterhelt, D., and Konings, W. N. (1992). Nucleotide sequence and functional properties of a sodium-dependent citrate transport system from *Klebsiella pneumoniae*. *J. Biol. Chem.* 267, 8971–8976.

Vandemoortele, G., Gevaert, K., and Eyckerman, S. (2016). Proteomics in the genome engineering era. *Proteomics* 16, 177–187.

Wang, W., Gu, W., Ding, Z., Ren, Y., Chen, J., and Hou, Y. (2005). A novel spiroplasma pathogen causing systemic infection in the crayfish *Procambarus clarkii* (Crustacea: Decapod), in China. *FEMS Microbiol. Lett.* 249, 131–137. doi: 10.1016/j.femsle.2005.06.005

Wang, W., Gu, W., Gasparich, G. E., Bi, K., Ou, J., Meng, Q., et al. (2011). *Spiroplasma eriocheiris* sp. nov., associated with mortality in the Chinese mitten crab, *Eriocheir sinensis*. *Int. J. Syst. Evol. Microbiol.* 61, 703–708. doi: 10.1099/ijs.0.020529-0

Wang, W., Wen, B., Gasparich, G. E., Zhu, N., Rong, L., Chen, J., et al. (2004). A spiroplasma associated with tremor disease in the Chinese mitten crab (*Eriocheir sinensis*). *Microbiology* 150(Pt 9), 3035–3040. doi: 10.1099/mic.0.26664-0

Waterworth, S. C., Flórez, L. V., Rees, E. R., Hertweck, C., Kaltenpoth, M., and Kwan, J. C. (2020). Horizontal gene transfer to a defensive symbiont with a reduced genome in a multipartite beetle microbiome. *mBio* 11:e02430-19. doi: 10.1128/mBio.02430-19

Watts, T., Haselkorn, T. S., Moran, N. A., and Markow, T. A. (2009). Variable incidence of Spiroplasma infections in natural populations of *Drosophila* species. *PLoS One* 4:e5703. doi: 10.1371/journal.pone.0005703

Westberg, J., Persson, A., Holmberg, A., Goesmann, A., Lundeberg, J., Johansson, K. E., et al. (2004). The genome sequence of *Mycoplasma mycoides* subsp. *mycoides* SC type strain PG1T, the causative agent of contagious bovine pleuropneumonia (CBPP). *Genome Res.* 14, 221–227. doi: 10.1101/gr.1673304

Xu, B., Liu, L., and Song, G. (2021). Functions and regulation of translation elongation factors. *Front. Mol. Biosci.* 8:816398. doi: 10.3389/fmolb.2021.816398

Yu, J., Wayadande, A. C., and Fletcher, J. (2000). *Spiroplasma citri* surface protein P89 implicated in adhesion to cells of the vector circulifer tenellus. *Phytopathology* 90, 716–722. doi: 10.1094/PHYTO.2000.90.7.716

frontiers | Frontiers in Microbiology

TYPE Original Research
PUBLISHED 28 November 2022
DOI 10.3389/fmicb.2022.1075199

OPEN ACCESS

EDITED BY
Chih-Horng Kuo,
Academia Sinica, Taiwan

REVIEWED BY
Steve Perlman,
University of Victoria, Canada
Matt Ballinger,
Mississippi State University,
United States

*CORRESPONDENCE
Hiroshi Arai
dazai39papilio@gmail.com
Daisuke Kageyama
kagymad@affrc.go.jp

SPECIALTY SECTION
This article was submitted to Microbial Symbioses, a section of the journal Frontiers in Microbiology

RECEIVED 20 October 2022
ACCEPTED 08 November 2022
PUBLISHED 28 November 2022

CITATION
Arai H, Inoue MN and Kageyama D (2022) Male-killing mechanisms vary between *Spiroplasma* species. *Front. Microbiol.* 13:1075199. doi: 10.3389/fmicb.2022.1075199

Male-killing mechanisms vary between *Spiroplasma* species

Hiroshi Arai[1,2]*, Maki N. Inoue[1] and Daisuke Kageyama[2]*

[1]United Graduate School of Agricultural Science, Tokyo University of Agriculture and Technology, Fuchu, Japan, [2]Institute of Agrobiological Sciences, National Agriculture and Food Research Organization (NARO), Tsukuba, Japan

Male-killing, a male-specific death of arthropod hosts during development, is induced by *Spiroplasma* (Mollicutes) endosymbionts of the Citri–Poulsonii and the Ixodetis groups, which are phylogenetically distant groups. *Spiroplasma poulsonii* induces male-killing in *Drosophila melanogaster* (Diptera) using the Spaid toxin that harbors ankyrin repeats, whereas little is known about the origin and mechanisms of male-killing induced by *Spiroplasma ixodetis*. Here, we analyzed the genome and the biological characteristics of a male-killing *S. ixodetis* strain *s*Hm in the moth *Homona magnanima* (Tortricidae, Lepidoptera). Strain *s*Hm harbored a 2.1 Mb chromosome and two potential plasmids encoding Type IV effectors, putatively involved in virulence and host–symbiont interactions. Moreover, *s*Hm did not harbor the *spaid* gene but harbored 10 ankyrin genes that were homologous to those in other *S. ixodetis* strains. In contrast to the predominant existence of *S. poulsonii* in hemolymph, our quantitative PCR assays revealed a systemic distribution of strain *s*Hm in *H. magnanima*, with particularly high titers in Malpighian tubules but low titers in hemolymph. Furthermore, transinfection assays confirmed that strain *s*Hm can infect cultured cells derived from distantly related insects, namely *Aedes albopictus* (Diptera) and *Bombyx mori* (Lepidoptera). These results suggest different origins and characteristics of *S. ixodetis*- and *S. poulsonii*-induced male-killing.

KEYWORDS

Spiroplasma*, male-killing, symbiosis, evolution, endosymbionts, *Homona magnanima*, *spaid

Introduction

In arthropods, maternally inherited endosymbiotic microbes frequently interact with the hosts in a mutualistic or a parasitic manner. Male-killing (MK), male-specific death in insects during development, is one of the reproductive manipulations induced by various intracellular bacteria, microsporidia, and viruses (Duron et al., 2008; Werren et al., 2008; Kageyama et al., 2012; Fujita et al., 2021). MK leads to the advantage of female siblings and is considered a selfish strategy of the intracellular microbes that promotes their spread and survival in nature (Hurst, 1991; Hurst and Jiggins, 2000;

Hornett et al., 2006). The genus *Spiroplasma* (class: Mollicutes) are the most studied bacteria that induce MK in diverse insects (Anbutsu and Fukatsu, 2011; Lo et al., 2015; Harumoto and Lemaitre, 2018; Binetruy et al., 2019). *Spiroplasma* are small, helical, and motile bacteria that include commensal, pathogenic, and mutualistic species and have a diverse host range, including plants and animals (Regassa and Gasparich, 2006; Duperron et al., 2013; Viver et al., 2017; He et al., 2018). Phylogenetically, the MK *Spiroplasma* strains are clustered into the Citri–Poulsonii group (harbored by *Drosophila* flies and lacewings) (Williamson and Poulson, 1979; Hayashi et al., 2016) and the Ixodetis clade (harbored by ladybugs, butterflies, moths, and aphids) (Hurst et al., 1999; Simon et al., 2011; Tabata et al., 2011; Smith et al., 2016).

The molecular mechanisms underlying *Spiroplasma*-induced MK have been mostly investigated using *S. poulsonii–Drosophila* systems (Harumoto and Lemaitre, 2018). *S. poulsonii* strain MSRO induces MK in *Drosophila melanogaster* by a toxic protein androcidin (Spaid) harboring ankyrin repeats that damage the male X chromosome (Harumoto and Lemaitre, 2018). In contrast, information regarding the mechanism underlying MK induced by the members of the Ixodetis group is limited. The *spaid* gene is conserved among *S. poulsonii* strains (Harumoto and Lemaitre, 2018; Gerth et al., 2021), whereas whether the *S. ixodetis* group uses Spaid as an MK factor is unknown. The genus *Spiroplasma* exhibits high genomic flexibility and dynamic evolution of various toxin loci, such as Spaid and ribosome-inactivating protein (RIP) (Hamilton et al., 2016; Ballinger et al., 2019; Gerth et al., 2021; Massey and Newton, 2022; Pollmann et al., 2022). Gnomic analyses have revealed dynamic *Spiroplasma* evolution driven by bacteriophage lysogenization (Ye et al., 1996; Carle et al., 2010; Ku et al., 2013) and by horizontal gene transfer (Mouches et al., 1984; Joshi et al., 2005). Virulence-associated genes are frequently exchanged between microbes sharing the same niche (Kent and Bordenstein, 2010; Wiedenbeck and Cohan, 2011). Likewise, *Spiroplasma* may have acquired MK genes by horizontal gene transfer because they often coexist with other endosymbionts, such as *Wolbachia* and *Rickettsia*, in the same host (Hurst et al., 1999; Majerus et al., 2000; Watanabe et al., 2012; Hayashi et al., 2016; Takamatsu et al., 2021). However, the horizontal gene transfer to *Spiroplasma* may be constrained by the unusual codon usage by *Spiroplasma* compared with other bacteria (notably, the use of UGA as a tryptophan rather than a stop codon; Lo et al., 2015). Although genomic studies on MK *S. poulsonii* have been done, comparative genomic analyses of other MK *Spiroplasma* species, such as *S. ixodetis*, are essential to infer the origin and evolution of the MK machinery.

In this study, we sequenced the genome of *S. ixodetis* strain *s*Hm that causes MK in the tea tortrix moth *Homona magnanima* (Tortricidae, Lepidoptera). Against the full-genome sequence of strain *s*Hm, we searched for genes encoding Spaid and RIP toxin homologs, as well as putative MK genes of other MK endosymbionts such as *Wolbachia* (Arai et al., 2020, 2022a) and Partiti-like virus Osugoroshivirus (OGVs) (Fujita et al., 2021) in *H. magnanima*. We also examined the propagation characteristics and infectivity of strain *s*Hm using quantification and transinfection assays. Finally, we argue that MK mechanisms and ecological characteristics are substantially different between *Spiroplasma* species.

Abbreviations: CI, cytoplasmic incompatibility; FBS, fetal bovine serum; HTH, helix-turn-helix; JSPS, Japan Society for the Promotion of Science; MK, male-killing; OGV, Osugoroshivirus; OTU, operational taxonomic unit; RIP, ribosome-inactivating protein; WGA, whole genome amplification.

Materials and methods

Rearing and sexing of *Homona magnanima*

To construct *S. ixodetis* sHm genome, we used the laboratory-maintained *Spiroplasma*-positive MK-inducing line (S+ line) of *H. magnanima* (Tsugeno et al., 2017). In the present study, we accidentally obtained a *Spiroplasma*-positive 1:1 sex ratio line (S+M+ line) as a subline of the S+ line. For every generation, the male moths picked up from the 1:1 sex ratio line, which had been confirmed negative for *Spiroplasma*, *Wolbachia*, and OGV (NSR line) (Takamatsu et al., 2021), were crossed with the female moths of the S+ and S+M+ lines as described by Arai et al. (2022a). The obtained larvae were reared using artificial diet SilkMate 2S (Nosan Co., Yokohama, Japan) at 25°C under a long photoperiod (16L:8D), i.e., till pupation. To eliminate *Spiroplasma* from the S+ line, the first instar larvae were reared with SilkMate 2S supplemented with 0.05% tetracycline (w/w) as described by Arai et al. (2019). Adult moths were sexed based on their morphology, and the hatched larvae and the unhatched pharate larvae (mature embryo) were sexed based on the presence or absence of the female-specific sex chromatin body (a condensed W chromosome), which was detected *via* lactic-acetic orcein staining (Arai et al., 2022a).

Spiroplasma detection and quantification in *Homona magnanima*

Total DNA was extracted from the abdomen of female adults (0-day post eclosion), the whole body of larvae and pupae (0-day post molting), and dissected tissues of *H. magnanima* larvae (0-day post molting) using cell lysis buffer, as described by Arai et al. (2019). To detect *Spiroplasma*, a pair of *Spiroplasma*-specific primers was used to amplify RNA polymerase β subunit gene (*RpoB*), which is a single copy conserved gene

in *Spiroplasma* spp., from the extracted DNA (adjusted to 50–100 ng/reaction) with EmeraldAmp MAX PCR Master Mix (TaKaRa Bio, Shiga, Japan); the primer sets are listed in **Table 1**. The PCR conditions were as follows: 35 cycles of 94°C for 30 s, 55°C for 30 s, and 72°C for 30 s, followed by 72°C for 7 min. β-Actin gene of *H. magnanima* was used as the control. To quantify *Spiroplasma* density, qPCR was performed using the extracted DNA, which was diluted to a concentration of 10 ng/μL with MilliQ water, *Spiroplasma RpoB* primers (**Table 1**), and KOD SYBR® qPCR Mix (Toyobo, Osaka, Japan) in a LightCycler® 96 system (Roche, Basel, Switzerland). The PCR consisted of 45 cycles of 98°C for 10 s, 60°C for 10 s, and 68°C for 30 s. Relative abundance of the gene was calculated using the expression of elongation factor 1a gene (*ef1a*) of *H. magnanima* as the control. *Spiroplasma* density (*RpoB* copies) and relative abundance (*RpoB*/*ef1a*) were calculated as described in Arai et al. (2019, 2022c).

Genome sequence of the *Spiroplasma* sHm strain

For genome sequencing of strain *s*Hm, high molecular weight DNA was extracted from the egg masses of S+ line moths using Nanobind Tissue Big DNA Kit (Circulomics Inc., MD, USA) and used for library construction using Ultra-Long DNA Sequencing Kit (Oxford Nanopore Technologies, Oxford, UK) following the manufacture's protocol. The constructed libraries were sequenced using ONT MinION flow cell (R 9.4.1) (Oxford Nanopore Technologies). The obtained reads were mapped to the *H. magnanima* reference genome (Jouraku et al., in preparation) with minimap2 (Li, 2018), and the non-mapped reads containing *Spiroplasma* reads were extracted with SAMtools v.1.9 (Li et al., 2009) and assembled using Canu 1.6 (Koren et al., 2017). The draft *Spiroplasma* genome (a circular main chromosome and plasmids) was annotated *via* BLASTn (NCBI nr database). The extracted DNA was also subjected to Illumina paired-end 150 bp sequencing (PE-150) at Novogene (Beijing, China). The Illumina data were used to polish the draft genome using minimap2 (Li, 2018) and Pilon v. 1.23 (Walker et al., 2014). Since no sequence changes were observed after the second polishing, the polished genome was considered as the complete genome of strain *s*Hm. The circularity of the *s*Hm genome was confirmed by BLASTn search, followed by manual deletion of overlapping sequence.

Resequencing of the sHm strain in the S+ and S+M+ moth lines

S+ (MK line) and S+M+ *H. magnanima* lines (non-MK line) were used for DNA extraction as described by Arai et al. (2022a). The DNA extracted from *Spiroplasma* cells was amplified using whole genome amplification (WGA) by REPLI-g Mini Kit (Qiagen, Hilden, Germany), following the manufacture's protocol. The WGA products, purified using AMPure XP beads (Beckman Coulter, Inc., CA, USA) and dissolved into TE buffer, were sequenced on Illumina platform (PE-150). The Illumina data assembled with unicycler (Wick et al., 2017) and Illumina raw read data were mapped to the *s*Hm reference genome using minimap2 (Li, 2018) to detect the genomic changes in the genome of *s*Hm in the S + M + line.

Genome annotations and homology surveys

The constructed *s*Hm genome was annotated *via* DFAST (Tanizawa et al., 2018). Effector genes were further annotated using EffectiveDB (Eichinger et al., 2016). Functional analysis of proteins (i.e., domain predictions and Gene ontology annotations) was conducted using InterPro.[1] Phage WO infections were annotated using PHASTER (Arndt et al., 2016). Protein homology between different *Spiroplasma* strains was analyzed using *S. apis* B31 (CP006682.1), *S. citri* strain BLH-MB (CP047437.1–CP047446.1), *S. syrphidicola* strain EA-1 (NC_021284.1), *D. melanogaster* endosymbiont *S. poulsonii* MSRO (CM020866.1–CM020867.1) (*s*Mel, MK strain, Masson et al., 2018), *Danaus chrysippus* (Nymphalidae) endosymbiont *S. ixodetis* (NZ_CADDIL010000001.1–NZ_CADDIL010000012.1) (*s*Da, MK strain, Martin et al., 2020), *Lariophagus distinguendus* (Pteromalidae) endosymbiont *S. ixodetis* (NZ_JALMUW010000001.1–NZ_JALMUW010000198.1) [*s*Dis, cytoplasmic incompatibility (CI) strain, Pollmann et al., 2022], and *Dactylopius coccus* (Dactylopiidae) endosymbiont *S. ixodetis* (JACSER010000001.1–JACSER010000358.1) (*s*Coc, non-MK strain, Vera-Ponce León et al., 2021) with OrthoVenn2.[2] Homology of *s*Hm genes and proteins with *spaid* from strain *s*Mel (Harumoto and Lemaitre, 2018), ankyrin genes from *S. ixodetis* (Yeoman et al., 2019; Martin et al., 2020; Vera-Ponce León et al., 2021), and the *Wolbachia* MK candidate factor responsible for WO-mediated killing (Wmk, presumed helix-turn-helix transcriptional regulator, Perlmutter et al., 2019; Arai et al., 2022b) was evaluated using both BLASTn and BLASTp. Moreover, to verify whether MK microbes of *H. magnanima* carried conserved genes, the genes on the MK-associated prophage region WO*w*Hm-t76 of MK *Wolbachia w*Hm-t (Arai et al., 2022b) and those of the Partiti-like virus OGVs (Fujita et al., 2021) were compared to the *s*Hm genes using both BLASTn and BLASTp. Unique genomic features of the *s*Hm strain in the MK S+ and non-MK S+M+ *H. magnanima* lines

1 https://www.ebi.ac.uk/interpro/

2 https://orthovenn2.bioinfotoolkits.net/home

TABLE 1 Sequences and related information of the primers used in this study.

Target	Gene	Primers sequences (5′–3′)	Product size (bp)	Annealing temperature (°C)	References
H. magnanima	β-Actin	297f:AACTGGGATGACATGGAGAAGATCTGGC	838	55	Tsugeno et al., 2017
		1139r: GAGATCCACATCTGCTGGAAGGTGGACAG			
	HmEf-1a	Hmef1a_F_val1_85: TTTCCAGGGTGGTTGAGCA	108	60	Arai et al., 2022c
		Hmef1a_R_val1_193: CCGTTAAGGAGCTGCGTCG			
	COI	LepF: ATTCAACCAATCATAAAGATATTGG	650	55	Hajibabaei et al., 2006
		LepR: TAAACTTCTGGATGTCCAAAAAATCA			
Spiroplasma	*RpoB*	HmSpiro_RpoB388qF: GCATACTCAACACCCGTACCA	95	60	This study
		HmSpiro_RpoB483qR: TGCTAACCGTGCTTTAATGGG			
		HmSpiro_RpoB155F: CGCCATCTTTCATCGAAGGTC	423	60	
		HmSpiro_RpoB578R ATTGTTGGACCAAACGAAGTTG			

were analyzed using GView[3] and BV-BRC variation analysis service.[4] Metabolic pathways of *S. poulsonii* sMel and *S. ixodetis* sHm were compared by using BV-BRC comparative analysis service (see text footnote 4). Phylogenetic trees of 16S rRNA gene and ankyrin genes of *Spiroplasma* strains were constructed by maximum likelihood with bootstrap re-sampling of 1,000 replicates using MEGA7 (Kumar et al., 2016). *Mycoplasma genitalium* G-37 (NR074611.1) was used as an outgroup.

Transinfection assays

A fifth instar female larva was sterilized in 50% bleach (ca. 3% sodium hypochlorite) for 10 min, in 70% ethanol for 10 min, and dissected in IPL-41 Insect Medium (Gibco, Carlsbad, CA, USA) supplemented with 10% (v/v) fetal bovine serum (FBS). Malpighian tubules of the dissected larva were transferred to flasks containing either the *Bombyx mori* NIAS-Bm-aff3 (aff3) cell line (Takahashi et al., 2006) or the *Aedes albopictus* NIAS-AeAl-2 (AeAl2) cell line (Mitsuhashi, 1981) maintained in IPL-41 Insect Medium (Gibco) with 10% (v/v) of FBS. The cells and Malpighian tubules were co-cultured at 23°C. Fresh medium was supplied to the flask every 10 d. Purified cells centrifuged at 1,000 *g* for 2 min were used to analyze infections and titers of the transinfected strain *s*Hm in the cells. DNA extraction, PCR, and qPCR assays were performed as mentioned in section "Spiroplasma detection and quantification in *Homona magnanima*."

Statistical analysis

Sex ratio bias was assessed using Fisher's exact test. *Spiroplasma* densities, male ratio in hatched larvae, and male ratio in unhatched pharate larvae were analyzed using either the Wilcoxon test or the Steel–Dwass test. All analyses were performed using R software v4.0[5].

Results and discussion

Spiroplasma ixodetis strain sHm induced embryonic male death in *Homona magnanima*

The S+ line moths harboring *s*Hm exhibited lower egg-hatching rates than the NSR line (**Figure 1A**), which is consistent with the results from previous studies that *Spiroplasma* infection halved the egg hatching rates of *H. magnanima* (Tsugeno et al., 2017; Takamatsu et al., 2021). Cytogenetic sexing based on the presence or absence of a sex chromatin body (W chromosome) revealed that the sex ratio of hatched larvae was strongly biased toward females in the S+ line moths but not in the NSR line moths ($P < 0.01$, **Figure 1B**). In contrast, the sex ratio of unhatched pharate larvae (late-stage embryos) were male-biased in the *s*Hm-infected line (S+) (Fisher's exact test, $P < 0.01$), confirming that *s*Hm killed male *H. magnanima* during embryogenesis. Moreover, the elimination of *Spiroplasma* by tetracycline treatment resulted in non-biased sex ratios in the subsequent generation ($P < 0.01$, **Figure 1C**).

Genome sequence and genetic characteristics of male-killing *Spiroplasma ixodetis* strain sHm

Both Illumina (816.3 Mb, 5,442,459 reads, and 150 bp average length) and Nanopore data (93.5 Mb, 25,820 reads,

3 https://server.gview.ca

4 https://www.bv-brc.org

5 https://www.r-project.org/

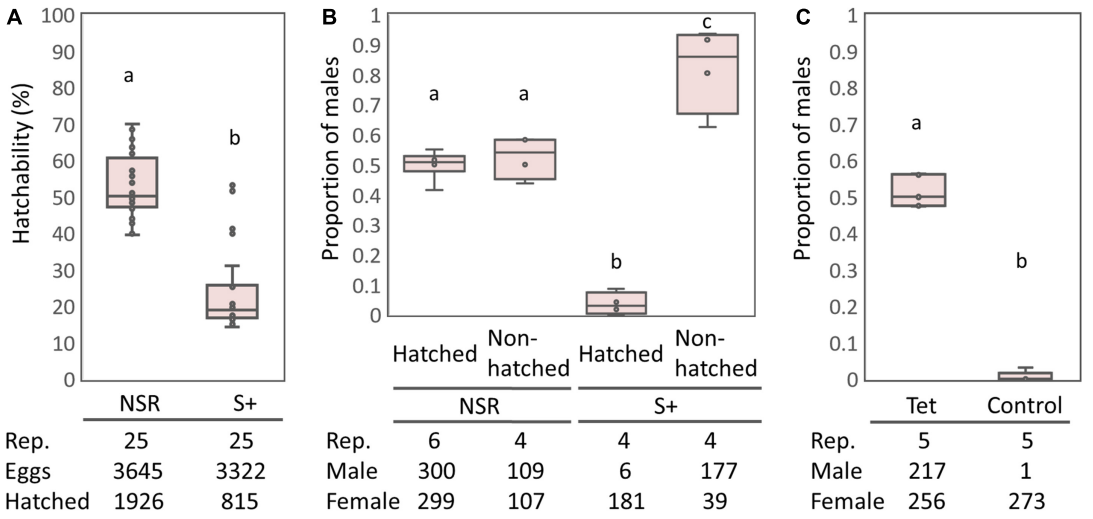


FIGURE 1
Egg hatching rates and sex ratios of the S+ line of *Homona magnanima*. **(A)** Egg hatching rates of 25 egg masses for each of the NSR and S+ lines (total 3,645 and 3,322 eggs, respectively). **(B)** Proportion of males among the hatched larvae and the unhatched pharate larvae. **(C)** Proportion of males of the subsequent generation after tetracycline treatment. Tet, tetracycline treatment; Control, non-treated control. The horizontal bar within the box represents the median. The upper and lower hinges of the box indicate upper quartile and lower quartile, respectively. Sample sizes are indicated below the panels. Different letters indicate significant differences between groups (Steel–Dwass test, $P < 0.05$).

and 3,624 bp average length) were used to reconstruct a complete genome consisting of a circular main chromosome (2, 102,039 bp in length) and two circular potential plasmids [20,119 bp (pSHM_1) and 16,408 bp (pSHM_2)]. Previously, Tsugeno et al. (2017) reported two 16S rRNA gene variants cloned from *Spiroplasma*-infected *H. magnanima*, but they did not elucidate whether the two sequences were interoperonic polymorphs of a single isolate or they were derived from two different strains. The present study confirmed that *H. magnanima* was infected with the MK *S. ixodetis* strain *s*Hm that harbored two distinct 16S rRNA gene sequences in its genome (**Figure 2**). Moreover, *Spiroplasma* strains often encode multiple ribosomal RNA gene sets in their genome (Chang et al., 2014; Tsai et al., 2018; Vera-Ponce León et al., 2021).

Strain *s*Hm harbored a higher number of coding sequences (CDS; 2,886 CDS) than strains *s*Mel (1.9 Mb in genome size; 2,405 CDS; Masson et al., 2018) and *s*Da (1.7 Mb in genome size; 1,813 CDS; Martin et al., 2020; **Table 2**). Although plasmids often contain key accessory genes such as *spaid* of *s*Mel (Harumoto and Lemaitre, 2018), genes on pSHM_1 ($n = 24$) and pSHM_2 ($n = 20$) mostly encoded hypothetical or uncharacterized proteins (**Supplementary Table 1**). In addition, strain *s*Hm harbored 12 prophage regions (size: 6.6–14.8 kb) in its genome, which is consistent with previous reports that several phage sequences are found in *Spiroplasma* genomes (Bai and Hogenhout, 2002; Lo et al., 2013; Ramirez et al., 2021). Bacteriophages frequently carry virulence-associated genes that encode toxins (Waldor and Mekalanos, 1996; Brüssow et al., 2004). Recently, the mechanistic bases of *Wolbachia*-induced cytoplasmic incompatibility (CI) and MK have been attributed to phages (Beckmann et al., 2017; LePage et al., 2017; Perlmutter et al., 2019; Arai et al., 2022b). Besides, phages have also been implicated in the defense phenotype exhibited by bacteria against parasitoids, such as *Hamiltonella defensa* (Brandt et al., 2017). Therefore, it is possible that the phages of *s*Hm contribute to the manifestation of MK phenotype or confer fitness advantage on hosts by protecting the hosts from natural enemies.

sHm harbored putative virulence-associated factors but did not harbor sMel spaid toxin

Recently, Yeoman et al. (2019) and Vera-Ponce León et al. (2021) reported that *D. coccus*-infecting *S. ixodetis* (sCoc) harbored a *spaid* homolog and *Cephus cinctus* (Cephidae)-infecting *S. ixodetis* harbored seven *spaid* homologs. Our BLAST searches confirmed that *s*Hm did not harbor the *spaid* gene (**Table 3**), however, some of the ankyrin genes of *s*Hm were homologous to the alleged gene sequences of *s*Coc and Cephidae-infecting *S. ixodetis* (**Table 4**). It is likely that *S. ixodetis* do not harbor the *spaid* gene. The superficial homology could be due to the presence of conserved ankyrin repeats (**Table 3**). Similarly, the amino acid sequences of ankyrin proteins of *s*Hm (such as SHM_18920) showed partial homology to the ankyrin domain of Spaid from strain *s*Mel (N-terminal 200 amino acids) as per BLASTp search, but the complete amino

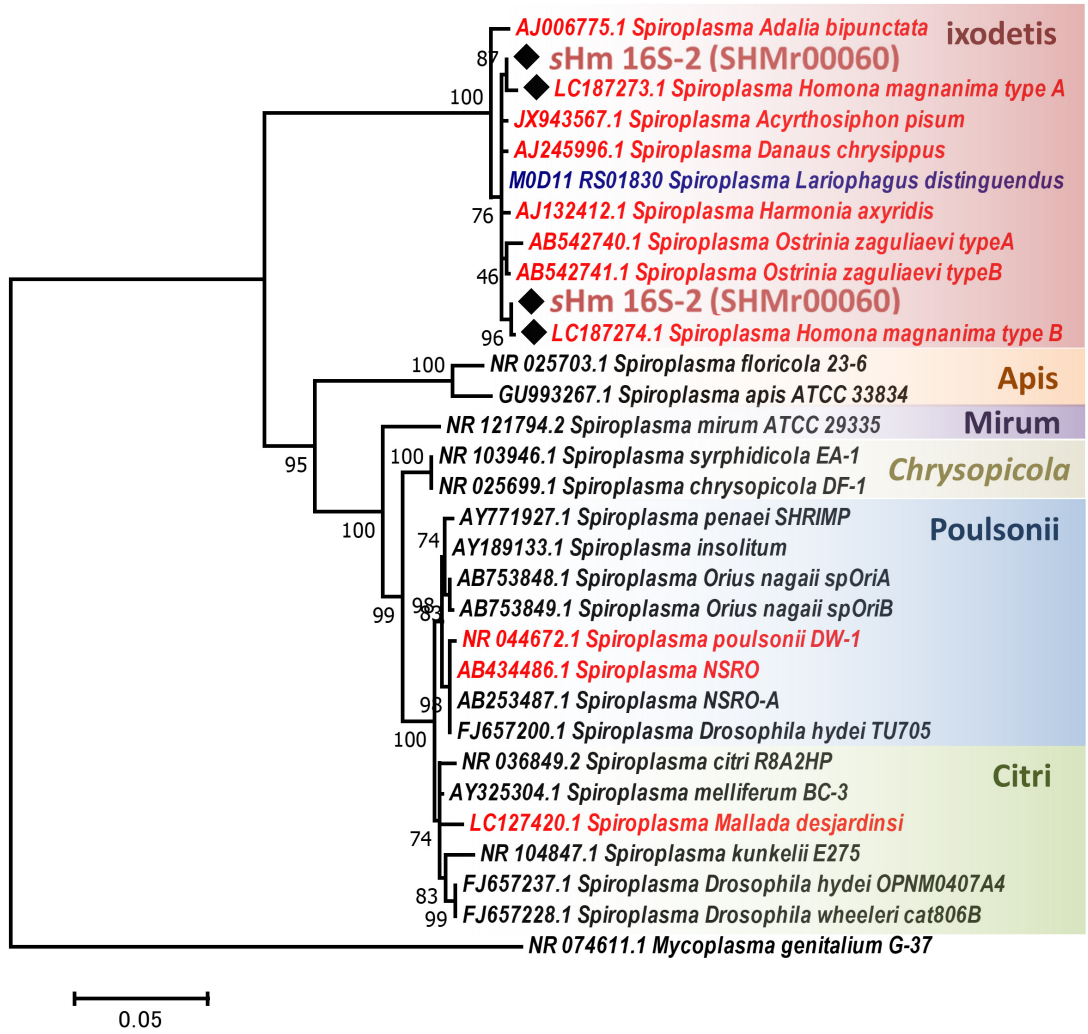


FIGURE 2
Phylogenetic tree of *Spiroplasma* strains based on 16S rRNA gene sequences. Phylogenetic tree based on 16S rRNA gene sequences of strain *s*Hm and other *Spiroplasma* strains (retrieved from NCBI database) using maximum likelihood method based on the Tamura–Nei model with 1,000 bootstrap replicates. Accession numbers are shown along with the operational taxonomic units (OTUs). Samples highlighted with red and blue color fonts are MK and CI strains, respectively. Black diamonds indicate 16S rRNA sequences of the *s*Hm. The classification of *Spiroplasma* is based on the study by Paredes et al. (2015). *Mycoplasma genitalium* was used as the outgroup.

acid sequences of the proteins of these two strains were not homologous (Table 3). Moreover, we also confirmed the absence of Spaid homologs in a MK *S. ixodetis* strain *s*Da by using BLASTp search. Gerth et al. (2021) reported that the Spaid homologs are conserved among *S. poulsonii* strains regardless of the MK phenotype. Because the spaid gene is not likely to be possessed by *S. ixodetis*, MK mechanisms may differ between *S. poulsonii* and *S. ixodetis* (i.e., having different causative genes).

We then focused on genes conserved among *Spiroplasma* strains. Distantly related *Spiroplasma* species such as *S. ixodetis* (*s*Hm), *S. poulsonii* (*s*Mel), *S. apis* B31, *S. citri* BLH-MB, and *S. syrphidicola* EA-1 shared 345 protein clusters (Figure 3A). For *S. ixodetis* strains, two MK strains (*s*Hm and *s*Da) and two non-MK strains (*s*Coc and *s*Dis) shared 595 protein clusters (Figure 3A). In addition, MK strains *s*Hm and *s*Da possessed additional 219 conserved protein clusters. *s*Hm also harbored strain-specific 77 protein clusters (470 genes) associated with metabolism and transposition (Figure 3B and Supplementary Table 1) as well as many putative Type IV secretory system effector genes (n = 144, based on T4SEpre prediction at EffectiveDB, Supplementary Table 1), some of which were located in the prophage regions. In *Spiroplasma*, RIP toxin irreversibly inactivates eukaryotic cytosolic ribosomes (Hamilton et al., 2016; Ballinger et al., 2019; Garcia-Arraez et al., 2019). Based on our blast searches, RIP-4 encoded by *Spiroplasma* endosymbiont of *Drosophila neotestacea* (ASM46790.1)

TABLE 2 Genomic features of strain *s*Hm and other *Spiroplasma* strains found in insects.

Genome ID	*Spiroplasma ixodetis* sHm	*Spiroplasma ixodetis* sDa	*Spiroplasma ixodetis* DCM	*Spiroplasma ixodetis* sDis	*Spiroplasma poulsonii* MSRO
Main chromosome/contigs	1 (closed/circular)	12	353	198	1 (closed/circular)
Plasmids	2 (closed/circular)	NA	NA	3 (closed/circular)	1 (closed/circular)
Estimated genome size (Mb)	2.14	1.75	1.32	1.16	1.96
N50	2,102,039	265,779	7,774	14,219	1,938,611
G + C content (%)	25.1	23.7	24.16	24.3	26.3
CDS genes	2,886	1,813	1,371	1,175	2,405
rRNA (16S, 5S, 23S)	6 (2,2,2)	4 (1,2,1)	3 (1,1,1)	3 (1,1,1)	3 (1,1,1)
tRNA	27	27	27	27	31
Phenotype	MK[1]	MK[2]	non-MK[3]	CI[4]	MK[5]
Insect associated	*Homona magnanima*[1]	*Danaus chrysippus*[2]	*Dactylopius coccus*[3]	*Lariophagus distinguendus*[4]	*Drosophila melanogaster*[5]

[1]Based on Tsugeno et al. (2017).
[2]Based on Martin et al. (2020).
[3]Based on Vera-Ponce León et al. (2021).
[4]Based on Pollmann et al. (2022).
[5]Based on Masson et al. (2018).

TABLE 3 Homology between Spaid [1,065 aa] of strain *s*Mel and proteins of *Spiroplasma ixodetis* strains based on BLASTp search.

Spiroplasma ixodetis proteins	Identity	Aligned length	*s*Mel spaid		*Spiroplasma ixodetis*		*e*-value	Bit score	References
			Start	End	Start	End			
sHm (SHM_18920)	36.0	205	217	409	60	261	2.99E-29	111	This study
sDa (SPD_05340)	40.9	220	54	266	28	237	5.56E-34	117	Martin et al., 2020
sCoc spaid-like (JACSEQ010000039.1)	48.7	80	137	216	21	100	5.71E-21	73.9	Vera-Ponce León et al., 2021
sWSS spaid-like (2132.146.peg.209)	43.0	179	45	223	18	186	2.08E-34	116	Yeoman et al., 2019
sWSS spaid-like (2132.146.peg.1)	40.9	105	126	229	56	158	4.11E-16	62.8	
sWSS spaid-like (2132.146.peg.21)	46.0	76	159	234	1	74	5.59E-16	59.3	
sWSS spaid-like (2132.146.peg.255)	33.3	120	107	220	17	126	2.93E-12	53.5	
sWSS spaid-like (2132.146.peg.305)	45.4	99	96	194	176	267	4.81E-20	78.6	
sWSS spaid-like (2132.146.peg.596)	34.3	233	69	266	16	248	2.33E-29	105	
sWSS spaid-like (2132.146.peg.469)	33.7	237	66	300	12	219	7.15E-25	96.7	

showed low homology to SHM_22560 (79–286 aa, *e*-value 6.7E-13, bit-score 60.8). Besides, SHM_22560 (hypothetical protein, 788 aa, **Supplementary Table 1**) was predicted to contain a RIP domain based on Interpro (hit: IPR016138, aligned length: 111–286 aa) and HHpred searches [hit: Sapolin (ID: 3HIQ), aligned length: 105–344 aa, *e*-value: 2.3E-29]. A homolog of an epsilon-like toxin (WP_252319264.1_36) encoded by CI-inducing *s*Dis was detected in the *s*Hm genome (SHM_25300, **Table 4**), while AbiEii abortive infection toxin (WP_252320055.1_19) and OTU-like cysteine protease (WP_252320277.1_1) were not detected.

Although *spaid* gene is the only ankyrin-coding gene identified in the genome of strain *s*Mel (Harumoto and Lemaitre, 2018), *s*Hm carried 10 ankyrin genes (**Figure 3C**). Some ankyrin genes harbored additional domains, such as those encoding for DnaJ (SHM_12110), cyclin-dependent kinase inhibitor (SHM_12110 and SHM_28080), and the protein ubiquitination associated FEM1A_DROME (SHM_21210); however, they did not encode for signal peptides, ovarian tumor-like deubiquitinase (OTU), or helix domains found in the *spaid* gene of strain *s*Me1 (Paredes et al., 2015; Harumoto and Lemaitre, 2018; Masson et al., 2018; Gerth et al., 2021). *Wolbachia* induces CI by the CI-inducing factors (Cif) harboring ankyrin repeats in insects (LePage et al., 2017). Pollmann et al. (2022) reported that CI-inducing *s*Dis strain did not harbor the *cif* gene. Similarly, *s*Hm-encoding 10 ankyrin genes has low homologies to

TABLE 4 Homology between ankyrin genes of two *Spiroplasma ixodetis* strains based on BLASTn search.

Spiroplasma ixodetis gene (length)	Strain sHm ankyrin gene (length)	Identity	Aligned length	*Spiroplasma ixodetis* genes		*s*Hm genes		*e*-value	Bit score	References
				Start (nt)	End (nt)	Start (nt)	End (nt)			
sCoc (JACSEQ010000039.1) (303 nt)	SHM_18920 (1,971 nt)	97	300	4	303	594	295	3.92E-144	505	Vera-Ponce León et al., 2021
sWSS (2132.146.peg.596) (894 nt)	SHM_18920 (1,971 nt)	97.6	894	1	894	1	894	0	1535	Yeoman et al., 2019
sWSS (2132.146.peg.469) (1656 nt)	SHM_21210 (1,370 nt)	96.2	974	114	1,087	414	1,380	0	1,587	
sWSS (2132.146.peg.209) (558 nt)	SHM_12270 (459 nt)	85.7	385	1	383	1	383	2.78E-113	403	
sWSS (2132.146.peg.21) (276 nt)	SHM_12270 (459 nt)	96.8	158	1	158	265	422	6.55E-72	265	
sWSS (2132.146.peg.255) (723 nt)	SHM_14030 (249 nt)	97.2	221	1	221	1	221	7.94E-105	375	
sWSS (2132.146.peg.305) (966 nt)	SHM_05900 (279 nt)	90.4	220	671	890	77	279	4.01E-74	274	
sDa (SDA_03750) (162 nt)	SHM_00770 (177 nt)	97.5	162	1	162	16	177	1.33E-75	278	Martin et al., 2020
sDa (SDA_06930) (438 nt)	SHM_05900 (279 nt)	87.9	241	269	506	56	279	4.73E-72	267	
sDa (SDA_19590) (348 nt)	SHM_08510 (1,113 nt)	94.6	546	1	546	1	544	0	846	
sDa (SDA_19580) (249 nt) *NANK	SHM_08510 (1,113 nt)	95.9	249	249	1	637	884	1.46E-112	403	
sDa (SDA_10430) (978 nt)	SHM_12110 (978 nt)	98.1	978	1	978	1	978	0	1,707	
sDa (SDA_08770) (723 nt)	SHM_12270 (459 nt)	94.1	292	1	292	1	290	3.45E-125	444	
sDa (SDA_12020) (723 nt)	SHM_14030 (249 nt)	96.8	221	1	221	1	221	3.10E-103	370	
sDa (SDA_05330) (591 nt)	SHM_18920 (1,971 nt)	87.7	236	50	285	1,082	1,314	7.66E-73	272	
sDa (SDA_05340) (538 nt)	SHM_21210 (1,370 nt)	94.9	736	1	736	1	730	0	1,149	
sDa (SDA_01840) (876 nt)	SHM_28080 (693 nt)	89.8	690	49	735	1	690	0	883	
sDa (SDA_01841) (231 nt)	SHM_28070 (162 nt)	93.8	162	1	162	1	161	4.40E-65	243	
sDis (WP_252318998.1_58) (738 nt)	SHM_12270 (459 nt)	85.974	385	1	383	1	383	7.99E-115	409	Pollmann et al., 2022
sDis (WP_252319959.1_15) (648 nt)	SHM_18920 (1,971 nt)	97.651	596	1	596	1	596	0	1,024	
sDis (WP_252320693.1_6) (975 nt)	SHM_08510 (1,113 nt)	87.514	913	1	904	1	910	0	1,044	
sDis (WP_252321112.1_1) (483 nt)	SHM_21210 (1,370 nt)	96.312	461	1	461	1	461	0	758	
sDis (WP_252319264.1_36, Epsilon-like toxin) (948 nt) *NANK	SHM_25300 (495 nt) *NAK	99.187	492	154	645	1	492	0.00E + 00	887	

*NANK, non-ankyrin genes.

10.3389/fmicb.2022.1075199

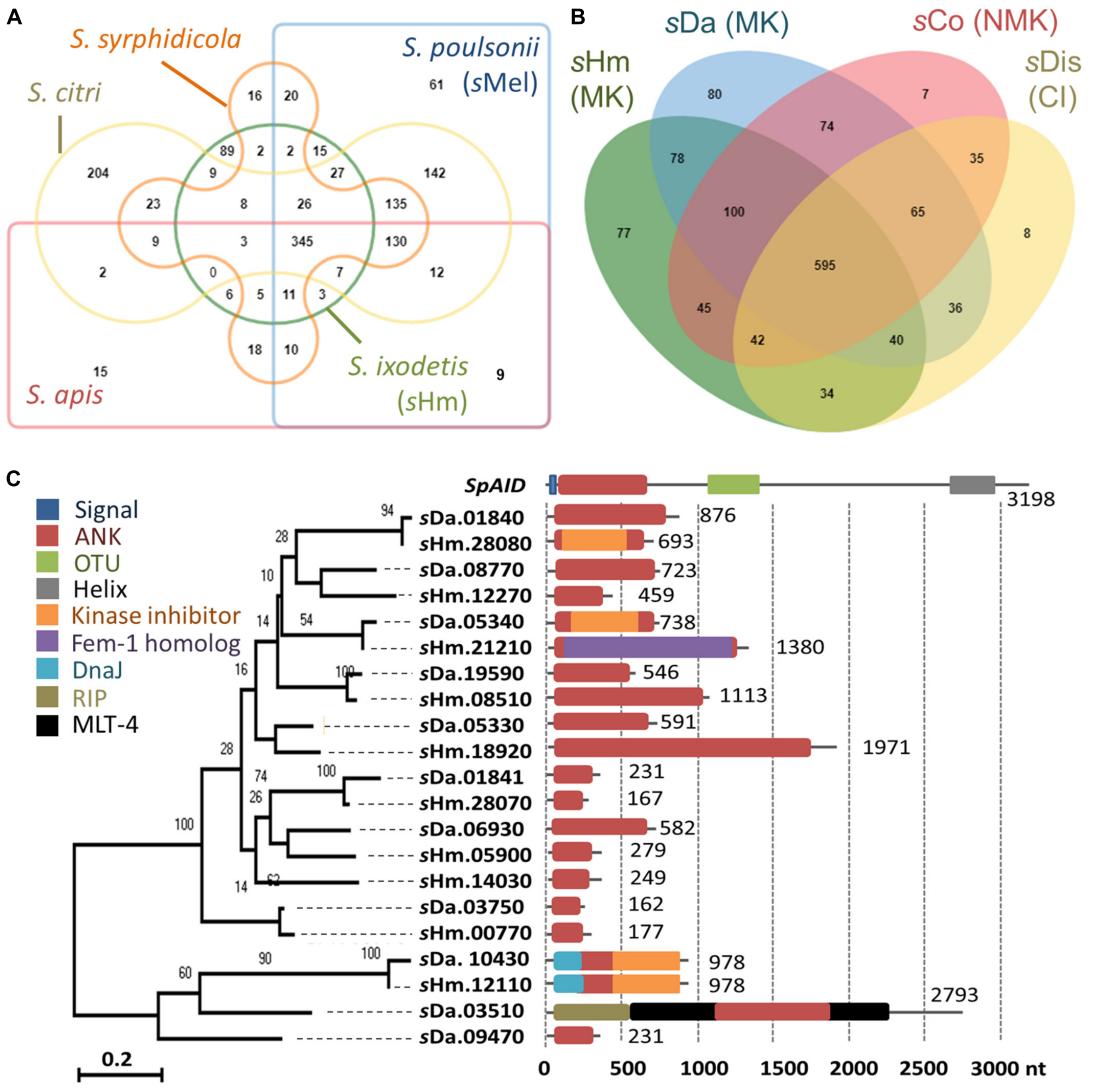


FIGURE 3
Comparison of *Spiroplasma* genomes and ankyrin genes. **(A–B)** Protein clusters conserved within *Spiroplasma* species (*S. apis, S. citri, S. ixodetis, S. poulsonii, and S. syrphidicola*) **(A)** and within *S. ixodetis* strains (*s*Hm, *s*Da, *s*Coc, and *s*Dis) **(B)**. Venn diagrams were visualized using OrthoVenn2. The *s*Hm-specific protein clusters and their functions are shown in **Supplementary Table 1**. MK, male-killing strain; NMK, non-male-killing strain; CI, cytoplasmic incompatibility strain. **(C)** Phylogeny of ankyrin genes in *Spiroplasma* strains *s*Hm and *s*Da. Numbers indicates nucleotide length (nt). Other domains of the genes are highlighted in different colors. Red: ankyrin; Blue: signal peptide; Green: OTU; Gray: transmembrane helix; Orange: cyclin-dependent kinase inhibitor; Purple: ubiquitination associated FEM1A_DROME; Sky blue: DNA J; Khaki: Ribosome–inactivating protein (RIP); Black: molting protein (MLT-4).

those of other bacteria such as *Wolbachia* and *Rickettsia* and were not homologous to the *cif* as well as *spaid* genes. Intriguingly, MK *s*Da and CI *s*Dis strains had 11 and 12 ankyrin genes, respectively. These findings suggest that *S. ixodetis* has similar characteristics to *Wolbachia* endosymbionts (Duplouy et al., 2013; Arai et al., 2022b) in terms of phenotypes (i.e., CI and MK) and genetic compositions (i.e., multiple ankyrin genes). Some ankyrin genes encoded by *S. poulsonii* and *Wolbachia* have been implicated in reproductive manipulation (LePage et al., 2017; Harumoto and Lemaitre, 2018), and the ankyrin genes found in the *s*Hm genome may also be involved in MK mechanisms.

Male-killing genes of sHm are different from those of other male-killers in *Homona magnanima*

Homona magnanima harbors three different types of MK endosymbionts (i.e., *Spiroplasma* *s*Hm, Partiti-like virus OGVs, and *Wolbachia* *w*Hm-t strain), some of which can

coinfect the same host (Arai et al., 2020; Takamatsu et al., 2021). Moreover, microbes sharing the same niche frequently exchange virulence-associated genes (Kent and Bordenstein, 2010; Wiedenbeck and Cohan, 2011). However, we found that strain *s*Hm did not harbor any gene homologous to those of MK Partiti-like virus OGVs (Fujita et al., 2021). Moreover, strain *s*Hm did not harbor *wmk* or effector genes (e.g., CifB-like) that are present on the MK-associated prophage WO*w*Hm-t76 region of strain *w*Hm-t (Arai et al., 2020, 2022b). The *wmk* gene, a candidate gene for *Wolbachia*-induced MK (Perlmutter et al., 2019, 2021; Arai et al., 2022b), possesses a helix-turn-helix (HTH) domain containing putative transcriptional regulator. Although no *wmk* homologs were identified, strain *s*Hm harbored 87 HTH domain-encoding genes, namely putative transposase (classified into IS-30, IS-3, and IS-5 type transposase, n = 83), a type II toxin-antitoxin system antitoxin *HipB* (SHM_ 03650), an AAA family ATPase (SHM_24830), an XRE family transcriptional regulator (SHM_17560), and a helix-turn-helix transcriptional regulator (SHM_05440). Notably, a putative transposase SHM_03660, encoded by a gene adjacent to *s*Hm-specific *HipB*-like SHM_3650, was homologous to the *Wolbachia* transcriptional regulator. Recently, Arai et al. (2022c) demonstrated that strains *s*Hm, *w*Hm-t, and OGVs affect *H. magnanima* males in different manners. Specifically, both strains *s*Hm and *w*Hm-t trigger abnormal apoptosis and interfere with sex determination in male embryos (manifested by the alteration of *doublesex* gene splicing), but only strain *w*Hm-t impairs the dosage-compensation system of the host (manifested by the alteration of the global gene expression on sex chromosomes). In contrast, the OGVs do not affect sex-determination cascades or dosage-compensation systems. These findings and our current results support the view that phylogenetically distinct microbes have independently developed different MK machinery even for the same host, i.e., *H. magnanima*. Therefore, an unknown factor in the *s*Hm genome may be responsible for the embryonic male death of *H. magnanima*.

sHm may require high infection density to kill *Homona magnanima* males

We observed that one of the sublines of the MK S+ line ceased to induce MK (**Figure 4A**). This subline, referred to as the S+M+ line, exhibited stable *s*Hm infections for at least four generations. We simultaneously re-sequenced the genome of strain *s*Hm from S+M+ and S+ lines at the second-generation stage since their divergence. We previously demonstrated from a genomic comparison of MK *Wolbachia* (*w*Hm-t) and non-MK *Wolbachia* (*w*Hm-c) that an MK-associated 76 kb prophage region was inserted only in *w*Hm-t (Arai et al., 2022b). Similarly, we mapped the MK and non-MK *s*Hm re-sequenced Illumina reads to the complete *s*Hm genome (main chromosome and two plasmids) but did not detect any large-scale structural variation (insertions or deletions) as observed in *w*Hm-t (**Figure 4B**). On the other hand, we found mutations specific to the non-MK *s*Hm mutant (i.e., frameshifts or insertion of stop codons) in 21 genes encoding hypothetical proteins (n = 4), tyrosine-tRNA ligase (n = 1), and transposase (n = 16) (**Supplementary Table 2**). The 21 genes were found on the main chromosome, not in plasmids. Moreover, the *s*Hm density in the S+M+ line was lower than that in the S+ line (Steel–Dwass test, $P < 0.05$, **Figure 4C**).

Rapid genetic evolution leading to resistance against MK *Spiroplasma* has been reported in various hosts, such as the lacewing (Hayashi et al., 2018) and the planthopper (Yoshida et al., 2021). However, we can exclude the host genetic changes from the possible cause of the observed loss of MK phenotype because females of the S+ and S+M+ lines were parallelly mated with the males of the genetically homogeneous NSR line that had been maintained *via* inbreeding in the laboratory for over 10 years (>120 generations).

Spiroplasma-induced phenotypic changes have been repeatedly observed in previous studies. For example, spontaneous loss of MK was found in *S. poulsonii* strains of *Drosophila* flies, wherein substitutions and deletions occurred in the MK gene *spaid* (Harumoto and Lemaitre, 2018). Moreover, the MK strain *S. poulsonii* NSRO and its non-MK variant NSRO-A exhibit difference in bacterial densities in *D. melanogaster* (Anbutsu and Fukatsu, 2003). Indeed, bacterial density is one of the crucial factors for *Spiroplasma*- and *Wolbachia*-induced MK phenotype (Hurst and Jiggins, 2000; Kageyama et al., 2007; Arai et al., 2020). Based on these results, we speculate that the loss of MK phenotype of *s*Hm-infected *H. magnanima* was due to (i) reduced *s*Hm density and/or (ii) mutations in *s*Hm MK gene(s) or factors regulating MK gene expression levels. However, we still do not know how the small genomic rearrangements (i.e., inversions and insertions) detected in this study are involved in the phenotypic changes of *s*Hm. Future *de novo* genome construction of *s*Hm from S+M+ lines and gene function analysis would help in elucidating MK mechanisms.

Population dynamics and tissue tropism of sHm

Strain *s*Hm was abundant at the late-developmental stages of *H. magnanima* (**Figure 5A**), and *s*Hm densities drastically increased from pupal to adult stages of the insect (**Figure 5B**). In *D. melanogaster*, *S. poulsonii* copy numbers gradually increase as the host larval development proceeds and are generally higher in pupae than in larvae (Anbutsu and Fukatsu, 2003). In contrast to *S. poulsonii*, which is reported to be the most

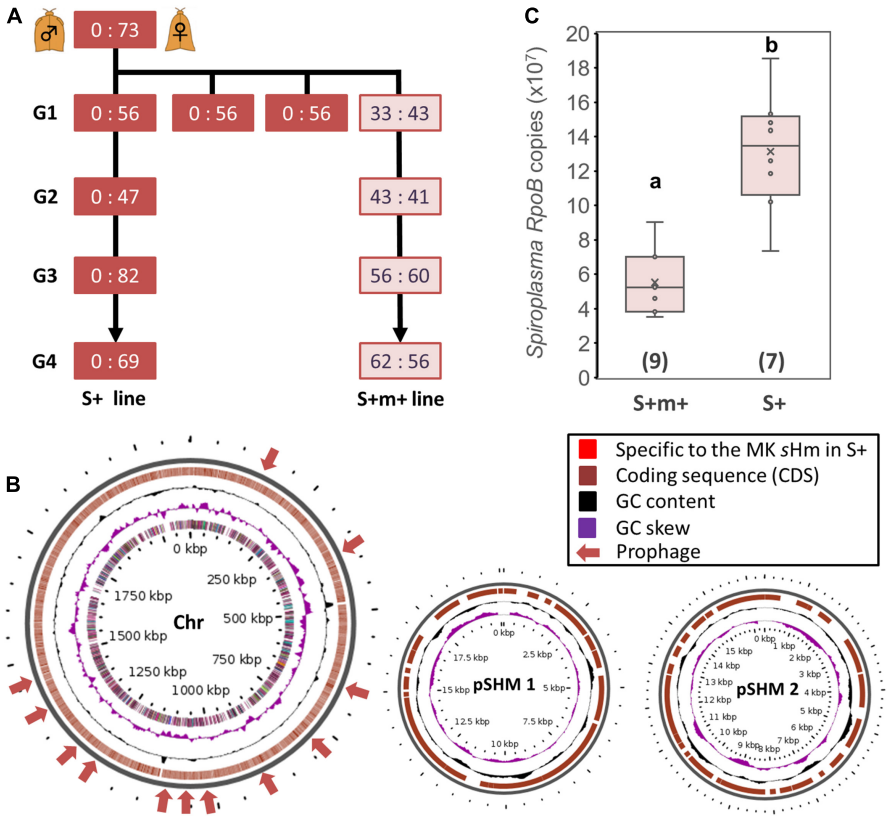


FIGURE 4
Comparison of *Spiroplasma* strain *s*Hm in MK and non-MK *Homona magnanima* lines. **(A)** *Spiroplasma*-positive 1:1 sex ratio line (S+M+ line) was maintained for over four generations (G1–G4), parallelly with the all-female S+ line. **(B)** Structural variations of the genome of *Spiroplasma* strain *s*Hm in the MK (S+) and non-MK (S+M+) *H. magnanima* lines were visualized *via* Gview software. No obvious structural variations (i.e., the red colored loci), specific to MK *s*Hm strain in the S+ matriline, were identified. Chr: main chromosome (2.1 Mb); pSHM 1: sHm plasmid 1 (20 Kb); pSHM 2: sHm plasmid 2 (16 Kb). **(C)** Abundance of *Spiroplasma* (based on copy numbers of *RpoB*) in adult females (0-day post eclosion) of S+M+ line and S+ line. The horizontal bar within the box represents the median. The upper and lower hinges of the box indicate upper quartile and lower quartile, respectively. Sample sizes (numbers of examined individuals) are indicated in parentheses. Different letters indicate significant differences between groups (Steel–Dwass test, $P < 0.05$).

abundant in hemolymph (Anbutsu and Fukatsu, 2003, 2006), strain *s*Hm exhibited low density in the hemolymph and high density in Malpighian tubules in the fifth instar larva stage (**Figure 5C**). High titers in Malpighian tubules are also a characteristic of *Wolbachia*; *Wolbachia* present in Malpighian tubules protects the host from RNA-virus infections and may constitute a secondary pool of vertically infected bacteria (Faria and Sucena, 2013; Pietri et al., 2016). The localization of strain *s*Hm in somatic tissues may have contributed to the fitness of *H. magnanima* Although there have been no reports of *S. ixodetis* localization patterns in insects, our findings suggest that *S. poulsonii* and *S. ixodetis* have distinct proliferation strategies. The hemolymph is a nutrient-rich environment but is likely an extreme habitat for microorganisms because it is well-defended by the immune system of the host (Blow and Douglas, 2019). Indeed, only a few microbial taxa are known to persist in the hemolymph of insects for extended periods without causing insect morbidity and death (Blow and Douglas, 2019). Intriguingly, *S. poulsonii* *s*Mel encoded more metabolic genes in its genome than *S. ixodetis* *s*Hm (**Supplementary Table 3**). Hemolymph-inhabiting *S. poulsonii* may have developed specific adaptations for its habitat, which are distinct from those of *S. ixodetis*. Further characterization of genomic features and localization patterns of *Spiroplasma* strains will clarify the distinct proliferation strategies of the two species (e.g., nutrient requirements).

Proliferation of sHm in insect cell culture

Tsugeno et al. (2017) reported that *s*Hm is horizontally transmitted by inoculating non-infected *H. magnanima* with concentrated hemolymph collected from *s*Hm-infected *H. magnanima*. Moreover, we revealed that *S. ixodetis* *s*Hm exhibited *Wolbachia*-like genetic characteristics (i.e., multiple

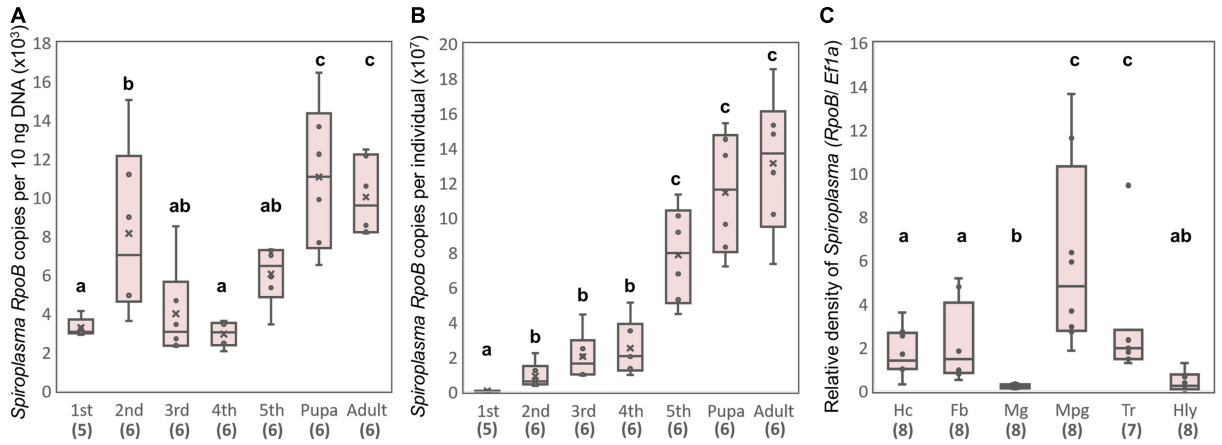


FIGURE 5
Propagation and localization of strain *s*Hm in *H. magnanima*. *Spiroplasma* densities deduced from *RpoB* copy numbers per 10 ng DNA **(A)** and *Spiroplasma* densities deduced from *RpoB* copy numbers per individual **(B)** at each developmental stage. **(C)** Relative *Spiroplasma* density in each organ deduced from *Spiroplasma RpoB* copy numbers per host *Ef1α* copy. Hc, hemocyte; Fb, fat body; Mg, midgut; Mpg, Malpighian tubules; Tr, trachea; Hly, hemolymph. Different letters indicate significant differences between groups (Steel–Dwass test, $P < 0.05$).

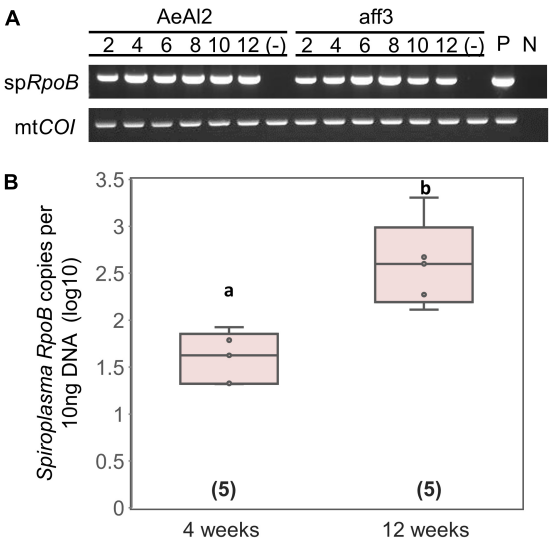


FIGURE 6
Transinfection of *Spiroplasma* strain *s*Hm into mosquito and silkworm cell lines. **(A)** Detections of strain *s*Hm in passaged cells. Numbers indicate periods (weeks) from the transinfection. P (positive control): S + female; N (negative control): NSR female. **(B)** *Spiroplasma* density in AeAl2 cells deduced from *RpoB* copy numbers per 10 ng DNA. Different letters indicate significant differences between groups (Steel–Dwass test, $P < 0.05$).

ankyrin genes) and localization patterns in somatic tissues. *Wolbachia* can infect and be maintained stably in insect cell lines derived from insect taxa that are distantly related to their native hosts (Fallon, 2021). To examine whether *s*Hm can infect insect cells, we transinfected *s*Hm to the cell lines of *A. albopictus* (AeAl2) and *B. mori* (aff3), which are known to be susceptible to *Wolbachia*. *s*Hm proliferated successfully by placing a piece of fat bodies or Malpighian tubules derived from an S+ female larva into a flask containing the AeAl2 or aff3 cells (**Figure 6**). *s*Hm was stably maintained in the cell lines for 12 weeks (**Figure 6A**) but not in cell-free medium IPL-41. qPCR revealed that *s*Hm titers in AeAl2 cells were significantly higher at 12 weeks than at 4 weeks after the introduction of *s*Hm (**Figure 6B**). This implies the potential of *s*Hm to survive in a wide host range besides *Homona* (Tsugeno et al., 2017), such as other lepidopteran and dipteran insects. *S. ixodetis* strains isolated from Japanese ticks were also shown to be culturable in the *A. albopictus* cell line C6/36 (Thu et al., 2019). We hypothesize that *S. ixodetis* strains have a broad host range like that of *Wolbachia*. It is not clear whether *S. poulsonii* has a broad host range because there is no but one report by Hackett et al. (1986) that showed the infectivity of strain WSRO (derived from *D. willistoni*) in the *Trichoplusia ni* cell line IPLB-TN-R^2. Several attempts to transinfect *S. poulsonii* (strain NSRO; derived from *D. nebulosa*) and MK *Spiroplasma* (derived from the lacewing *Mallada desjardinsi*) into AeAl2 and aff3 cells failed (personal observation by DK). It has been shown by hemolymph injection that *S. poulsonii* can infect drosophilid flies but not houseflies, suggesting its narrow host range (Williamson and Poulson, 1979).

Summary and perspectives

In this study, we sequenced and analyzed the genome of an MK *S. ixodetis* strain *s*Hm. *S. poulsonii* possesses the Spaid toxin as the MK factor, whereas our study revealed that MK *S. ixodetis* did not harbor *spaid* homologs. We speculate that

MK *S. ixodetis* strains found in a diverse range of insects (Hurst et al., 1999; Jiggins et al., 2000; Simon et al., 2011; Tabata et al., 2011; Sanada-Morimura et al., 2013) harbor yet-unknown MK gene(s), other than *spaid*; thus, future studies should focus on the identification of these MK genes. Besides, high infection efficiencies of strain *s*Hm in other insect cells led us to speculate that MK *S. ixodetis* has been horizontally transmitted among insect species, like *Wolbachia*, which has expanded its host range (Zhou et al., 1998; Baldo et al., 2006). Further studies would be required to understand whether closely related MK *Spiroplasma* strains (i.e., the *S. ixodetis* group) share common or different MK mechanisms, which will answer evolutionary questions such as how frequent novel MK genes arose, how MK genes moved between different *Spiroplasma* strains (if it did), and whether MK genes are associated with host sex determining systems.

Data availability statement

All sequence data are available at DRA under BioProject: PRJDB14468, Biosamples: SAMD00547685, SAMD00547900, and DRA014961. Spiroplasma genome data are available in the DDBJ database under the following accession numbers: AP026933–AP026935.

Author contributions

HA conducted all experiments, data analysis, and wrote the original manuscripts. MI assisted insect rearing, experiments, and discussions. DK managed the experiments and revised the original manuscript. All authors approved the final version of the manuscript.

Funding

This study was supported by the Japan Society for the Promotion of Science (JSPS) Research Fellowships for Young Scientists (grant numbers: 19J13123 and 21J00895), JSPS Grant-in-Aid for Scientific Research (grant number: 22K14902), and Cabinet Office, Government of Japan, Cross-ministerial Moonshot Agriculture, Forestry and Fisheries Research and Development Program (grant number: JPJ009237).

Acknowledgments

We thank Dr. Akiya Jouraku [National Agriculture and Food Research Organization (NARO), 1-2 Owashi, Tsukuba, Ibaraki, Japan] for providing *H. magnanima* genome data.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2022.1075199/full#supplementary-material

References

Anbutsu, H., and Fukatsu, T. (2003). Population dynamics of male-killing and non-male-killing *Spiroplasmas* in *Drosophila melanogaster*. *Appl. Environ. Microbiol.* 69, 1428–1434. doi: 10.1128/AEM.69.3.1428-1434.2003

Anbutsu, H., and Fukatsu, T. (2006). Tissue-specific infection dynamics of male-killing and nonmale-killing *Spiroplasmas* in *Drosophila melanogaster*. *FEMS Microbiol. Ecol.* 57, 40–46. doi: 10.1111/j.1574-6941.2006.00087.x

Anbutsu, H., and Fukatsu, T. (2011). *Spiroplasma* as a model insect endosymbiont. *Environ. Microbiol. Rep.* 3, 144–153. doi: 10.1111/j.1758-2229.2010.00240.x

Arai, H., Ishitsubo, Y., Nakai, M., and Inoue, M. N. (2022a). Mass-rearing and molecular studies in *Tortricidae* Pest Insects. *J. Vis. Exp.* 181, doi: 10.3791/63737

Arai, H., Takamatsu, T., Lin, S. R., Mizutani, T., Omatsu, T., Katayama, Y., et al. (2022c). Distinct effects of the male-killing bacteria *Wolbachia* and *Spiroplasma* and a partiti-like virus in the tea pest moth, *Homona magnanima*. *Biorxiv* [Preprint]. doi: 10.1101/2022.04.29.490121

Arai, H., Anbutsu, H., Nishikawa, Y., Kogawa, M., Ishii, K., Hosokawa, M., et al. (2022b). Male-killing-associated bacteriophage WO identified from comparisons of *Wolbachia* endosymbionts of *Homona magnanima*. *Biorxiv* [Preprint]. doi: 10.1101/2022.06.12.495854v2

Arai, H., Hirano, T., Akizuki, N., Abe, A., Nakai, M., Kunimi, Y., et al. (2019). Multiple infection and reproductive manipulations of *Wolbachia* in *Homona magnanima* (*Lepidoptera*: *Tortricidae*). *Microb. Ecol.* 77, 257–266. doi: 10.1007/s00248-018-1210-4

Arai, H., Lin, S. R., Nakai, M., Kunimi, Y., and Inoue, M. N. (2020). Closely related male-killing and nonmale-killing *Wolbachia* strains in the Oriental tea tortrix *Homona magnanima*. *Microb. Ecol.* 79, 1011–1020. doi: 10.1007/s00248-019-01469-6

Arndt, D., Grant, J. R., Marcu, A., Sajed, T., Pon, A., Liang, Y., et al. (2016). PHASTER: A better, faster version of the PHAST phage search tool. *Nucleic Acids Res.* 44, W16–W21. doi: 10.1093/nar/gkw387

Bai, X., and Hogenhout, S. A. (2002). A genome sequence survey of the mollicute corn stunt *Spiroplasma Spiroplasma kunkelii*. *FEMS Microbiol. Lett.* 210, 7–17. doi: 10.1111/j.1574-6968.2002.tb11153.x

Baldo, L., Dunning Hotopp, J. C., Jolley, K. A., Bordenstein, S. R., Biber, S. A., Choudhury, R. R., et al. (2006). Multilocus sequence typing system for the endosymbiont *Wolbachia pipientis*. *Appl. Environ. Microbiol.* 72, 7098–7110. doi: 10.1128/AEM.00731-06

Ballinger, M. J., Gawryluk, R. M., and Perlman, S. J. (2019). Toxin and genome evolution in a *Drosophila* defensive symbiosis. *Genome Biol. Evol.* 11, 253–262. doi: 10.1093/gbe/evy272

Beckmann, J. F., Ronau, J. A., and Hochstrasser, M. (2017). A *Wolbachia* deubiquitylating enzyme induces cytoplasmic incompatibility. *Nat. Microbiol.* 2, 1–7. doi: 10.1038/nmicrobiol.2017.7

Binetruy, F., Bailly, X., Chevillon, C., Martin, O. Y., Bernasconi, M. V., and Duron, O. (2019). Phylogenetics of the *Spiroplasma ixodetis* endosymbiont reveals past transfers between ticks and other arthropods. *Ticks Tick Borne Dis.* 10, 575–584. doi: 10.1016/j.ttbdis.2019.02.001

Blow, F., and Douglas, A. E. (2019). The hemolymph microbiome of insects. *J. Insect Physiol.* 115, 33–39. doi: 10.1016/j.jinsphys.2019.04.002

Brandt, J. W., Chevignon, G., Oliver, K. M., and Strand, M. R. (2017). Culture of an aphid heritable symbiont demonstrates its direct role in defence against parasitoids. *Proc. R. Soc. Lond. B Biol. Soc.* 284:20171925. doi: 10.1098/rspb.2017.1925

Brüssow, H., Canchaya, C., and Hardt, W. D. (2004). Phages and the evolution of bacterial pathogens: From genomic rearrangements to lysogenic conversion. *Microbiol. Mol. Biol. Rev.* 68, 560–602. doi: 10.1128/MMBR.68.3.560-602.2004

Carle, P., Saillard, C., Carrère, N., Carrère, S., Duret, S., Eveillard, S., et al. (2010). Partial chromosome sequence of *Spiroplasma citri* reveals extensive viral invasion and important gene decay. *Appl. Environ. Microbiol.* 76, 3420–3426. doi: 10.1128/AEM.02954-09

Chang, T. H., Lo, W. S., Ku, C., Chen, L. L., and Kuo, C. H. (2014). Molecular evolution of the substrate utilization strategies and putative virulence factors in mosquito-associated *Spiroplasma* species. *Genome Biol. Evol.* 6, 500–509. doi: 10.1093/gbe/evu033

Duperron, S., Pottier, M. A., Léger, N., Gaudron, S. M., Puillandre, N., Le Prieur, S. L., et al. (2013). A tale of two chitons: Is habitat specialisation linked to distinct associated bacterial communities? *FEMS Microbiol. Ecol.* 83, 552–567. doi: 10.1111/1574-6941.12014

Duplouy, A., Iturbe-Ormaetxe, I., Beatson, S. A., Szubert, J. M., Brownlie, J. C., McMeniman, C. J., et al. (2013). Draft genome sequence of the male-killing *Wolbachia* strain wBol1 reveals recent horizontal gene transfers from diverse sources. *BMC Genom.* 14:20. doi: 10.1186/1471-2164-14-20

Duron, O., Bouchon, D., Boutin, S., Bellamy, L., Zhou, L., Engelstädter, J., et al. (2008). The diversity of reproductive parasites among arthropods: *Wolbachia* do not walk alone. *BMC Biol.* 6:27. doi: 10.1186/1741-7007-6-27

Eichinger, V., Nussbaumer, T., Platzer, A., Jehl, M. A., Arnold, R., and Rattei, T. (2016). EffectiveDB—updates and novel features for a better annotation of bacterial secreted proteins and Type III, IV, VI secretion systems. *Nucleic Acids Res.* 44, D669–D674. doi: 10.1093/nar/gkv1269

Fallon, A. M. (2021). Growth and maintenance of *Wolbachia* in insect cell lines. *Insects* 12:706. doi: 10.3390/insects12080706

Faria, V. G., and Sucena, E. (2013). *Wolbachia* in the Malpighian tubules: Evolutionary dead-end or adaptation?. *J. Exp. Zool. B Mol. Dev. Evol.* 320, 195–199. doi: 10.1002/jez.b.22498

Fujita, R., Inoue, M. N., Takamatsu, T., Arai, H., Nishino, M., Abe, N., et al. (2021). Late male-killing viruses in *Homona magnanima* identified as Osugoroshi viruses, novel members of *Partitiviridae*. *Front. Microbiol.* 11:620623. doi: 10.3389/fmicb.2020.620623

Garcia-Arraez, M. G., Masson, F., Escobar, J. C. P., and Lemaitre, B. (2019). Functional analysis of RIP toxins from the *Drosophila* endosymbiont *Spiroplasma poulsonii*. *BMC Microbiol.* 19:46. doi: 10.1186/s12866-019-1410-1

Gerth, M., Martinez-Montoya, H., Ramirez, P., Masson, F., Griffin, J. S., Aramayo, R., et al. (2021). Rapid molecular evolution of *Spiroplasma* symbionts of *Drosophila*. *Microb. Genom.* 7:000503. doi: 10.1099/mgen.0.000503

Hackett, K. J., Lynn, D. E., Williamson, D. L., Ginsberg, A. S., and Whitcomb, R. F. (1986). Cultivation of the *Drosophila* sex-ratio *Spiroplasma*. *Science* 232, 1253–1255. doi: 10.1126/science.232.4755.1253

Hajibabaei, M., Janzen, D. H., Burns, J. M., Hallwachs, W., and Hebert, P. D. (2006). DNA barcodes distinguish species of tropical *Lepidoptera*. *Proc. Natl. Acad. Sci. U. S. A.* 103, 968–971. doi: 10.1073/pnas.051046610

Hamilton, P. T., Peng, F., Boulanger, M. J., and Perlman, S. J. (2016). A ribosome-inactivating protein in a *Drosophila* defensive symbiont. *Proc. Natl. Acad. Sci. U. S. A.* 113, 350–355. doi: 10.1073/pnas.1518648113

Harumoto, T., and Lemaitre, B. (2018). Male-killing toxin in a bacterial symbiont of *Drosophila*. *Nature* 557, 252–255. doi: 10.1038/s41586-018-0086-2

Hayashi, M., Nomura, M., and Kageyama, D. (2018). Rapid comeback of males: Evolution of male-killer suppression in a green lacewing population. *Proc. R. Soc. Lond. B Biol. Sci.* 285:20180369. doi: 10.1098/rspb.2018.0369

Hayashi, M., Watanabe, M., Yukuhiro, F., Nomura, M., and Kageyama, D. (2016). A nightmare for males? A maternally transmitted male-killing bacterium and strong female bias in a green lacewing population. *PLoS One* 11:e0155794. doi: 10.1371/journal.pone.0155794

He, L. S., Zhang, P. W., Huang, J. M., Zhu, F. C., Danchin, A., and Wang, Y. (2018). The enigmatic genome of an obligate ancient *Spiroplasma* symbiont in a Hadal Holothurian. *Appl. Environ. Microbiol.* 84, e01965–17. doi: 10.1128/AEM.01965-17

Hornett, E. A., Charlat, S., Duplouy, A. M. R., Davies, N., Roderick, G. K., Wedell, N., et al. (2006). Evolution of male-killer suppression in a natural population. *PLoS Biol.* 4:e283. doi: 10.1371/journal.pbio.0040283

Hurst, G. D. D., Graf von der Schulenburg, J. H., Majerus, T. M. O., Bertrand, D., Zakharov, I. A., Baungaard, J., et al. (1999). Invasion of one insect species, *Adalia bipunctata*, by two different male-killing bacteria. *Insect Mol. Biol.* 8, 133–139. doi: 10.1046/j.1365-2583.1999.810133.x

Hurst, G. D., and Jiggins, F. M. (2000). Male-killing bacteria in insects: Mechanisms, incidence, and implications. *Emerg. Infect. Dis.* 6, 329–336. doi: 10.3201/eid0604.000402

Hurst, L. D. (1991). The incidences and evolution of cytoplasmic male killers. *Proc. R. Lond. Soc. B Biol. Soc.* 244, 91–99. doi: 10.1098/rspb.1991.0056

Jiggins, F. M., Hurst, G. D., Jiggins, C. D., v d Schulenburg, J. H., and Majerus, M. E. (2000). The butterfly *Danaus chrysippus* is infected by a male-killing *Spiroplasma* bacterium. *Parasitology* 120, 439–446. doi: 10.1017/s0031182099005867

Joshi, B. D., Berg, M., Rogers, J., Fletcher, J., and Melcher, U. (2005). Sequence comparisons of plasmids pBJS-O of *Spiroplasma citri* and pSKU146 of *S. kunkelii*: Implications for plasmid evolution. *BMC Genom.* 6:175. doi: 10.1186/1471-2164-6-175

Kageyama, D., Anbutsu, H., Shimada, M., and Fukatsu, T. (2007). *Spiroplasma* infection causes either early or late male killing in *Drosophila*, depending on maternal host age. *Naturwissenschaften* 94, 333–337. doi: 10.1007/s00114-006-0195-x

Kageyama, D., Narita, S., and Watanabe, M. (2012). Insect sex determination manipulated by their endosymbionts: Incidences, mechanisms and implications. *Insects* 3, 161–199. doi: 10.3390/insects3010161

Kent, B. N., and Bordenstein, S. R. (2010). Phage WO of *Wolbachia*: Lambda of the endosymbiont world. *Trends Microbiol.* 18, 173–181. doi: 10.1016/j.tim.2009.12.011

Koren, S., Walenz, B. P., Berlin, K., Miller, J. R., Bergman, N. H., and Phillippy, A. M. (2017). Canu: Scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res.* 27, 722–736. doi: 10.1101/gr.215087.116

Ku, C., Lo, W. S., Chen, L. L., and Kuo, C. H. (2013). Complete genomes of two dipteran-associated *Spiroplasmas* provided insights into the origin, dynamics, and impacts of viral invasion in *Spiroplasma*. *Genome Biol. Evol.* 5, 1151–1164. doi: 10.1093/gbe/evt084

Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* 33, 1870–1874. doi: 10.1093/molbev/msw054

LePage, D. P., Metcalf, J. A., Bordenstein, S. R., On, J., Perlmutter, J. I., Shropshire, J. D., et al. (2017). Prophage WO genes recapitulate and enhance *Wolbachia*-induced cytoplasmic incompatibility. *Nature* 543, 243–247. doi: 10.1038/nature21391

Li, H. (2018). Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics* 34, 3094–3100. doi: 10.1093/bioinformatics/bty191

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* 25, 2078–2079. doi: 10.1093/bioinformatics/btp352

Lo, W. S., Chen, L. L., Chung, W. C., Gasparich, G. E., and Kuo, C. H. (2013). Comparative genome analysis of *Spiroplasma melliferum* IPMB4A, a honeybee-associated bacterium. *BMC Genom.* 14:22. doi: 10.1186/1471-2164-14-22

Lo, W. S., Gasparich, G. E., and Kuo, C. H. (2015). Found and lost: The fates of horizontally acquired genes in arthropod-symbiotic *Spiroplasma*. *Genome Biol. Evol.* 7, 2458–2472. doi: 10.1093/gbe/evv160

Majerus, M. E., Hinrich, J., Schulenburg, G. V. D., and Zakharov, I. A. (2000). Multiple causes of male-killing in a single sample of the two-spot ladybird, *Adalia bipunctata* (*Coleoptera*: *Coccinellidae*) from Moscow. *Heredity* 84, 605–609. doi: 10.1046/j.1365-2540.2000.00710.x

Martin, S. H., Singh, K. S., Gordon, I. J., Omufwoko, K. S., Collins, S., Warren, I. A., et al. (2020). Whole-chromosome hitchhiking driven by a male-killing endosymbiont. *PLoS Biol.* 18:e3000610. doi: 10.1371/journal.pbio.3000610

Massey, J. H., and Newton, I. L. (2022). Diversity and function of arthropod endosymbiont toxins. *Trends Microbiol.* 30, 185–198. doi: 10.1016/j.tim.2021.06.008

Masson, F., Calderon Copete, S., Schüpfer, F., Garcia-Arraez, G., and Lemaitre, B. (2018). In vitro culture of the insect endosymbiont *Spiroplasma poulsonii* highlights bacterial genes involved in host-symbiont interaction. *Mbio* 9, e00024–18. doi: 10.1128/mBio.00024-18

Mitsuhashi, J. (1981). A new continuous cell line from larvae of the mosquito *Aedes albopictus* (*Diptera*, *Culicidae*). *Biomed. Res.* 2, 599–606. doi: 10.2220/biomedres.2.599

Mouches, C., Barroso, G., Gadeau, A., and Bové, J. M. (1984). Characterization of two cryptic plasmids from *Spiroplasma citri* and occurrence of their DNA sequences among various *Spiroplasmas*. *Ann. Inst. Pasteur Microbiol.* 135A, 17–24. doi: 10.1016/s0769-2609(84)80054-x

Paredes, J. C., Herren, J. K., Schüpfer, F., Marin, R., Claverol, S., Kuo, C. H., et al. (2015). Genome sequence of the *Drosophila melanogaster* male-killing *Spiroplasma* strain MSRO endosymbiont. *Mbio* 6, e02437–14. doi: 10.1128/mBio.02437-14

Perlmutter, J. I., Bordenstein, S. R., Unckless, R. L., LePage, D. P., Metcalf, J. A., Hill, T., et al. (2019). The phage gene *wmk* is a candidate for male killing by a bacterial endosymbiont. *PLoS Pathog.* 15:e1007936. doi: 10.1371/journal.ppat.1007936

Perlmutter, J. I., Meyers, J. E., and Bordenstein, S. R. (2021). A single synonymous nucleotide change impacts the male-killing phenotype of prophage WO gene *wmk*. *Elife* 10:e67686. doi: 10.7554/eLife.67686

Pietri, J. E., DeBruhl, H., and Sullivan, W. (2016). The rich somatic life of *Wolbachia*. *Microbiologyopen* 5, 923–936. doi: 10.1002/mbo3.390

Pollmann, M., Moore, L. D., Krimmer, E., D'Alvise, P., Hasselmann, M., Perlman, S. J., et al. (2022). Highly transmissible cytoplasmic incompatibility by the extracellular insect symbiont *Spiroplasma*. *Iscience* 25:104335. doi: 10.1016/j.isci.2022.104335

Ramirez, P., Leavitt, J. C., Gill, J. J., and Mateos, M. (2021). Preliminary characterization of phage-like particles from the male-killing mollicute *Spiroplasma poulsonii* (an endosymbiont of *Drosophila*). *Biorxiv* [Preprint]. doi: 10.1101/2021.12.09.471767

Regassa, L. B., and Gasparich, G. E. (2006). *Spiroplasmas*: Evolutionary relationships and biodiversity. *Front. Biosci.* 11:2983–3002. doi: 10.2741/2027

Sanada-Morimura, S., Matsumura, M., and Noda, H. (2013). Male killing caused by a *Spiroplasma* symbiont in the small brown planthopper, *Laodelphax striatellus*. *J. Hered.* 104, 821–829. doi: 10.1093/jhered/est052

Simon, J. C., Boutin, S., Tsuchida, T., Koga, R., Le Gallic, J. F., Frantz, A., et al. (2011). Facultative symbiont infections affect aphid reproduction. *PLoS One* 6:e21831. doi: 10.1371/journal.pone.0021831

Smith, D. A., Gordon, I. J., Traut, W., Herren, J., Collins, S., Martins, D. J., et al. (2016). A neo-W chromosome in a tropical butterfly links colour pattern, male-killing, and speciation. *Proc. R. Soc. Lond. B Biol. Sci.* 283:20160821. doi: 10.1098/rspb.2016.0821

Tabata, J., Hattori, Y., Sakamoto, H., Yukuhiro, F., Fujii, T., Kugimiya, S., et al. (2011). Male killing and incomplete inheritance of a novel *Spiroplasma* in the moth *Ostrinia zaguliaevi*. *Microb. Ecol.* 61, 254–263. doi: 10.1007/s00248-010-9799-y

Takahashi, T., Murakami, H., Imanishi, S., Miyazaki, M., Kamiie, K., Suzuki, K., et al. (2006). Calreticulin is transiently induced after immunogen treatment in the fat body of the silkworm *Bombyx mori*. *J. Insect Biotechnol. Sericol.* 75, 79–84. doi: 10.11416/jibs.75.79

Takamatsu, T., Arai, H., Abe, N., Nakai, M., Kunimi, Y., and Inoue, M. N. (2021). Coexistence of two male-killers and their impact on the development of oriental tea tortrix *Homona magnanima*. *Microb. Ecol.* 81, 193–202. doi: 10.1007/s00248-020-01566-x

Tanizawa, Y., Fujisawa, T., and Nakamura, Y. (2018). DFAST: A flexible prokaryotic genome annotation pipeline for faster genome publication. *Bioinformatics* 34, 1037–1039. doi: 10.1093/bioinformatics/btx713

Thu, M. J., Qiu, Y., Kataoka-Nakamura, C., Sugimoto, C., Katakura, K., Isoda, N., et al. (2019). Isolation of *Rickettsia*, *Rickettsiella*, and *Spiroplasma* from questing ticks in Japan using arthropod cells. *Vector Borne Zoonotic Dis.* 19, 474–485. doi: 10.1089/vbz.2018.2373

Tsai, Y. M., Chang, A., and Kuo, C. H. (2018). Horizontal gene acquisitions contributed to genome expansion in insect-symbiotic *Spiroplasma clarkii*. *Genome Biol. Evol.* 10, 1526–1532. doi: 10.1093/gbe/evy113

Tsugeno, Y., Koyama, H., Takamatsu, T., Nakai, M., Kunimi, Y., and Inoue, M. N. (2017). Identification of an early male-killing agent in the oriental tea tortrix, *Homona magnanima*. *J. Hered.* 108, 553–560. doi: 10.1093/jhered/esx049

Vera-Ponce León, A., Dominguez-Mirazo, M., Bustamante-Brito, R., Higareda-Alvear, V., Rosenblueth, M., and Martínez-Romero, E. (2021). Functional genomics of a *Spiroplasma* associated with the carmine cochineals *Dactylopius coccus* and *Dactylopius opuntiae*. *BMC Genom.* 22:240. doi: 10.1186/s12864-021-07540-2

Viver, T., Orellana, L. H., Hatt, J. K., Urdiain, M., Díaz, S., Richter, M., et al. (2017). The low diverse gastric microbiome of the jellyfish *Cotylorhiza tuberculata* is dominated by four novel taxa. *Environ. Microbiol.* 19, 3039–3058. doi: 10.1111/1462-2920.13763

Waldor, M. K., and Mekalanos, J. J. (1996). Lysogenic conversion by a filamentous phage encoding cholera toxin. *Science* 272, 1910–1914. doi: 10.1126/science.272.5270.1910

Walker, B. J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., et al. (2014). Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* 9:e112963. doi: 10.1371/journal.pone.0112963

Watanabe, M., Tagami, Y., Miura, K., Kageyama, D., and Stouthamer, R. (2012). Distribution patterns of *Wolbachia* endosymbionts in the closely related flower bugs of the genus *Orius*: Implications for coevolution and horizontal transfer. *Microb. Ecol.* 64, 537–545. doi: 10.1007/s00248-012-0042-x

Werren, J. H., Baldo, L., and Clark, M. E. (2008). *Wolbachia*: Master manipulators of invertebrate biology. *Nat. Rev. Microbiol.* 6, 741–751. doi: 10.1038/nrmicro1969

Wick, R. R., Judd, L. M., Gorrie, C. L., and Holt, K. E. (2017). Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Comp. Biol.* 13:e1005595. doi: 10.1371/journal.pcbi.1005595

Wiedenbeck, J., and Cohan, F. M. (2011). Origins of bacterial diversity through horizontal genetic transfer and adaptation to new ecological niches. *FEMS Microbiol. Rev.* 35, 957–976. doi: 10.1111/j.1574-6976.2011.00292.x

Williamson, D. L., and Poulson, D. F. (1979). "Plant and insect mycoplasmas," in *The Mycoplasmas, Volume III*, eds R. F. Whitcomb and J. G. Tully (New York, NY: Academic Press), 175–208.

Ye, F., Melcher, U., Rascoe, J. E., and Fletcher, J. (1996). Extensive chromosome aberrations in *Spiroplasma citri* strain BR3. *Biochem. Genet.* 34, 269–286. doi: 10.1007/BF02399947

Yeoman, C. J., Brutscher, L. M., Esen, Ö. C., Ibaoglu, F., Fowler, C., Eren, A. M., et al. (2019). Genome-resolved insights into a novel *Spiroplasma* symbiont of the wheat stem sawfly (*Cephus cinctus*). *Peerj* 7:e7548. doi: 10.7717/peerj.7548

Yoshida, K., Sanada-Morimura, S., Huang, S. H., and Tokuda, M. (2021). Silence of the killers: Discovery of male-killing suppression in a rearing strain of the small brown planthopper, *Laodelphax striatellus*. *Proc. R. Lond. B Biol. Sci.* 288:20202125. doi: 10.1098/rspb.2020.2125

Zhou, W., Rousset, F., and O'Neil, S. (1998). Phylogeny and PCR–based classification of *Wolbachia* strains using wsp gene sequences. *Proc. R. Lond. B Biol. Sci.* 265, 509–515. doi: 10.1098/rspb.1998.0324

frontiers | Frontiers in Microbiology

TYPE Original Research
PUBLISHED 18 May 2023
DOI 10.3389/fmicb.2023.1148263

OPEN ACCESS

EDITED BY
Akiko Sugio,
Institut National de la Recherche Agronomique, France

REVIEWED BY
Hiroshi Arai,
National Agriculture and Food Research Organization (NARO), Japan
Emily Hornett,
University of Liverpool, United Kingdom

*CORRESPONDENCE
Logan D. Moore
✉ ldm427@msstate.edu

RECEIVED 19 January 2023
ACCEPTED 19 April 2023
PUBLISHED 18 May 2023

CITATION
Moore LD and Ballinger MJ (2023) The toxins of vertically transmitted *Spiroplasma*. *Front. Microbiol.* 14:1148263. doi: 10.3389/fmicb.2023.1148263

The toxins of vertically transmitted *Spiroplasma*

Logan D. Moore* and Matthew J. Ballinger

Department of Biological Sciences, Mississippi State University, Mississippi State, MS, United States

Vertically transmitted (VT) microbial symbionts play a vital role in the evolution of their insect hosts. A longstanding question in symbiont research is what genes help promote long-term stability of vertically transmitted lifestyles. Symbiont success in insect hosts is due in part to expression of beneficial or manipulative phenotypes that favor symbiont persistence in host populations. In *Spiroplasma*, these phenotypes have been linked to toxin and virulence domains among a few related strains. However, these domains also appear frequently in phylogenetically distant *Spiroplasma,* and little is known about their distribution across the *Spiroplasma* genus. In this study, we present the complete genome sequence of the *Spiroplasma* symbiont of *Drosophila atripex*, a non-manipulating member of the Ixodetis clade of *Spiroplasma*, for which genomic data are still limited. We perform a genus-wide comparative analysis of toxin domains implicated in defensive and reproductive phenotypes. From 12 VT and 31 non-VT *Spiroplasma* genomes, ribosome-inactivating proteins (RIPs), OTU-like cysteine proteases (OTUs), ankyrins, and ETX/MTX2 domains show high propensity for VT *Spiroplasma* compared to non-VT *Spiroplasma*. Specifically, OTU and ankyrin domains can be found only in VT-*Spiroplasma*, and RIP domains are found in all VT *Spiroplasma* and three non-VT *Spiroplasma*. These domains are frequently associated with *Spiroplasma* plasmids, suggesting a possible mechanism for dispersal and maintenance among heritable strains. Searching insect genome assemblies available on public databases uncovered uncharacterized *Spiroplasma* genomes from which we identified several *spaid*-like genes encoding RIP, OTU, and ankyrin domains, suggesting functional interactions among those domain types. Our results suggest a conserved core of symbiont domains play an important role in the evolution and persistence of VT *Spiroplasma* in insects.

KEYWORDS

***Spiroplasma*, Ixodetis, ETX/MTX2, OTU-like cysteine protease, ribosome-inactivating proteins, ankyrin, vertical transmission**

Introduction

Gene duplications, losses, and horizontal transfers can facilitate dramatic shifts in bacterial lifestyle and capabilities (Romero and Palacios, 1997; Moran, 2002; Arnold et al., 2022). Gene loss is a dominant feature of symbiont evolution due to the selective benefits of removing metabolically costly genes (McCutcheon and Moran, 2012; McCutcheon et al., 2019). Alternatively, gene gain via horizontal transfer and duplication can lead to rapid adaptation across symbiotic species. For example, several mutualistic soil bacteria with nitrogen-fixing capabilities have benefited from receiving "symbiont islands" and plasmids enriched with nitrogen-fixing genes (Sullivan and Ronson, 1998; Andrews et al., 2018). Vertically transmitted symbionts are no exception to this phenomenon and can acquire novel phenotypes that benefit

a heritable lifestyle via the transfer of symbiosis-supporting genes (Ballinger et al., 2019; Manzano-Marin et al., 2020; Waterworth et al., 2020).

Vertically transmitted microbial symbionts can strongly influence the life history traits of their insect hosts to maintain stable vertical transmission (VT) (Moran et al., 2008). For example, the facultative endosymbiont *Wolbachia* can defend against viral infections in their *Drosophila* hosts (Teixeira et al., 2008) but can also be penetrant reproductive parasites of *Drosophila*, using manipulative phenotypes like male-killing and cytoplasmic incompatibility to favor their transmission to subsequent generations. Heritable symbionts are ubiquitous among insects (Moran et al., 2008), and yet, a central question remains for many heritable microbes—what gene effectors promote vertical transmission and how do they arise?

Spiroplasma is a genus of helical, cell wall-less bacteria estimated to infect up to 7% of terrestrial arthropod species (Duron et al., 2008). The genus is phylogenetically structured into three large clades—the Apis clade, the Citri-Chrysopicola-Mirum clade, and the Ixodetis clade (Gasparich et al., 2004). Despite their rich diversity, all *Spiroplasma* are non-free-living bacteria, infecting the gut or hemolymph of invertebrates and/or the phloem of plants. At ecological timescales, *Spiroplasma* infections are transmitted horizontally by ingestion or maternally via the cytoplasm of mature oocytes (Herren et al., 2013). The Apis clade contains several insect pathogens, including *Spiroplasma taiwanense* and *Spiroplasma culicicola* of mosquitoes and *Spiroplasma apis* of European honey-bees (Mouches et al., 1982; Humphery-Smith et al., 1991; Bolaños et al., 2015). No VT *Spiroplasma* have been described in the Apis clade, while both VT and non-VT *Spiroplasma* have been reported in the Citri-Chrysopicola-Mirum clade (hereafter referred to as the Citri clade) and the Ixodetis clade (Jiggins et al., 2000; Haselkorn, 2010; Ballinger et al., 2018; Binetruy et al., 2019; Vera-Ponce León et al., 2021).

Heritable *Spiroplasma* have been described across insect orders, and are well-studied in vinegar flies, butterflies, beetles, aphids and parasitoid wasps. Some members of the Citri clade defend *Drosophila* hosts against parasitic nematodes and parasitoid wasps (Xie et al., 2010; Jaenike et al., 2010bcc Haselkorn and Jaenike, 2015). Others in this clade are both male-killers and defensive symbionts of *Drosophila* (Montenegro et al., 2005; Xie et al., 2014). Reproductive manipulation is also common among Ixodetis clade *Spiroplasma*. Ixodetis *Spiroplasma* of the butterfly *Danaus chrysippus* (*sChrys*) and the ladybird beetle *Anisosticta novemdecimpunctata* are effective VT male-killers (Jiggins et al., 2000; Tinsley and Majerus, 2006) and the recently discovered *Spiroplasma* symbiont that causes cytoplasmic incompatibility in the parasitoid wasp *Lariophagus distinguendus* (*sDis*) is also an Ixodetis clade member (Pollmann et al., 2022). Cytoplasmic incompatibility is a form of reproductive manipulation whereby toxic or modified sperm of symbiont-infected fathers kills the offspring of uninfected mothers, thus driving symbiont spread through the population. As defensive symbionts, Ixodetis *Spiroplasma* have been shown to confer resistance to fungal infections in their aphid hosts (Łukasik et al., 2013). Indeed, VT strains are present in divergent *Spiroplasma* clades, yet little is known about the genetic mechanisms supporting their host-associated phenotypes beyond those of the Poulsonii group in the Citri clade. Within the Poulsonii group, symbiont toxins have been implicated in both defensive and reproductive phenotypes (Hamilton et al., 2016; Ballinger and Perlman, 2017; Harumoto and Lemaitre, 2018).

Toxic peptides are known effectors of defensive and reproductive phenotypes across symbiotic bacterial taxa (Gillespie et al., 2018; Oliver and Perlman, 2020; Massey and Newton, 2022). *Hamiltonella defensa* utilizes phage-encoded toxins to defend aphid hosts against parasitoid wasps (Degnan and Moran, 2008). *Wolbachia* symbionts employ a toxin-antitoxin or host modification system to inflict cytoplasmic incompatibility in their hosts (Beckmann et al., 2017; LePage et al., 2017; Horard et al., 2022; Kaur et al., 2022). Within the Citri-based Poulsonii group, including *Spiroplasma* strains infecting *Drosophila hydei, Drosophila neotestacea,* and *Drosophila melanogaster* (*sHyd*, *sNeo*, and MSRO respectively), defensive strains utilize ribosome-inactivating protein toxins (RIPs) to contribute to defense against parasitoid wasps and parasitic nematodes (Hamilton et al., 2016; Ballinger and Perlman, 2017). RIPs cause irreversible damage to a key enzymatic residue of ribosomes, resulting in cell death (Stirpe, 2004). RIP gene families are remarkably diverse in Citri clade *Spiroplasma*, ostensibly due to extensive horizontal transmission, duplication, loss, and recombination (Ballinger et al., 2019; Gerth et al., 2021). *Spiroplasma poulsonii* strain MSRO is a reproductive manipulator that encodes a multidomain toxin called Spaid that causes male-killing in *Drosophila* (Harumoto and Lemaitre, 2018). Spaid consists of three integral domains including ankyrin repeat domains, an OTU-like cysteine protease domain (OTU) with predicted function in deubiquitination, and a C-terminal transmembrane domain. OTUs and ankyrin domains are also associated with type V cifB-like proteins of CI-inducing *Wolbachia*, suggesting they can be adapted for different reproductive phenotypes in diverse systems (Martinez et al., 2021). While toxins have been identified to play an essential role in maintaining symbiosis in these heritable Citri clade *Spiroplasma*, far less is known about effector genes supporting the diverse Ixodetis *Spiroplasma* phenotypes. However, recent advances in genomic representation among these strains (Yeoman et al., 2019; Martin et al., 2020; Vera-Ponce León et al., 2021; Pollmann et al., 2022) facilitates an in-depth comparative analysis of toxin gene content in this enigmatic clade.

In this study, we report the genome sequence of a *Drosophila*-associated Ixodetis clade member, the *Spiroplasma* symbiont of *Drosophila atripex* (*sAtri*). This strain does not induce male-killing or cytoplasmic incompatibility in *Drosophila* (Haselkorn et al., 2009) but does provide protection against parasitoid wasps in the family Figitidae (T. Chris Amuwa, unpublished results). We identify a diverse set of toxin and virulence domains in the genome, including the same domains contributing to defense and reproductive manipulation in the Citri clade. We conducted a genus-wide survey of 43 *Spiroplasma* genomes across all three clades to characterize the distribution of toxin domains between heritable and non-heritable *Spiroplasma*. Our results uncover a striking association of RIPs, OTUs, ankyrin, and ETX/MTX2 domains with heritable strains, suggesting that these domains play an important role in promoting VT symbiosis across *Spiroplasma*.

Results

Genome of *sAtri*

Using a hybrid long and short read sequencing strategy, we produced a closed genome assembly of 1.27 Mb for *sAtri*. The

genome is covered to 29X depth by short reads (Supplementary Figure S1). Phylogenomic analysis nests *sAtri* with strong support among other Ixodetis clade endosymbiont members that infect a wide range of insect hosts (Figure 1). Genome completeness based on single copy orthologs was estimated by BUSCO at 98% (Table 1). Missing orthologs from the genome completeness analysis include the genes *rplL* and *rsmL*. We identified *rplL* in the *sAtri* genome with a manual BLAST search but were unable to identify *rsmL*. While *rsmL* (rRNA ribose-2'-O-methyltransferase) has a conserved role in bacterial translation, it is absent from all currently available Ixodetis clade genomes (Pollmann et al., 2022). The hybrid genome assembly includes eight putative plasmids, which were assigned as such based on gene content, coverage depth, and circularity (Supplementary Tables S1, S2). Repeat content is 10.73%, but only nine transposable elements could be annotated. Of the 1,313 putative protein coding genes, 639 could be assigned a functional annotation by prokka. *sAtri* encodes complete ATP synthase and glycolysis pathways, as well as transport components for glucose, fructose, GlcNAc, glycerol, and possibly mannose. Summaries of functional predictions and carbon metabolism pathway components are shown in Supplementary Figure S2A and Supplementary Table S3. Amino acid and lipid biosynthesis pathways are incomplete or absent, as reported previously for hemolymph-dwelling *Spiroplasma* species, in both the Citri and Ixodetis clades (Paredes et al., 2015; Yeoman et al., 2019; Vera-Ponce León et al., 2021). In addition to the PTS genes for carbon import (Supplementary Figure S2B), ABC membrane proteins for transport of phosphate (*ptsABCS*), nucleosides (bmpA, *nupABC*), and micronutrients (*ecfA1, ecfA2*) could be identified. A comparative summary of membrane transport, biosynthesis, and metabolism genes in *sAtri* and two other Ixodetis clade genomes is present in Supplementary Table S3. *sAtri* also encodes a Type II-C CRISPR/Cas9 system which has not been previously reported in a heritable *Spiroplasma*. No matches between *sAtri* spacers and known phages or phage regions were identified. However, several of the spacers in *sAtri* are exact matches to sequences present on *sAtri* plasmids, suggesting a possible interaction with CRISPR/Cas9 (e.g., plasmid copy number suppression). *sAtri* spacers also match to non-phage genomic sequences of other Ixodetis *Spiroplasma*, though it is unclear if those sequences are associated with plasmids.

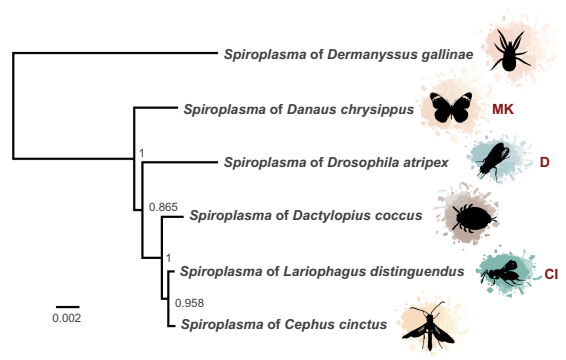


FIGURE 1
Phylogeny of heritable Ixodetis *Spiroplasma*. Midpoint-rooted phylogeny of 67 genes in ten syntenous blocks (72kb) from available heritable Ixodetis genomes shows relationship of *sAtri* to its closest neighbors. Five gene loci from the Ixodetis *Spiroplasma* of *Dermanyssus gallinae* were used in alignment to provide additional context for phylogeny grouping. Known phenotypes are indicated with abbreviations: CI=cytoplasmic incompatibility and MK=male killing. Node labels indicate ML-like FastTree support values.

sAtri encodes a diverse set of toxins and virulence genes

The *sAtri* genome assembly reveals similarities to and departures from other *Spiroplasma* with regard to toxin and virulence gene content. Three RIP domain-containing proteins are present, two of which encode a signal peptide for secretion. RIP toxins contribute to defense against natural enemies in some Citri clade members (Hamilton et al., 2016; Ballinger and Perlman, 2017). The *sAtri* RIPs are phylogenetically distant from one another (Figure 2) and from other non-Ixodetis *Spiroplasma* RIPs with less than 30% similarity at the amino acid level. We also identified RIPs in the genome assemblies of the *Spiroplasma* symbiont of *Danaus chrysippus* (*sChrys*) and the *Spiroplasma* symbiont of *Cephus cinctus* (*sCinc*), and retrieved the RIPs reported in the *Spiroplasma* symbiont of *Lariophagus distinguendus* (*sDis*) and the *Spiroplasma* symbiont of *Dactylopius coccus* (*sCoccus*). Examining the phylogenetic relationships of Ixodetis clade RIPs reveals them to be polyphyletic, grouping separately with distinct clades of Citri RIPs. This is consistent with ancient ancestry or horizontal transfer within the genus (Ballinger et al., 2019). *sAtri* encodes two ETX/MTX2 toxins, both of which possess signal peptides for secretion. ETX/MTX2 toxins are β pore-forming toxins that inflict cytocidal activity against their target cells and are noted for their insecticidal activity (Palma et al., 2014; Moar et al., 2017). Like *sAtri* RIPs, these toxins are diverse, grouping separately with other Ixodetis clade-encoded ETX/MTX2 toxins (Supplementary Figure S3). ETX/MTX2 toxins are numerous and widespread in *Spiroplasma* but their potential role in host-symbiont relationships is unknown. *sAtri* encodes 11 total genes containing ankyrin domains (Supplementary Figure S4) and a separate gene with an OTU-like cysteine protease. While not inherently toxic, ankyrin and OTU domains are associated with genes involved in host manipulation, and work in conjunction to cause male-killing by MSRO and possibly cytoplasmic incompatibility by *Wolbachia* (Pan et al., 2008; Nguyen et al., 2014; Harumoto and Lemaitre, 2018; Martinez et al., 2021). The OTU of *sAtri* is highly conserved among other heritable Ixodetis members (Supplementary Figure S5) and does not share an open reading frame with ankyrin domains. Despite the presence of these domains, no sex-ratio distortion has been observed in previous studies (Haselkorn et al., 2009) or in our lab stock (T. Chris Amuwa, unpublished results).

sAtri also encodes virulence genes that are scarcely found or not yet described in the *Spiroplasma* genus. These include an adenylate cyclase domain annotated by HMMER as anthrax toxin LF subunit. This domain is also found on the edema factor of *Bacillus anthracis* anthrax toxin. The edema factor is itself an adenylate cyclase toxin that causes large increases in the ubiquitous signaling molecule cAMP, resulting in severe disruptions to cellular processes (Tang and Guo, 2009).

sAtri encodes two metalloproteases of the M60 family which are implicated in virulence and infection outcomes in some bacteria and viruses (Nakjang et al., 2012; Belousov et al., 2018). M60-like

TABLE 1 Genome assembly statistics of *Spiroplasma*.

Genome	Size (bp)	Contigs	%GC	N50	CDS	tRNAs
Spiroplasma of *Drosophila atripex*	1,271,056	1	23.7	Closed	1,506	27
Spiroplasma of *Lariophagus distinguendus*††	1,163,832	198	24.3	14,219	1,175	27
Spiroplasma of *Cephus cinctus*	713,566	145	24.9	5,160	754	23
Spiroplasma of *Danaus chrysippus*†	1,745,430	12	23.7	215,399	1,782	27
Spiroplasma of *Dactylopius coccus* (DCF)	1,195,508	286	23.7	6,014	1,253	27
Spiroplasma poulsonii strain MSRO†,‡	1,883,005	1	26.4	Closed	2,217	31

†Male-killing strain.
††CI-inducing strain.
‡Citri clade.

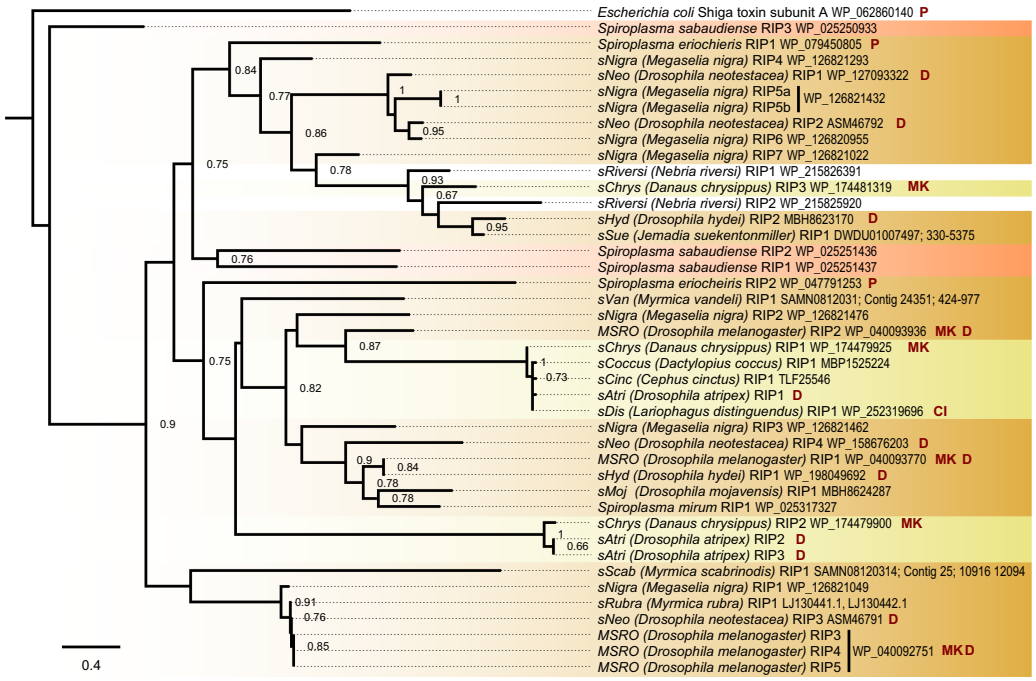


FIGURE 2
Ixodetis *Spiroplasma* encode diverse RIPs. FastTree phylogeny built from a MAFFT alignment of *Spiroplasma* RIP domain amino acid sequences and rooted to *Escherichia coli* shiga toxin subunit A. Ixodetis clade *Spiroplasma* possess multiple RIP-domain bearing proteins with a distribution explained by gains, losses, duplications, and horizontal transfers. Red shading indicates Apis clade, orange shading indicates Citri clade and yellow shading indicates Ixodetis clade. Abbreviations indicate known phenotypes of the strain from which each domain was extracted: (P: pathogenic, MK: male-killing, CI: cytoplasmic incompatibility-inducing, D: defensive). *sAtri* RIP domains are highlighted in bold. Support values greater than or equal to 0.60 are shown.

peptidases, also commonly referred to as enhancins, are well studied in, but not restricted to, insect pathogenic dsDNA viruses in the family Baculoviridae and function by binding to and degrading mucin layers that protect epithelial tissues (Ishimwe et al., 2015).

A conserved core of toxin domains in heritable *Spiroplasma*

The discovery that the *Drosophila* symbiont *sAtri* harbors multiple functional domains previously reported in distantly related heritable Citri clade symbionts motivated a more detailed analysis of domain distributions across the genus *Spiroplasma*. We focused the scope of this analysis on RIP, OTU and ankyrin domains due to their role in defensive and reproductive manipulation phenotypes. We also analyzed ETX/MTX2 domains across *Spiroplasma* which are found widespread across this genus with little understanding of their role. VT-capable strains were determined based on a number of qualifying characteristics including evidence of transovarial transmission, PCR detection in hemolymph, ovaries and/or eggs, and presence of reproductive manipulation phenotypes (Table 2). VT *Spiroplasma* strains investigated in this study and their associated hosts, taxonomy and phenotypes are listed in Table 3. OTU and ankyrin domains can be found only in VT *Spiroplasma* across the Citri and Ixodetis clade (Figure 3). RIP domains are found in all VT *Spiroplasma* and in three non-VT *Spiroplasma*. ETX/MTX2 domains are more widely

TABLE 2 Evidence of vertical transmission.

Strain (host)	Evidence of vertical transmission	References
sAtri (Drosophila atripex)	Transgenerational transmission PCR detection in hemolymph Genome sequenced from host ovaries	T. Chris Amuwa (unpublished results) and this study
MSRO (Drosophila melanogaster)	Transgenerational transmission Detection in hemolymph via transinfection experiments FISH visualization in eggs FISH visualization of ovaries Reproductive manipulation	Pool et al. (2006) and Harumoto and Lemaitre (2018) Herren et al. (2013)
sNeo (Drosophila neotestacea)	Transgenerational transmission Detection in hemolymph via transinfection experiments	Jaenike et al. (2010a) and Haselkorn and Jaenike (2015)
sHyd (Drosophila hydei)	Transgenerational transmission Detection in hemolymph via transinfection PCR detection in eggs (unpublished)	Mateos et al. (2006) and Xie et al. (2010)
sMoj (Drosophila mojavensis)	Transgenerational transmission	Haselkorn et al. (2013)
sVan (Myrmica vandeli)	PCR detection in hemolymph High infection frequency (92.3%) Host specificity in sympatry	Ballinger et al. (2018)
sScab (Myrmica scabrinodis)	PCR detection in hemolymph High infection frequency (91.9–97.6%) Host specificity in sympatry	Ballinger et al. (2018)
sNigra (Megaselia nigra)	Transgenerational transmission (unpublished) PCR detection in hemolymph (unpublished) High infection frequency (73.3–80.2%)	Ballinger et al. (2019)
sRiversi (Nebria riversi)	PCR detection in egg, larval and adult stages	Weng et al. (2021)
sChrys (Danaus chrysippus)	Transgenerational transmission Reproductive manipulation	Martin et al. (2020)
sCoccus (Dactylopius coccus)	PCR detection in ovaries	Vera-Ponce León et al. (2021)
sDis (Lariophagus distinguendus)	Transgenerational transmission FISH visualization in ovaries Reproductive manipulation	Pollmann et al. (2022)

distributed among members of the Citri and Ixodetis clades, including most VT *Spiroplasma* and a few non-VT *Spiroplasma*. All four domains are almost entirely absent from the Apis clade except for three RIPs in *Spiroplasma sabaudiense* and a single ETX/MTX2 copy in *Spiroplasma culicicola*. Interestingly, no VT strains have been described in the Apis clade (Bolaños et al., 2015). We used BayesTraits software to identify correlations between distributions of toxin domains and heritability under two different models—one that assumes independent evolution and one that assumes dependent evolution. We find that all four domains are significantly better represented under a dependent model of evolution compared to an independent model (Table 4). For comparison, we used the same approach to determine if a correlation existed between heritability and other putative virulent *Spiroplasma* genes including *glpO* (glycerol 3 phosphate oxidase), *chiA* (chitinase), and *spi* (spiralin) (Alexeev et al., 2012; Chang et al., 2014; Duret et al., 2014). We find little to no support that the distribution of these domains across *Spiroplasma* are associated with heritability (Table 4; Supplementary Figure S6).

VT-specific domains vary in copy number and origin

While RIP, OTU, ankyrin, and ETX/MTX2 domains are distributed across diverse heritable *Spiroplasma*, the number of genes possessing these domains can vary drastically, even between closely related strains (Figure 4). For example, the number of genes possessing RIP domains varies both within and between clades, likely due to gene duplications and horizontal gene transfers. The number of genes possessing ankyrin domains are consistent within clade but show extreme copy number variation between clades characterized by ankyrin enrichment in Ixodetis *Spiroplasma*. Conversely, the number of OTU and ETX/MTX2 copies is relatively consistent within and between clades. All four of these domains can be found on plasmids (Harumoto and Lemaitre, 2018; Ballinger et al., 2019). In some cases, a given strain's entire domain repertoire can only be found on plasmids such as ankyrins and OTUs in MSRO, and ETX/MTX2 in *Spiroplasma citri*. In contrast, only one domain could be found within an endogenous phage region (ETX/MTX2 of MSRO), suggesting

TABLE 3 Vertically transmitted strains of *Spiroplasma* in study.

Host	*Spiroplasma* strain	Strain taxonomy	Phenotype
Drosophila melanogaster	*MSRO*	Citri-Poulsonii	Male-killing; defensive
Drosophila neotestacea	*sNeo*	Citri-Poulsonii	Defensive
Drosophila hydei	*sHyd*	Citri-Poulsonii	Defensive
Myrmica vandeli	*sVan*	Citri	–
Myrmica scabrinodis	*sScab*	Citri	–
Drosophila mojavensis	*sMoj*	Citri	–
Megaselia nigra	*sNigra*	Citri	–
Nebria riversi	*sRiversi*	Unplaced	–
Dactylopius coccus	*sCoccus*	Ixodetis	–
Danaus chrysippus	*sChrys*	Ixodetis	Male-killing
Lariophagus distinguendus	*sDis*	Ixodetis	Cytoplasmic incompatibility
Drosophila atripex	*sAtri*	Ixodetis	Defensive

Dash (–) indicates a phenotype has not yet been published.

Spiroplasma phages do not explain the distribution of RIP, OTU, ankyrin, or ETX/MTX2 domains across the genus.

VT-specific domains source diverse recombinant genes

Across Citri and Ixodetis-clade *Spiroplasma*, we observe multiple structural variants of RIPs, OTUs, ankyrins, and ETX/MTXs within the same open reading frame (Figure 5). In a particularly interesting case, we uncovered two Spaid-like toxins from the genome of the *Spiroplasma* endosymbiont of *Drosophila hydei* (strain *sHyd1*, NCBI BioProject PRJNA274591) and a *Spiroplasma* genome extracted from whole-genome sequence data of the South American butterfly *Jemadia suekentonmiller* (*sSue*). These open reading frames include ankyrin domains, an OTU domain, and the Spaid C-terminal transmembrane domain. Toward the N-terminus, these proteins are also equipped with two RIP domains which together possess all the conserved active site residues of a single RIP domain. These Spaid-like toxins are present across several *Spiroplasma* genomes extracted from hesperid butterfly assemblies (Supplementary Figure S7).

Spiroplasma with VT-associated domains are widespread across Hesperiidae butterflies and other insects

We further investigated the presence of *Spiroplasma* in insect genome assemblies using MSRO Spaid, and FtsZ and rpoB proteins from Apis, Citri, and Ixodetis clade taxa as queries against whole genome shotgun (WGS) sequence databases on NCBI and identified Citri clade *Spiroplasma* genomic sequences in several South American members of the Hesperiidae butterfly family. A few of these *Spiroplasma* infections are also reported in an unpublished preprint (Medina et al., 2020; Manzano-Marin et al., 2020). These Hesperiidae-infecting *Spiroplasma* strains group closely to the heritable *Spiroplasma* strains infecting *Drosophila mojavensis* and *Megaselia nigra* (*sMoj* and *sNigra*, respectively) (Figure 6). Metagenomic binning of *Spiroplasma* contigs reveals this novel group of *Spiroplasma* is also equipped with RIPs, OTUs, ankyrins, and ETX/MTX2 domains. Phylogenetic analysis of Ixodetis clade *Spiroplasma* extracted from WGS sequences reveals an association with diverse insects including two ants, a butterfly, a springtail and a drosophilid (Figure 6). The presence of RIPs, OTUs, ankyrins, and ETX/MTX2 domains is variable among these novel Ixodetis strains (Figure 6). For example, the *Spiroplasma* genome extracted from a *Monomorium pharaonic* ant assembly (*sPharaoh*) is estimated to be 98% complete (Supplementary Table S4) and only encodes a single ETX/MTX2 domain. Alternatively, Ixodetis clade *Spiroplasma* genomes extracted from the assemblies of the butterfly *Colias croceus* (*sCroceus*) and the drosophilid *Zaprionus kolodkinae* (*sKolod*) both encode RIPs, OTUs, ankyrins, and ETX/MTX2 domains.

Discussion

In this study, we report a conserved core of toxin and virulence domains associated with heritable *Spiroplasma*. As one of the most successful and diverse arthropod-associated heritable symbionts on Earth (Duron et al., 2008), the mechanisms that support *Spiroplasma* persistence across ecological timescales are of great interest and yet largely unknown. Previous work has demonstrated that *Spiroplasma* exploit conserved yolk protein import machinery to invade maturing oocytes, providing a possible explanation for their effective vertical transmission in diverse arthropods (Herren et al. 2013). But how *Spiroplasma* navigates numerous other obstacles required for growth and persistence in host hemolymph throughout host development is not well understood. While toxins are often viewed through the lens of pathogenesis, they can also mediate and manipulate host processes to maintain symbiont persistence. For instance, RIPs and an OTU-ankyrin-bearing Spaid toxin promote *Spiroplasma* persistence by supporting defensive and reproductive phenotypes, respectively (Hamilton et al., 2016; Ballinger and Perlman, 2017; Harumoto and Lemaitre, 2018). Our assembly of the Ixodetis *Spiroplasma* genome from *Drosophila atripex* and accompanying comparative analyses reveal that the presence of these domains is highly conserved across

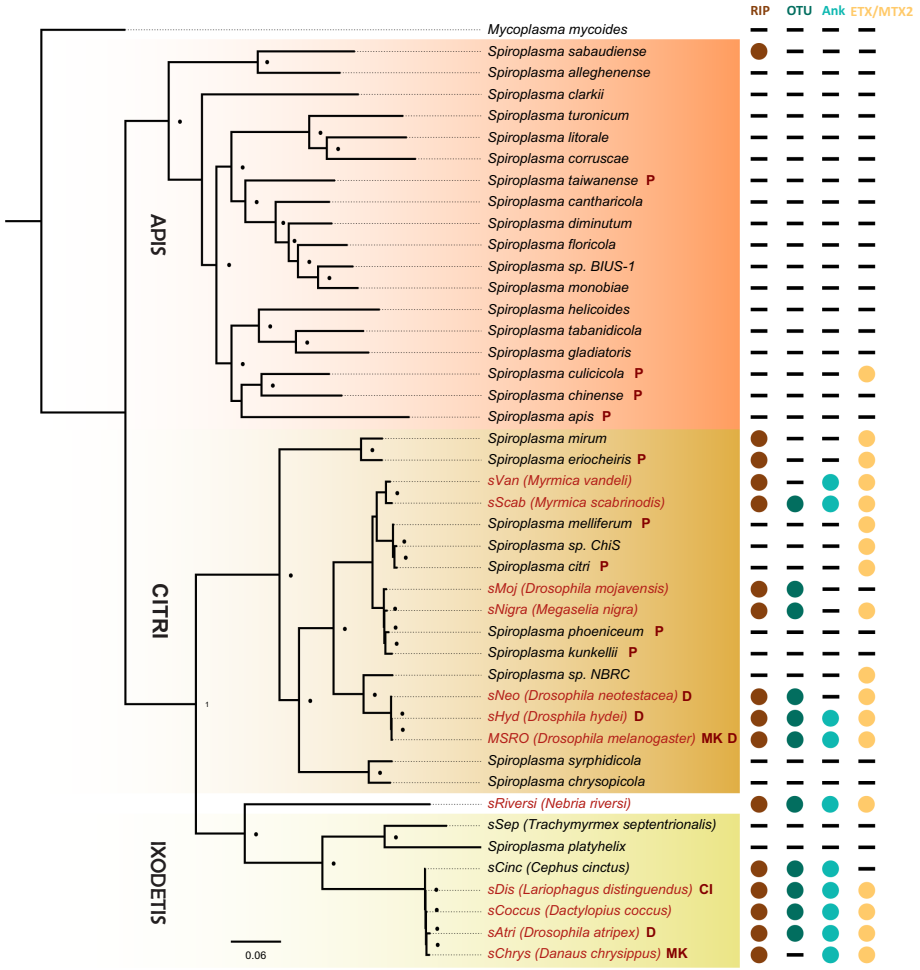


FIGURE 3
Phylogenetic distribution of select toxins and virulence domains in the genus *Spiroplasma*. FastTree phylogeny built from MAFFT nucleotide alignments of concatenated *Spiroplasma* ftsZ, rpoB and gyrB and rooted to *M. mycoides*. RIP, OTU, ankyrin, and ETX/MTX domains are preferentially distributed among VT *Spiroplasma*. Small black dots next to nodes indicate FastTree support values greater than 75%. Large colored circles to the right of branch labels indicate at least one domain copy is present in the genome. Black dash indicates no domain copies present in the genome. Red text is used for VT *Spiroplasma* and black text is used for non-VT *Spiroplasma*. Abbreviations indicate known phenotypes of the strain from which each domain was extracted: (P: pathogenic, MK: male-killing, CI: cytoplasmic incompatibility-inducing, D: defensive).

TABLE 4 Model comparisons for describing toxin and virulence domains across heritable *Spiroplasma*.

Domain	Dependent model likelihood score	Independent model likelihood score	Chi score	df	value of *p**
RIP	−29.855283	−40.389039	21.067512	4	0.000307
OTU	−26.546300	−37.764235	22.43587	4	0.000164
ankyrin	−26.036137	−35.577156	19.082038	4	0.000757
ETX/MTX2	−33.930099	−44.915624	21.97105	4	0.000203
glpO	−44.902777	−46.963108	4.056606	4	0.398347
chiA	−37.674589	−40.678499	6.00782	4	0.198552
spiralin	−32.183710	−37.138576	9.91	4	0.041971

*Value of *p* threshold has been set at 0.001.

diverse heritable *Spiroplasma*. RIPs, OTUs and ankyrin domains are specific to VT *Spiroplasma*, and are rare (RIPs) or entirely absent (OTUs and ankyrins) in non-VT *Spiroplasma* genomes. These toxin and virulence domains may play a role in facilitating and maintaining a vertically transmissible lifestyle across heritable *Spiroplasma* as has been demonstrated for a small number of Citri clade members. Finally, ETX/MTX2 toxins are more widespread in non-VT Citri and Ixodetis *Spiroplasma* compared to RIPs, OTUs, and ankyrins but are

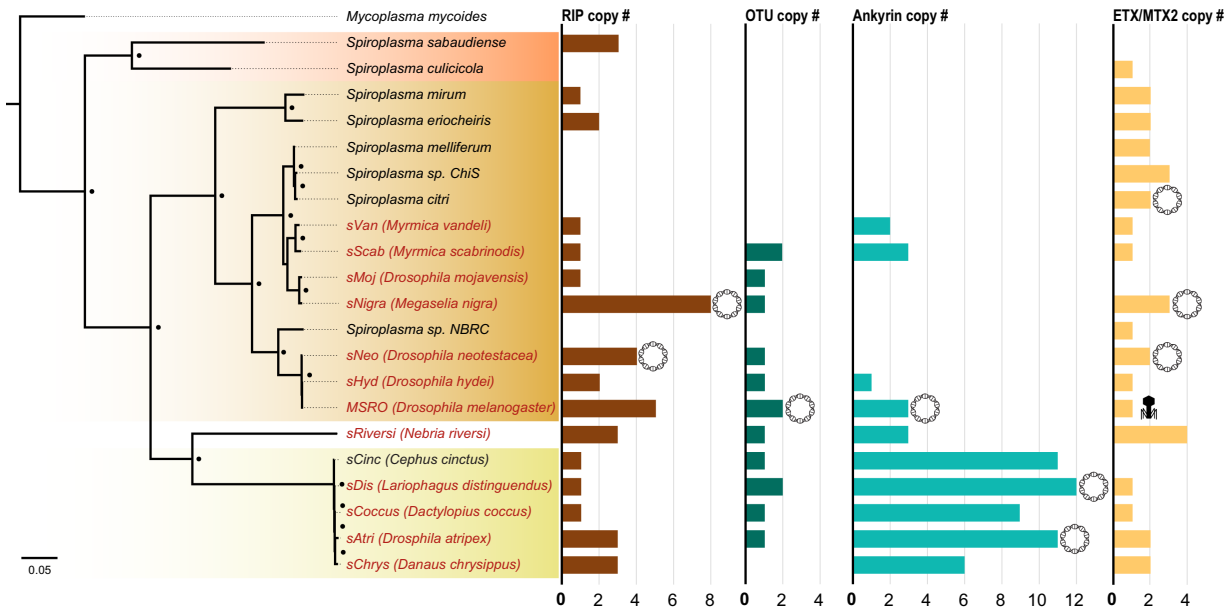


FIGURE 4
Number of toxin and virulence domains varies across *Spiroplasma* genus. FastTree phylogeny built from MAFFT nucleotide alignment of concatenated *Spiroplasma* ftsZ, rpoB, and gyrB sequences and rooted to *M. mycoides*. All *Spiroplasma* present in the phylogeny encode at least one domain copy of RIP, OTU, ankyrin, or ETX/MTX2. Bar graphs show the number of domain copies present in each genome for each domain type. Double-helix circles indicate at least one of the domains are found on a plasmid. Phage illustrations indicate domains are found within a phage region of the genome. Red shading represents Apis clade, orange shading represents Citri clade and yellow shading represents Ixodetis clade. Red color text is used for VT *Spiroplasma* and black color text is used for non-VT *Spiroplasma*. Small black dots indicate FastTree support values above 0.75.

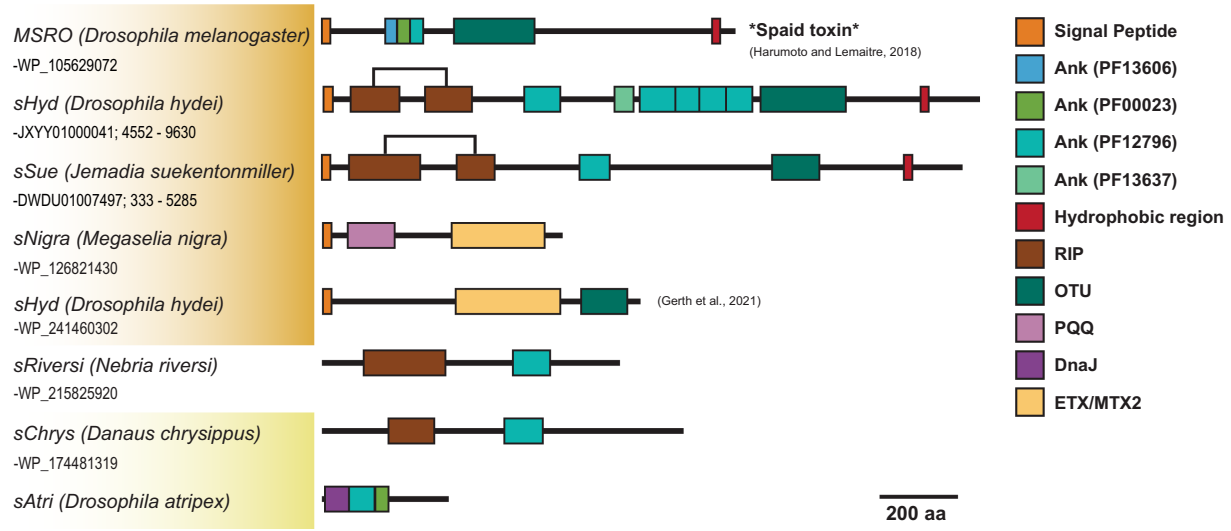


FIGURE 5
Dynamic variants of toxin and virulence genes across *Spiroplasma*. Scaled representation showing diverse array of *Spiroplasma* toxin-bearing proteins identified in this study. Orange shading indicates Citri clade and yellow shading indicates Ixodetis clade. A bracket is used for *sHyd* and *sSue* to highlight that the two separated RIP domains make up a whole RIP domain. Spaid toxin responsible for male-killing by MSRO is shown above.

still highly correlated with a VT lifestyle based on BayesTraits analysis. As previously stated, the role of ETX/MTX2 toxins in host-*Spiroplasma* interactions is unknown but their propensity for VT-*Spiroplasma* may suggest a role in maintaining a VT lifestyle similar to RIPs, OTUs and ankyrin domains (i.e., defense or host manipulation).

We find that similar toxin evolutionary processes previously reported in Poulsonii clade strains (Ballinger et al., 2019; Gerth et al., 2021) are mirrored in toxin domain-containing proteins of the Ixodetis clade, suggesting conservation of dominant toxin evolution mechanisms across heritable strains throughout the genus. Citri clade RIPs experience duplications, losses, and domain

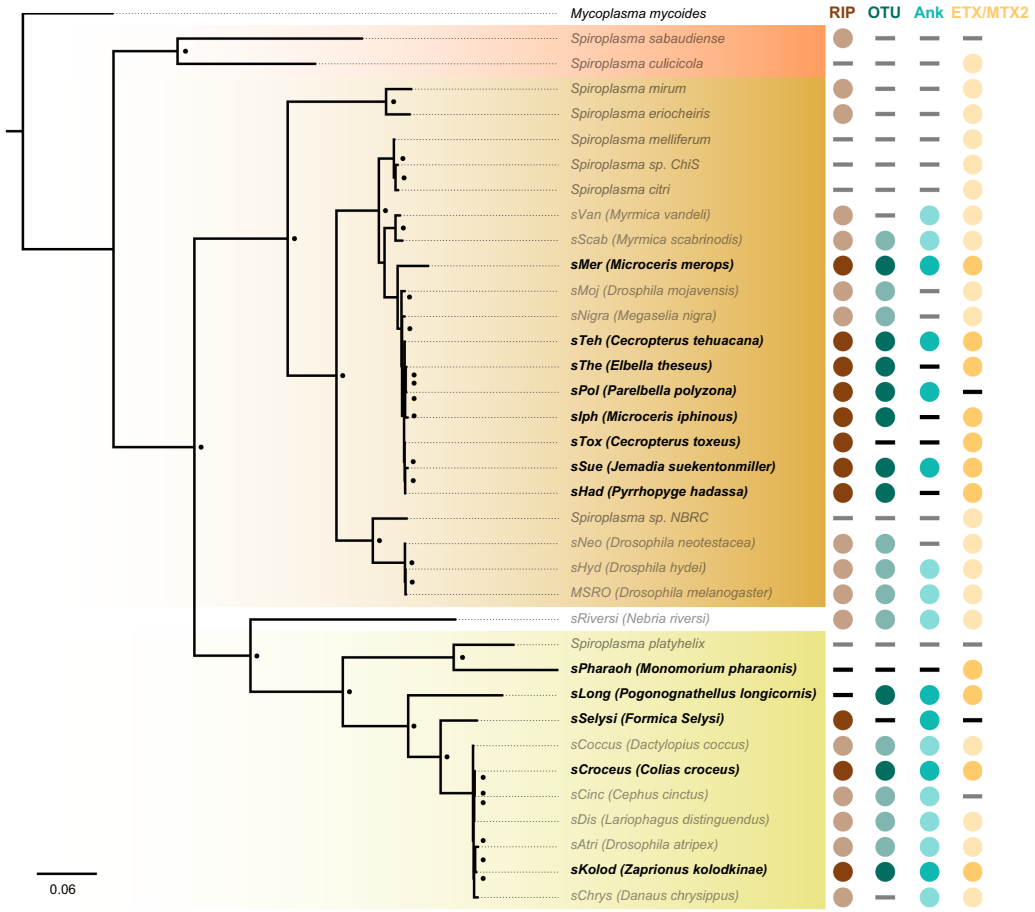


FIGURE 6
Cryptic *Spiroplasma* carry domains associated with a heritable lifestyle. FastTree phylogeny built from MAFFT nucleotide alignment of concatenated *Spiroplasma* ftsZ, rpoB, and gyrB and rooted to *M. mycoides*. This phylogram shows the placement of novel *Spiroplasma* strains (black bold) extracted from insect WGS sequences. Many of these *Spiroplasma* strains carry RIPs, OTUs, and/or ankyrin domains that are strongly associated with a heritable lifestyle. Circles indicate the presence of a given domain and dashes indicate their absence. Red shading indicates Apis clade, orange shading indicates Citri clade and yellow shading indicates Ixodetis clade. WGS-extracted taxa are shown in bold typeface. Black dots adjacent to nodes indicate a FastTree support value greater than 0.75.

swapping events that have accompanied an explosion in RIP diversity. Similarly, Ixodetis RIPs are polyphyletic and RIP representation differs between strains. For example, a RIP from the Ixodetis endosymbiont of the butterfly *Danaus chrysippus* groups closely with the enigmatic double RIP domain-possessing proteins recently uncovered from *sHyd1* and *sSue* (Figure 5; Gerth et al., 2021). Ankyrin domain-containing proteins are especially numerous in members of the Ixodetis clade, ranging from six to fourteen copies across members of this clade. Believed to originate in eukaryotes, ankyrin domains are more commonly found in heritable microbes than other bacterial taxa (Jernigan and Bordenstein, 2014), and may facilitate a variety of microbe-host interactions. This makes the expansion of ankyrin domains in Ixodetis particularly interesting, especially given the broad range of insect hosts they infect (Figure 1).

Dynamic evolution is also demonstrated in the varied number of genes that possess these domains (i.e., RIP, OTU, ankyrin, and ETX/MTX2)—even among closely related strains—indicating that duplications, gains, and losses are common generators of their diversity in *Spiroplasma*. RIP, OTU, ankyrin, and ETX/MTX2 domains are also present on plasmids, possibly contributing functional roles that favor their dispersal and persistence through the genus. Plasmids often confer adaptive phenotypes to their bacterial hosts to help drive their own spread among bacterial populations (Dimitriu et al., 2021). Interestingly, no RIP, OTU, or ankyrin domains are present within the diverse plasmids of non-VT Citri clade *Spiroplasma*. For example, some *Spiroplasma citri* strains can have upwards of nine unique plasmids (Rattner et al., 2021). Given the promiscuity of plasmids across *Spiroplasma*, the lack of RIP, OTU, and ankyrin domains on plasmids in non-heritable strains suggests that they may have little adaptive function in that lifestyle.

Not only do RIPs, OTUs, and ankyrin domains exist within diverse proteins across *Spiroplasma*, but many VT *Spiroplasma* species are equipped with several structural variants of these domains on the same protein. In one notable example, our analysis revealed a *Spiroplasma* genome within the WGS assembly of the South American butterfly *J. suekentonmiller* that encodes a protein with RIP domains, an OTU domain, an ankyrin domain and a C-terminal transmembrane domain, i.e., a RIP-Spaid fusion protein.

These VT-associated domains appear to frequently recombine to create novel protein configurations. For instance, Spaid and Spaid-like proteins encode multiple ankyrin domains each and these ankyrin domains vary greatly in copy number and protein family (Figure 5; Supplementary Figure S7B), suggesting that ankyrin-spanning region is especially prone to domain losses and gains. The close association of RIPs, OTUs, and ankyrins on the same proteins across Citri and Ixodetis *Spiroplasma* suggests functional evolutionary ties between these domains.

Extracting symbiont genomes from publicly available nucleotide databases is a powerful and convenient approach to studying symbiont evolution (Salzberg et al., 2005; Gerth and Hurst, 2017; Pascar and Chandler, 2018; Scholz et al., 2020; Pilgrim et al., 2021). We explored the presence of RIP, OTU, ankyrin, and ETX/MTX2 domains in other insect assemblies by mining partial to near complete genomes of both Ixodetis and Citri clade *Spiroplasma* from a diverse array of insect WGS assemblies. Within Citri, there is a large group of *Spiroplasma* strains infecting various members of the South American butterfly family Hesperiidae, and are equipped with RIP, OTU, and ankyrin domain-possessing proteins. Within Ixodetis, we identified five novel *Spiroplasma* genomes extracted from a diverse variety of insect hosts with a varied presence of RIP, OTU, ankyrin, and ETX/MTX2 domains. This analysis provides promising candidates for future studies on heritable *Spiroplasma* and the identification of RIP, OTU, ankyrin, and ETX/MTX2 domains within insect WGS assemblies may be useful for developing early hypotheses on *Spiroplasma*-host interactions.

Methods

Genome sequencing, assembly, and annotation

Nucleic acids were extracted from ovaries of eight adult female *Drosophila atripex* using the phenol-chloroform method. Paired end 150 bp Illumina genomicDNA reads were sequenced by Novogene (CA, United States). Long reads were sequenced on an Oxford Nanopore MinION device and R9.4.1 flow cell following library preparation using the Ligation Sequencing Kit (SQK-LSK109). Guppy 6.1.5 implemented in MinKNOW 22.05.5 was used to generate high accuracy basecalls. Only long reads above 10 kb were used in hybrid assembly. Illumina reads were trimmed for adapter sequences and quality (qtrim = r trimq = 10) with BBMap 38.35 (Bushnell, 2014), and a metagenome was hybrid assembled using Unicycler 0.5.0 (Wick et al., 2017). *Spiroplasma*-derived contigs were retained through a combination of metagenomic binning (Laczny et al., 2017) and manual filtering. Manual filtering was performed through targeted searches of sequences likely to contaminate the bin based on similar nucleotide content and was guided by kmer abundance analyses performed with KAT (Mapleson et al., 2016). The draft genome was annotated through Prokka 1.14.5 (Seemann, 2014) and KEGG numbers were assigned with BLASTKoala (Kanehisa et al., 2016). Genome coverage and plasmid coverage were determined using BBMAP (Bushnell, 2014). Genome completeness was estimated with BUSCO v5 (Simão et al., 2015) on the gVolante webserver (Nishimura et al., 2017). *Spiroplasma* homologs of interest, including those used for phylogenetics, were identified using HMMER (Finn et al., 2011) and PfamScan (Mistry et al., 2021) with a threshold of E-value = 0.01. Proteins of interest that could not be annotated by the previously mentioned approaches were investigated further with HHpred (Zimmermann et al., 2018), BLASTp and alignments. Protein and nucleotide sequences were aligned with MAFFT 7.388 (Katoh and Standley, 2013) and phylogenies were built using FastTree 2.1.11 (Price et al., 2010) unless stated otherwise. BLASTp searches with relaxed significance thresholds (expect values ≤0.01) were also performed to ensure detection of more distantly related protein domains if present.

Characterizing distribution of heritability, toxin, and virulence domains across *Spiroplasma* genus

All *Spiroplasma* sp. genomes, plasmids, and reads were downloaded from NCBI, and their accession numbers are listed in Supplementary Table S5. Heritable *Spiroplasma* were identified based on evidence provided from transgenerational screenings, ovarian tissue and egg screens, high infection frequencies among populations, multi-year screenings and systemic infections (Table 2). We searched for RIP, OTU, ankyrin, and ETX/MTX2 domains in all *Spiroplasma* genomes using tBLASTn (threshold E-value = 0.01) with a curated list of *Spiroplasma* domains that have been identified in this study and extracted using HMMER domain annotations as a guide (Supplementary Table S6). Protein domains were confirmed with HMMER 3.3 and pfamscan with a.01 e-value threshold, and through alignments to confirm presence of conserved residues. Domains present on open reading frames and on pseudogenes are both included in the total count for domains of interest. Phage regions were determined using Phaster webserver (Zhou et al., 2011; Arndt et al., 2016). Annotated phage regions were extracted from the genomes of *Spiroplasma* strains known to contain RIP, ankyrin, OTU, or ETX/MTX2 domains. These phage regions were investigated to determine if they included any of these domains.

We determined whether a correlation existed between heritability and specific encoded domains using the software BayesTraits V4.0.0 (Pagel and Meade, 2006). Heritability was treated a discrete binary trait (heritable or non-heritable) and domains were also treated as a binary trait (domain present domain absent in genome). Likelihood scores were calculated for both a dependent model (i.e., assumes heritability and domain acquisition/retention evolved dependently) and an independent model (i.e., assumes heritability and domain acquisition/retention evolved independently). Likelihood scores were compared with a chi-square analysis as recommended by BayesTraits documentation. We rejected the null hypothesis of independent evolution at $p < 0.001$.

Identification and extraction of *Spiroplasma* genomes from insect whole genome shotgun assemblies

Using Spaid toxin from MSRO, and ftsZ and rpoB from Ixodetis, Citri, and Apis clade taxa as queries, we performed tBLASTn searches of whole genome shotgun contigs (WGS) located on the NCBI database. Default search parameters were used, and we limited

our organism search to Insecta (taxid:50557). Assemblies with matches to these genes were submitted to BusyBee web server for genomic binning (Laczny et al., 2017). Bins outputted by BusyBee were manually inspected for the presence of core *Spiroplasma* genes including ftsZ, rpoB, and gyrB. These core genes were also used to construct the phylogeny in Figure 6. We performed tBLASTn against the extracted *Spiroplasma* genomes using a curated list of *Spiroplasma* RIP, OTU and ankyrin peptide sequences. If a domain was determined to be missing from an extracted *Spiroplasma* genome, we searched for it in the original insect assembly to ensure it wasn't present on a *Spiroplasma* contig missed in the binning process. Genome completeness was determined using BUSCO through gVolante webserver, and coding DNA sequences and tRNA content was determined with Prokka 1.14.5. Genome statistics and insect assembly accession numbers are available in Supplementary Table S4.

Constructing phylogenies

MAFFT 7.388 was used to create all alignments and FastTree 2.1.11 was used to construct *Spiroplasma* phylogenies. Jukes-Cantor model was used to build phylogenies from nucleotide sequence alignments. RIP and OTU proteins are often flanked with diverse, nonhomologous accessory domains. For this reason, RIP and OTU regions immediately outside of the conserved active site residues were trimmed out of alignments manually; conserved active site residues span the majority of both domains. *sSue* RIP, *sHyd* RIP2, and *sChrys* RIP3 all possess peptide insertion sequences ranging from 54 to 157 aa that split their RIP domain. This appears to be a conserved feature among this clade of RIPs. Due to their large size and highly divergent sequence variation, we identified these insertion sequence regions using alignments to other RIPs and removed them manually. *sHyd* RIP2 and *sRiversi* RIP1 also group among these RIPs, however they are small proteins encoding only a partial RIP domain and therefore did not require additional trimming. The trimmed RIP and OTU sequences were then aligned with MAFFT 7.388 and submitted to MEGA software (Kumar et al., 2018) to determine the best substitution model for PhyML tree building. WAG+G+I substitution model was used to construct a RIP phylogeny and cpREV+G substitution model was used to construct an OTU phylogeny. ETX/MTX2 are divergent toxins that lack conserved active site residues to help guide trimming. We created an alignment of ETX/MTX2-possessing proteins with MAFFT 7.388 and confirmed that the annotated ETX/MTX2 domains were aligned. The alignment was then trimmed using ClipKIT software (Steenwyk et al., 2020) in kpi-gappy mode. The trimmed alignment output was submitted to MEGA software to determine the best substitution model for PhyML tree building. WAG+G+F substitution model was used to construct an ETX/MTX2 phylogeny.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found at: The Bioproject accession is "PRJNA928682" and the Biosample accession is "SAMN32939082" and the SRA accessions are "SRX19199603" and "SRX19199602."

Author contributions

LM and MB conceived and designed the study, collected data and performed analyses, and revised the manuscript and approved the final submission. LM prepared figures and wrote the first draft of the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by National Science Foundation award 2144270 to MB.

Acknowledgments

We are grateful to Luciano Matzkin for sharing *Drosophila atripex* lines.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2023.1148263/full#supplementary-material

References

Alexeev, D., Kostrjukova, E., Aliper, A., Popenko, A., Bazaleev, N., Tyakht, A., et al. (2012). Application of *Spiroplasma melliferum* proteogenomic profiling for the discovery of virulence factors and pathogenicity mechanisms in host-associated *Spiroplasmas*. *J. Proteome Res.* 11, 224–236. doi: 10.1021/pr2008626

Andrews, M., De Meyer, S., James, E., Stępkowski, T., Hodge, S., Simon, M., et al. (2018). Horizontal transfer of symbiosis genes within and between Rhizobial genera: occurrence and importance. *Genes* 9:321. doi: 10.3390/genes9070321

Arndt, D., Grant, J. R., Marcu, A., Sajed, T., Pon, A., Liang, Y., et al. (2016). PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Res.* 44, W16–W21. doi: 10.1093/nar/gkw387

Arnold, B. J., Huang, I.-T., and Hanage, W. P. (2022). Horizontal gene transfer and adaptive evolution in bacteria. *Nat. Rev. Microbiol.* 20, 206–218. doi: 10.1038/s41579-021-00650-4

Ballinger, M. J., Gawryluk, R. M. R., and Perlman, S. J. (2019). Toxin and genome evolution in a *Drosophila* defensive symbiosis. *Genome Biol. Evol.* 11, 253–262. doi: 10.1093/gbe/evy272

Ballinger, M. J., Moore, L. D., and Perlman, S. J. (2018). Evolution and diversity of inherited *Spiroplasma* symbionts in myrmica ants. *Appl. Environ. Microbiol.* 84:e02299-17. doi: 10.1128/AEM.02299-17

Ballinger, M. J., and Perlman, S. J. (2017). Generality of toxins in defensive symbiosis: ribosome-inactivating proteins and defense against parasitic wasps in *Drosophila*. *PLoS Pathog.* 13:e1006431. doi: 10.1371/journal.ppat.1006431

Beckmann, J. F., Ronau, J. A., and Hochstrasser, M. (2017). A Wolbachia deubiquitylating enzyme induces cytoplasmic incompatibility. *Nat. Microbiol.* 2:17007. doi: 10.1038/nmicrobiol.2017.7

Belousov, M. V., Bondarev, S. A., Kosolapova, A. O., Antonets, K. S., Sulatskaya, A. I., Sulatsky, M. I., et al. (2018). M60-like metalloprotease domain of the Escherichia coli YghJ protein forms amyloid fibrils.*PLoS ONE* 13, e0191317. doi: 10.1371/journal.pone.0191317

Binetruy, F., Bailly, X., Chevillona, C., Martin, O. Y., Bernasconi, M. V., and Duron, O. (2019). Phylogenetics of the *Spiroplasma ixodetis* endosymbiont reveals past transfers between ticks and other arthropods. *Ticks Tick-Borne Dis.* 10, 575–584. doi: 10.1016/j.ttbdis.2019.02.001

Bolaños, L. M., Servín-Garcidueñas, L. E., and Martínez-Romero, E. (2015). Arthropod–*Spiroplasma* relationship in the genomic era. *FEMS Microbiol. Ecol.* 91, 1–8. doi: 10.1093/femsec/fiu008

Bushnell, B. (2014). *BBMap: A Fast, Accurate, Splice-Aware Aligner*. Lawrence Berkeley National Laboratory. LBNL Report #: LBNL– 7065E.

Chang, T.-H., Lo, W.-S., Ku, C., Chen, L.-L., and Kuo, C.-H. (2014). Molecular evolution of the substrate utilization strategies and putative virulence factors in mosquito-associated *Spiroplasma* species. *Genome Biol. Evol.* 6, 500–509. doi: 10.1093/gbe/evu033

Degnan, P. H., and Moran, N. A. (2008). Diverse phage-encoded toxins in a protective insect Endosymbiont. *Appl. Environ. Microbiol.* 74, 6782–6791. doi: 10.1128/AEM.01285-08

Dimitriu, T., Matthews, A. C., and Buckling, A. (2021). Increased copy number couples the evolution of plasmid horizontal transmission and plasmid-encoded antibiotic resistance. *Proc. Natl. Acad. Sci. U. S. A.* 118:e2107818118. doi: 10.1073/pnas.2107818118

Duret, S., Batailler, B., Dubrana, M.-P., Saillard, C., Renaudin, J., Béven, L., et al. (2014). Invasion of insect cells by S piroplasma citri involves spiralin relocalization and lectin/glycoconjugate-type interactions: Spiralin relocalization and insect cell adhesion. *Cell. Microbiol.* 16, 1119–1132. doi: 10.1111/cmi.12265

Duron, O., Bouchon, D., Boutin, S., Bellamy, L., Zhou, L., Engelstädter, J., et al. (2008). The diversity of reproductive parasites among arthropods: Wolbachiado not walk alone. *BMC Biol.* 6:27. doi: 10.1186/1741-7007-6-27

Finn, R. D., Clements, J., and Eddy, S. R. (2011). HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res.* 39, W29–W37. doi: 10.1093/nar/gkr367

Gasparich, G. E., Whitcomb, R. F., Dodge, D., French, F. E., Glass, J., and Williamson, D. L. (2004). The genus *Spiroplasma* and its non-helical descendants: phylogenetic classification, correlation with phenotype and roots of the *Mycoplasma mycoides* clade. *Int. J. Syst. Evol. Microbiol.* 54, 893–918. doi: 10.1099/ijs.0.02688-0

Gerth, M., and Hurst, G. D. D. (2017). Short reads from honey bee (Apis sp.) sequencing projects reflect microbial associate diversity. *Peer J* 5:e3529. doi: 10.7717/peerj.3529

Gerth, M., Martinez-Montoya, H., Ramirez, P., Masson, F., Griffin, J. S., Aramayo, R., et al. (2021). Rapid molecular evolution of *Spiroplasma* symbionts of *Drosophila*. *Microbial Genomics* 7, 1–15. doi: 10.1099/mgen.0.000503

Gillespie, J. J., Driscoll, T. P., Verhoeve, V. I., Rahman, M. S., Macaluso, K. R., and Azad, A. F. (2018). A tangled web: origins of reproductive parasitism. *Genome Biol. Evol.* 10, 2292–2309. doi: 10.1093/gbe/evy159

Hamilton, P. T., Peng, F., Boulanger, M. J., and Perlman, S. J. (2016). A ribosome-inactivating protein in a *Drosophila* defensive symbiont. *Proc. Natl. Acad. Sci. U. S. A.* 113, 350–355. doi: 10.1073/pnas.1518648113

Harumoto, T., and Lemaitre, B. (2018). Male-killing toxin in a bacterial symbiont of *Drosophila*. *Nature* 557, 252–255. doi: 10.1038/s41586-018-0086-2

Haselkorn, T. S. (2010). The *Spiroplasma* heritable bacterial endosymbiont of *Drosophila*. *Fly* 4, 80–87. doi: 10.4161/fly.4.1.10883

Haselkorn, T. S., and Jaenike, J. (2015). Macroevolutionary persistence of heritable endosymbionts: acquisition, retention and expression of adaptive phenotypes in *Spiroplasma*. *Mol. Ecol.* 24, 3752–3765. doi: 10.1111/mec.13261

Haselkorn, T. S., Markow, T. A., and Moran, N. A. (2009). Multiple introductions of the *Spiroplasma* bacterial endosymbiont into *Drosophila*. *Mol. Ecol.* 18, 1294–1305. doi: 10.1111/j.1365-294X.2009.04085.x

Haselkorn, T. S., Watts, T. D., and Markow, T. A. (2013). Density dynamics of diverse *Spiroplasma* strains naturally infecting different species of *Drosophila*. *Fly* 7, 204–210. doi: 10.4161/fly.25469

Herren, J. K., Paredes, J. C., Schüpfer, F., and Lemaitre, B. (2013). Vertical transmission of a *Drosophila* Endosymbiont via cooption of the yolk transport and internalization machinery. *MBio* 4, e00532–e00512. doi: 10.1128/mBio.00532-12

Horard, B., Terretaz, K., Gosselin-Grenet, A.-S., Sobry, H., Sicar, M., Landmann, F., et al. (2022). Paternal transmission of the Wolbachia CidB toxin underlies cytoplasmic incompatibility. *Curr. Biol.* 32, 1319–1331.e5. doi: 10.1016/j.cub.2022.01.052

Humphery-Smith, I., Grulet, O., Le Goff, F., and Chastel, C. (1991). *Spiroplasma* (mollicutes: spiroplasmataceae) pathogenic for Aedes aegypti and Anopheles stephensi (diptera: culicidae). *J. Med. Entomol.* 28, 219–222. doi: 10.1093/jmedent/28.2.219

Ishimwe, E., Hodgson, J. J., Clem, R. J., and Passarelli, A. L. (2015). Reaching the melting point: degradative enzymes and protease inhibitors involved in baculovirus infection and dissemination. *Virology* 479-480, 637–649. doi: 10.1016/j.virol.2015.01.027

Jaenike, J., Stahlhut, J. K., Boelio, L. M., and Unckless, R. L. (2010a). Association between Wolbachia and *Spiroplasma* within *Drosophila* neotestacea: an emerging symbiotic mutualism?: wolbachia-*Spiroplasma* mutualism. *Mol. Ecol.* 19, 414–425. doi: 10.1111/j.1365-294X.2009.04448.x

Jaenike, J., Unckless, R., Cockburn, S. N., Boelio, L. M., and Perlman, S. J. (2010b). Adaptation via symbiosis: recent spread of a *Drosophila* defensive symbiont. *Science* 329, 212–215. doi: 10.1126/science.1188235

Jernigan, K. K., and Bordenstein, S. R. (2014). Ankyrin domains across the tree of life. *PeerJ* 2:e264. doi: 10.7717/peerj.264

Jiggins, F. M., Hurst, G. D. D., Jiggins, C. D. V. D., Schulenburg, J. H. G., and Majerus, M. E. N. (2000). The butterfly Danaus chrysippus is infected by a male-killing *Spiroplasma* bacterium. *Parasitology* 120, 439–446. doi: 10.1017/S0031182099005867

Kanehisa, M., Sato, Y., and Morishima, K. (2016). BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and Metagenome sequences. *J. Mol. Biol.* 428, 726–731. doi: 10.1016/j.jmb.2015.11.006

Katoh, K., and Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780. doi: 10.1093/molbev/mst010

Kaur, R., Leigh, B. A., Ritchie, I. T., and Bordenstein, S. R. (2022). The Cif proteins from Wolbachia prophage WO modify sperm genome integrity to establish cytoplasmic incompatibility. *PLoS Biol.* 20:e3001584. doi: 10.1371/journal.pbio.3001584

Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. (2018). MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* 35, 1547–1549. doi: 10.1093/molbev/msy096

Laczny, C. C., Kiefer, C., Galata, V., Fehlmann, T., Backes, C., and Keller, A. (2017). BusyBee web: metagenomic data analysis by bootstrapped supervised binning and annotation. *Nucleic Acids Res.* 45, W171–W179. doi: 10.1093/nar/gkx348

LePage, D. P., Metcalf, J. A., Bordenstein, S. R., On, J., Perlmutter, J. I., Shropshire, J. D., et al. (2017). Prophage WO genes recapitulate and enhance Wolbachia-induced cytoplasmic incompatibility. *Nature* 543, 243–247. doi: 10.1038/nature21391

Łukasik, P., van Asch, M., Guo, H., Ferrari, J., Charles, J., and Godfray, H. (2013). Unrelated facultative endosymbionts protect aphids against a fungal pathogen. *Ecol. Lett.* 16, 214–218. doi: 10.1111/ele.12031

Manzano-Marin, A., Coeur Dacier, A., Clamens, A. L., Orvain, C., Cruaud, C., Barbe, V., et al. (2020). Serial horizontal transfer of vitamin-biosynthetic genes enables the establishment of new nutritional symbionts in aphids' di-symbiotic systems. *ISME J.* 14, 259–273. doi: 10.1038/s41396-019-0533-6

Mapleson, D., Garcia Accinelli, G., Kettleborough, G., Wright, J., and Clavijo, B. J. (2016). KAT: a K-mer analysis toolkit to quality control NGS datasets and genome assemblies. *Bioinformatics* 33, 574–576. doi: 10.1093/bioinformatics/btw663

Martin, S. H., Singh, K. S., Gordon, I. J., Omufwoko, K. S., Collins, S., Warren, I. A., et al. (2020). Whole-chromosome hitchhiking driven by a male-killing endosymbiont. *PLoS Biol.* 18:e3000610. doi: 10.1371/journal.pbio.3000610

Martinez, J., Klasson, L., Welch, J. J., and Jiggins, F. M. (2021). Life and death of selfish genes: comparative genomics reveals the dynamic evolution of cytoplasmic incompatibility. *Mol. Biol. Evol.* 38, 2–15. doi: 10.1093/molbev/msaa209

Massey, J. H., and Newton, I. L. G. (2022). Diversity and function of arthropod endosymbiont toxins. *Trends Microbiol.* 30, 185–198. doi: 10.1016/j.tim.2021.06.008

Mateos, M., Castrezana, S. J., Nankivell, B. J., Estes, A. M., Markow, T. A., and Moran, N. A. (2006). Heritable endosymbionts of *Drosophila*. *Genetics* 174, 363–376. doi: 10.1534/genetics.106.058818

McCutcheon, J. P., Boyd, B. M., and Dale, C. (2019). The life of an insect endosymbiont from the cradle to the grave. *Curr. Biol.* 29, R485–R495. doi: 10.1016/j.cub.2019.03.032

McCutcheon, J. P., and Moran, N. A. (2012). Extreme genome reduction in symbiotic bacteria. *Nat. Rev. Microbiol.* 10, 13–26. doi: 10.1038/nrmicro2670

Medina, P., Russell, S. L., Aswadhati, K., and Corbett-Detig, R. (2020). Deep data mining reveals variable abundance and distribution of microbial reproductive manipulators within and among diverse host species. *bioRxiv.* doi: 10.1101/679837 [Pre-print].

Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G. A., Sonnhammer, E. L. L., et al. (2021). Pfam: the protein families database in 2021. *Nucleic Acids Res.* 49, D412–D419. doi: 10.1093/nar/gkaa913

Moar, W. J., Evans, A. J., Kessenich, C. R., Baum, J. A., Bowen, D. J., Edrington, T. C., et al. (2017). The sequence, structural, and functional diversity within a protein family and implications for specificity and safety: the case for ETX_MTX2 insecticidal proteins. *J. Invertebr. Pathol.* 142, 50–59. doi: 10.1016/j.jip.2016.05.007

Montenegro, H., Solferini,V. N., Klaczko, L. B., and Hurst, G. D. D. (2005). Male-killing Spiroplasma naturally infecting Drosophila melanogaster: Male-killing Spiroplasma in D. melanogaster. *Insect. Mol. Biol.* 14, 281–287. doi: 10.1111/j.1365-2583.2005.00558.x

Moran, N. A. (2002). Microbial Minimalism. *Cells* 108, 583–586. doi: 10.1016/S0092-8674(02)00665-7

Moran, N. A., McCutcheon, J. P., and Nakabachi, A. (2008). Genomics and evolution of heritable bacterial Symbionts. *Annu. Rev. Genet.* 42, 165–190. doi: 10.1146/annurev.genet.41.110306.130119

Mouches, C., Bové, J. M., Albisetti, J., Clark, T. B., and Tully, J. G. (1982). A *Spiroplasma* of serogroup IV causes a may-disease-like disorder of honeybees in southwestern France. *Microb. Ecol.* 8, 387–399. doi: 10.1007/BF02010677

Nakjang, S., Ndeh, D. A., Wipat, A., Bolam, D. N., and Hirt, R. P. (2012). A Novel Extracellular Metallopeptidase Domain Shared by Animal Host-Associated Mutualistic and Pathogenic Microbes. *PLoS ONE* 7, e30287. doi: 10.1371/journal.pone.0030287

Nguyen, M. T. H. D., Liu, M., and Thomas, T. (2014). Ankyrin-repeat proteins from sponge symbionts modulate amoebal phagocytosis. *Mol. Ecol.* 23, 1635–1645. doi: 10.1111/mec.12384

Nishimura, O., Hara, Y., and Kuraku, S. (2017). gVolante for standardizing completeness assessment of genome and transcriptome assemblies. *Bioinformatics* 33, 3635–3637. doi: 10.1093/bioinformatics/btx445

Oliver, K. M., and Perlman, S. J. (2020). Toxin-mediated protection against natural enemies by insect defensive symbionts. *Adv Insect Physiol* 58, 277–316. doi: 10.1016/bs.aiip.2020.03.005

Pagel, M., and Meade, A. (2006). Bayesian Analysis of Correlated Evolution of Discrete Characters by Reversible-Jump Markov Chain Monte Carlo. *The American Naturalist* 167, 808–825. doi: 10.1086/503444

Palma, L., Muñoz, D., Berry, C., Murillo, J., and Caballero, P. (2014). *Bacillus thuringiensis* toxins: an overview of their Biocidal activity. *Toxins* 6, 3296–3325. doi: 10.3390/toxins6123296

Pan, X., Lührmann, A., Satoh, A., Laskowski-Arce, M. A., and Roy, C. R. (2008). Ankyrin repeat proteins comprise a diverse family of bacterial type IV effectors. *Science* 320, 1651–1654. doi: 10.1126/science.1158160

Paredes, J. C., Herren, J. K., Schüpfer, F., Marin, R., Claverol, S., Kuo, C.-H., et al. (2015). Genome sequence of the *Drosophila melanogaster* male-killing *Spiroplasma* strain MSRO endosymbiont. *mBio* 6, e02437–e02414. doi: 10.1128/mBio.02437-14

Pascar, J., and Chandler, C. H. (2018). A bioinformatics approach to identifying Wolbachia infections in arthropods. *PeerJ* 6:e5486. doi: 10.7717/peerj.5486

Pilgrim, J., Thongprem, P., Davison, H. R., Siozios, S., Baylis, M., Zakharov, E. V., et al. (2021). Torix Rickettsia are widespread in arthropods and reflect a neglected symbiosis. *GigaScience* 10:giab021. doi: 10.1093/gigascience/giab021

Pollmann, M., Moore, L. D., Krimmer, E., D'Alvise, P., Hasselmann, M., Perlman, S. J., et al. (2022). Highly transmissible cytoplasmic incompatibility by the extracellular insect symbiont *Spiroplasma*. *iScience* 25:104335. doi: 10.1016/j.isci.2022.104335

Pool, J. E., Wong, A., and Aquadro, C. F. (2006). Finding of male-killing *Spiroplasma* infecting *Drosophila melanogaster* in Africa implies transatlantic migration of this endosymbiont. *Heredity* 97, 27–32. doi: 10.1038/sj.hdy.6800830

Price, M. N., Dehal, P. S., and Arkin, A. P. (2010). FastTree 2 – approximately maximum-likelihood trees for large alignments. *PLoS One* 5:e9490. doi: 10.1371/journal.pone.0009490

Rattner, R., Thapa, S. P., Dang, T., Osman, F., Selvaraj, V., Maheshwari, Y., et al. (2021). Genome analysis of *Spiroplasma citri* strains from different host plants and its leafhopper vectors. *BMC Genomics* 22:373. doi: 10.1186/s12864-021-07637-8

Romero, D., and Palacios, R. (1997). Gene amplification and genomic plasticity in prokaryotes. *Annu. Rev. Genet.* 31, 91–111. doi: 10.1146/annurev.genet.31.1.91

Salzberg, S. L., Hotopp, J., Delcher, A. L., Pop, M., Smith, D. R., Eisen, M. B., et al. (2005). Serendipitous discovery of Wolbachia genomes in multiple *Drosophila* species. *Genome Biol.* 6:R23. doi: 10.1186/gb-2005-6-3-r23

Scholz, M., Albanese, D., Tuohy, K., Donati, C., Segata, N., and Rota-Stabelli, O. (2020). Large scale genome reconstructions illuminate Wolbachia evolution. *Nat. Commun.* 11:5235. doi: 10.1038/s41467-020-19016-0

Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30, 2068–2069. doi: 10.1093/bioinformatics/btu153

Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., and Zdobnov, E. M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31, 3210–3212. doi: 10.1093/bioinformatics/btv351

Steenwyk, J. L., Buida, T. J., Li, Y., Shen, X.-X., and Rokas, A. (2020). ClipKIT: a multiple sequence alignment trimming software for accurate phylogenomic inference. *PLoS Biol.* 18:e3001007. doi: 10.1371/journal.pbio.3001007

Stirpe, F. (2004). Ribosome-inactivating proteins. *Toxicon* 44, 371–383. doi: 10.1016/j.toxicon.2004.05.004

Sullivan, J. T., and Ronson, C. W. (1998). Evolution of rhizobia by acquisition of a 500-kb symbiosis island that integrates into a phe-tRNA gene. *Proc. Natl. Acad. Sci. U. S. A.* 95, 5145–5149. doi: 10.1073/pnas.95.9.5145

Tang, W.-J., and Guo, Q. (2009). The adenylyl cyclase activity of anthrax edema factor. *Mol. Asp. Med.* 30, 423–430. doi: 10.1016/j.mam.2009.06.001

Teixeira, L., Ferreira, Á., and Ashburner, M. (2008). The bacterial Symbiont Wolbachia induces resistance to RNA viral infections in *Drosophila melanogaster*. *PLoS Biol.* 6:e1000002:e2. doi: 10.1371/journal.pbio.1000002

Tinsley, M. C., and Majerus, M. E. N. (2006). A new male-killing parasitism: *Spiroplasma* bacteria infect the ladybird beetle Anisosticta novemdecimpunctata (Coleoptera: Coccinellidae). *Parasitology* 132, 757–765. doi: 10.1017/S0031182005009789

Vera-Ponce León, A., Dominguez-Mirazo, M., Bustamante-Brito, R., Higareda-Alvear, V., Rosenblueth, M., and Martínez-Romero, E. (2021). Functional genomics of a *Spiroplasma* associated with the carmine cochineals Dactylopius coccus and *Dactylopius opuntiae*. *BMC Genomics* 22:240. doi: 10.1186/s12864-021-07540-2

Waterworth, S. C., Flórez, L. V., Rees, E. R., Hertweck, C., Kaltenpoth, M., and Kwan, J. C. (2020). Horizontal gene transfer to a defensive Symbiont with a reduced genome in a multipartite beetle microbiome. *mBio* 11:e02430-19. doi: 10.1128/mBio.02430-19

Weng, Y., Francoeur, C. B., Currie, C. R., Kavanaugh, D. H., and Schoville, S. D. (2021). A high-quality carabid genome assembly provides insights into beetle genome evolution and cold adaptation. *Mol. Ecol. Resour.* 21, 2145–2165. doi: 10.1111/1755-0998.13409

Wick, R. R., Judd, L. M., Gorrie, C. L., and Holt, K. E. (2017). Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Comput. Biol.* 13:e1005595. doi: 10.1371/journal.pcbi.1005595

Xie, J., Butler, S., Sanchez, G., and Mateos, M. (2014). Male killing Spiroplasma protects Drosophila melanogaster against two parasitoid wasps. *Heredity.* 112, 399–408. doi: 10.1038/hdy.2013.118

Xie, J., Vilchez, I., and Mateos, M. (2010). *Spiroplasma* bacteria enhance survival of *Drosophila hydei* attacked by the parasitic wasp Leptopilina heterotoma. *PLoS One* 5:e12149. doi: 10.1371/journal.pone.0012149

Yeoman, C. J., Brutscher, L. M., Esen, Ö. C., Ibaoglu, F., Fowler, C., Eren, A. M., et al. (2019). Genome-resolved insights into a novel *Spiroplasma* symbiont of the wheat stem sawfly (*Cephus cinctus*). *PeerJ* 7:e7548. doi: 10.7717/peerj.7548

Zhou, Y., Liang, Y., Lynch, K. H., Dennis, J. J., and Wishart, D. S. (2011). PHAST: a fast phage search tool. *Nucleic Acids Res.* 39, W347–W352. doi: 10.1093/nar/gkr485

Zimmermann, L., Stephens, A., Nam, S.-Z., Rau, D., Kübler, J., Lozajic, M., et al. (2018). A completely Reimplemented MPI bioinformatics toolkit with a new HHpred server at its Core. *J. Mol. Biol.* 430, 2237–2243. doi: 10.1016/j.jmb.2017.12.007

frontiers | Frontiers in Microbiology

TYPE Original Research
PUBLISHED 02 October 2023
DOI 10.3389/fmicb.2023.1252709

OPEN ACCESS

EDITED BY
Chih-Horng Kuo,
Academia Sinica, Taiwan

REVIEWED BY
Amit Yadav,
National Centre for Cell Science, India
Kenro Oshima,
Hosei University, Japan

*CORRESPONDENCE
Ying-Kun Huang
✉ huangyk64@163.com
Xia-Hong He
✉ hexiahong@hotmail.com

RECEIVED 04 July 2023
ACCEPTED 12 September 2023
PUBLISHED 02 October 2023

CITATION
Zhang R-Y, Wang X-Y, Li J, Shan H-L, Li Y-H, Huang Y-K and He X-H (2023) Complete genome sequence of "*Candidatus* Phytoplasma sacchari" obtained using a filter-based DNA enrichment method and Nanopore sequencing. *Front. Microbiol.* 14:1252709. doi: 10.3389/fmicb.2023.1252709

Complete genome sequence of "*Candidatus* Phytoplasma sacchari" obtained using a filter-based DNA enrichment method and Nanopore sequencing

Rong-Yue Zhang[1,2], Xiao-Yan Wang[1], Jie Li[1], Hong-Li Shan[1], Yin-Hu Li[1], Ying-Kun Huang[1]* and Xia-Hong He[2,3]*

[1]Yunnan Key Laboratory of Sugarcane Genetic Improvement, Sugarcane Research Institute, Yunnan Academy of Agricultural Sciences, Kaiyuan, China, [2]State Key Laboratory for Conservation and Utilization of Bio-Resources in Yunnan, Yunnan Agricultural University, Kunming, China, [3]School of Landscape and Horticulture, Southwest Forestry University, Kunming, China

Phytoplasmas are phloem-limited plant pathogens, such as sugarcane white leaf (SCWL) phytoplasma, which are responsible for heavy economic losses to the sugarcane industry. Characterization of phytoplasmas has been limited because they cannot be cultured *in vitro*. However, with the advent of genome sequencing, different aspects of phytoplasmas are being investigated. In this study, we developed a DNA enrichment method for sugarcane white leaf (SCWL) phytoplasma, evaluated the effect of DNA enrichment *via* Illumina sequencing technologies, and utilized Illumina and Nanopore sequencing technologies to obtain the complete genome sequence of the "*Candidatus* Phytoplasma sacchari" isolate SCWL1 that is associated with sugarcane white leaf in China. Illumina sequencing analysis elucidated that only 1.21% of the sequencing reads from total leaf DNA were mapped to the SCWL1 genome, whereas 40.97% of the sequencing reads from the enriched DNA were mapped to the SCWL1 genome. The genome of isolate SCWL1 consists of a 538,951 bp and 2976 bp long circular chromosome and plasmid, respectively. We identified 459 protein-encoding genes, 2 complete 5S-23S-16S rRNA gene operons, 27 tRNA genes, and an incomplete potential mobile unit (PMU) in the circular chromosome. Phylogenetic analyses and average nucleotide identity (ANI) and digital DNA–DNA hybridization (dDDH) values based on the sequenced genome revealed that SCWL phytoplasma and sugarcane grassy shoot (SCGS) phytoplasma belonged to the same phytoplasma species. This study provides a genomic DNA enrichment method for phytoplasma sequencing. Moreover, we report the first complete genome of a "*Ca.* Phytoplasma sacchari" isolate, thus contributing to future studies on the evolutionary relationships and pathogenic mechanisms of "*Ca.* Phytoplasma sacchari" isolates.

KEYWORDS

phytoplasma, sugarcane white leaf, sugarcane grassy shoot, genome sequencing, DNA enrichment

Introduction

Phytoplasmas are phloem-limited bacterial plant pathogens that were discovered in 1967 and were initially classified as mycoplasma-like organisms (MLOs) (Doi et al., 1967). MLO being replaced with "phytoplasma" was initially suggested in 1992 at the meeting on the taxonomy of Mollicutes (Tully, 1993). In 2004, different species of phytoplasma were included in the provisional genus "*Candidatus* Phytoplasma" by the IRPCM Phytoplasma/Spiroplasma Working Team-Phytoplasma taxonomy group (2004). However, limited information could be obtained regarding these pathogens because it is difficult to culture them *in vitro*. With the advent of genome sequencing technologies and comparative genome analysis, our understanding of the genetic structure, phylogeny, evolution, metabolic pathways, and possible virulence factors of phytoplasmas has enhanced. The first complete genome sequence of the genus was reported for "*Candidatus* Phytoplasma asteris" OY-M isolate in 2004 (Oshima et al., 2004); 12 complete phytoplasma genomes and 35 draft phytoplasma genomes have been reported so far (Bertaccini et al., 2022; Wei and Zhao, 2022; Kirdat et al., 2023). However, in a previous study, phytoplasma genome sequencing using total DNA generated only 0.17% of Illumina sequencing reads (Cho et al., 2019) because it is difficult to obtain pure phytoplasma genomic DNA due to its unculturable nature, indicating that the enrichment of phytoplasma genomic DNA is essential for sequencing. The main enrichment methods used in previous studies are density gradient centrifugation, pulse field gel electrophoresis (PFGE) (Oshima et al., 2004; Bai et al., 2006; Kube et al., 2008; Chen et al., 2014), methyl-CpG-binding domain-mediated method (Kirdat et al., 2020, 2021; Nijo et al., 2021; Debonneville et al., 2022), and immunoprecipitation-based method (Tan et al., 2021).

Phytoplasmas are naturally transmitted by phloem-feeding insects and cause more than 1,000 types of plant diseases worldwide, resulting in significant economic losses in the agriculture industry (Wang et al., 2022). The varied symptoms of phytoplasma infections include plant dwarfing, leaf yellowing, phyllody, witches' broom, stunting, proliferation, and phloem tissue necrosis (Oshima, 2021; Wang et al., 2022). Phytoplasmas are also responsible for two diseases in sugarcane, namely, sugarcane white leaf (SCWL) and sugarcane grassy shoot (SCGS), which cause heavy losses in several sugarcane-growing countries (Zhang et al., 2020; Kirdat et al., 2021). The symptoms of SCWL are similar to those of SCGS, namely leaf whitening, increased tillering, and dwarfing (Viswanathan et al., 2011; Kirdat et al., 2021). The phytoplasma isolates associated with SCWL and SCGS belong to the 16SrXI group based on the high similarity of their 16S rRNA gene sequences (Viswanathan et al., 2011; Abeysinghe et al., 2016). The group 16SrXI includes the subgroups 16SrXI-B, 16SrXI-D, and 16SrXI-F (Zhang et al., 2016; Yadav et al., 2017). In 2020, the draft SCGS phytoplasma genome was published (Kirdat et al., 2020), and based on its analysis, the phytoplasma isolates associated with SCGS have been classified as a novel taxon "*Candidatus* Phytoplasma sacchari" (Kirdat et al., 2021). Multilocus sequence typing revealed that SCWL and SCGS phytoplasmas belong to two different populations of "*Ca.* Phytoplasma sacchari" (Abeysinghe et al., 2016; Zhang et al., 2023).

With the development of sequencing technology, whole-genome sequencing of phytoplasmas has become feasible for many laboratories. Genome analysis is an efficient and effective approach to generate a significantly large amount of data for the biological characterization of unculturable bacteria. In this study, we developed a method for the enrichment of SCWL phytoplasma DNA for performing genome sequencing, and we obtained the complete genome sequence of the isolate by combining Illumina and Nanopore technologies. Our study will provide a simple method for the enrichment of phytoplasma genomic DNA and enhance our understanding of the genetic characteristics of the "*Ca.* Phytoplasma sacchari" species, thus providing a basis for research on its pathogenic mechanisms and other aspects.

Materials and methods

Source of phytoplasma

Sugarcane (*Saccharum officinarum* L.) samples exhibiting SCWL symptoms were collected from Lincang, Yunnan province, China, in 2018. They were maintained and propagated in an insect-proof greenhouse at the Sugarcane Research Institute, Yunnan Academy of Agricultural Sciences. We used the ROC22 sugarcane variety in this study.

Extraction of genomic DNA from leaves

Genomic DNA from sugarcane leaves was extracted using the SDS method (Lim et al., 2016). The extracted DNA was detected using 1% agarose gel electrophoresis and quantified using a Qubit® 3.0 Fluorometer (Invitrogen, USA).

Enrichment of SCWL phytoplasma DNA

Approximately 5 g of sugarcane leaves were cut into small pieces using scissors and ground to obtain homogenate in 1× PBS buffer (Sangon Biotech Co., Ltd., Shanghai, China). The homogenates were placed in 50 ml centrifuge tubes and centrifuged at 12,000 rpm for 5 min. The supernatant was discarded, and the pellet was resuspended in 50 ml of 1× PBS buffer; this step was repeated thrice. The suspension was sequentially filtered through 100, 70, 40, 10, and 5 μm filters (Erwu Industrial Co., Ltd., Shanghai, China). The filtrate was centrifuged at 12,000 rpm for 5 min, and the supernatant was discarded; 20 μl of DNase I (3 units/μl) (TransGen Biotech Co., LTD, Beijing, China), 20 μl of 10×DNase I Reaction Buffer, and 200 μl of ddH_2O were added to the pellet and mixed well. Next, the pellet was incubated at 37°C for 10 min, and then 40 μl of EDTA (25 mmol/L) was added and incubated at 65°C for 10 min. The obtained solution was centrifuged at 12,000 rpm for 5 min, the supernatant was discarded, and the pellet was used to extract DNA using an Ezup Column Bacteria Genomic DNA Purification Kit (Sangon Biotech Co., Ltd., Shanghai, China), according to the manufacturer's instructions. Three biological replicates were performed.

10.3389/fmicb.2023.1252709

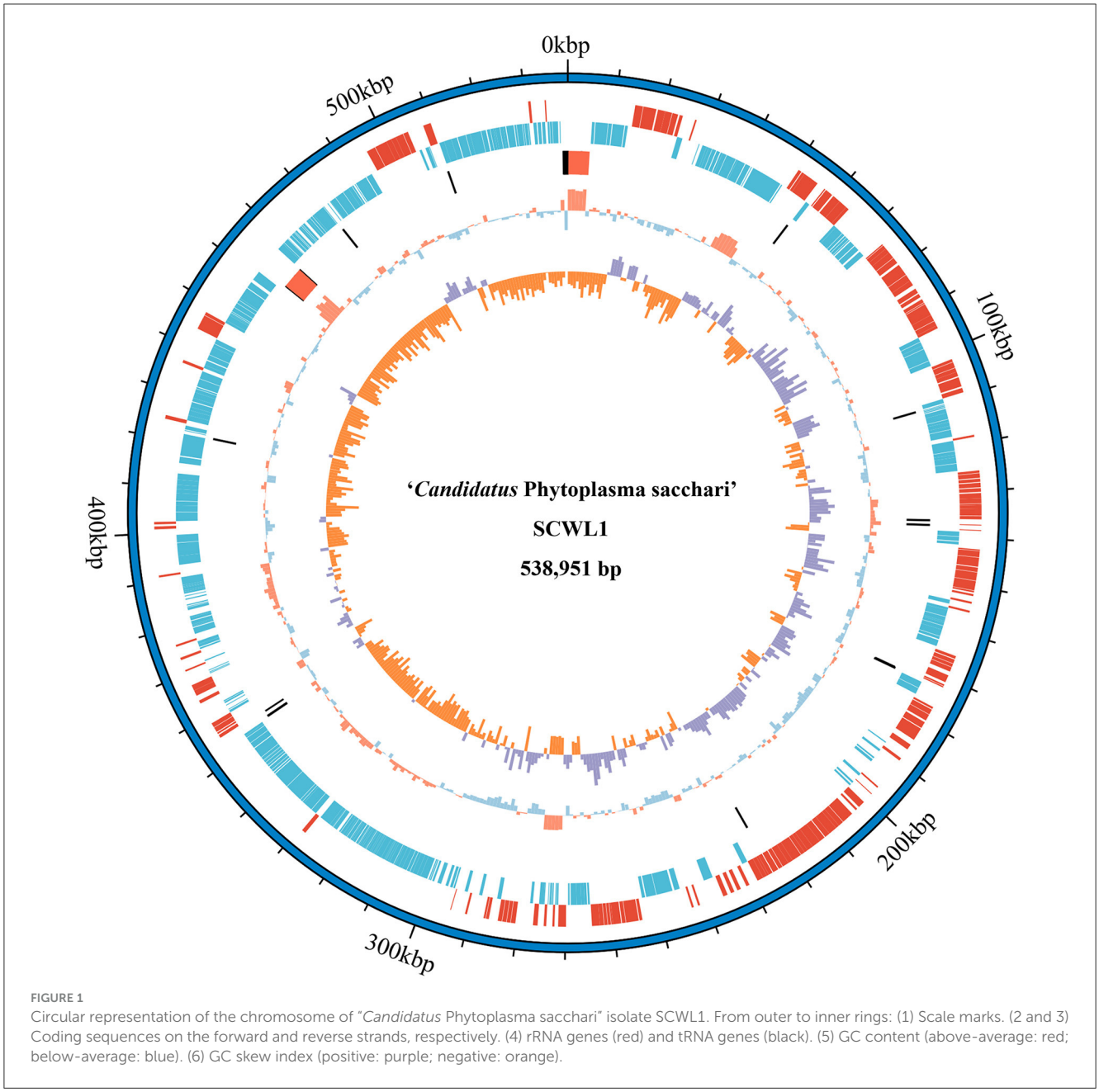


FIGURE 1
Circular representation of the chromosome of "*Candidatus* Phytoplasma sacchari" isolate SCWL1. From outer to inner rings: (1) Scale marks. (2 and 3) Coding sequences on the forward and reverse strands, respectively. (4) rRNA genes (red) and tRNA genes (black). (5) GC content (above-average: red; below-average: blue). (6) GC skew index (positive: purple; negative: orange).

TABLE 1 General features of the genome of "*Candidatus* Phytoplasma sacchari" isolate SCWL1.

Gene type	Number	Total length	Average length	Percentage of genome (%)
Gene	500	425,685	851	78.98
CDS	459	413,403	901	76.71
tRNA	27	2,152	80	0.40
23S rRNA	2	5,723	2,862	1.06
16S rRNA	2	3,046	1,523	0.57
5S rRNA	2	232	116	0.04
misc RNA	8	1,129	141	0.21

misc RNA, miscellaneous RNA.

Library preparation and sequencing

Both Illumina short-read and Nanopore long-read sequencing technologies were used for genome sequencing. For Illumina, 0.2 μg of enriched DNA was used as the input material for DNA library preparations. The sequencing library was generated using a NEBNext® Ultra™ DNA Library Prep Kit for Illumina (NEB, USA), according to the manufacturer's instructions. The DNA libraries were sequenced on an Illumina NovaSeq 6000 platform (Illumina, San Diego, USA), and 150 bp paired-end reads were generated. For Oxford Nanopore Technology (ONT) sequencing, 2.5 μg of total DNA was used as the input material for the DNA library preparations. The sequencing library was prepared using an ONT Ligation Kit (SQK-LSK109), followed by PromethION sequencing (ONT, Oxford, UK).

Genome assembly and annotation

The Unicycler v 0.5.0 software (Wick et al., 2017) was used to assemble the filtered reads. First, highly accurate Illumina data (Q30 > 85%) were used for assembly to obtain high-quality genome contigs. Second, the Nanopore data were used to connect the high-quality contigs with a complete genome. Finally, the Pilon software (Walker et al., 2014) was used to correct the assembled genome using the Illumina data to obtain the final genome sequence with higher accuracy. The Illumina sequences were mapped to the SCWL1 genome using BWA v.0.7.17 (Li and Durbin, 2009) to evaluate the effect of SCWL phytoplasma DNA enrichment. Bamdst was used to analyze the depth of sequencing. Genome annotation was performed using Prokka v1.14.6 (Seemann, 2014), which comprises Prodigal, Aragorn, RNAmmer, and Infernal that predict open reading frames (ORFs), tRNAs, rRNAs, and ncRNA, respectively. KEGG (Kyoto Encyclopedia of Genes and Genomes), COG (Cluster of Orthologous Groups of proteins), NR (Non-Redundant Protein), UniProt (Unified Protein), GO (Gene Ontology), Pfam (Protein families), RefSeq (Reference Sequence), and TIGRFAMs databases were used for functional annotation of the genome.

Phylogenetic analysis

For phylogenetic analysis, 14 complete phytoplasma genomes and the draft genome of "*Ca*. Phytoplasma sacchari" isolate SCGS (Supplementary Table 1) were compared. The homologous gene clusters were identified using OrthoMCL (Li et al., 2003). Multiple sequence alignments of single-copy homologous gene clusters were prepared using MUSCLE (Edgar, 2004) and concatenated to produce one super alignment matrix. The resulting multiple sequence alignment was used to build a phylogenetic tree using the maximum likelihood method implemented in MEGA X (Kumar et al., 2018). The average nucleotide identity (ANI) was calculated using the orthoANI tool of EzBioCloud (https://www.ezbiocloud.

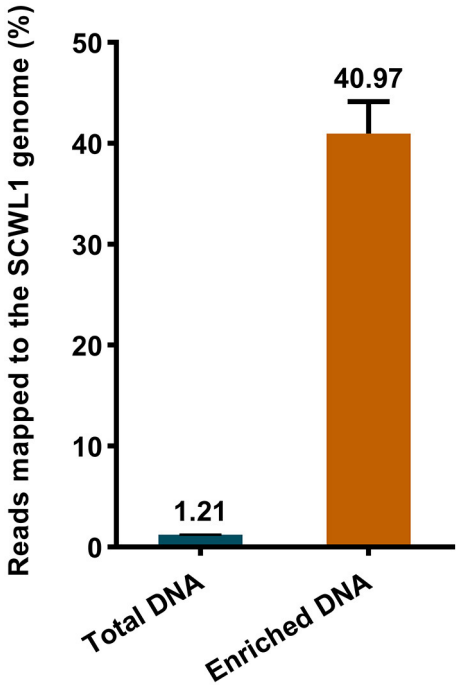


FIGURE 2
Percentage of sequence reads mapped to the genome of "*Candidatus* Phytoplasma sacchari" isolate SCWL1.

TABLE 2 Illumina sequencing data.

Run[a]	Raw reads	Raw base (G)	Clean reads	Clean base (G)	Effective rate (%)	Sequencing depth (fold)	Reads mapped to SCWL1 genome	Coverage (%)
C1	18,032,029	5.41	17,930,551	5.38	99.44	61.48	220,891	99.13
C2	15,799,153	4.74	15,687,968	4.71	99.30	52.04	186,987	99.06
C3	17,098,873	5.13	16,989,440	5.08	99.36	57.16	205,372	99.11
S1	6,060,382	1.82	5,235,443	1.57	86.39	733.85	2,322,713	100
S2	3,929,251	1.18	3,528,420	1.06	89.80	364.14	1,429,359	100
S3	4,428,736	1.33	3,894,516	1.17	87.94	358.57	1,481,357	100

[a]Runs C1–C3: Illumina sequencing of total leaf DNA, with three biological replicates; Runs S1–S3: Illumina sequencing of enriched DNA, with three biological replicates.

net/tools/ani) (Yoon et al., 2017). The digital DNA–DNA hybridization values were calculated using the Genome-to-Genome Distance Calculator (GGDC 3.0; https://ggdc.dsmz.de/ggdc.php#) (Meier-Kolthoff et al., 2022).

Results

General features of the genome of "*Ca*. phytoplasma sacchari" isolate SCWL1

Our analysis revealed that the genome of "*Ca*. Phytoplasma sacchari" isolate SCWL1 was composed of a circular chromosome and a plasmid comprising 538,951 bp with 20.54% G+C content (Figure 1) and 2,976 bp with 21.00% G+C content, respectively. The chromosome contained 459 coding sequences (CDSs), two complete 5S-23S-16S rRNA gene operons, and 27 tRNA genes (Table 1 and Figure 1). The sequence identity between the two 16S rRNA gene sequences was 100%. The total length of the CDS was 413,403 bp, and the average length was 901 bp, accounting for 76.71% of the total length of the chromosome.

Evaluation of isolate SCWL DNA enrichment method

We evaluated the efficacy of the SCWL phytoplasma DNA enrichment method for Illumina sequencing. The enriched DNA and total DNA from leaves were sequenced using Illumina sequencing. After quality-control assessment of enriched DNA and total DNA sequencing reads, an average of 4,219,460 and 16,869,320 clean reads were obtained, respectively (Table 2). Only an average of 204,417 reads from the total DNA were mapped to the SCWL1 genome, accounting for only 1.21% of all clean reads, whereas an average of 1,744,476 reads from the enriched DNA were mapped to the SCWL1 genome, accounting for 40.97% of all clean

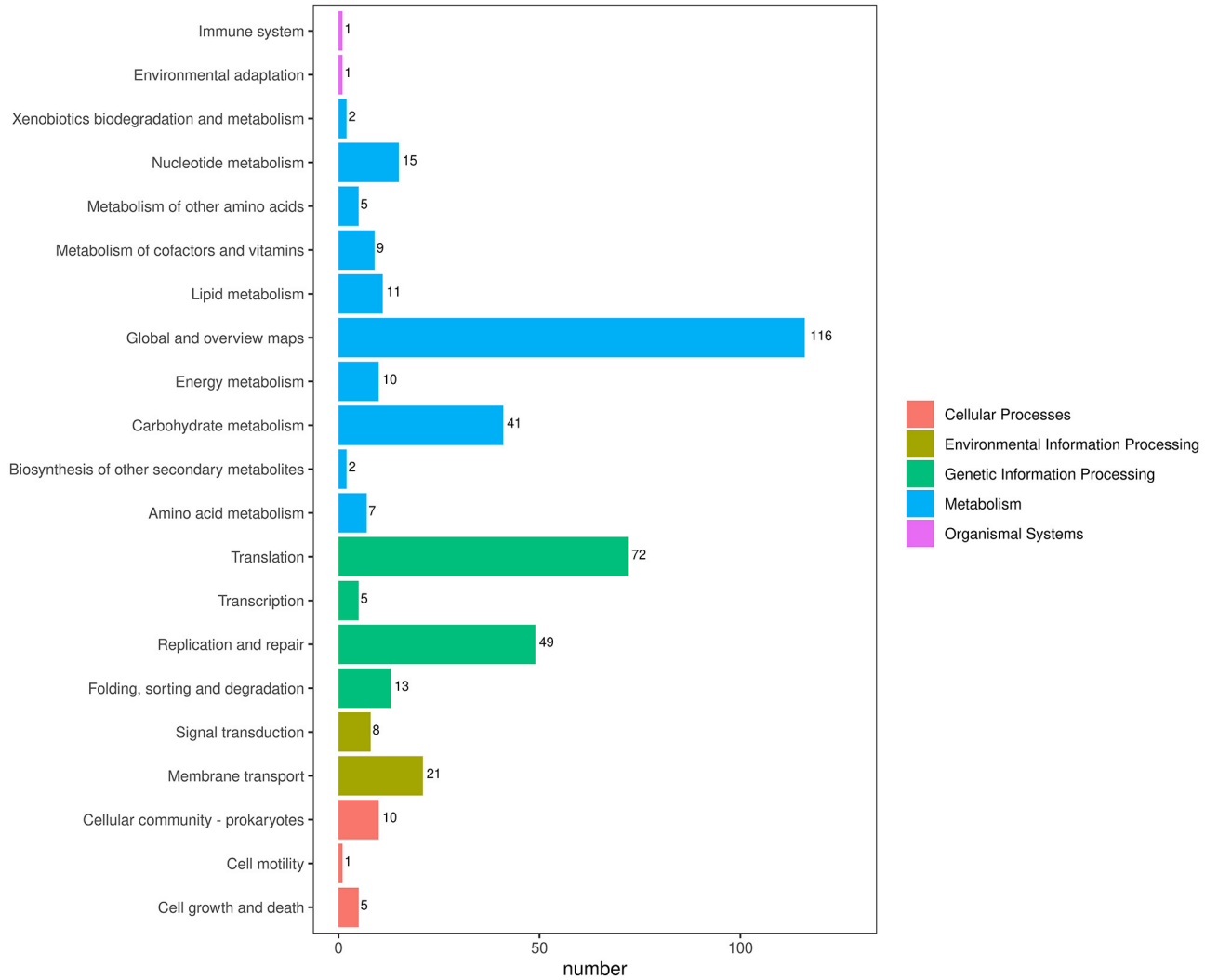


FIGURE 3
Classification map of Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway annotation analysis of the genome of "*Candidatus* Phytoplasma sacchari" isolate SCWL1.

reads (Figure 2). The highest sequence coverage from total DNA was 99.13% and that from enriched DNA was 100% (Table 2).

Functional annotation for the protein-coding genes

To obtain comprehensive information on gene function, the protein-coding genes in the SCWL1 genome were annotated using eight databases (Supplementary Table 2). Two hundred genes were annotated using the KEGG database and classified according to the KEGG pathway (Figure 3). The maximum number of genes in metabolism were enriched in global and overview maps (116 genes) and carbohydrate metabolism (41 genes); in genetic information processing, they were enriched in translation (72 genes) and replication and repair (49 genes). Three hundred and forty-eight genes were annotated using the COG database and assigned to 21 functional categories (Figure 4). The most abundant functional class was COG class J (translation, ribosomal structure, and biogenesis). Based on the GO database, we annotated 378 genes, which were categorized into three functional categories (biological process, cellular component, and molecular function). The top 20 GO terms with the most annotations of each functional category are shown in Figure 5. The most enriched biological process, cellular component, and molecular function terms were translation, integral component of plasma membrane and cytoplasm, and ATP binding, respectively. The highest number of genes was annotated in the Nr database (424 genes) and the RefSeq database (424 genes). In the NR database, 389 genes were annotated to the genome of "*Ca*. Phytoplasma sacchari," accounting for 91.75% of all annotated genes.

Metabolic pathways

Like other phytoplasma isolates, isolate SCWL1 lacked many genes encoding the tricarboxylic acid cycle, epoxidative phosphorylation, pentose phosphate pathway, and F1F0 ATP synthase. The genes encoding the phosphoenol pyruvate-dependent sugar phosphotransferase system (PTS), hexokinase, and sugar transport system (*malE*, *malG*, and *malF*) were also

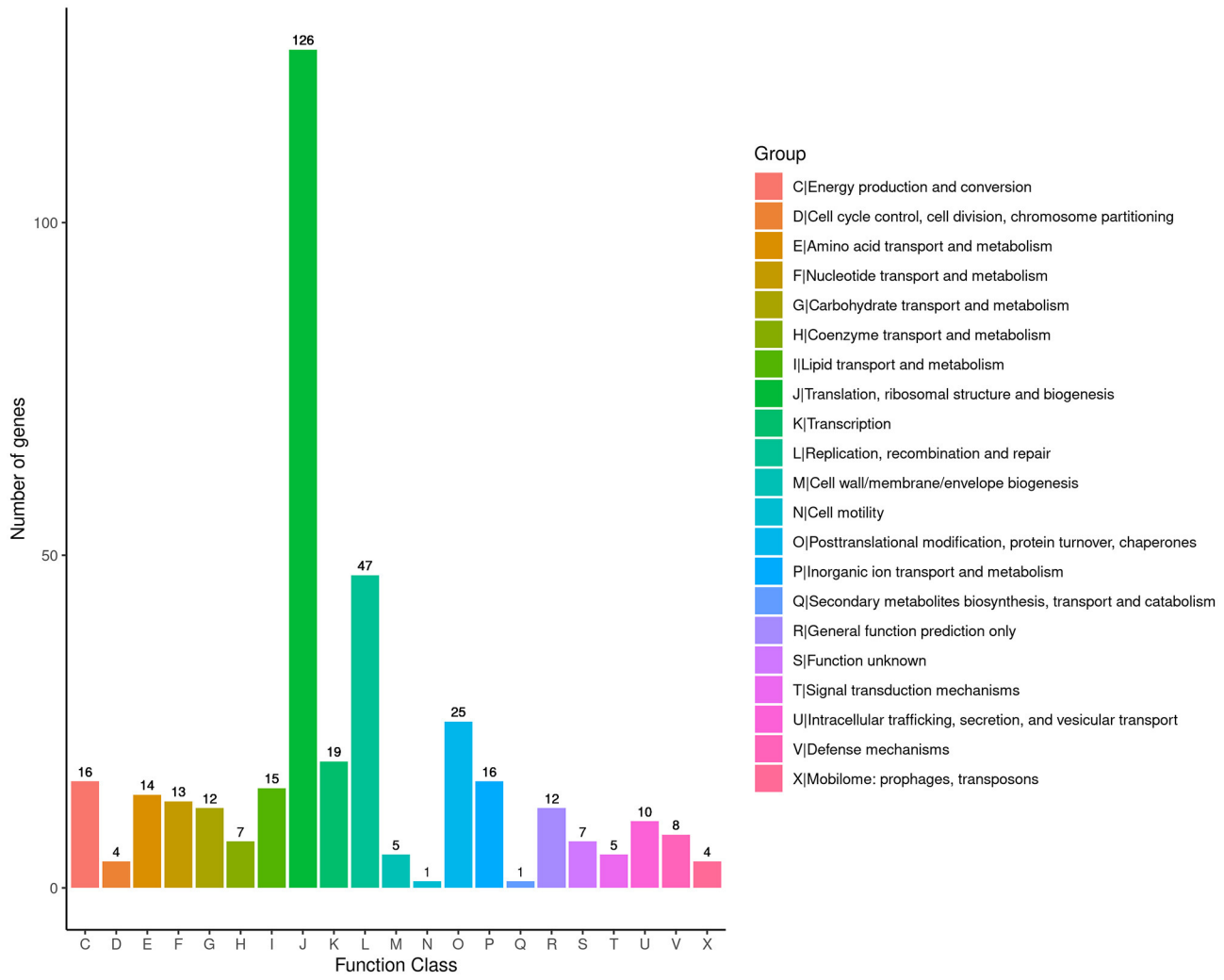


FIGURE 4
Classification map of Cluster of Orthologous Groups of proteins (COG) annotation analysis of the genome of "*Candidatus* Phytoplasma sacchari" isolate SCWL1.

absent. Similar to that in the genome of "*Ca*. Phytoplasma mali" isolate AT, only five glycolysis-related genes (*Pgi*, *PfkA*, *FbaA*, *TpiA*, and *PykF*) were present in the genome of isolate SCWL1 (Figure 6). Although the SCWL1 genome did not possess glycolytic pathway-related genes, the genes encoding malate or citrate transporter protein (*citS*), malic enzyme (*sfcA*), pyruvate dehydrogenase multienzyme complex (*pdhA*, *pdhB*, *pdhC*, and *pdhD*), and a putative phosphate propanoyl transferase (*pduL*) were present (Figure 6). In addition, the citrate lyase gene clusters *(citXFEDG)* encoding the apo-citrate lyase phosphoribosyl-dephospho-CoA transferase *(citX)*, the α-subunit *(citF)*, the β-subunit *(citE)* and the γ-subunit *(citD)* of citrate lyase, and 2-(5′-triphosphoribosyl)-3′-dephosphocoenzyme-A synthase (*citG*) is found in SCWL1 genomes.

Potential mobile units (PMUs) and effector genes

PMUs are commonly found in phytoplasma genomes. A PMU, with a size of 23.6 kb, consisting of *tra5*, *tmk*, *dnaB*, and *dnaG* was found in the genome of isolate SCWL1 (Figure 7 and Supplementary Table 3). Other core genes of the phytoplasma PMU region, such as *ssb*, *rpoD*, and *himA*, were scattered throughout the genome of the isolate. In the PMU region, two incomplete *hflB* genes and one incomplete *dnaG* gene were annotated. Proteins homologous to phytoplasma effectors, such as TENGU, SAP05, SAP11, and SAP54, were not found in the genome of isolate SCWL1.

Phylogenetic relationships

The comparative analysis of isolate SCWL1 and 14 phytoplasma genomes revealed the presence of 191 single-copy orthologous proteins. The phylogenetic tree constructed based on the concatenated sequences of these single-copy proteins elucidated that isolate SCWL1 was most closely related to isolate SCGS (Figure 8). Comparison analysis of 16S rRNA gene sequences (full length) indicated that isolates SCWL1 and SCGS shared 99.87% sequence identity. At the whole-genome level, the ANI value for isolate SCGS against isolate SCWL1 was 98.80%, and the digital DNA–DNA hybridization (dDDH) value was 89.50%.

Discussion

In the present study, we developed a novel method for enriching the DNA of phytoplasma isolate SCWL, which is simpler and faster than the previously established methods. In brief, the method is as follows: First, the SCWL phytoplasma

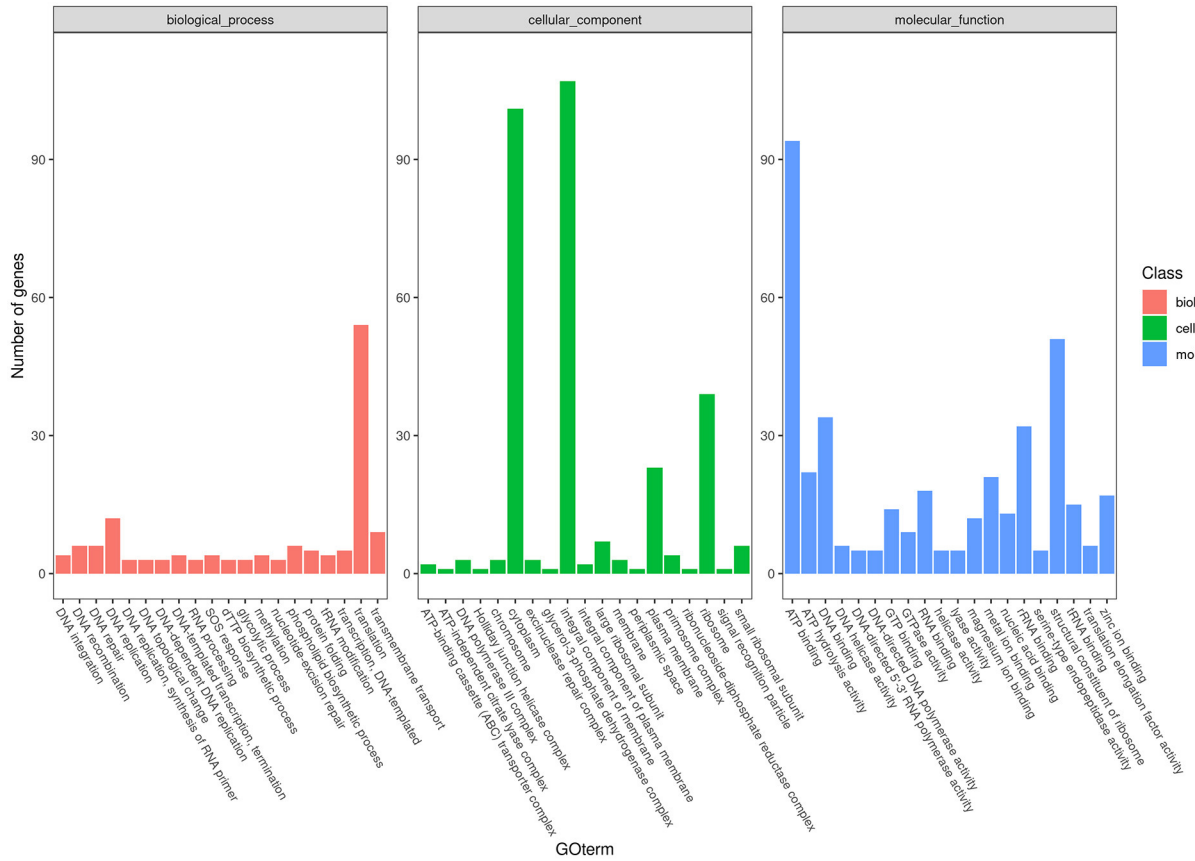


FIGURE 5
Classification map of Gene Ontology (GO) annotation analysis of the genome of "*Candidatus* Phytoplasma sacchari" isolate SCWL1.

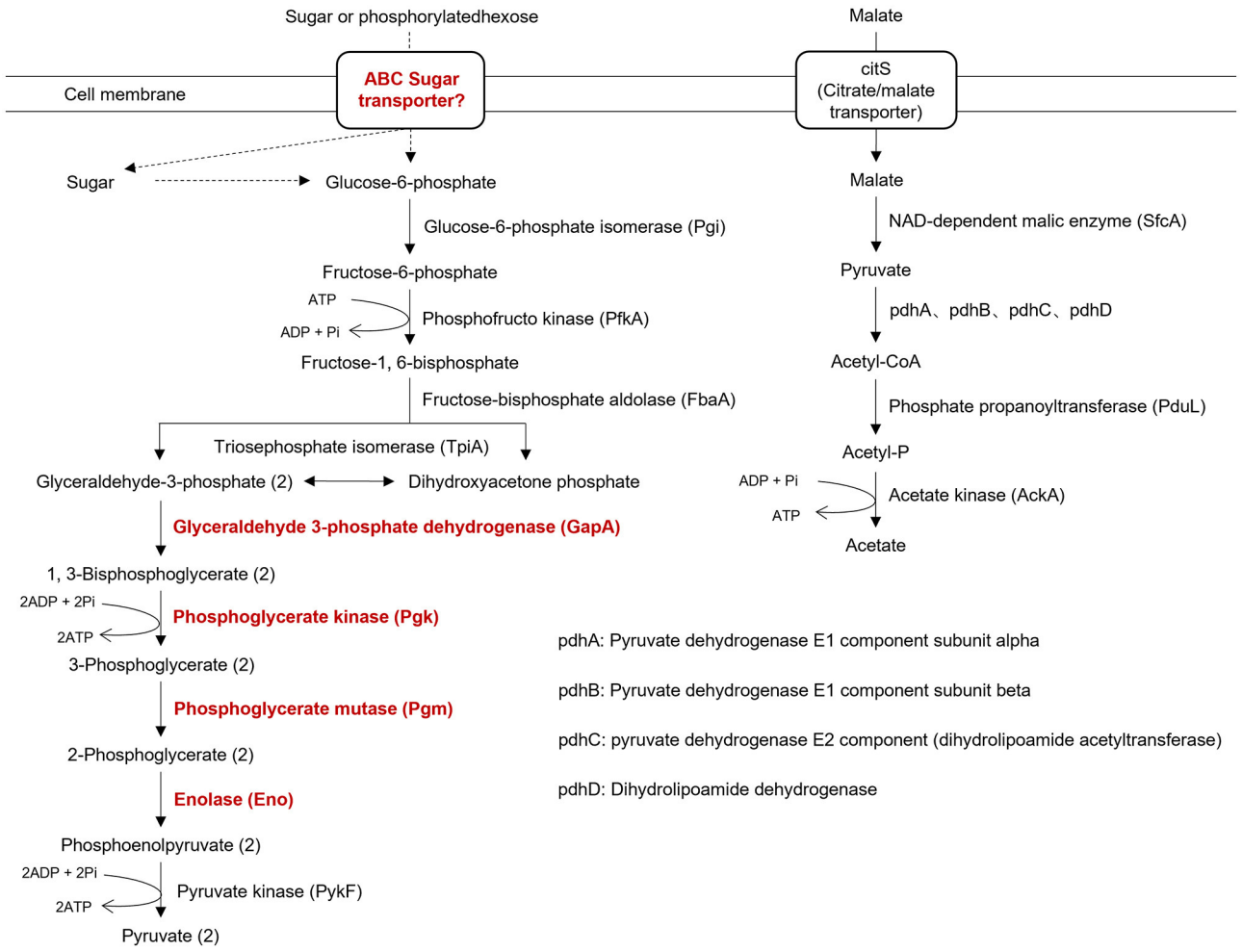


FIGURE 6
Annotation of genes related to energy-yielding pathways in the genome of isolate SCWL1. The proteins marked in red indicate that they are not absent in the genome of isolate SCWL1 genome; the dotted line represents the absence of the enzyme involved in the reaction.

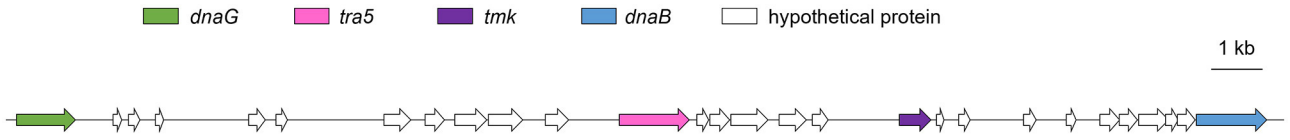


FIGURE 7
Potential mobile units (PMUs) in the genome of "*Candidatus* Phytoplasma sacchari" isolate SCWL1. dnaG, DNA primase; tra5, IS3 family transposase; hflB, ATP-dependent Zn protease; tmk, thymidylate kinase; dnaB, replicative DNA helicase.

was released by grinding the sugarcane leaves. Second, the host DNA released during the grinding process was removed by washing several times; the host tissues and cells were removed *via* serial filtration; and the residual host DNA was digested using DNase I. Finally, the genomic DNA from the filtered SCWL phytoplasma cells was extracted. The results of Illumina sequencing revealed that the number of SCWL phytoplasma reads for enriched DNA was significantly increased, and data such as average sequencing depth and coverage were better than those obtained using total leaf DNA. Although up to 40.97% of the reads obtained *via* the sequencing of enriched DNA were mapped to the genome of isolate SCWL1, non-SCWL phytoplasma reads still accounted for a large fraction of the total reads. The reason for this result could be that many endophytic microorganisms and host plant organelles were present during the final DNA extraction step, i.e., the filtration process. Although this is a limitation of the enrichment method developed in this study, the method does not require expensive equipment and reagents and is convenient and fast, and the enriched DNA can meet Illumina sequencing requirements.

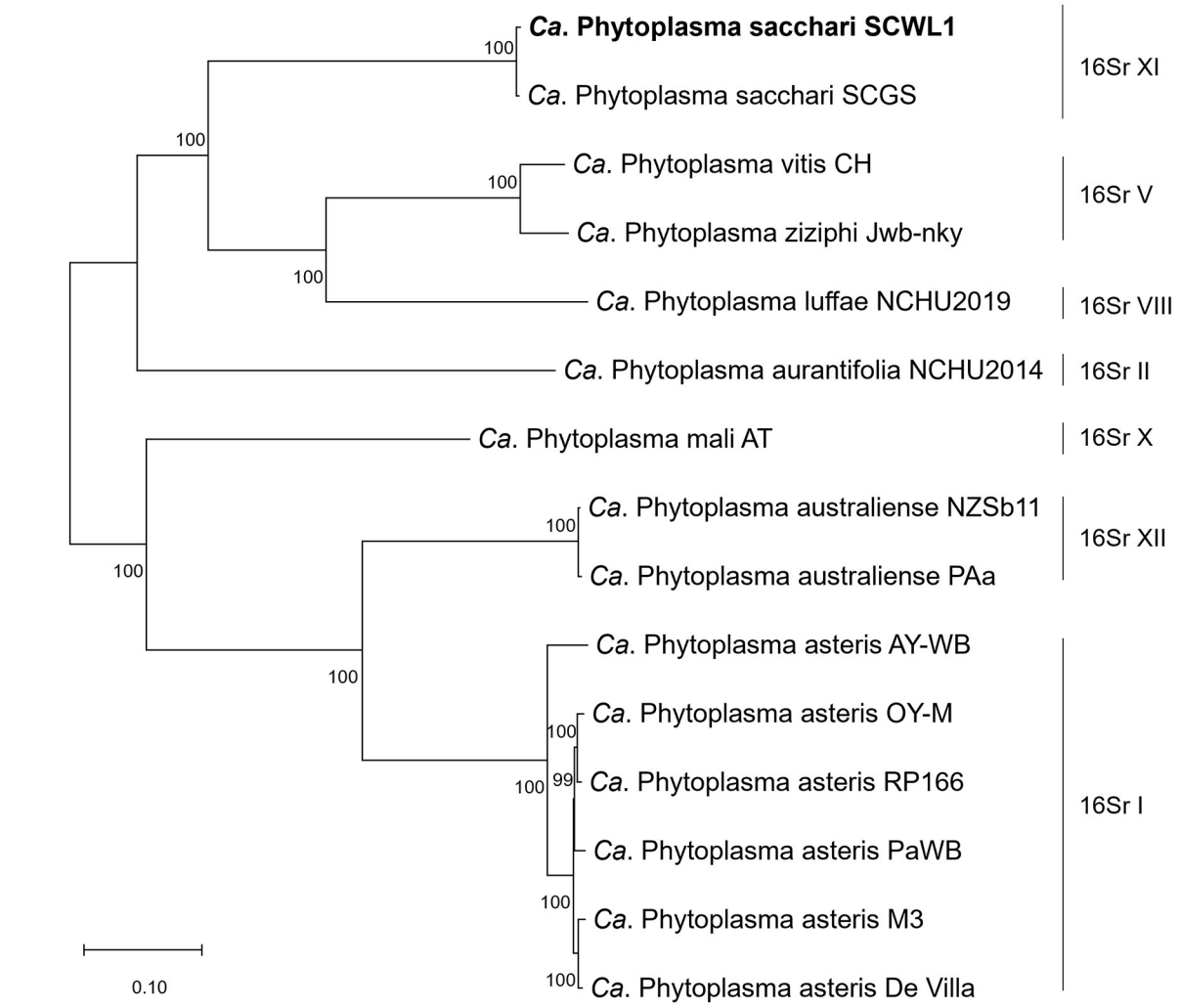


FIGURE 8
Maximum likelihood phylogeny of phytoplasmas inferred based on a concatenated alignment of 191 single-copy orthologous proteins.

Recently, with the reduction in sequencing costs and the development of numerous sequencing technologies, such as NGS, phytoplasma genomes can be easily sequenced and larger sequencing data can be generated with limited funds, thereby achieving higher coverage and generating a more complete genome draft. The emergence of third-generation sequencing technology has enabled the generation of longer read lengths, making genome assembly easier. Phytoplasma genomes are rich in repeated DNA sequences, thus making genome assembly difficult using only second-generation sequencing data. In this study, although the coverage of Illumina sequencing using enriched DNA was 100%, the assembly of the genome of isolate SCWL1 was unsuccessful using only Illumina sequencing data. Recently, several complete phytoplasma genomes have been generated by combining second- and third-generation sequencing technologies (Wang et al., 2018; Debonneville et al., 2022; Huang et al., 2022). In this study, although enriched DNA was used for second-generation sequencing, it was not suitable for third-generation library preparation and sequencing due to low DNA concentration; therefore, we performed Nanopore sequencing using total leaf DNA. In this study, although the combination of second- and third-generation sequencing of phytoplasma genomes does not require genome enrichment, Illumina sequencing reads using the total leaf DNA did not completely cover the genome of isolate SCWL1. The accuracy of Nanopore sequencing is lower than that of Illumina sequencing, suggesting that the accuracy of the assembled genome assembly will be reduced if only Nanopore sequencing data are used. Therefore, appropriate enrichment of phytoplasma DNA is essential for phytoplasma genome sequencing.

Initially, the size of phytoplasma genomes was estimated to be 530–1350 kb with 21–33% GC content (Neimark and Kirkpatrick, 1993; Marcone et al., 1999; IRPCM Phytoplasma/Spiroplasma Working Team-Phytoplasma Taxonomy Group, 2004). Recent studies have reported the size of complete phytoplasma genomes to be 576–960 kb (Wei and Zhao, 2022). In this study, the size of the SCWL1 chromosome is 538,951 bp, which is the smallest complete phytoplasma chromosome reported, and the predicted GC content and the number of coding genes are also the least among all complete phytoplasma genomes that have been reported. Similar to that of the "*Ca*. Phytoplasma mali" isolate AT, the chromosome of isolate SCWL1 exhibited a regular cumulative GC-skew pattern. Due to the lack of all glycolysis-related genes in the genome of

"*Ca*. Phytoplasma mali" isolate AT, it is proposed that malate is utilized as carbon and energy sources (Kube et al., 2008). In this study, the genome of isolate SCWL1 also lacked the glycolysis-related enzymes, but the enzymes encoding the conversion pathway for malate conversion to acetate were present. Therefore, it is possible that isolate SCWL1 does not rely on glycolysis for energy production, and the malate-to-acetate pathway is an alternative to glycolysis and the main pathway for isolate SCWL1 to obtain carbon sources and produce energy.

With advancements in genome sequencing technology, the classification and phylogeny of phytoplasmas based on whole-genome sequence can be elucidated. The accepted minimum threshold for taxon assignment in prokaryotes using genomic data is that the same species isolates should have ANI values >95–96% and DDH values >70% (Richter and Rosselló-Móra, 2009; Chun et al., 2018). In 2022, the revised version of the guidelines for defining "*Ca*. phytoplasma" species proposed a whole-genome ANI standard of 95% for "*Ca*. phytoplasma" species delineation (Bertaccini et al., 2022). Previous studies have proposed isolate SCGS as a novel taxon "*Ca*. Phytoplasma sacchari" (Kirdat et al., 2021). Recently, multilocus sequence typing revealed that SCGS and SCWL phytoplasma isolates belonged to different populations of "*Ca*. Phytoplasma sacchari," but the classification was made without any genomic-level evidence (Abeysinghe et al., 2016; Zhang et al., 2023). In this study, we analyzed the phylogenetic relationships between isolates SCGS and SCWL at the genomic level and found that the ANI and dDDH values between their genomes were higher than the threshold values for taxon assignment of "*Ca*. phytoplasma" species. Since the genomes of only two "*Ca*. Phytoplasma sacchari" isolates are available, for further elucidation of the evolutionary relationship and population structure of "*Ca*. Phytoplasma sacchari," genome sequencing of more isolates belonging to this genus is required; this can be performed using the enrichment method developed in this study.

Conclusion

To improve the efficiency of phytoplasma sequencing, a filter-based enrichment method for the genome of phytoplasma isolate SCWL was developed. The method increased the number of phytoplasma sequences obtained *via* Illumina sequencing. This method will not only help in the initiation of more "*Ca*. Phytoplasma sacchari" genome sequencing projects but also act as an important reference for the enrichment of the genome DNA of other phytoplasma species. The genome sequence of isolate SCWL1 is the first complete genome sequence of a phytoplasma isolate belonging to the 16SrXI group, thus promoting an in-depth understanding of the genomic characteristics of the 16SrXI group. Moreover, the chromosome of "*Ca*. Phytoplasma sacchari" isolate SCWL1 is the smallest circular chromosome among the phytoplasma with complete genome sequences available. This study also provides genomic evidence that isolates SCGS phytoplasma and SCWL belong to the same phytoplasma species. The availability of the complete genome of isolate SCWL1 will contribute to future studies on the molecular evolution and pathogenesis of "*Ca*. Phytoplasma sacchari."

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary material.

Author contributions

R-YZ and X-HH conceived and designed the experiment, analyzed the data, and wrote the manuscript. X-YW, JL, H-LS, and Y-HL performed the experiment and participated in the data analysis. Y-KH acquired the funding, supervised the project, and revised the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This study was supported by the National Natural Science Foundation of China (Grant Numbers 32260645 and 31760504), the Yunnan Fundamental Research Project (Grant Number 202101AT070240), the China Agriculture Research System of MOF and MARA (Grant Number CARS-170303), and the Yunling Industry and Technology Leading Talent Training Program Prevention and Control of Sugarcane Pests (Grant Number 2018LJRC56).

Acknowledgments

The authors would like to thank Yin-Zhan Cui for technical assistance.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2023.1252709/full#supplementary-material

References

Abeysinghe, S., Abeysinghe, P. D., Silva, C. K., Udagama, P., Warawichanee, K., Aljafar, N., et al. (2016). Refinement of the taxonomic structure of 16SrXI and 16SrXIV phytoplasmas of gramineous plants using multilocus sequence typing. *Plant Dis.* 100, 2001–2010. doi: 10.1094/PDIS-02-16-0244-RE

Bai, X., Zhang, J., Ewing, A., Miller, S. A., Jancso Radek, A., Shevchenko, D. V., et al. (2006). Living with genome instability: the adaptation of phytoplasmas to diverse environments of their insect and plant hosts. *J. Bacteriol.* 188, 3682–3696. doi: 10.1128/JB.188.10.3682-3696.2006

Bertaccini, A., Arocha-Rosete, Y., Contaldo, N., Duduk, B., Fiore, N., Montano, H. G., et al. (2022). Revision of the '*Candidatus* Phytoplasma' species description guidelines. *Int. J. Syst. Evol. Microbiol.* 72, 005353. doi: 10.1099/ijsem.0.005353

Chen, W., Li, Y., Wang, Q., Wang, N., and Wu, Y. (2014). Comparative genome analysis of wheat blue dwarf phytoplasma, an obligate pathogen that causes wheat blue dwarf disease in China. *PLoS ONE* 9, e96436. doi: 10.1371/journal.pone.0096436

Cho, S. T., Lin, C. P., and Kuo, C. H. (2019). Genomic characterization of the periwinkle leaf yellowing (PLY) phytoplasmas in Taiwan. *Front. Microbiol.* 10, 2194. doi: 10.3389/fmicb.2019.02194

Chun, J., Oren, A., Ventosa, A., Christensen, H., Arahal, D. R., da Costa, M. S., et al. (2018). Proposed minimal standards for the use of genome data for the taxonomy of prokaryotes. *Int. J. Syst. Evol. Microbiol.* 68, 461–466. doi: 10.1099/ijsem.0.002516

Debonneville, C., Mandelli, L., Brodard, J., Groux, R., Roquis, D., and Schumpp, O. (2022). The complete genome of the "Flavescence Dorée" phytoplasma reveals characteristics of low genome plasticity. *Biology.* 11, 953. doi: 10.3390/biology11070953

Doi, Y., Teranaka, M., Yora, K., and Asuyama, H. (1967). Mycoplasma- or PLT group-like microorganisms found in the phloem elements of plants infected with mulberry dwarf, potato witches' broom, aster yellows or paulownia witches' broom. *Ann. Phytopaththol. Soc. Jpn.* 33, 259–266. doi: 10.3186/jjphytopath.33.259

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792–1797. doi: 10.1093/nar/gkh340

Huang, C. T., Cho, S. T., Lin, Y. C., Tan, C. M., Chiu, Y. C., Yang, Y. J., et al. (2022). Comparative genome analysis of '*Candidatus* Phytoplasma luffae' reveals the influential roles of potential mobile units in phytoplasma evolution. *Front. Microbiol.* 13, 773608. doi: 10.3389/fmicb.2022.773608

IRPCM Phytoplasma/Spiroplasma Working Team-Phytoplasma Taxonomy Group (2004). '*Candidatus* Phytoplasma', a taxon for the wall-less, non-helical prokaryotes that colonize plant phloem and insects. *Int. J. Syst. Evol. Microbiol.* 54, 1243–1255. doi: 10.1099/ijs.0.02854-0

Kirdat, K., Tiwarekar, B., Sathe, S., and Yadav, A. (2023). From sequences to species: Charting the phytoplasma classification and taxonomy in the era of taxogenomics. *Front. Microbiol.* 14, 1123783. doi: 10.3389/fmicb.2023.1123783

Kirdat, K., Tiwarekar, B., Thorat, V., Narawade, N., Dhotre, D., Sathe, S., et al. (2020). Draft genome sequences of two phytoplasma strains associated with Sugarcane Grassy Shoot (SCGS) and Bermuda Grass White Leaf (BGWL) diseases. *Mol. Plant Microbe Interact.* 33, 715–717. doi: 10.1094/MPMI-01-20-0005-A

Kirdat, K., Tiwarekar, B., Thorat, V., Sathe, S., Shouche, Y., and Yadav, A. (2021). '*Candidatus* Phytoplasma sacchari', a novel taxon-associated with Sugarcane Grassy Shoot (SCGS) disease. *Int. J. Syst. Evol. Microbiol.* 71, 004591. doi: 10.1099/ijsem.0.004591

Kube, M., Schneider, B., Kuhl, H., Dandekar, T., Heitmann, K., Migdoll, A. M., et al. (2008). The linear chromosome of the plant-pathogenic mycoplasma '*Candidatus* Phytoplasma mali'. *BMC Genomics* 9, 306. doi: 10.1186/1471-2164-9-306

Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. (2018). MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* 35, 1547–1549. doi: 10.1093/molbev/msy096

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754–1760. doi: 10.1093/bioinformatics/btp324

Li, L., Stoeckert, C. J., and Roos, D. S. (2003). OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res.* 13, 2178–2189. doi: 10.1101/gr.1224503

Lim, H. J., Lee, E. H., Yoon, Y., Chua, B., and Son, A. (2016). Portable lysis apparatus for rapid single-step DNA extraction of Bacillus subtilis. *J. Appl. Microbiol.* 120, 379–387. doi: 10.1111/jam.13011

Marcone, C., Neimark, H., Ragozzino, A., Lauer, U., and Seemuller, E. (1999). Chromosome sizes of phytoplasmas composing major phylogenetic groups and subgroups. *Phytopathology* 89, 805–810. doi: 10.1094/PHYTO.1999.89.9.805

Meier-Kolthoff, J. P., Carbasse, J. S., Peinado-Olarte, R. L., and Göker, M. (2022). TYGS and LPSN: a database tandem for fast and reliable genome-based classification and nomenclature of prokaryotes. *Nucleic Acids Res.* 50, D801–D807. doi: 10.1093/nar/gkab902

Neimark, H., and Kirkpatrick, B. C. (1993). Isolation and characterization of full-length chromosomes from non-culturable plant-pathogenic Mycoplasma-like organisms. *Mol. Microbiol.* 7, 21–28. doi: 10.1111/j.1365-2958.1993.tb01093.x

Nijo, T., Iwabuchi, N., Tokuda, R., Suzuki, T., Matsumoto, O., Miyazaki, A., et al. (2021). Enrichment of phytoplasma genome DNA through a methyl-CpG binding domain-mediated method for efficient genome sequencing. *J. Gen. Plant Pathol.* 87, 154–163. doi: 10.1007/s10327-021-00993-z

Oshima, K. (2021). Molecular biological study on the survival strategy of phytoplasma. *J. Gen. Plant Pathol.* 87, 403–407. doi: 10.1007/s10327-021-01027-4

Oshima, K., Kakizawa, S., Nishigawa, H., Jung, H. Y., Wei, W., Suzuki, S., et al. (2004). Reductive evolution suggested from the complete genome sequence of a plant-pathogenic phytoplasma. *Nat. Genet.* 36, 27–29. doi: 10.1038/ng1277

Richter, M., and Rosselló-Móra, R. (2009). Shifting the genomic gold standard for the prokaryotic species definition. *Proc. Natl. Acad. Sci. USA.* 106, 19126–19131. doi: 10.1073/pnas.0906412106

Seemann, T. (2014). Prokka: Rapid prokaryotic genome annotation. *Bioinformatics* 30, 2068–2069. doi: 10.1093/bioinformatics/btu153

Tan, C. M., Lin, Y. C., Li, J. R., Chien, Y. Y., Wang, C. J., Chou, L., et al. (2021). Accelerating complete phytoplasma genome assembly by immunoprecipitation-based enrichment and MinION-based DNA sequencing for comparative analyses. *Front. Microbiol.* 12, 766221. doi: 10.3389/fmicb.2021.766221

Tully, J. G. (1993). International committee on systematic bacteriology subcommittee on the taxonomy of mollicutes. Minutes of the interim meetings, 1 and 2 Aug. 1992, Ames, Iowa. *Int. J. Syst. Bacteriol.* 43, 394–397. doi: 10.1099/00207713-43-2-394

Viswanathan, R., Chinnaraja, C., Karuppaiah, R., Ganesh, K. V., Jenshi, R. J., and Malathi, P. (2011). Genetic diversity of sugarcane grassy shoot (SCGS)-phytoplasmas causing grassy shoot disease in India. *Sugar Tech* 13, 220–228. doi: 10.1007/s12355-011-0084-2

Walker, B. J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., et al. (2014). Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS ONE* 9, e112963. doi: 10.1371/journal.pone.0112963

Wang, J., Song, L., Jiao, Q., Yang, S., Gao, R., Lu, X., et al. (2018). Comparative genome analysis of jujube witches'-broom phytoplasma, an obligate pathogen that causes jujube witches'- broom disease. *BMC Genomics* 19, 689. doi: 10.1186/s12864-018-5075-1

Wang, X. Y., Zhang, R. Y., Li, J., Li, Y. H., Shan, H. L., Li, W. F., et al. (2022). The diversity, distribution and status of phytoplasma diseases in China. *Front. Sustain. Food Syst.* 6, 943080. doi: 10.3389/fsufs.2022.943080

Wei, W., and Zhao, Y. (2022). Phytoplasma taxonomy: nomenclature, classification, and identification. *Biology.* 11, 1119. doi: 10.3390/biology11081119

Wick, R. R., Judd, L. M., Gorrie, C. L., and Holt, K. E. (2017). Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Comput. Biol.* 13, e1005595. doi: 10.1371/journal.pcbi.1005595

Yadav, A., Thorat, V., Deokule, S., Shouche, Y., and Prasad, D. T. (2017). New subgroup 16SrXI-F phytoplasma strain associated with sugarcane grassy shoot (SCGS) disease in India. *Int. J. Syst. Evol. Microbiol.* 67, 374–378. doi: 10.1099/ijsem.0.001635

Yoon, S. H., Ha, S. M., Kwon, S., Lim, J., Kim, Y., Seo, H., et al. (2017). Introducing EzBioCloud: a taxonomically united database of 16S rRNA gene sequences and whole-genome assemblies. *Int. J. Syst. Evol. Microbiol.* 67, 1613–1617. doi: 10.1099/ijsem.0.001755

Zhang, R. Y., Li, W. F., Huang, Y. K., Wang, X. Y., Shan, H. L., Luo, Z. M., et al. (2016). Group 16SrXI phytoplasma strains, including subgroup 16SrXI-B and a new subgroup, 16SrXI-D, are associated with sugar cane white leaf. *Int. J. Syst. Evol. Microbiol.* 66, 487–491. doi: 10.1099/ijsem.0.000712

Zhang, R. Y., Shan, H. L., Huang, Y. K., Wang, X. Y., Li, J., Li, W. F., et al. (2020). Survey of incidence and nested PCR detection of sugarcane white leaf in different varieties. *Plant Dis.* 104, 2665–2668. doi: 10.1094/PDIS-11-19-2482-RE

Zhang, R. Y., Wang, X. Y., Shan, H. L., Li, J., Li, Y. H., Li, W. F., et al. (2023). Multilocus sequence typing reveals two distinct populations of "*Candidatus* Phytoplasma sacchari" in China. *Trop. Plant Pathol.* 48, 199–206. doi: 10.1007/s40858-023-00562-z

frontiers | Frontiers in Microbiology

TYPE Mini Review
PUBLISHED 22 March 2024
DOI 10.3389/fmicb.2024.1330660

OPEN ACCESS

EDITED BY
Michal Letek,
University of León, Spain

REVIEWED BY
Dimitri Poddighe,
Nazarbayev University, Kazakhstan

*CORRESPONDENCE
Peng Ling
lp5501211@139.com

RECEIVED 21 November 2023
ACCEPTED 27 February 2024
PUBLISHED 22 March 2024

CITATION
Yang Z, Zhou J, Su N, Zhang Z, Chen J, Liu P and Ling P (2024) Insights into the defensive roles of lncRNAs during *Mycoplasma pneumoniae* infection. *Front. Microbiol.* 15:1330660. doi: 10.3389/fmicb.2024.1330660

Insights into the defensive roles of lncRNAs during *Mycoplasma pneumoniae* infection

Zhujun Yang[1,2], Junjun Zhou[1,2], Nana Su[1,2], Zifan Zhang[1,2], Jiaxin Chen[1,2], Peng Liu[1,2] and Peng Ling[1]*

[1]Department of Critical Care Medicine, The Central Hospital of Shaoyang City and Affiliated Shaoyang Hospital, Hengyang Medical College, University of South China, Shaoyang, China, [2]Institute of Pathogenic Biology, Basic Medical School, Hengyang Medical School, University of South China, Hunan Provincial Key Laboratory for Special Pathogens Prevention and Control, Hengyang, China

Mycoplasma pneumoniae causes respiratory tract infections, affecting both children and adults, with varying degrees of severity ranging from mild to life-threatening. In recent years, a new class of regulatory RNAs called long non-coding RNAs (lncRNAs) has been discovered to play crucial roles in regulating gene expression in the host. Research on lncRNAs has greatly expanded our understanding of cellular functions involving RNAs, and it has significantly increased the range of functions of lncRNAs. In lung cancer, transcripts associated with lncRNAs have been identified as regulators of airway and lung inflammation in a process involving protein complexes. An excessive immune response and antibacterial immunity are closely linked to the pathogenesis of *M. pneumoniae*. The relationship between lncRNAs and *M. pneumoniae* infection largely involves lncRNAs that participate in antibacterial immunity. This comprehensive review aimed to examine the dysregulation of lncRNAs during *M. pneumoniae* infection, highlighting the latest advancements in our understanding of the biological functions and molecular mechanisms of lncRNAs in the context of *M. pneumoniae* infection and indicating avenues for investigating lncRNAs-related therapeutic targets.

KEYWORDS

lncRNAs, *Mycoplasma pneumoniae*, *Mycoplasma pneumoniae* infection, lncRNAs functions, regulatory mechanism of lncRNAs

1 Introduction

Mycoplasma pneumoniae, an atypical bacterium, is one of the smallest prokaryotic microorganisms without a cell wall (Shimizu, 2015). There are 200 known mycoplasma species, including six main species, which can cause human respiratory and reproductive tract diseases, among other diseases (Combaz-Söhnchen and Kuhn, 2017; Gómez Rufo et al., 2021). *M. pneumoniae* is one of the main pathogenic mycoplasmas, and it is a significant cause of respiratory tract infections. It causes endemic and epidemic primary atypical pneumonia, tracheobronchitis, pharyngitis, and asthma worldwide. *M. pneumoniae* pneumonia is the most significant disease associated with *M. pneumoniae* infection (Shimizu, 2016; Waites et al., 2017; Tsai et al., 2021). In addition, *M. pneumoniae* can cause infections outside the lungs (de Groot et al., 2017) by penetrating host cell membranes and invading respiratory tract mucous membranes, resulting in a pronounced inflammatory response outside the respiratory system (Poddighe, 2018). The severity of the diseases caused by *M. pneumoniae* ranges from mild to

life-threatening (Waites et al., 2017). The dominant pathogenic mechanisms of *M. pneumoniae* are direct cytotoxicity and adhesion to host cells, immune evasion, and inflammation-induced damage (Jiang et al., 2021). The pathogenic mechanisms of extrapulmonary manifestations also involve direct injury mediated by inflammatory factors, indirect injury caused by the host immune response, and vascular occlusion (Hu et al., 2022).

Genes, which direct an organism's development and function, include sequences with and without protein-coding functions (García-Andrade et al., 2022). Long non-coding RNAs (lncRNAs) comprise >200 nucleotides that do not code for proteins (Mattick et al., 2023). LncRNAs are widely expressed and play key roles in gene expression regulation. LncRNAs mainly interact with microRNAs (miRNAs), mRNAs, DNAs, and proteins, and they can thereby modulate gene expression in a variety of ways, e.g., by modulating chromatin function or regulating membraneless nuclear body assembly and function (Zhang et al., 2019; Statello et al., 2021a). LncRNAs are newly discovered regulators in many diseases, and there is a growing body of literature suggesting a relationship between lncRNAs and *M. pneumoniae* infection (Gu et al., 2020; Sun et al., 2022).

LncRNAs can be used by the host to modulate immune-related gene expression in order to resist *M. pneumoniae* invasion or decrease the damage caused by *M. pneumoniae* invasion, and *M. pneumoniae* can evade immune clearance by modulating the host lncRNAs (Wen et al., 2020).

This review summarizes the broad categories and common regulatory mechanisms of lncRNAs, the roles of lncRNAs in various diseases, and the defense mechanisms involving host cells' lncRNAs against *M. pneumoniae* infection. It also provides an overview that indicates avenues for investigating lncRNAs-related therapeutic targets in *M. pneumoniae* infection and other diseases.

2 Category of lncRNAs

LncRNAs encompass a wide range of transcripts (Djebali et al., 2012) that exhibit significant diversity in terms of the presence of initiation codons, genomic location, and functional roles, making it difficult to easily characterize them. They can be broadly categorized into three types based on their mechanisms of action: (1) transcriptional regulation, (2) post-transcriptional regulation, and (3) other (Table 1) (Ma et al., 2013). The mechanisms of action of lncRNAs involved in transcriptional regulation can be further classified as (i) transcriptional interference, (ii) chromatin remodeling, and (iii) regulation effect. (The latter involves eRNAs, ncRNA-a1, Evf-2RNA, and Alpha-250/Alpha-280) (Table 1) (Ma et al., 2013). The mechanisms of action of lncRNAs involved in post-transcriptional regulation can be divided into (i) splicing regulation, (ii) translational control, lncRNAs that participate in translational control may function through binding to translation factors citation or ribosome (Ma et al., 2013), and (iii) other (the latter involves siRNA, 1/2-sbsRNA1, 21A, linc-MD1, IPS1, HULC, and BACE1-AS) (Table 1) (Rintala-Maki and Sutherland, 2009). The remaining lncRNAs can be classified into five categories based on other regulatory mechanisms: (i) protein localization (Watanabe and Yamamoto, 1994), (ii) telomere replication (Feng et al., 1995), (iii) RNA interference (Hellwig and Bass, 2008; Smekalova et al., 2016), (iv) regulation beyond transcription; unlike many other lncRNAs, promoter antisense RNAs (PAS RNAs) were initially considered to be merely passive transcription by-products of active promoters (Yang, 2022), and (v) translation regulation (Table 1).

3 Conventional lncRNAs regulatory mechanism

LncRNAs were initially thought to be merely interfering factors in gene transcription, (i.e., acting as accessory products that impede gene transcription involving RNA polymerase II), but they were later found to play essential roles in various biological activities. Notably, lncRNAs participate in transcription but prevent transcription by other chromosomes (Cabili et al., 2015).

The lncRNA LINC02159 (which is highly expressed in non-small cell lung cancer) forms a complex with Aly/REF export factor (ALYREF) through its 5-methylcytosine m^5C modified sites and then binds to YAP1 mRNA, thereby increasing its stability (Chen et al., 2023). The lncRNA ADPGK-AS1, which mainly exists in mitochondria, is upregulated in artificially induced human M2 macrophages, and it binds to mitochondrial ribosomal protein MRPL35 and thereby promotes the tricarboxylic acid cycle and mitochondrial division, resulting in tumor growth (Karger et al., 2023).

The lncRNA MALAT1, also known as non-coding nuclear-enriched abundant transcript 2 (NEAT2), epigenetically regulates gene expression. Highly efficient knockdown of MALAT1 (using zinc finger nuclease-based technology) in extensive organization tumor cells confirmed that MALAT1 promotes *in vitro* and *in vivo* metastasis without affecting tumor cell proliferation (Gutschner et al., 2013). During extensive tumor cell proliferation, MALAT1 is regulated by multiple signaling pathways and has important roles in invasion and metastasis (Chen et al., 2022). MALAT1 regulates the activity of serine/arginine (SR) splicing factors, thereby influencing gene expression via alternative splicing (Tripathi et al., 2010). MALAT1 is also involved in cell cycle regulation, interacting with and promoting the cytoplasmic transport of heterogeneous nuclear ribonucleoprotein C (hnRNP C) in the G2/M phase, thereby controlling gene expression (Yang et al., 2011). Seven novel lncRNAs have been identified as competitive endogenous RNAs. Their abnormal expression leads to the widespread expression of tumorigenic genes (Figure 1A).

As shown in the schematic in Figure 1, there are eight conventional lncRNAs regulatory mechanisms: (1) transcription interference involving transcription from the upstream promoter region of a target protein-coding gene (Figure 1A); (2) inhibiting RNA polymerase II or inducing chromatin remodeling or histone modification, which interferes with target gene transcription (Figure 1B); (3) generating complementary double strands involving mRNAs, which interferes with mRNA cleavage (Figure 1C); (4) generating endogenous short interfering RNAs (endo-siRNAs), which target specific mRNAs (RNA interference) (Figure 1D); (5) binding to a specific protein to modulate its activity (Figure 1E); (6) forming a ribozyme-protein complex, which catalyzes specific reactions (Figure 1F); (7) binding to a specific protein to alter its cellular localization (Figure 1G); and (8) producing small RNA precursors (Figure 1H).

Some upregulated lncRNAs play a tumor-promoting role, while downregulated lncRNAs in gastric cancer play a tumor-inhibitory role (Figure 1B) (Ahmed Shehata et al., 2021). Some lncRNAs can regulate protein activity (Figure 1C). SiRNAs or overexpression plasmids were transfected (with adequate transfection efficiency) into cells and

TABLE 1 Conventional functions of lncRNAs.

Function of lncRNAs		lncRNAs	Reference
Transcriptional regulation	Transcriptional interference	DHFR upstream transcripts, SRG1 RNAs, 7SK snRNA, B2 SINE RNA, chromatin remodeling	Ma et al. (2013)
	Chromatin remodeling	*fbp1*, promoter RNAs, *Xist*, *MEG3*, *GAL10*-ncRNA, *HOTAIR*, *HOTTIP*, *COLDAIR*	
	Regulation effect	eRNAs, ncRNA-a1, Evf-2RNA, Alpha-250/Alpha-280	
Post-transcriptional regulation	Splicing regulation	MIAT, Malat 1, LUST	Rintala-Maki and Sutherland (2009) and Ma et al. (2013)
	Translational control	BC1, BC200, snaR, Gadd7, Zeb2, Zeb2NAT	
	Other	siRNA, 1/2-sbsRNA1, 21A, linc-MD1, IPS1, HULC, BACE1-AS	
Other	Protein localization	MeiRNA, ENOD40 RNA	Watanabe and Yamamoto (1994)
	Telomere replication	TERC	Feng et al. (1995)
	RNA interference	shRNAs or sgRNAs	Hellwig and Bass (2008) and Smekalova et al. (2016)
	Beyond transcription	PAS RNA	Yang (2022)
	Translation regulation	rncs-1	Ma et al. (2013)

verified using fluorescent markers (Figure 1D) (Cao et al., 2019). Some lncRNAs can form a complementary double strand with mRNA (which interferes with mRNA cleavage), and some lncRNAs can produce endo-siRNAs under the action of the Dicer enzyme (Figure 1E). Many lncRNAs are characteristically expressed in polarized tissues and specific cancer types (Xing et al., 2021). They form nucleic acid protein complexes with the proteins acting as structural components (Figure 1F) (Zhou et al., 2016), thereby altering protein localization (Figure 1G). LncRNAs (which are >200 nucleotides in length) have no protein-coding potential (Figure 1H).

4 LncRNAs in *M. pneumoniae* infection

4.1 LncRNAs in intrapulmonary *M. pneumoniae* manifestations

LncRNAs have the function of modifying cell biology (Statello et al., 2021b). LncRNAs can act with mRNAs, DNAs, proteins, and miRNAs to adjust gene expression at the epigenetic, transcriptional, post-transcriptional, translational, and post-translational levels in a variety of ways (Zhang et al., 2019). LncRNAs have many functions, including in *M. pneumoniae* infection, involving both: (1) transcriptional regulation, (2) post-transcriptional regulation, and (3) others (Table 1) (Wright et al., 2013). The interaction of these three regulatory mechanisms plays an important role in the *M. pneumoniae* infection of host cells (Dykes and Emanueli, 2017).

4.1.1 Acute respiratory distress syndrome

LncRNAs are key regulators in respiratory diseases, and they can modulate cell growth arrest. The lncRNA GAS5 plays a significant role in many inflammatory diseases, including acute lung injury, idiopathic pulmonary fibrosis, and *M. pneumoniae* infection (Yang et al., 2021). GAS5 overexpression enhances cellular energy production and downregulates the pro-inflammatory cytokines IL-1β and IL-6 in human acute monocytic leukemia THP-1 cells (Figure 2A). The overexpression of miR-222-3p, which targets and reverses *M. pneumoniae*-induced THP-1 cell energy production, reduces *M. pneumoniae*-induced THP-1 cell viability, and accelerates the inflammatory response. GAS5 silencing reduces *M. pneumoniae*-induced chondrocyte activity and exacerbates *M. pneumoniae*-induced host cell inflammatory injuries. These findings offer new targets for treating *M. pneumoniae* infection (Yang et al., 2021). When exposed to host cells, *M. pneumoniae* upregulates the community-acquired respiratory distress syndrome acute respiratory distress syndrome (ARDS) toxin protein (encoded by the MPN372 gene), which is involved in host-cell interactions (Medina et al., 2012).

GAS5 was downregulated in lung epithelial cells treated with lipopolysaccharide (which can cause ARDS), suggesting that GAS5 is involved in the development of ARDS. The GAS5/miR-200c-3p/ACE2 signaling axis is involved in the apoptosis of ARDS lung epithelial cells. These findings offer new therapeutic targets for ARDS and enrich our understanding of the GAS5-mediated regulation of lung injury, which is of great significance for understanding the pathogenesis of ARDS (Li et al., 2018).

4.1.2 Acute pneumonia

M. pneumoniae lipopolysaccharide can enter human embryonic lung WI-38 fibroblasts, induce inflammatory damage, and destroy the cells by triggering lncRNA HAGLROS upregulation. *M. pneumoniae* can induce inflammatory damage in WI-38 cells by modulating the miR-100/NF-κB axis. HAGLROS upregulation inhibits miRNA-100 (which therefore no longer targets and suppresses NF-κB3), thereby increasing NF-κB activity. HAGLROS knockout prevents NF-κB activation and thereby enhances WI-38 cell viability, inhibits apoptosis, and mitigates *M. pneumoniae*-induced cell damage (Figure 2A). Reducing the expression of miR-100 activates NF-κB3 and thereby causes WI-38 cell damage and apoptosis; this effect of reducing the expression of miR-100 can be prevented by NF-κB3

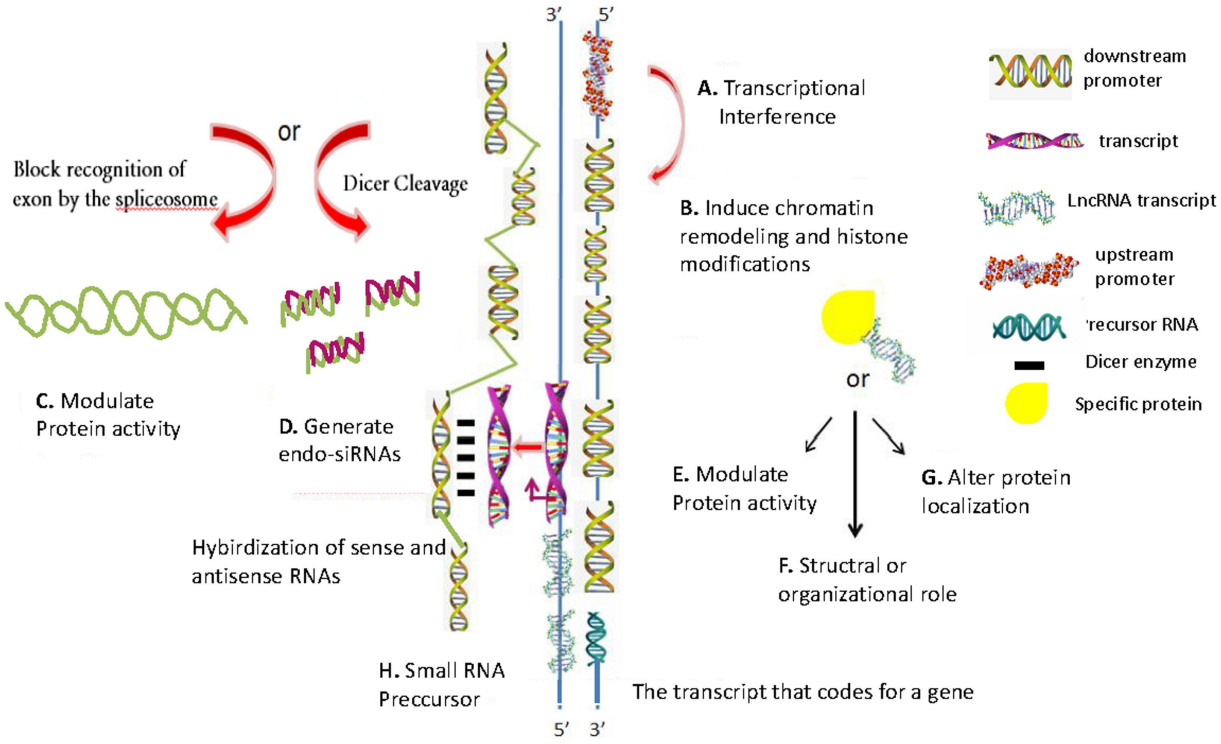


FIGURE 1
Schematic of conventional lncRNA regulatory mechanisms. **(A)** Transcription interference involves transcription from the upstream promoter region of a target protein-coding gene. **(B)** Inhibiting RNA polymerase II or inducing chromatin remodeling or histone modification, which interferes with target gene transcription. **(C)** Modulate protein activity. **(D)** Generating endogenous short interfering RNAs (endo-siRNAs), which target specific mRNAs (RNA interference). **(E)** Binding to a specific protein to modulate its activity. **(F)** Structural or organizational role, which catalyzes specific reactions. **(G)** Binding to a specific protein to alter its cellular localization. **(H)** Producing small RNA precursors. The diagrams are original and free of copyright restrictions.

knockout (Figure 2A) (Liu et al., 2018). During *M. pneumoniae*-induced pneumonia, type I alveolar epithelial cells defend against *M. pneumoniae* infection by activating the innate immune response (Yamamoto et al., 2012), including the PI3K/AKT/NFκB pathway (Yang et al., 2021). The HAGLROS/miR-100/NF-κB axis may be a new target for the treatment of *M. pneumoniae* infection (Liu et al., 2018).

Another lncRNA that can regulate NF-κB activity is NKILA, which can exert an anti-inflammatory effect in airway epithelial cells. NKILA can mask the phosphorylation motif of I-κB (an inhibitor of NF-κB) and thereby prevent I-κB degradation and NF-κB translocation (Figure 2B) (Liu et al., 2015; Ke et al., 2018; Wang et al., 2018; Yu et al., 2018), inhibiting *M. pneumoniae*-induced inflammatory response genes (Peschke et al., 2014). NKILA is decreased and cytokines (IL-8 and TNF-α) are increased in bronchoalveolar lavage fluid from children infected with *M. pneumoniae* and NKILA knockdown in airway epithelial cells promotes *M. pneumoniae*-induced cytokine secretion NKILA exerts its anti-inflammatory effect by weakening the negative feedback loop of NF-κB signaling that regulates cytokine secretion (Figure 2B) (Zhang et al., 2021).

Moreover, downregulation of the lncRNA CRNDE and upregulation of miR-141 (which inhibits NF-κB and is targeted by CRNDE) inhibit the *M. pneumoniae* endotoxin-induced apoptosis and inflammatory response of human embryonic lung MRC-5 fibroblasts, thereby promoting cell survival (Figure 2A) (Zúñiga et al., 2012).

4.1.3 Asthma

During *M. pneumoniae* infection, the host can regulate certain lncRNAs to inhibit inflammation and apoptosis. Activation of the transcription factor NF-κB, which regulates various inflammatory response genes, plays a significant role in *M. pneumoniae*-induced airway inflammation. Under normal conditions, NF-κB is bound to its inhibitor, I-κB, and remains in the cytoplasm. When stimulated by *M. pneumoniae* lipoproteins, which are recognized by toll-like receptors (TLRs), I-κB is phosphorylated and degraded, releasing NF-κB; activated phosphorylated NF-κB p65 then enters the nucleus and upregulates inflammation-related genes (Zhu et al., 2010). The pro-inflammatory mechanism of action of the lncRNA MALAT1 partially relies on it increasing NF-κB activation (by directly binding to it or indirect regulation, i.e., acting as a competitive endogenous RNA, competing with miRNAs that target NF-κB mRNA, and thus enhancing NF-κB activity). It thereby regulates the *M. pneumoniae*-induced inflammatory response (Figure 2B) (Dai et al., 2018; Lei et al., 2018). NF-κB upregulates TNF-α, which can damage capillary endothelial cells, thereby promoting microthrombosis and leading to ischemic necrosis, so TNF-α is associated with pneumonia severity (Figure 2B) (Salvatore et al., 2007). MALAT1 knockdown inhibits *M. pneumoniae*-induced NF-κB p65 phosphorylation in mouse airway epithelial cells and mouse lung tissue. Thus, the regulatory role of MALAT1 in *M. pneumoniae* infection-induced inflammation is closely related to NF-κB activation (Zhang et al., 2021).

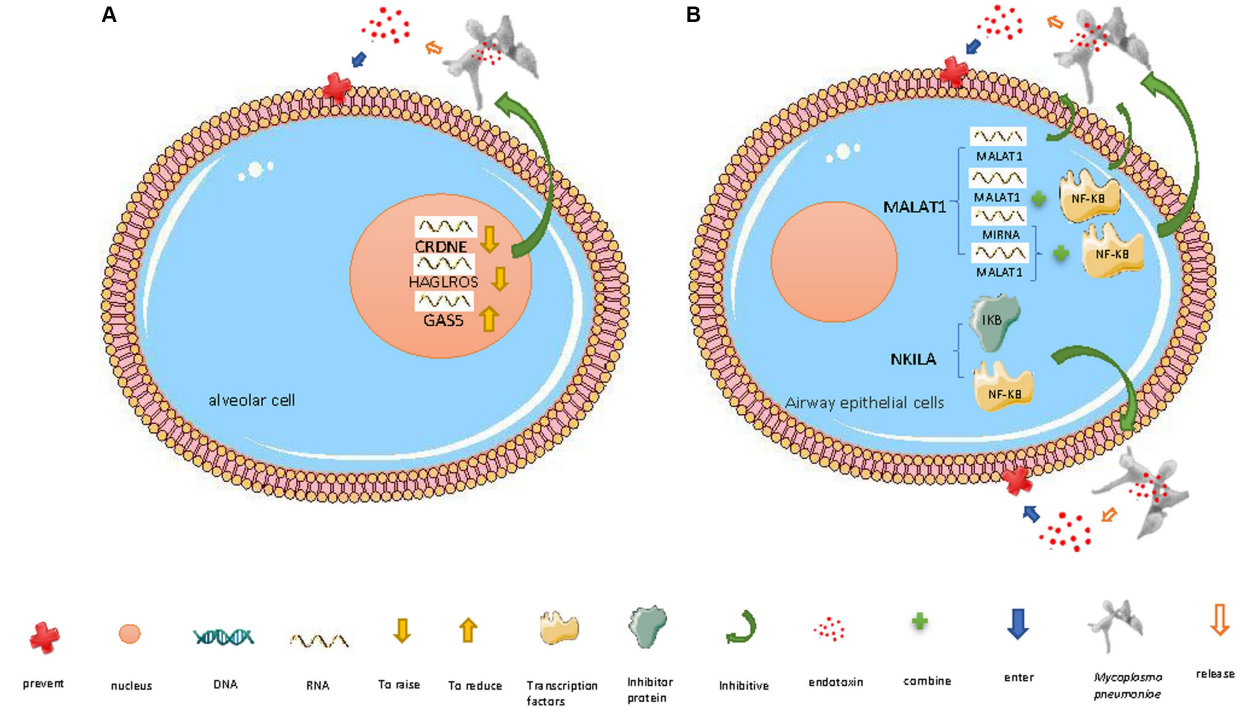


FIGURE 2
Mechanisms of lncRNAs defending against *M. pneumoniae* invasion. **(A)** In the nucleus: (i) lncRNA CRNDE downregulation, (ii) lncRNA HAGLROS downregulation, and (iii) lncRNA GAS5 upregulation inhibit *M. pneumoniae*-induced inflammatory factor release, thereby preventing damage to host cells. **(B)** In the cytoplasm: (i) downregulated lncRNA MALAT1 inhibits the inflammatory response triggered by *M. pneumoniae* endotoxin release by altering NF-κB activation; MALAT1 regulates NF-κB via three pathways: (a) competing with miRNAs that target NF-κB mRNA, (b) binding to NF-κB, and (c) directly resistance to endotoxins (ii) lncRNA NKILA inhibits the inflammatory response triggered by *M. pneumoniae* endotoxin release by preventing the dissociation of NF-κB from its inhibitor I-κB and thereby preventing NF-κB translocation to the nucleus. The diagrams are original and free of copyright restrictions.

The mechanism by which lower respiratory tract *M. pneumoniae* infections trigger or worsen asthma in children is not completely clear (Kumar et al., 2019). Following *M. pneumoniae* infection in children, a small percentage of individuals present with recurrent wheezing episodes, and the prevalence of *M. pneumoniae* infection in children with acute asthma has been reported to be 46% (Kassisse et al., 2018). *M. pneumoniae* can induce mucin overproduction by inhibiting the transcription suppressor FOXA2 Lung function is improved by restoring FOXA2's transcription suppressor function and downregulating goblet cell hyperplasia and metaplasia (GCHM)-promoting pathways in *M. pneumoniae*-infected airways in asthma patients with abnormal mucin secretion and accumulation in airway lumens, which are clinical markers of asthma (Hao et al., 2014). In addition, asthma is associated with upregulated MALAT1 and downregulated miRNA-216a (which is inhibited by MALAT1, acting as a molecular sponge), while the opposite (MALAT1 downregulation and/or miRNA-216a upregulation) significantly increases apoptosis while significantly decreasing cell proliferation, migration, and invasion (Huang J. et al., 2021).

4.2 Immune-mediated mechanisms of *M. pneumoniae* extrapulmonary manifestations

M. pneumoniae can cause various extrapulmonary manifestations, including those that affect the cardiovascular system, skin, and liver.

The cardiovascular manifestations of *M. pneumoniae* infection (Bakshi et al., 2006) include aortic thrombosis (Flateau et al., 2013) and pulmonary thrombosis. *M. pneumoniae* can directly spread via the blood to distant organs and induce local production of cytokines and chemokines (TNF-α and IL-8), eventually leading to local vasculitis or thrombosis. *M. pneumoniae* can also indirectly lead to systemic hypercoagulability by activating chemical mediators, complement, and fibrin D-dimer, which increase the risk of thrombotic vascular occlusion (Hu et al., 2022).

The dermatological manifestations of *M. pneumoniae* infection include erythema nodosum [an immune complex-mediated disease that primarily affects young women (Kakourou et al., 2001)] and cutaneous leucocytic vasculitis [characterized by perivascular neutrophilia reported to be caused by circulating immune complexes (Kakourou et al., 2001; Perez and Montes, 2002)]. Although *M. pneumoniae* cannot infect the squamous cell epithelium, it may produce inflammatory bullous lesions due to the transfer of cytokines from the respiratory tract to the skin via the blood (Narita, 2016).

The hepatic manifestations of *M. pneumoniae* infection can arise as a result of modulation of T-cell-mediated immune responses by T cell immunoglobulin and mucin domain-containing proteins (TIMs) expressed on T cells, which can regulate T cell cytokine differentiation (Wang et al., 2008). Liver damage can also be caused by inflammatory cell activation induced by signaling involving TLR2 and TLR4, which are expressed on cell surfaces and can detect and initiate responses to extracellular pathogens (Kawasaki and Kawai, 2014; Shimizu et al., 2014). *M. pneumoniae* causes acute and severe hepatitis in children,

which is likely to be immune-mediated and involve both innate and adaptive immune responses (Poddighe, 2020).

In summary, the detailed mechanisms underlying the three abovementioned types of *M. pneumoniae* extrapulmonary infection are unclear, but it is clear that they generally involve inflammatory immune responses (Poddighe et al., 2022).

4.3 LncRNAs and immune-mediated mechanisms of *M. pneumoniae* infection

4.3.1 Intrapulmonary *M. pneumoniae* manifestation

M. pneumoniae adhesion molecules and metabolites can cause immune damage to respiratory epithelial cells. *M. pneumoniae* infection decreases $CD4^+$T cell function, which is the primary cause of immune dysfunction in patients with *M. pneumoniae* infection, impairing antigen presentation, B-cell maturation, and antibody production. *M. pneumoniae* also disrupts other humoral and cellular immune responses (Hu et al., 2022). During host cells' non-specific immune defense against *M. pneumoniae*, lncRNAs regulate reactive oxygen species production by NADPH oxidase to fight *M. pneumoniae* (Lee et al., 2020). LncRNAs can also be exploited by *M. pneumoniae* to evade the immune system (Hu et al., 2022).

4.3.2 Extrapulmonary in children infected (neurological) *M. pneumoniae* manifestations

M. pneumoniae infection-induced neurological diseases are likely to be a result of immune responses to the infection, based on indirect immunofluorescence and PCR analysis of cerebrospinal fluid samples from patients with these neurological diseases (Poddighe, 2018).

The lncRNA NKILA was downregulated while IL-8 and TNF-α were upregulated in children infected with *M. pneumoniae*. NKILA knockdown *in vitro* promotes the inflammatory effect of *M. pneumoniae* on A549 and BEAS-2B respiratory epithelial cells (Zhang et al., 2021). IL-8 and TNF-α are two well-known pro-inflammatory cytokines that play crucial roles in airway inflammation and chemotaxis caused by *M. pneumoniae* (Martin et al., 1997).

4.4 LncRNAs/circRNAs in drug-resistant *M. pneumoniae* infection

Both macrolide-resistant and refractory *M. pneumoniae* infections complicate the clinical management of *M. pneumoniae* pneumonia (Tsai et al., 2021). Macrolide-resistant *M. pneumoniae* harbors a point mutation in 23S rRNA domain V (with substitutions mainly detected at positions 2063 and 2064) (Yang et al., 2017). Circular RNAs (circRNAs), which are like lncRNAs but form a closed loop (Ashekyan et al., 2022), play important roles in gene expression regulation by sequestering miRNA targets (acting as molecular sponges) (Meng et al., 2017). The miRNA targets of circRNAs (detected by high-throughput sequencing) could be utilized as biomarkers for the diagnosis of early-stage refractory *M. pneumoniae* pneumonia (Huang F. et al., 2021).

5 LncRNAs in other diseases

LncRNAs affect cardiovascular development, including the embryonic development of the heart and vascular system (Kohlmaier et al., 2023). The lncRNA CARMEN can regulate the fate, differentiation, and homeostasis of human cardiac progenitor cells (Ounzain et al., 2015). Additionally, lncRNAs serve as key regulators in cardiovascular diseases such as arterial hypertension, coronary heart disease, and acute myocardial infarction (Correia et al., 2021). For example, overexpression of lnc-Ang362 indirectly activates the (NF-κB) signaling pathway, which promotes vascular smooth muscle cell proliferation and migration, thereby aggravating arterial hypertension (Wang et al., 2020). Additionally, upregulation of the lncRNA cardiac hypertrophy-related factor (CHRF) in cardiomyocytes can upregulate myeloid differentiation primary response 88 (MYD88), inducing cardiomyocyte hypertrophy and apoptosis, leading to heart failure (Wang et al., 2014).

LncRNAs also regulate the development and differentiation of neurons and the nervous system, and they play various pathological roles, leading to various neurodegenerative diseases (Nadhan et al., 2022). In Alzheimer's disease, the highly upregulated antisense lncRNA BACE1-AS stably binds to BACE1, enhancing the production of β-amyloid plaques (Zeng et al., 2019). In schizophrenia, the reduced expression of the lncRNA MIAT is associated with behavioral changes (Ip et al., 2016). In autism spectrum disorder, the lncRNA SYNGAP-AS1 can downregulate SYNGAP1, causing cortical functional impairment (Velmeshev et al., 2013). In ischemic stroke, the reduced expression of the lncRNA MEG3 activates the Notch signaling pathway and thereby promotes angiogenesis (Yan et al., 2016).

In cancer, some lncRNAs have been identified as oncogenes, while others have been identified as tumor suppressors (Nadhan et al., 2022). The lncRNA HOTTIP acts as an oncogene in acute myeloid leukemia, where it is abnormally elevated and functions as an epigenetic regulator, modulating hematopoietic gene-associated chromatin signatures and transcription (Luo et al., 2019). The p53-dependent lncRNA PVT1 inhibits lung cancer by downregulating c-Myc (Olivero et al., 2020). The abovementioned lncRNA CHRF plays a crucial role in the progression of various tumors, such as prostate cancer, by miRNA binding (Gai et al., 2019). The lncRNA LUCAT1 is associated with various cancers, including cervical cancer, where it exerts oncogenic functions by sequestering miR-181a (Xing et al., 2021). Finally, the highly expressed lncRNA NEAT1 sequesters miR-155 and upregulates TIM3, which promotes CD8 T cell apoptosis and thereby facilitates hepatocellular carcinoma immune evasion and development (Yan et al., 2019).

In endocrine diseases such as diabetes and related conditions (including diabetic nephropathy, diabetic retinopathy, and diabetic neuropathic pain), dysregulated lncRNAs have significant effects (Alipoor et al., 2021). For example, the downregulation of the lncRNA H19 disrupts mitochondrial fatty acid β-oxidation and leads to fatty acid accumulation and insulin resistance (Gui et al., 2020). The lncRNA PVT1 is overexpressed in diabetic nephropathy. PVT1 silences FOXA1 by directly binding to and stabilizing the histone methyltransferase EZH2 to induce trimethylation-based silencing (Liu D. W. et al., 2019). The reduced expression of FOXA1 induces podocyte apoptosis, contributing to the progression of diabetic

nephropathy. The lncRNA MALAT1 sequesters miR-125b and thereby upregulates target genes, promotes neovascularization, and impairs vision (Liu P. et al., 2019). The lncRNA NONRATT021972 is upregulated in diabetic neuropathic pain, which it exacerbates by upregulating TNF-α and purinergic receptors (P2X) 3 and 7 It increases the expression of TNFα as well as purinergic receptors (P2X) 3 and 7 (Peng et al., 2017).

5.1 Small molecule response induced by lncRNAs in *M. pneumoniae* infection

Neutrophils are one of the cells that respond to inflammation sites and play a vital role in killing pathogens (Schenten et al., 2018). The inflammatory response caused by neutrophil activation can be triggered by endogenous ligands called damage-associated molecular patterns (DAMPs) or actively aerated alarmins (Chan et al., 2012). Recently, S100A8/9 proteins have been identified as DAMPs released by neutrophils and monocytes [which has been proposed to be an active process dependent on the microtubule network (Schiopu and Cotoi, 2013) or a process involving NETosis (Ehrchen et al., 2009; Bianchi et al., 2011)]. The elevation of S100A8/9 increases neutrophils in the blood, which can promote the occurrence of atherosclerotic disease due to neutrophil accumulations in artery walls (Schiopu and Cotoi, 2013).

6 LncRNAs as targets for treating *M. pneumoniae* infection

LncRNAs in the nucleus (e.g., CRNDE, HAGLROS, and GAS5) and cytoplasm (e.g., MALAT1 and NKILA) work together to resist *M. pneumoniae* invasion. Downregulation of CRNDE can upregulate miR-141 and thereby inhibit lipopolysaccharide-induced MRC-5 fibroblast apoptosis and the associated inflammatory response (Meng et al., 2019). HAGLROS downregulation ameliorates lipopolysaccharide-induced PI3K/AKT/NF-κB pathway activation and inflammatory damage in WI-38 cells by causing a lack of HAGLROS to compete with miRNA-100, leading to NF-κB3 downregulation (Torrealba et al., 2020). The HAGLROS/miR-100/NF-κB axis may provide a new target for the treatment of acute-phase *M. pneumoniae* pneumonia (Fang and Shi, 2022). Highly expressed lncRNA GAS5 reduces the inflammatory response and the viability of LAMP-1-induced human acute monocytic leukemia THP-1 cells by targeting the miR-222-3p/TIMP3 axis (Yang et al., 2021). Downregulated MALAT1 plays a key regulatory role in reducing *M. pneumoniae*-induced inflammation (Zhao et al., 2016) by downregulating NF-κB signaling (Shimizu et al., 2008). NKILA inhibits the *M. pneumoniae*-induced inflammatory response of airway epithelial cells by modulating NF-κB (Zhu et al., 2019).

The findings that lncRNAs/circRNAs carried by exosomes in breast cancer (BC) regulate breast cancer-related target genes (Ashekyan et al., 2022) prompt the question of whether the lncRNAs/circRNAs/target genes are related to *M. pneumoniae* infection and whether they may represent novel targets for the treatment of *M. pneumoniae* (Tang et al., 2020). LncRNAs have been shown to have broad clinical applications, including cancer diagnosis and prognosis biomarkers (Ashekyan et al., 2022).

7 Perspectives

Although recent lncRNA sequencing analyses have identified potentially key lncRNAs associated with *M. pneumoniae* pneumonia (Huang et al., 2016), their biological roles and function mechanisms remain largely unknown (Chen et al., 2018). It is important to determine the pivotal molecular mechanisms underlying *M. pneumoniae* pneumonia in order to develop effective treatment strategies (Chen et al., 2018). Studying lncRNAs may provide an academic foundation for more comprehensive understanding of the molecular mechanisms underlying *M. pneumoniae* pneumonia and for identifying effective treatment targets, thereby identifying unconventional strategies for the treatment of acute-phase *M. pneumoniae* pneumonia.

LncRNA regulates cardiovascular development (Correia et al., 2021) and the development and differentiation of neurons and the nervous system (Nadhan et al., 2022). In cancer, some lncRNAs have been identified as oncogenes, while others have been identified as tumor suppressors (Nadhan et al., 2022). In endocrine diseases such as diabetes and related conditions, dysregulated lncRNAs have significant effects (Alipoor et al., 2021). LncRNAs play a variety of roles in these diseases, which may provide insights into the currently unknown roles of lncRNAs in various *M. pneumoniae* infection states.

8 Conclusion

LncRNAs encompass a wide range of transcripts with significant diversity in terms of the presence of initiation codons, genomic location, and functional roles. They are newly discovered regulators in many diseases, and there is a growing body of literature suggesting a relationship between lncRNAs and *M. pneumoniae* infection. In this review, we broadly classified lncRNAs' mechanisms of action as transcriptional regulation, post-transcriptional regulation, and others, and detailed the conventional mechanisms of action of lncRNAs. We also discussed lncRNAs' roles in the pathogenesis of four major disease types (cardiovascular diseases, neurological disorders, cancers, and the endocrine disease diabetes). Furthermore, we provided insights into lncRNAs' key protective roles against intrapulmonary, extrapulmonary, and drug-resistant *M. pneumoniae* infections. This review serves as a succinct overview and indicates avenues for investigating lncRNAs' roles as novel therapeutic targets.

Author contributions

ZY: Writing – original draft, Project administration. JZ: Writing – original draft. NS: Writing – original draft. ZZ: Writing – original draft. JC: Writing – original draft. PLiu: Writing – review & editing, Writing – original draft, Supervision, Resources. PLin: Supervision, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This research was supported by the Natural Science Foundation of Hunan Province, China (2023JJ30503); Research Foundation of Education Bureau of Hunan Province, China (22A0297); Research Foundation of University of South

China (190XQD015); and Hunan Provincial College Students' Innovation and Entrepreneurship Training Program (2022X10555197).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Ahmed Shehata, W., Maraee, A., Abd El Monem Ellaithy, M., Tayel, N., Abo-Ghazala, A., and Mohammed El-Hefnawy, S. (2021). Circulating long noncoding RNA growth arrest-specific transcript 5 as a diagnostic marker and indicator of degree of severity in plaque psoriasis. *Int. J. Dermatol.* 60, 973–979. doi: 10.1111/ijd.15494

Alipoor, B., Nikouei, S., Rezaeinejad, F., Malakooti-Dehkordi, S. N., Sabati, Z., and Ghasemi, H. (2021). Long non-coding RNAs in metabolic disorders: pathogenetic relevance and potential biomarkers and therapeutic targets. *J. Endocrinol. Investig.* 44, 2015–2041. doi: 10.1007/s40618-021-01559-8

Ashekyan, O., Abdallah, S., Shoukari, A. A., Chamandi, G., Choubassy, H., Itani, A. R. S., et al. (2022). Spotlight on exosomal non-coding RNAs in breast cancer: an in silico analysis to identify potential lncRNA/circRNA-miRNA-target axis. *Int. J. Mol. Sci.* 23:8351. doi: 10.3390/ijms23158351

Bakshi, M., Khemani, C., Vishwanathan, V., Anand, R. K., and Khubchandani, R. P. (2006). Mycoplasma pneumonia with antiphospholipid antibodies and a cardiac thrombus. *Lupus* 15, 105–106. doi: 10.1191/0961203306lu2258cr

Bianchi, M., Niemiec, M. J., Siler, U., Urban, C. F., and Reichenbach, J. (2011). Restoration of anti-*Aspergillus* defense by neutrophil extracellular traps in human chronic granulomatous disease after gene therapy is calprotectin-dependent. *J. Allergy Clin. Immunol.* 127, 1243–52.e7. doi: 10.1016/j.jaci.2011.01.021

Cabili, M. N., Dunagin, M. C., McClanahan, P. D., Biaesch, A., Padovan-Merhar, O., Regev, A., et al. (2015). Localization and abundance analysis of human lncRNAs at single-cell and single-molecule resolution. *Genome Biol.* 16:20. doi: 10.1186/s13059-015-0586-4

Cao, H. L., Liu, Z. J., Huang, P. L., Yue, Y. L., and Xi, J. N. (2019). lncRNA-RMRP promotes proliferation, migration and invasion of bladder cancer via miR-206. *Eur. Rev. Med. Pharmacol. Sci.* 23, 1012–1021. doi: 10.26355/eurrev_201902_16988

Chan, J. K., Roth, J., Oppenheim, J. J., Tracey, K. J., Vogl, T., Feldmann, M., et al. (2012). Alarmins: awaiting a clinical response. *J. Clin. Invest.* 122, 2711–2719. doi: 10.1172/JCI62423

Chen, J. F., Wu, P., Xia, R., Yang, J., Huo, X. Y., Gu, D. Y., et al. (2018). STAT3-induced lncRNA HAGLROS overexpression contributes to the malignant progression of gastric cancer cells via mTOR signal-mediated inhibition of autophagy. *Mol. Cancer* 17:6. doi: 10.1186/s12943-017-0756-y

Chen, L., Xie, Y., Yu, M., and Gou, Q. (2022). Long noncoding RNAs in lung cancer: from disease markers to treatment roles. *Cancer Manag. Res.* 14, 1771–1782. doi: 10.2147/CMAR.S365762

Chen, P., Yu, J., Luo, Q., Li, J., and Wang, W. (2023). Construction of disulfidptosis-related lncRNA signature for predicting the prognosis and immune escape in colon adenocarcinoma. *BMC Gastroenterol.* 23:382. doi: 10.1186/s12876-023-03020-x

Combaz-Söhnchen, N., and Kuhn, A. (2017). A systematic review of mycoplasma and ureaplasma in urogynaecology. *Geburtshilfe Frauenheilkd.* 77, 1299–1303. doi: 10.1055/s-0043-119687

Correia, C. C. M., Rodrigues, L. F., de Avila Pelozin, B. R., Oliveira, E. M., and Fernandes, T. (2021). Long non-coding RNAs in cardiovascular diseases: potential function as biomarkers and therapeutic targets of exercise training. *Noncoding RNA* 7:65. doi: 10.3390/ncrna7040065

Dai, L., Zhang, G., Cheng, Z., Wang, X., Jia, L., Jing, X., et al. (2018). Knockdown of lncRNA MALAT1 contributes to the suppression of inflammatory responses by up-regulating miR-146a in LPS-induced acute lung injury. *Connect. Tissue Res.* 59, 581–592. doi: 10.1080/03008207.2018.1439480

de Groot, R. C. A., Meyer Sauteur, P. M., Unger, W. W. J., and van Rossum, A. M. C. (2017). Things that could be *Mycoplasma pneumoniae*. *J. Infect.* 74, S95–s100. doi: 10.1016/S0163-4453(17)30198-6

Djebali, S., Davis, C. A., Merkel, A., Dobin, A., Lassmann, T., Mortazavi, A., et al. (2012). Landscape of transcription in human cells. *Nature* 489, 101–108. doi: 10.1038/nature11233

Dykes, I. M., and Emanueli, C. (2017). Transcriptional and post-transcriptional gene regulation by long non-coding RNA. *Genomics Proteomics Bioinformatics* 15, 177–186. doi: 10.1016/j.gpb.2016.12.005

Ehrchen, J. M., Sunderkötter, C., Foell, D., Vogl, T., and Roth, J. (2009). The endogenous Toll-like receptor 4 agonist S100A8/S100A9 (calprotectin) as innate amplifier of infection, autoimmunity, and cancer. *J. Leukoc. Biol.* 86, 557–566. doi: 10.1189/jlb.1008647

Fang, X. L., and Shi, S. G. (2022). lncRNA FGD5-AS1 acts AS a ceRNA to regulate lipopolysaccharide-induced injury via the miR-223-3p-3p/GAS5 axis in cardiomyocytes. *Hum. Exp. Toxicol.* 41:9603271221138969. doi: 10.1177/09603271221138969

Feng, J., Funk, W. D., Wang, S. S., Weinrich, S. L., Avilion, A. A., Chiu, C. P., et al. (1995). The RNA component of human telomerase. *Science* 269, 1236–1241. doi: 10.1126/science.7544491

Flateau, C., Asfalou, I., Deman, A. L., Ficko, C., Andriamanantena, D., Fontan, E., et al. (2013). Aortic thrombus and multiple embolisms during a *Mycoplasma pneumoniae* infection. *Infection* 41, 867–873. doi: 10.1007/s15010-013-0475-2

Gai, H. Y., Wu, C., Zhang, Y., and Wang, D. (2019). Long non-coding RNA CHRF modulates the progression of cerebral ischemia/reperfusion injury via miR-126/SOX6 signaling pathway. *Biochem. Biophys. Res. Commun.* 514, 550–557. doi: 10.1016/j.bbrc.2019.04.161

García-Andrade, F., Vigueras-Villaseñor, R. M., Chávez-Saldaña, M. D., Rojas-Castañeda, J. C., Bahena-Ocampo, I. U., Aréchaga-Ocampo, E., et al. (2022). The role of microRNAs in the gonocyte theory as target of malignancy: looking for potential diagnostic biomarkers. *Int. J. Mol. Sci.* 23:10526. doi: 10.3390/ijms231810526

Gómez Rufo, D., García Sánchez, E., García Sánchez, J. E., and García Moro, M. (2021). Clinical implications of the genus *Mycoplasma*. *Rev. Esp. Quimioter.* 34, 169–184. doi: 10.37201/req/014.2021

Gu, H., Zhu, Y., Zhou, Y., Huang, T., Zhang, S., Zhao, D., et al. (2020). LncRNA MALAT1 affects *Mycoplasma pneumoniae* pneumonia via NF-κB regulation. *Front. Cell Dev. Biol.* 8:563693. doi: 10.3389/fcell.2020.563693

Gui, W., Zhu, W. F., Zhu, Y., Tang, S., Zheng, F., Yin, X., et al. (2020). LncRNAH19 improves insulin resistance in skeletal muscle by regulating heterogeneous nuclear ribonucleoprotein A1. *Cell Commun. Signal* 18:173. doi: 10.1186/s12964-020-00654-2

Gutschner, T., Hämmerle, M., Eißmann, M., Hsu, J., Kim, Y., Hung, G., et al. (2013). The noncoding RNA MALAT1 is a critical regulator of the metastasis phenotype of lung cancer cells. *Cancer Res.* 73, 1180–1189. doi: 10.1158/0008-5472.CAN-12-2850

Hao, Y., Kuang, Z., Jing, J., Miao, J., Mei, L. Y., Lee, R. J., et al. (2014). *Mycoplasma pneumoniae* modulates STAT3-STAT6/EGFR-FOXA2 signaling to induce overexpression of airway mucins. *Infect. Immun.* 82, 5246–5255. doi: 10.1128/IAI.01989-14

Hellwig, S., and Bass, B. L. (2008). A starvation-induced noncoding RNA modulates expression of dicer-regulated genes. *Proc. Natl. Acad. Sci. U.S.A.* 105, 12897–12902. doi: 10.1073/pnas.0805118105

Hu, J., Ye, Y., Chen, X., Xiong, L., Xie, W., and Liu, P. (2022). Insight into the pathogenic mechanism of *Mycoplasma pneumoniae*. *Curr. Microbiol.* 80:14. doi: 10.1007/s00284-022-03103-0

Huang, F., Fan, H., Yang, D., Zhang, J., Shi, T., Zhang, D., et al. (2021). Ribosomal RNA-depleted RNA sequencing reveals the pathogenesis of refractory *Mycoplasma pneumoniae* pneumonia in children. *Mol. Med. Rep.* 24:761. doi: 10.3892/mmr.2021.12401

Huang, S., Feng, C., Chen, L., Huang, Z., Zhou, X., Li, B., et al. (2016). Identification of potential key long non-coding RNAs and target genes associated with pneumonia using long non-coding RNA sequencing (lncRNA-Seq): a preliminary study. *Med. Sci. Monit.* 22, 3394–3408. doi: 10.12659/MSM.900783

Huang, J., Wang, F. H., Wang, L., Li, Y., Lu, J., and Chen, J. Y. (2021). LncRNA MALAT1 promotes proliferation and migration of airway smooth muscle cells in asthma by downregulating microRNA-216a. *Saudi J. Biol. Sci.* 28, 4124–4131. doi: 10.1016/j.sjbs.2021.03.076

Ip, J. Y., Sone, M., Nashiki, C., Pan, Q., Kitaichi, K., Yanaka, K., et al. (2016). Gomafu lncRNA knockout mice exhibit mild hyperactivity with enhanced responsiveness to the psychostimulant methamphetamine. *Sci. Rep.* 6:27204. doi: 10.1038/srep27204

Jiang, Z., Li, S., Zhu, C., Zhou, R., and Leung, P. H. M. (2021). *Mycoplasma pneumoniae* infections: pathogenesis and vaccine development. *Pathogens* 10:119. doi: 10.3390/pathogens10020119

Kakourou, T., Drosatoua, P., Psychou, F., Aroni, K., and Nicolaidou, P. (2001). Erythema nodosum in children: a prospective study. *J. Am. Acad. Dermatol.* 44, 17–21. doi: 10.1067/mjd.2001.110877

Karger, A., Mansouri, S., Leisegang, M. S., Weigert, A., Günther, S., Kuenne, C., et al. (2023). ADPGK-AS1 long noncoding RNA switches macrophage metabolic and phenotypic state to promote lung cancer growth. *EMBO J.* 42:e111620. doi: 10.15252/embj.2022111620

Kassisse, E., García, H., Prada, L., Salazar, I., and Kassisse, J. (2018). Prevalence of *Mycoplasma pneumoniae* infection in pediatric patients with acute asthma exacerbation. *Arch. Argent. Pediatr.* 116, 179–185. doi: 10.5546/aap.2018.eng.179

Kawasaki, T., and Kawai, T. (2014). Toll-like receptor signaling pathways. *Front. Immunol.* 5:461. doi: 10.3389/fimmu.2014.00461

Ke, S., Li, R. C., Meng, F. K., and Fang, M. H. (2018). NKILA inhibits NF-κB signaling and suppresses tumor metastasis. *Aging* 10, 56–71. doi: 10.18632/aging.101359

Kohlmaier, A., Holdt, L. M., and Teupser, D. (2023). Long noncoding RNAs in cardiovascular disease. *Curr. Opin. Cardiol.* 38, 179–192. doi: 10.1097/HCO.0000000000001041

Kumar, S., Roy, R. D., Sethi, G. R., and Saigal, S. R. (2019). *Mycoplasma pneumoniae* infection and asthma in children. *Trop. Dr.* 49, 117–119. doi: 10.1177/0049475518816591

Lee, D., Lal, N. K., Lin, Z. J. D., Ma, S., Liu, J., Castro, B., et al. (2020). Regulation of reactive oxygen species during plant immunity through phosphorylation and ubiquitination of RBOHD. *Nat. Commun.* 11:1838. doi: 10.1038/s41467-020-15601-5

Lei, L., Chen, J., Huang, J., Lu, J., Pei, S., Ding, S., et al. (2018). Functions and regulatory mechanisms of metastasis-associated lung adenocarcinoma transcript 1. *J. Cell. Physiol.* 234, 134–151. doi: 10.1002/jcp.26759

Li, H. B., Zi, P. P., Shi, H. J., Gao, M., and Sun, R. Q. (2018). Role of signaling pathway of long non-coding RNA growth arrest-specific transcript 5/microRNA-200c-3p/angiotensin converting enzyme 2 in the apoptosis of human lung epithelial cell A549 in acute respiratory distress syndrome. *Zhonghua Yi Xue Za Zhi* 98, 3354–3359. doi: 10.3760/cma.j.issn.0376-2491.2018.41.013

Liu, M., Han, T., Shi, S., and Chen, E. (2018). Long noncoding RNA HAGLROS regulates cell apoptosis and autophagy in lipopolysaccharides-induced WI-38 cells via modulating miR-100/NF-κB axis. *Biochem. Biophys. Res. Commun.* 500, 589–596. doi: 10.1016/j.bbrc.2018.04.109

Liu, P., Jia, S.-B., Shi, J.-M., Li, W.-J., Tang, L.-S., Zhu, X.-H., et al. (2019). LncRNA-MALAT1 promotes neovascularization in diabetic retinopathy through regulating miR-125b/VE-cadherin axis. *Biosci. Rep.* 39:BSR20181469. doi: 10.1042/BSR20181469

Liu, B., Sun, L., Liu, Q., Gong, C., Yao, Y., Lv, X., et al. (2015). A cytoplasmic NF-κB interacting long noncoding RNA blocks IκB phosphorylation and suppresses breast cancer metastasis. *Cancer Cell* 27, 370–381. doi: 10.1016/j.ccell.2015.02.004

Liu, D. W., Zhang, J. H., Liu, F. X., Wang, X. T., Pan, S. K., Jiang, D. K., et al. (2019). Silencing of long noncoding RNA PVT1 inhibits podocyte damage and apoptosis in diabetic nephropathy by upregulating FOXA1. *Exp. Mol. Med.* 51, 1–15. doi: 10.1038/s12276-019-0259-6

Luo, H., Zhu, G., Xu, J., Lai, Q., Yan, B., Guo, Y., et al. (2019). HOTTIP lncRNA promotes hematopoietic stem cell self-renewal leading to AML-like disease in mice. *Cancer Cell* 36, 645–659.e8. doi: 10.1016/j.ccell.2019.10.011

Ma, L., Bajic, V. B., and Zhang, Z. (2013). On the classification of long non-coding RNAs. *RNA Biol.* 10, 925–933. doi: 10.4161/rna.24604

Martin, L. D., Rochelle, L. G., Fischer, B. M., Krunkosky, T. M., and Adler, K. B. (1997). Airway epithelium as an effector of inflammation: molecular regulation of secondary mediators. *Eur. Respir. J.* 10, 2139–2146. doi: 10.1183/09031936.97.10092139

Mattick, J. S., Amaral, P. P., Carninci, P., Carpenter, S., Chang, H. Y., Chen, L. L., et al. (2023). Long non-coding RNAs: definitions, functions, challenges and recommendations. *Nat. Rev. Mol. Cell Biol.* 24, 430–447. doi: 10.1038/s41580-022-00566-8

Medina, J. L., Coalson, J. J., Brooks, E. G., Winter, V. T., Chaparro, A., Principe, M. F. R., et al. (2012). *Mycoplasma pneumoniae* CARDS toxin induces pulmonary eosinophilic and lymphocytic inflammation. *Am. J. Respir. Cell Mol. Biol.* 46, 815–822. doi: 10.1165/rcmb.2011-0135OC

Meng, J., Chen, Y., and Zhang, C. (2019). Protective impacts of long noncoding RNA taurine-upregulated 1 against lipopolysaccharide-evoked injury in MRC-5 cells through inhibition of microRNA-127. *J. Cell. Biochem.* 120, 14928–14935. doi: 10.1002/jcb.28755

Meng, S., Zhou, H., Feng, Z., Xu, Z., Tang, Y., Li, P., et al. (2017). CircRNA: functions and properties of a novel potential biomarker for cancer. *Mol. Cancer* 16:94. doi: 10.1186/s12943-017-0663-2

Nadhan, R., Isidoro, C., Song, Y. S., and Dhanasekaran, D. N. (2022). Signaling by lncRNAs: structure, cellular homeostasis, and disease pathology. *Cells* 11:2517. doi: 10.3390/cells11162517

Narita, M. (2016). Classification of Extrapulmonary manifestations due to *Mycoplasma pneumoniae* infection on the basis of possible pathogenesis. *Front. Microbiol.* 7:23. doi: 10.3389/fmicb.2016.00023

Olivero, C. E., Martínez-Terroba, E., Zimmer, J., Liao, C., Tesfaye, E., Hooshdaran, N., et al. (2020). p53 activates the long noncoding RNA Pvt1b to inhibit myc and suppress tumorigenesis. *Mol. Cell* 77, 761–774.e8. doi: 10.1016/j.molcel.2019.12.014

Ounzain, S., Micheletti, R., Arnan, C., Plaisance, I., Cecchi, D., Schroen, B., et al. (2015). CARMEN, a human super enhancer-associated long noncoding RNA controlling cardiac specification, differentiation and homeostasis. *J. Mol. Cell. Cardiol.* 89, 98–112. doi: 10.1016/j.yjmcc.2015.09.016

Peng, H., Zou, L., Xie, J., Wu, H., Wu, B., Zhu, G., et al. (2017). lncRNA NONRATT021972 siRNA decreases diabetic neuropathic pain mediated by the $P2X_3$ receptor in dorsal root ganglia. *Mol. Neurobiol.* 54, 511–523. doi: 10.1007/s12035-015-9632-1

Perez, C., and Montes, M. (2002). Cutaneous leukocytoclastic vasculitis and encephalitis associated with *Mycoplasma pneumoniae* infection. *Arch. Intern. Med.* 162, 352–354. doi: 10.1001/archinte.162.3.352

Peschke, K., Weitzmann, A., Heger, K., Behrendt, R., Schubert, N., Scholten, J., et al. (2014). IκB kinase 2 is essential for IgE-induced mast cell de novo cytokine production but not for degranulation. *Cell Rep.* 8, 1300–1307. doi: 10.1016/j.celrep.2014.07.046

Poddighe, D. (2018). Extra-pulmonary diseases related to *Mycoplasma pneumoniae* in children: recent insights into the pathogenesis. *Curr. Opin. Rheumatol.* 30, 380–387. doi: 10.1097/BOR.0000000000000494

Poddighe, D. (2020). *Mycoplasma pneumoniae*-related hepatitis in children. *Microb. Pathog.* 139:103863. doi: 10.1016/j.micpath.2019.103863

Poddighe, D., Demirkaya, E., Sazonov, V., and Romano, M. (2022). *Mycoplasma pneumoniae* infections and primary immune deficiencies. *Int. J. Clin. Pract.* 2022:6343818. doi: 10.1155/2022/6343818

Rintala-Maki, N. D., and Sutherland, L. C. (2009). Identification and characterisation of a novel antisense non-coding RNA from the RBM5 gene locus. *Gene* 445, 7–16. doi: 10.1016/j.gene.2009.06.009

Salvatore, C. M., Fonseca-Aten, M., Katz-Gaynor, K., Gomez, A. M., Mejias, A., Somers, C., et al. (2007). Respiratory tract infection with *Mycoplasma pneumoniae* in interleukin-12 knockout mice results in improved bacterial clearance and reduced pulmonary inflammation. *Infect. Immun.* 75, 236–242. doi: 10.1128/IAI.01249-06

Schenten, V., Plançon, S., Jung, N., Hann, J., Bueb, J. L., Bréchard, S., et al. (2018). Secretion of the phosphorylated form of S100A9 from neutrophils is essential for the proinflammatory functions of extracellular S100A8/A9. *Front. Immunol.* 9:447. doi: 10.3389/fimmu.2018.00447

Schiopu, A., and Cotoi, O. S. (2013). S100A8 and S100A9: DAMPs at the crossroads between innate immunity, traditional risk factors, and cardiovascular disease. *Mediat. Inflamm.* 2013:828354. doi: 10.1155/2013/828354

Shimizu, T. (2015). Pathogenic factors of mycoplasma. *Nihon Saikingaku Zasshi* 70, 369–374. doi: 10.3412/jsb.70.369

Shimizu, T. (2016). Inflammation-inducing factors of *Mycoplasma pneumoniae*. *Front. Microbiol.* 7:414. doi: 10.3389/fmicb.2016.00414

Shimizu, T., Kida, Y., and Kuwano, K. (2008). *Mycoplasma pneumoniae*-derived lipopeptides induce acute inflammatory responses in the lungs of mice. *Infect. Immun.* 76, 270–277. doi: 10.1128/IAI.00955-07

Shimizu, T., Kimura, Y., Kida, Y., Kuwano, K., Tachibana, M., Hashino, M., et al. (2014). Cytadherence of *Mycoplasma pneumoniae* induces inflammatory responses through autophagy and toll-like receptor 4. *Infect. Immun.* 82, 3076–3086. doi: 10.1128/IAI.01961-14

Smekalova, E. M., Kotelevtsev, Y. V., Leboeuf, D., Shcherbinina, E. Y., Fefilova, A. S., Zatsepin, T. S., et al. (2016). lncRNA in the liver: prospects for fundamental research and therapy by RNA interference. *Biochimie* 131, 159–172. doi: 10.1016/j.biochi.2016.06.007

Statello, L., Guo, C. J., Chen, L. L., and Huarte, M. (2021a). Gene regulation by long non-coding RNAs and its biological functions. *Nat. Rev. Mol. Cell Biol.* 22, 96–118. doi: 10.1038/s41580-020-00315-9

Statello, L., Guo, C. J., Chen, L. L., and Huarte, M. (2021b). Author correction: gene regulation by long non-coding RNAs and its biological functions. *Nat. Rev. Mol. Cell Biol.* 22:159. doi: 10.1038/s41580-021-00330-4

Sun, Y., Wang, Y., Zou, M., Wang, T., Wang, L., and Peng, X. (2022). Lnc90386 sponges miR-33-5p to mediate *Mycoplasma gallisepticum*-induced inflammation and apoptosis in chickens via the JNK pathway. *Front. Immunol.* 13:887602. doi: 10.3389/fimmu.2022.887602

Tang, X., Wang, T., Qiu, C., Zheng, F., Xu, J., and Zhong, B. (2020). Long non-coding RNA (lncRNA) CRNDE regulated lipopolysaccharides (LPS)-induced MRC-5 inflammation injury through targeting MiR-141. *Med. Sci. Monit.* 26:e920928. doi: 10.12659/MSM.920928

Torrealba, N., Vera, R., Fraile, B., Martínez-Onsurbe, P., Paniagua, R., and Royuela, M. (2020). TGF-β/PI3K/AKT/mTOR/NF-kB pathway. clinicopathological features in prostate cancer. *Aging Male* 23, 801–811. doi: 10.1080/13685538.2019.1597840

Tripathi, V., Ellis, J. D., Shen, Z., Song, D. Y., Pan, Q., Watt, A. T., et al. (2010). The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation. *Mol. Cell* 39, 925–938. doi: 10.1016/j.molcel.2010.08.011

Tsai, T. A., Tsai, C. K., Kuo, K. C., and Yu, H. R. (2021). Rational stepwise approach for *Mycoplasma pneumoniae* pneumonia in children. *J. Microbiol. Immunol. Infect.* 54, 557–565. doi: 10.1016/j.jmii.2020.10.002

Velmeshev, D., Magistri, M., and Faghihi, M. A. (2013). Expression of non-protein-coding antisense RNAs in genomic regions related to autism spectrum disorders. *Mol. Autism.* 4:32. doi: 10.1186/2040-2392-4-32

Waites, K. B., Xiao, L., Liu, Y., Balish, M. F., and Atkinson, T. P. (2017). *Mycoplasma pneumoniae* from the respiratory tract and beyond. *Clin. Microbiol. Rev.* 30, 747–809. doi: 10.1128/CMR.00114-16

Wang, M., Jiang, Y. M., Xia, L. Y., Wang, Y., Li, W. Y., and Jin, T. (2018). LncRNA NKILA upregulation mediates oxygen glucose deprivation/re-oxygenation-induced neuronal cell death by inhibiting NF-κB signaling. *Biochem. Biophys. Res. Commun.* 503, 2524–2530. doi: 10.1016/j.bbrc.2018.07.010

Wang, K., Liu, F., Zhou, L. Y., Long, B., Yuan, S. M., Wang, Y., et al. (2014). The long noncoding RNA CHRF regulates cardiac hypertrophy by targeting miR-489. *Circ. Res.* 114, 1377–1388. doi: 10.1161/CIRCRESAHA.114.302476

Wang, Y., Meng, J., Wang, X., Liu, S., Shu, Q., Gao, L., et al. (2008). Expression of human TIM-1 and TIM-3 on lymphocytes from systemic lupus erythematosus patients. *Scand. J. Immunol.* 67, 63–70. doi: 10.1111/j.1365-3083.2007.02038.x

Wang, H., Qin, R., and Cheng, Y. (2020). LncRNA-Ang362 promotes pulmonary arterial hypertension by regulating miR-221 and miR-222. *Shock* 53, 723–729. doi: 10.1097/SHK.0000000000001410

Watanabe, Y., and Yamamoto, M. (1994). *S. pombe* mei2+ encodes an RNA-binding protein essential for premeiotic DNA synthesis and meiosis I, which cooperates with a novel RNA species meiRNA. *Cell* 78, 487–498. doi: 10.1016/0092-8674(94)90426-X

Wen, Y., Chen, H., Luo, F., Zhou, H., and Li, Z. (2020). Roles of long noncoding RNAs in bacterial infection. *Life Sci.* 263:118579. doi: 10.1016/j.lfs.2020.118579

Wright, A. A., Howitt, B. E., Myers, A. P., Dahlberg, S. E., Palescandolo, E., van Hummelen, P., et al. (2013). Oncogenic mutations in cervical cancer: genomic differences between adenocarcinomas and squamous cell carcinomas of the cervix. *Cancer* 119, 3776–3783. doi: 10.1002/cncr.28288

Xing, C., Sun, S. G., Yue, Z. Q., and Bai, F. (2021). Role of lncRNA LUCAT1 in cancer. *Biomed. Pharmacother.* 134:111158. doi: 10.1016/j.biopha.2020.111158

Yamamoto, K., Ferrari, J. D., Cao, Y., Ramirez, M. I., Jones, M. R., Quinton, L. J., et al. (2012). Type I alveolar epithelial cells mount innate immune responses during pneumococcal pneumonia. *J. Immunol.* 189, 2450–2459. doi: 10.4049/jimmunol.1200634

Yan, K., Fu, Y., Zhu, N., Wang, Z., Hong, J. L., Li, Y., et al. (2019). Repression of lncRNA NEAT1 enhances the antitumor activity of $CD8^+$T cells against hepatocellular carcinoma via regulating miR-155/Tim-3. *Int. J. Biochem. Cell Biol.* 110, 1–8. doi: 10.1016/j.biocel.2019.01.019

Yan, H., Yuan, J., Gao, L., Rao, J., and Hu, J. (2016). Long noncoding RNA MEG3 activation of p53 mediates ischemic neuronal death in stroke. *Neuroscience* 337, 191–199. doi: 10.1016/j.neuroscience.2016.09.017

Yang, F. (2022). Promoter antisense RNAs: beyond transcription by-products of active promoters. *RNA Biol.* 19, 533–540. doi: 10.1080/15476286.2022.2062177

Yang, L., Lin, C., Liu, W., Zhang, J., Ohgi, K. A., Grinstein, J. D., et al. (2011). ncRNA- and Pc2 methylation-dependent gene relocation between nuclear structures mediates gene activation programs. *Cell* 147, 773–788. doi: 10.1016/j.cell.2011.08.054

Yang, H. J., Song, D. J., and Shim, J. Y. (2017). Mechanism of resistance acquisition and treatment of macrolide-resistant *Mycoplasma pneumoniae* pneumonia in children. *Korean J. Pediatr.* 60, 167–174. doi: 10.3345/kjp.2017.60.6.167

Yang, L., Zhang, X., and Liu, X. (2021). Long non-coding RNA GAS5 protects against *Mycoplasma pneumoniae* pneumonia by regulating the microRNA-222-3p/TIMP3 axis. *Mol. Med. Rep.* 23:380. doi: 10.3892/mmr.2021.12019

Yu, X., Tang, W., Yang, Y., Tang, L., Dai, R., Pu, B., et al. (2018). Long noncoding RNA NKILA enhances the anti-cancer effects of baicalein in hepatocellular carcinoma via the regulation of NF-κB signaling. *Chem. Biol. Interact.* 285, 48–58. doi: 10.1016/j.cbi.2018.02.027

Zeng, T., Ni, H., Yu, Y., Zhang, M., Wu, M., Wang, Q., et al. (2019). BACE1-AS prevents BACE1 mRNA degradation through the sequestration of BACE1-targeting miRNAs. *J. Chem. Neuroanat.* 98, 87–96. doi: 10.1016/j.jchemneu.2019.04.001

Zhang, X., Wang, W., Zhu, W., Dong, J., Cheng, Y., Yin, Z., et al. (2019). Mechanisms and functions of long non-coding RNAs at multiple regulatory levels. *Int. J. Mol. Sci.* 20:5573. doi: 10.3390/ijms20225573

Zhang, F., Zhang, J., Liu, F., Zhou, Y., Guo, Y., Duan, Q., et al. (2021). Attenuated lncRNA NKILA enhances the secretory function of airway epithelial cells stimulated by *Mycoplasma pneumoniae* via NF-κB. *Biomed. Res. Int.* 2021:6656298. doi: 10.1155/2021/6656298

Zhao, G., Su, Z., Song, D., Mao, Y., and Mao, X. (2016). The long noncoding RNA MALAT1 regulates the lipopolysaccharide-induced inflammatory response through its interaction with NF-κB. *FEBS Lett.* 590, 2884–2895. doi: 10.1002/1873-3468.12315

Zhou, M., Diao, Z., Yue, X., Chen, Y., Zhao, H., Cheng, L., et al. (2016). Construction and analysis of dysregulated lncRNA-associated ceRNA network identified novel lncRNA biomarkers for early diagnosis of human pancreatic cancer. *Oncotarget* 7, 56383–56394. doi: 10.18632/oncotarget.10891

Zhu, X., du, J., Yu, J., Guo, R., Feng, Y., Qiao, L., et al. (2019). LncRNA NKILA regulates endothelium inflammation by controlling a NF-κB/KLF4 positive feedback loop. *J. Mol. Cell. Cardiol.* 126, 60–69. doi: 10.1016/j.yjmcc.2018.11.001

Zhu, C., Zhang, A., Huang, S., Ding, G., Pan, X., and Chen, R. (2010). Interleukin-13 inhibits cytokines synthesis by blocking nuclear factor-κB and c-Jun N-terminal kinase in human mesangial cells. *J. Biomed. Res.* 24, 308–316. doi: 10.1016/S1674-8301(10)60043-7

Zúñiga, J., Buendía-Roldán, I., Zhao, Y., Jiménez, L., Torres, D., Romo, J., et al. (2012). Genetic variants associated with severe pneumonia in A/H1N1 influenza infection. *Eur. Respir. J.* 39, 604–610. doi: 10.1183/09031936.00020611

frontiers | Frontiers in Microbiology

TYPE Original Research
PUBLISHED 31 May 2024
DOI 10.3389/fmicb.2024.1411609

OPEN ACCESS

EDITED BY
Atsushi Nakabachi,
Toyohashi University of Technology, Japan

REVIEWED BY
Rosario Gil,
University of Valencia, Spain
Qingguo Meng,
Nanjing Normal University, China

*CORRESPONDENCE
Shigeyuki Kakizawa
✉ s.kakizawa@aist.go.jp

RECEIVED 03 April 2024
ACCEPTED 08 May 2024
PUBLISHED 31 May 2024

CITATION
Mizutani M, Omori S, Yamane N, Suzuki Y, Glass JI, Chuang R-Y, Fukatsu T and Kakizawa S (2024) Cloning and sequencing analysis of whole *Spiroplasma* genome in yeast. *Front. Microbiol.* 15:1411609. doi: 10.3389/fmicb.2024.1411609

Cloning and sequencing analysis of whole *Spiroplasma* genome in yeast

Masaki Mizutani[1], Sawako Omori[1], Noriko Yamane[1], Yo Suzuki[2], John I. Glass[2], Ray-Yuan Chuang[2,3], Takema Fukatsu[1,4,5] and Shigeyuki Kakizawa[1]*

[1]Bioproduction Research Institute, National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba, Japan, [2]Synthetic Biology Group, J. Craig Venter Institute, La Jolla, CA, United States, [3]Telesis Bio, San Diego, CA, United States, [4]Department of Biological Sciences, Graduate School of Science, The University of Tokyo, Tokyo, Japan, [5]Graduate School of Life and Environmental Sciences, University of Tsukuba, Tsukuba, Japan

Cloning and transfer of long-stranded DNA in the size of a bacterial whole genome has become possible by recent advancements in synthetic biology. For the whole genome cloning and whole genome transplantation, bacteria with small genomes have been mainly used, such as mycoplasmas and related species. The key benefits of whole genome cloning include the effective maintenance and preservation of an organism's complete genome within a yeast host, the capability to modify these genome sequences through yeast-based genetic engineering systems, and the subsequent use of these cloned genomes for further experiments. This approach provides a versatile platform for in-depth genomic studies and applications in synthetic biology. Here, we cloned an entire genome of an insect-associated bacterium, *Spiroplasma chrysopicola*, in yeast. The 1.12 Mbp whole genome was successfully cloned in yeast, and sequences of several clones were confirmed by Illumina sequencing. The cloning efficiency was high, and the clones contained only a few mutations, averaging 1.2 nucleotides per clone with a mutation rate of 4×10^{-6}. The cloned genomes could be distributed and used for further research. This study serves as an initial step in the synthetic biology approach to *Spiroplasma*.

KEYWORDS

***Spiroplasma*, whole genome cloning, synthetic biology, yeast artificial chromosome vector, transformation-associated recombination (TAR) cloning**

Introduction

The synthetic biology has expanded greatly in recent years, making it possible to handle long DNA and conduct research using it (Kelwick et al., 2014; Venter et al., 2022). Bacteria with small genomes, such as *Mycoplasma* species and allied bacteria, have been used as targets in synthetic biology. Several studies have reported that their entire genomes were synthesized, cloned, modified, and transplanted into other bacterial cells. For example, whole genome cloning of several bacteria were reported including *Mycoplasma genitalium* (Gibson et al., 2008), *Mycoplasma mycoides* (Benders et al., 2010; Gibson et al., 2010), *Mycoplasma capricolum* and related species (Labroussaa et al., 2016), *Acholeplasma laidlawii* (Karas et al., 2012), *Synechococcus elongatus* PCC 7942 (Noskov et al., 2012), *Prochlorococcus marinus* MED4 (Tagwerker et al., 2012), and *Mesoplasma florum* (Baby et al., 2018). The entire genome of *M. mycoides* and its related species were transplanted into *M. capricolum* recipient cells (Lartigue et al., 2007, 2009; Labroussaa et al., 2016; Baby et al., 2018). The entire genome of *M. mycoides* was chemically synthesized, cloned into yeast, and transplanted into *M. capricolum* recipient cell to create JCVI-syn1.0, a bacterium

with a chemically synthesized genome (Gibson et al., 2010). Based on JCVI-syn1.0, a minimal cell JCVI-syn3.0, whose genome retains only essential genes, was constructed (Hutchison et al., 2016). Techniques for editing the entire genome of bacteria cloned into yeast have also been reported (Chandran et al., 2014; Tsarmpopoulos et al., 2016; Zhao et al., 2023). The advantages of enabling whole genome cloning encompass the ability to preserve the entire genome of an organism, the possibility of genome modification using the genetic engineering system in yeast, and its subsequent use in further research.

Members of the genus *Spiroplasma* comprise a bacterial group closely related to *Mycoplasma* (Davis et al., 1972). *Spiroplasma* species, characterized by their spiral cell shape and rotational swimming motility, infect a wide range of hosts including plants, insects, crustaceans, and mammals (Whitcomb, 1980; Cisak et al., 2015). *Spiroplasma citri* and *Spiroplasma kunkelii* are notorious to cause significant damage to citrus and corn, respectively (Whitcomb et al., 1986). *Spiroplasma mirum* is reported to cause cataracts and neurological damage in suckling mice (Tully et al., 1982). *Spiroplasma* species are detected in a wide variety of insects (Regassa and Gasparich, 2006; Cisak et al., 2015; Kakizawa et al., 2022), some of which cause male-killing phenotypes in fruit flies and other insects. In *Drosophila* fruit flies, *Spiroplasma poulsonii* induces male-killing, wherein an effector protein named Spaid was identified to induce male-specific apoptosis during embryogenesis (Harumoto and Lemaitre, 2018; Harumoto, 2023). Thus far, many *Spiroplasma* species remain unculturable, making it extremely difficult to elucidate their detailed characteristics. Furthermore, the absence of genetic knockdown or overexpression systems in most *Spiroplasma* strains has impeded experimental studies on their functional aspects. Recently, however, this difficulty was overcome, at least partially, by adoption of heterologous expression system in *Mycoplasma* cells. When cytoskeletal genes *mreB*s of *Spiroplasma* were expressed in *Mycoplasma* cells, the transformed *Mycoplasma* cells showed spiral cell shape and swimming motility (Kiyama et al., 2022; Lartigue et al., 2022). Such a new approach to investigate the genetic and functional aspects of *Spiroplasma* is anticipated.

In this study, we report cloning of the whole genome of *Spiroplasma chrysopicola* isolated from a deer fly *Chrysops* sp. (Diptera: Tabanidae) (Whitcomb et al., 1997; Ku et al., 2013) and the whole genome resequencing of the obtained clones. Although the size of the entire genome as large as 1.12 Mbp, the cloning efficiency was high, and the cloned sequences contained only a small number of mutations (1.2 nucleotides per clone), confirming that the whole *Spiroplasma* genome can be cloned in yeast cells with little alteration to the original genome information. This study serves as an initial step in the synthetic biology approach to *Spiroplasma*.

Materials and methods

Spiroplasma and yeast strains, cultivation, and culture media

The bacterial strain *S. chrysopicola* DF-1 was obtained from the American Type Culture Collection (ATCC 43209), which was cultured statically at 30°C using SP-4 medium (Tully et al., 1979). The cell growth was judged by the color of phenol red, a pH indicator. Cell morphology of *S. chrysopicola* was observed by optical microscope (IX71; Olympus). The yeast strain *Saccharomyces cerevisiae* VL6-48 was obtained from the ATCC (ATCC MYA-3666), which was cultured at 30°C using YPDA medium or SD-His medium (synthetic defined media lacking histidine, Clontech 630312) (Noskov et al., 2010).

Preparation of genomic *Spiroplasma* DNA for TAR cloning

Preparation of circular genomic DNA from *S. chrysopicola* cells was performed according to a previously reported method (Lartigue et al., 2007). Briefly, *S. chrysopicola* cells were cultured, collected by centrifugation, suspended in a buffer (10 mM Tris-HCl, pH 6.5, 500 mM sucrose, 50 mM EDTA), mixed with an equal volume of 2% UltraPure low melting point agarose (Invitrogen), and poured into plug molds (BioRad) to prepare agarose gel plugs. The cells were digested for 2 days using 1 mg/ml Proteinase K and 1% SDS solution at 55°C, thoroughly washed with a washing buffer (20 mM Tris-HCl, pH 8.0, 50 mM EDTA), and stored at 4°C. The circular genomes within the plugs were digested with restriction enzymes *Asc*I, *Sfi*I, and I-CeuI (New England Biolabs), and then the agar plugs were melted using thermostable beta-agarase (Nippon Gene). Pulsed-field gel electrophoresis (PFGE) was performed as described (Gibson et al., 2010). The condition of electrophoresis was 6 V, 50–90 sec pulse time for 22 h at 14°C. DNA bands were visualized with GelRed (Biotium Inc.).

Preparation of yeast artificial chromosome cloning vector

A summary of the plasmid YAC vector construction is shown in Supplementary Figure S1. PCR was performed using pRS313 plasmid as a template and primers (ROC800 and ROC801) to amplify a 4.2 kbp fragment containing *his3* gene, a selection marker gene in yeast. PCR was performed using the *M. mycoides* JCVI-syn1.0 genome as a template and primers (ROC802 and ROC803) to amplify a 6.2 kbp fragment containing yeast centromere. These PCR products were purified using PCR purification kit (QIAquick PCR Purification Kit; Qiagen), introduced into yeast VL6-48 cells, and assembled by *in vivo* homologous recombination to create pRC65 vector (10,415 bp). The pRC65 plasmid could multiply both in yeast and *Escherichia coli*, and be used as YAC cloning vector. The pRC65 plasmid was extracted from the yeast cells using Miniprep kit (QIAprep Spin Miniprep Kit; Qiagen) and introduced into *E. coli* DH5-alpha cells. The plasmid samples were extracted from the *E. coli* cells using Miniprep kit (Qiagen) and then sequenced by the Sanger sequencing method. PCR was performed using the pRC65 as a template and SChry_TAR primers to obtain a 6.8 kbp PCR product. These primers include flanking sequences homologous to the terminal regions of each genomic fragment within the *S. chrysopicola* genome. The PCR products were purified using PCR purification kit (Qiagen) and used as

vectors for the whole genome cloning. Primer sequences are shown in the Supplementary Table S1.

Primer design for colony PCRs

Primers for colony PCR to confirm the inserts were designed based on single-copy genes in the *S. chrysopicola* genome (Ku et al., 2013) to ensure specific amplification of certain genomic locations. To detect single-copy genes, BLAST analysis was performed on the *S. chrysopicola* genome using all *S. chrysopicola* genes as queries, and genes for which only one homolog was detected were designated as single-copy genes. Among detected single-copy genes, those that were evenly distributed across the genome were selected. In total 28 primer sets were designed on the genome, with seven sets for each genomic fragment. The positions of the designed primers and names of selected single-copy genes are listed in the Supplementary Figure S2, and sequences of the designed primers are shown in Supplementary Table S1.

Transformation-associated recombination cloning

The cloning method followed previously reported methods (Lartigue et al., 2009; Benders et al., 2010; Kouprina and Larionov, 2016). The circular S. *chrysopicola* genome was digested with the same restriction enzymes, and a mixture of 100 ng of the vector fragment and 1 μg of digested *S. chrysopicola* genome was used for yeast transformation.

In brief, the yeast strain was cultured in 50 ml of YPDA medium at 30°C with shaking until the OD_{600} reached 6.0–7.0. After harvesting by centrifugation (1,900×g, 4°C, 5 min), the cells were suspended in a 1 M sorbitol solution and incubated at 4°C overnight. The cells were then treated with the cell wall lytic enzyme Zymolyase (NACALAI TESQUE) to prepare spheroplasts in a phosphate buffer (pH 7.5) containing 1 M sorbitol, 10 mM EDTA, and 0.2% (v/v) beta-mercaptoethanol. After washing with the 1 M sorbitol solution, the cells were resuspended in STC buffer (1 M sorbitol, 10 mM Tris-HCl, pH 7.5, 10 mM $CaCl_2$) and incubated at room temperature for 10 min. The cells were then mixed with vector and genomic DNA inserts and incubated at room temperature for 10 min. A 20% polyethylene glycol (PEG) 8000 solution was added, mixed gently, and incubated at room temperature for 20 min. The cells were collected by centrifugation (3,200×g, 5 min), resuspended in SOS solution (1 M sorbitol, 6 mM $CaCl_2$, 0.3% yeast extract, 0.6% peptone), and incubated at 30°C for 30 min. Finally, the cells were mixed with the heat-melted SD-His TOP agar medium containing 1 M sorbitol, 2% glucose, and 3% agar, then spread onto SD-His agar plates. Selection of transformed yeast strains was performed on SD-His medium.

Analysis of YAC clones

To visualize insert bands in YAC clones, combination of conventional agarose gel electrophoresis and PFGE were performed. To exclude yeast chromosomes, conventional agarose gel electrophoresis was performed. The yeast clones containing the *S. chrysopicola* genome were cultured in SD-His medium until the OD_{600} reached 1.0 – 1.5 and embedded in agarose plugs. After treatment of the agarose plugs with Zymolyase, Proteinase K, and SDS for 2 days at 55°C, linear yeast chromosomes were flushed out from the agarose plugs by conventional electrophoresis (1% agarose gel, 100 V for 3 h at 4°C). This method is based on the phenomenon that only linear DNA migrates from the agarose plugs during electrophoresis, while circular, large DNA persists in the agarose plugs (Lartigue et al., 2009). After electrophoresis, agarose plugs were picked up from the agarose gel, washed with the washing buffer, and treated with *Not*I restriction enzyme for 2 h at 37°C. Since the cloning vector has two *Not*I recognition sites (Supplementary Table S1), the circular YAC vector including the genome insert was digested between the vector-insert boundaries. Then, the agarose plugs were washed with the washing buffer, and subjected to PFGE under the conditions of 6 V, 50–90 sec pulse time for 22 h at 14°C. DNA bands were visualized with GelRed.

The yeast clones containing the *S. chrysopicola* genome were cultured and embedded in agarose plugs, and liner DNAs were removed as described above. Then, from the agarose plugs, remaining circular DNA consisting of YAC including the *S. chrysopicola* genome was extracted using NucleoSpin Tissue kit (Macherey-Nagel). The extracted DNA was amplified using phi29 DNA polymerase (Thermo Scientific) and subjected to Illumina sequencing. The sequence reads were mapped onto the original genome sequence (GenBank accession number NC_021280.1), then mutations in cloned genomes were detected using CLC Genomics Workbench (Filgen).

Results

Preparation of whole *Spiroplasma* genome

Intact circular genomic DNA of *S. chrysopicola* was prepared from cultured bacterial cells embedded in agarose plugs. The agarose gel plugs were digested with restriction enzymes *Asc*I, *Sfi*I, and I-CeuI, and then subjected to PFGE. Length of all the fragments cut by the enzymes *Asc*I, *Sfi*I, and I-CeuI were similar with the expected band patterns (Figure 1). Note that it has been estimated that PFGE has an uncertainty in size determination ranging from 5% to 27% (Huang et al., 1999; Duck et al., 2003; Ferris Matthew et al., 2004). These digested fragments were used for further cloning experiments. By digestion with three enzymes (*Asc*I, *Sfi*I, and I-CeuI), four fragments were generated, all of which were used for genome cloning. Additionally, two types of whole genomes, each digested with either *Sfi*I or I-CeuI, were also used for genome cloning (Supplementary Table S2). Figure 2 illustrates the entire process from the isolation of the *S. chrysopicola* genome to cloning.

Preparation of cloning vector and whole genome cloning

Four partial genome fragments and two whole genome fragments of *S. chrysopicola* were used for cloning. As a result,

10.3389/fmicb.2024.1411609

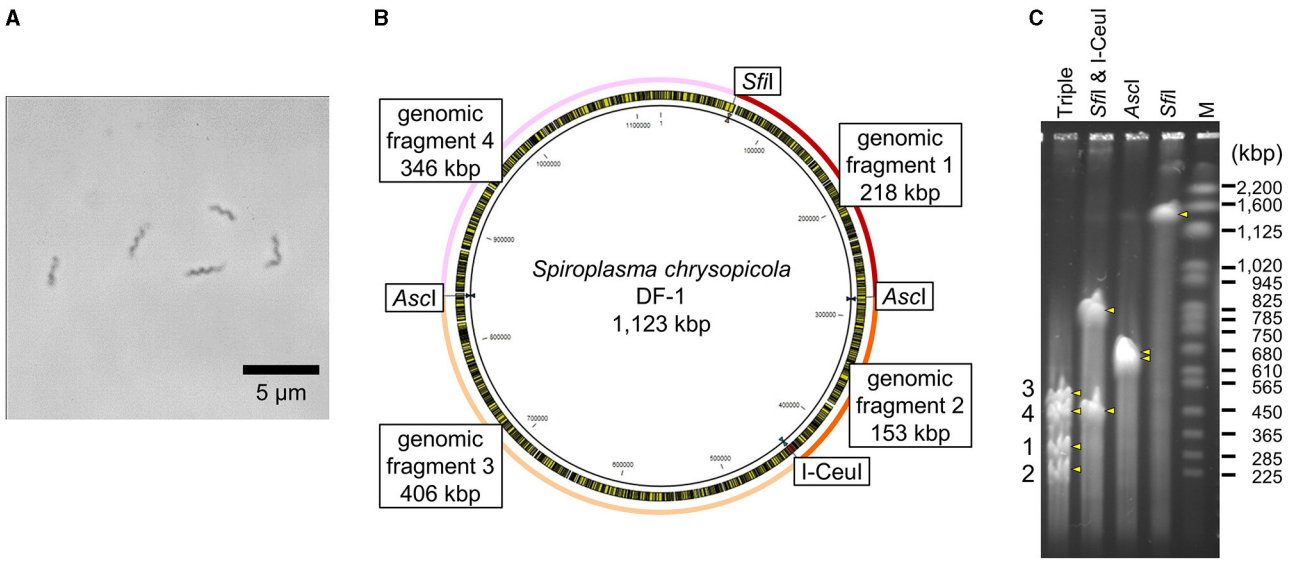


FIGURE 1
Genomic features of *S. chrysopicola*. **(A)** Spiral cell morphology of *S. chrysopicola*. **(B)** Genome map of *S. chrysopicola* DF-1 strain (1,123 kbp). Recognition sites for restriction enzymes *Asc*I, *Sfi*I and I-CeuI used, and four genome fragments produced by digestion with the enzymes are shown. **(C)** PFGE profiles of the *S. chrysopicola* genome after digestion with the restriction enzymes. Bands are indicated by yellow arrowheads. Four genome fragments, which correspond to those shown in **(B)**, are labeled on the left. Lane labels: Triple, triple digestion with *Asc*I, *Sfi*I and I-CeuI; *Sfi*I & I-CeuI, double digestion with *Sfi*I and I-CeuI; *Asc*I, single digestion with *Asc*I; *Sfi*I, single digestion with *Sfi*I; M, *Saccharomyces cerevisiae* chromosome marker (BioRad). Sizes of marker bands are shown on the right.

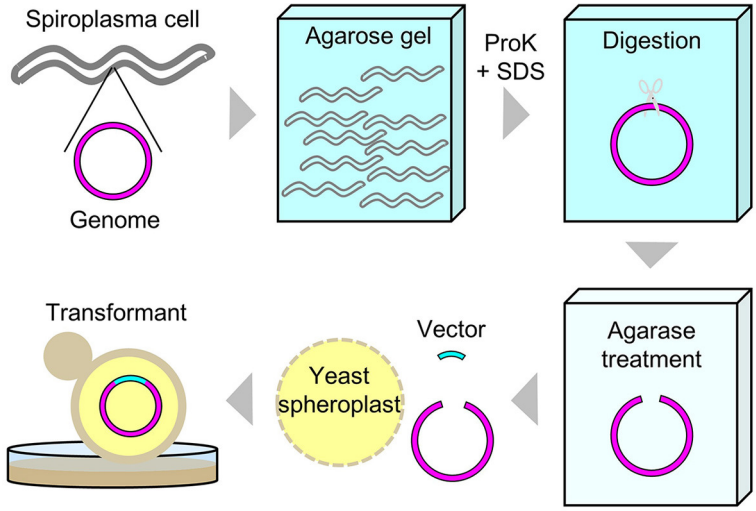


FIGURE 2
Schematic diagram of experimental procedures for the whole genome cloning of *S. chrysopicola*.

a large number of transformed yeast colonies were obtained for all the genome fragments (Supplementary Table S2). Sixteen clones were isolated for each of the genome fragments, the presence of the insert was checked by colony PCRs, and positive clones were obtained for most of the genome fragments (Supplementary Table S2). It was observed that the smaller the insert size, the higher the proportion of positive clones. Particularly in the fragment No. 2 (153 kbp), 100% of the clones (16/16) were judged positive by colony PCRs. Subsequent PFGE showed that most clones exhibited the expected fragment size (Figure 3). To exclude yeast chromosomes, the agarose plugs were subjected to conventional agarose gel electrophoresis in advance, but the removal was not complete and the remnant chromosomal DNAs persisted as background. Notwithstanding this, the sizes of the inserted genome fragments were clearly recognizable on the PFGE gels (Figure 3).

Notably, for two clones (3-2 and 5-1), their insert sizes were different from the expected ones. Particularly for the clone 5-1

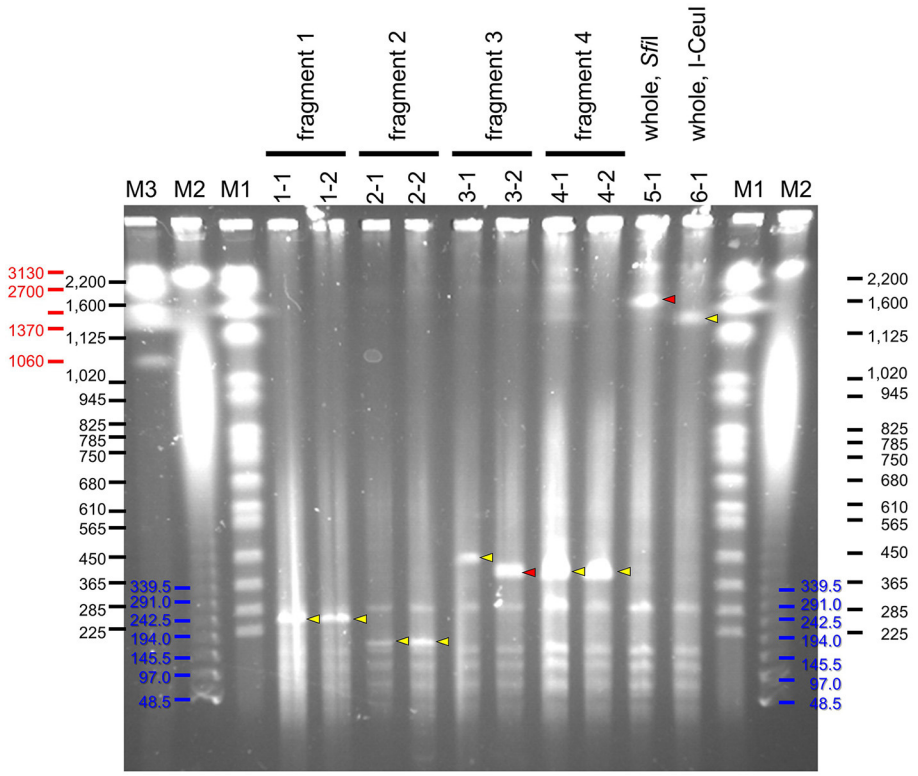


FIGURE 3
PFGE profiles of YAC clones inserted with genome fragments of *S. chrysopicola*. Yellow arrowheads indicate specific bands of expected size, whereas red arrowheads show specific bands of unexpected size. Lane labels: M1, *Saccharomyces cerevisiae* chromosome marker (BioRad), whose band sizes of bands are shown in black; M2, λ ladder marker (BioRad), whose band sizes are shown in blue; M3, *Hansenula wingei* chromosome marker (BioRad), whose band sizes are shown in red. For the other labels, see fragment numbers and clone numbers shown in Figure 1 and Supplementary Tables S1, S2.

(whole genome cut with *Sfi*I), although some colony PCRs failed to yield expected products presumably due to the absence of some genomic regions, PFGE results suggested that the insert size was longer than expected (Figure 3). For the clone 3-2, the insert size was shorter than expected, suggesting that the insert sequence would be partially missing.

Sequence analysis of clones

The obtained clones were subjected to Illumina sequencing analysis. The same clones used for PFGE were also used for the sequencing analysis. After purification by PFGE, the DNA samples, which were expected to be inserted YAC clones, were subjected to whole genome amplification using phi29 DNA polymerase, and then to Illumina short read sequencing. For each of all the 10 clones analyzed, a sufficient number of reads were obtained and mapped onto the target *S. chrysopicola* genome (Figure 4; Supplementary Table S3). The proportion of sequence reads mapped onto the *S. chrysopicola* genome ranged from 2.9% to 36.4%, with an average of 18.8%. The remaining reads were mapped onto yeast chromosomes, mitochondrial DNA, and vector sequence. Coverage of the cloned inserts varied from 137 to 591-fold, with an average of 347-fold, indicating sufficient read quantity. Mapping results showed that most clones were covered by the mapped reads almost evenly, in general, across the entire length of the insert, as expected (Figure 4). For two clones (clone 3-2 and 5-1) whose insert sizes were estimated by PFGE as different from the expected sizes (Figure 3), resequencing results also confirmed differences in length from the reference sequence, indicating partial loss of the insert. The insert size of the clone 5-1 was estimated by PFGE as larger than expected (Figure 3), but resequencing results uncovered some missing regions, suggesting that the insert sequence might have partially duplicated during the cloning process.

Comparison with the reference genome revealed that an extremely low number of mutations, with 0 to 3 mutations observed in each clone, and some clones had no mutations at all (Supplementary Table S3). We judged them as mutations when more than 90% of the mapped reads did not match the reference genome but supported the mutation (>90% mapped frequency). Lowering this threshold to 30% detected some additional mutations (Supplementary Table S4), most of which were in regions with consecutive A or T bases, with an increase or decrease in the number of the consecutive bases. This might be due to the inaccuracy of the phi29 DNA polymerase on the consecutive bases.

The sum of the sizes of all clones (total number of bases cloned), excluding duplicated regions, was 2,963,696 bases, and the total number of mutations in these clones was 12 nucleotides, resulting in an average mutation rate of 4.0 nucleotides per Mb, highlighting highly accurate genome cloning.

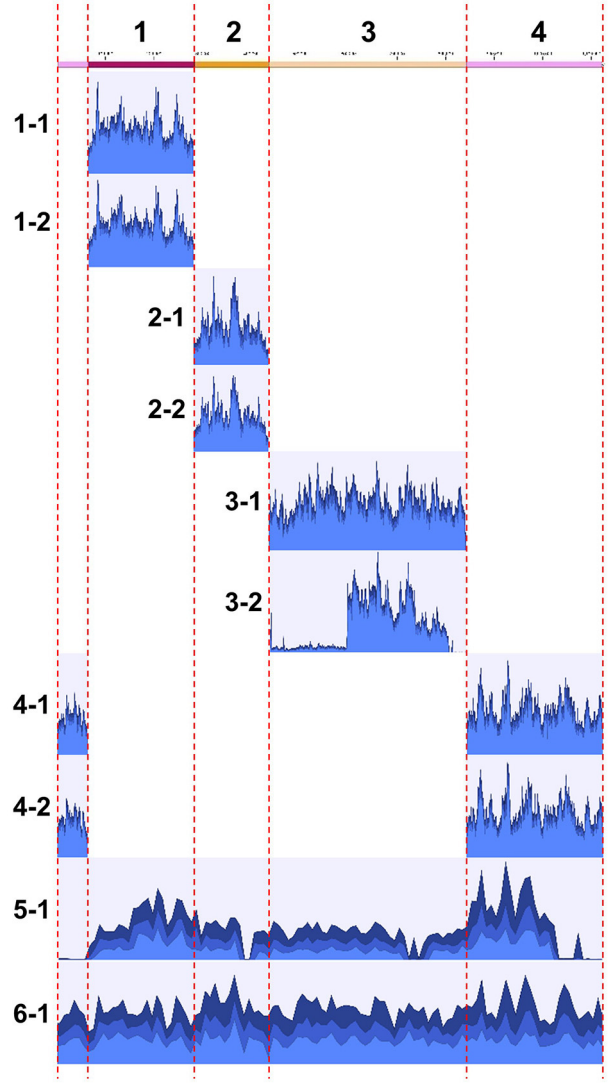


FIGURE 4
Mapping of Illumina short reads to the genome sequence of *S. chrysopicola*. On the left, clone numbers from Supplementary Table S2 are displayed. Note that the sizes of clones 3-2 and 5-1 are not consistent with the estimation by PFGE (see Figure 3).

Discussion

Cloning efficiency

We adopted the transformation-associated recombination (TAR) cloning method using the yeast *S. cerevisiae* as host cells (Kouprina and Larionov, 2016). The cloning efficiency observed in this study appears to be high. A large number of colonies were obtained with fewer experimental runs, and many clones contained inserts with the correct sequences. This might be related to the extent of repeat sequences in the insert genome fragments, since large parts of the genome could be lost through homologous recombination during TAR cloning procedures. It might also be related to the amount and concentration of the insert genome fragments (Rideau et al., 2017). It seems that the shorter the length of the insert, the higher the cloning efficiency, which is consistent with the characteristics of conventional cloning procedures. The cloning efficiency for inserts shorter than 200 kbp was exceptionally high, and cloning was possible even for inserts exceeding 1 Mbp. Circular YAC has been reported to clone inserts up to 1.66 Mbp (Tagwerker et al., 2012). Since many bacteria, archaea, and organelles with small genome sizes fall within this range, cloning their entire genomes seems feasible, in case that non-sheared, sufficient quantity of genomic DNA can be prepared.

Sequence analysis of YAC clones

In this study, the YAC insert fragments after cloning were subjected to Illumina short read sequencing. In order to purify the inserted YAC vector DNA, we attempted to exclude linear yeast chromosomes from the agarose plugs by conventional agarose gel electrophoresis, whereby circular YAC DNA was expected to persist in the agarose plugs. Then, DNA was recovered from the agarose plugs, and subjected to phi29 polymerase amplification and Illumina sequencing. Unexpectedly, however, on average, only 18.8% of sequence reads were mapped onto the insert sequences, with the remaining reads mapped onto yeast chromosomes and mitochondrial DNA. Given that YAC behaves as one of the 16 chromosomes in yeast cells, theoretically, if total DNA was extracted and sequenced directly from yeast cells without any treatment, ~1% of the reads are expected to originate from the YAC sequence. Hence, the method adopted in this study could enrich YAC sequences by an average of 18-fold. Here it should be noted that, with the recent advancement in next-generation sequencing analysis, sufficient coverage for large DNA inserts can be achieved relatively easily by obtaining a large number of reads, even when insert ratio in the DNA sample is <1%.

Mutation rate of clones

The resequencing of YAC clones revealed only a small number of mutations, and several clones were identical to the original genome sequence. Several single base deletions or insertions were observed in consecutive A or T bases, which were presumably introduced by amplification errors of phi29 DNA polymerase, but frequency of such mutations was low. In this study, circular whole genomes were extracted directly from cultured bacterial cells of *S. chrysopicola*, and theoretically, there are almost no steps where mutations are introduced into the bacterial genome, which may account for why the number of mutations in the inserts was at such a low level. A previous study reported the possibility of large degradation events in YAC-cloned large genome inserts after around 60 generations of yeast cultivation (Rideau et al., 2017). In this study, we used yeast cells cultured for 2-3 passages (12-18 generations) and many clones showed no mutations at all, indicating that the YAC-cloned large genome inserts are sufficiently stable at least in such a small number of yeast passages.

Future prospects

The entire *Spiroplasma* genome cloned in this study could be utilized for a variety of future studies.

First, genetic modifications of the *Spiroplasma* genome using yeast genetic engineering tools are possible. In yeast, a variety of genetic engineering tools are available, e.g., CRISPR/Cas9 system (Tsarmpopoulos et al., 2016; Ruiz et al., 2019), TREC (Noskov et al., 2010), and TREC-In (Chandran et al., 2014).

Second, functional analysis through genome transplantation could also be possible. Thus far, successful whole genome transplantation has been reported only in a very limited number of *Mycoplasma* species in the *M. mycoides* group (Lartigue et al., 2007; Labroussaa et al., 2016). Certainly transplantation of the entire *Spiroplasma* genome must be challenging, but, considering the close phylogenetic relationship between *M. mycoides* group and *Spiroplasma* (Lo et al., 2013), it would become feasible in the future. To achieve this goal, various relevant factors should be examined systematically, including selection of recipient cells, modification of genome sequences, improvement of the transplantation methods, and verification and elimination of inhibitory effects of restriction enzymes or nucleases (Lartigue et al., 2009). For example, many bacteria, including *Spiroplasma* and *Mycoplasma* are known to possess restriction modification systems that confer resistance to phage invasion by cleaving foreign DNAs. The cloned genome in yeast is not methylated, therefore, it is likely to be digested when transplanted into bacterial cells. The efficiency of genome transplantation could be enhanced by using bacterial methylases to methylate the donor genome extracted from yeast cells (Lartigue et al., 2007), or by using recipient cells that lack nucleases (Labroussaa et al., 2023). These approaches might be also effective in *Spiroplasma* genome transplantation. If genome transplantation in *Spiroplasma* becomes possible, various genetic modifications would also be feasible, including the knockout or overexpression of certain genes, as well as the introduction of complete metabolic pathways or genetic systems. Furthermore, synthetic biology approaches could facilitate large-scale genomic deletions, insertions, or replacements. These techniques might enhance our understanding of *Spiroplasma* biology.

Third, our sequencing results showed a very limited number of mutations in the cloned *Spiroplasma* genomes, indicating that the genetic information of the cloned bacterial genome in yeast cells can be stably preserved with minimal alterations. This observation highlights the possibility that this technique could be potentially utilized as a tool for preserving and storing the whole undamaged microbial genomes. For example, preservation, storage, and usage upon necessity of such microbial genomes that are with extremely slow growth rates, requiring complex media or specific conditions for cultivation, or difficult to access, would be enabled by retaining their entire genomes within yeast cells. Yeast grows easily and rapidly, its culture medium is inexpensive and easy to prepare, it does not require specific facilities for cultivation, and yeast cells can be stably preserved in freezers for long periods. By making use of sophisticated genetic tools available for yeast, the whole microbial genomes cloned in yeast can be subjected to large-scale genetic manipulation of the original genomic DNA, allowing the cloned genomes to be utilized as genetic resources. Most of *Spiroplasma* isolates are cultured in SP-4 medium whose composition is complex and contains a considerable amount of expensive fetal bovine serum (Tully et al., 1979). Therefore, the yeast clones obtained in this study are considered useful for preparing large quantities of *Spiroplasma* genomic DNA. The whole-genome cloning technology is also expected to be useful for utilizing environmentally inaccessible microorganisms as genetic resources. We expect that, as the yeast-mediated bacterial whole-genome cloning technology becomes easier and more accessible, it will be applied to diverse microbial species and research purposes, thereby facilitating further utilization of the cloned microbial genomes.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found at: https://www.ddbj.nig.ac.jp/, DRA018010-DRA018019.

Author contributions

MM: Formal analysis, Investigation, Data curation, Writing—review & editing. SO: Investigation, Writing—review & editing. NY: Investigation, Writing—review & editing. YS: Conceptualization, Writing—review & editing. JG: Conceptualization, Writing—review & editing. R-YC: Conceptualization, Methodology, Writing—review & editing. TF: Project administration, Writing—original draft, Writing—review & editing. SK: Conceptualization, Formal analysis, Investigation, Data curation, Visualization, Project administration, Writing—original draft, Writing—review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This study was supported by the JST ERATO Grant Number JPMJER1902 to SK and TF, AMED Grant Number JP23gm1610002 to SK, and the JSPS KAKENHI Grant 18H02433, 26710015, 26106004, 15KK0266 to SK, and JP17H06388 to TF. MM was supported by the JSPS Research Fellowships for Young Scientists (22KJ318 to MM).

Acknowledgments

We thank Rumi Numazaki (National Institute of Advanced Industrial Science and Technology) for technical support on molecular experiments.

Conflict of interest

R-YC was employed by Telesis Bio.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission.

This had no impact on the peer review process and the final decision.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2024.1411609/full#supplementary-material

SUPPLEMENTARY FIGURE S1
Schematic diagram of plasmid vector construction and whole genome cloning. Plasmid pRC65 was derived from pRS313 (available at Addgene) and a yeast centromere plasmid YCp. As a vector for whole genome cloning, regions homologous to the insert were added as flanking sequences of PCR primers. The schematic size of each gene does not correspond to its actual size.

SUPPLEMENTARY FIGURE S2
Genome map of *S. chrysopicola* with target genes of colony PCR. Seven single-copy genes from each of four genomic regions were selected for PCR amplification. Length of PCR fragments were designed to be differentiated by agarose gel electrophoresis.

SUPPLEMENTARY TABLE S1
List of primers used in this study.

SUPPLEMENTARY TABLE S2
Summary of whole genome cloning of *S. chrysopicola*.

SUPPLEMENTARY TABLE S3
Summary of resequencing of YAC clones.

SUPPLEMENTARY TABLE S4
Details of mutations in YAC clones.

References

Baby, V., Labroussaa, F., Brodeur, J., Matteau, D., Gourgues, G., Lartigue, C., et al. (2018). Cloning and transplantation of the *Mesoplasma florum* genome. *ACS Synth. Biol.* 7, 209–217. doi: 10.1021/acssynbio.7b00279

Benders, G. A., Noskov, V. N., Denisova, E. A., Lartigue, C., Gibson, D. G., Assad-Garcia, N., et al. (2010). Cloning whole bacterial genomes in yeast. *Nucleic Acids Res.* 38, 2558–2569. doi: 10.1093/nar/gkq119

Chandran, S., Noskov, V. N., Segall-Shapiro, T. H., Ma, L., Whiteis, C., Lartigue, C., et al. (2014). TREC-IN: gene knock-in genetic tool for genomes cloned in yeast. *BMC Genomics* 15:1180. doi: 10.1186/1471-2164-15-1180

Cisak, E., Wójcik-Fatla, A., Zajac, V., Sawczyn, A., Sroka, J., and Dutkiewicz, J. (2015). Spiroplasma – an emerging arthropod-borne pathogen? *Ann. Agric. Environ. Med.* 22, 589–593. doi: 10.5604/12321966.1185758

Davis, R. E., Worley, J. F., Whitcomb, R. F., Ishijima, T., and Steere, R. L. (1972). Helical filaments produced by a mycoplasma-like organism associated with corn stunt disease. *Science* 176, 521–523. doi: 10.1126/science.176.4034.521

Duck, W. M., Steward, C. D., Banerjee, S. N., McGowan, J. E. Jr., and Tenover, F. C. (2003). Optimization of computer software settings improves accuracy of pulsed-field gel electrophoresis macrorestriction fragment pattern analysis. *J. Clin. Microbiol.* 41, 3035–3042. doi: 10.1128/JCM.41.7.3035-3042.2003

Ferris Matthew, M., Yan, X., Habbersett Robbert, C., Shou, Y., Lemanski Cheryl, L., Jett James, H., et al. (2004). Performance assessment of DNA fragment sizing by high-sensitivity flow cytometry and pulsed-field gel electrophoresis. *J. Clin. Microbiol.* 42, 1965–1976. doi: 10.1128/JCM.42.5.1965-1976.2004

Gibson, D. G., Benders, G. A., Andrews-Pfannkoch, C., Denisova, E. A., Baden-Tillson, H., Zaveri, J., et al. (2008). Complete chemical synthesis, assembly, and cloning of a *Mycoplasma genitalium* genome. *Science* 319, 1215–1220. doi: 10.1126/science.1151721

Gibson, D. G., Glass, J. I., Lartigue, C., Noskov, V. N., Chuang, R. Y., Algire, M. A., et al. (2010). Creation of a bacterial cell controlled by a chemically synthesized genome. *Science* 329, 52–56. doi: 10.1126/science.1190719

Harumoto, T. (2023). Self-stabilization mechanism encoded by a bacterial toxin facilitates reproductive parasitism. *Curr. Biol.* 33, 4021–4029. doi: 10.1016/j.cub.2023.08.032

Harumoto, T., and Lemaitre, B. (2018). Male-killing toxin in a bacterial symbiont of *Drosophila*. *Nature* 557, 252–255. doi: 10.1038/s41586-018-0086-2

Huang, Z., Jett, J. H., and Keller, R. A. (1999). Bacteria genome fingerprinting by flow cytometry. *Cytometry* 35, 169–175. doi: 10.1002/(SICI)1097-0320(19990201)35:2<169::AID-CYTO9>3.0.CO;2-K

Hutchison, C. A. 3rd., Chuang, R. Y., Noskov, V. N., Assad-Garcia, N., Deerinck, T. J., Ellisman, M. H., et al. (2016). Design and synthesis of a minimal bacterial genome. *Science* 351:aad6253. doi: 10.1126/science.aad6253

Kakizawa, S., Hosokawa, T., Oguchi, K., Miyakoshi, K., and Fukatsu, T. (2022). *Spiroplasma* as facultative bacterial symbionts of stinkbugs. *Front. Microbiol.* 13:1044771. doi: 10.3389/fmicb.2022.1044771

Karas, B. J., Tagwerker, C., Yonemoto, I. T., Hutchison, C. A. 3rd., and Smith, H. O. (2012). Cloning the *Acholeplasma laidlawii* PG-8A genome in *Saccharomyces cerevisiae* as a yeast centromeric plasmid. *ACS Synth. Biol.* 1, 22–28. doi: 10.1021/sb200013j

Kelwick, R., MacDonald, J. T., Webb, A. J., and Freemont, P. (2014). Developments in the tools and methodologies of synthetic biology. *Front. Bioeng. Biotechnol.* 2:60. doi: 10.3389/fbioe.2014.00060

Kiyama, H., Kakizawa, S., Sasajima, Y., Tahara, Y. O., and Miyata, M. (2022). Reconstitution of a minimal motility system based on *Spiroplasma* swimming by two bacterial actins in a synthetic minimal bacterium. *Sci. Adv.* 8:eabo7490. doi: 10.1126/sciadv.abo7490

Kouprina, N., and Larionov, V. (2016). Transformation-associated recombination (TAR) cloning for genomics studies and synthetic biology. *Chromosoma* 125, 621–632. doi: 10.1007/s00412-016-0588-3

Ku, C., Lo, W. S., Chen, L. L., and Kuo, C. H. (2013). Complete genomes of two dipteran-associated spiroplasmas provided insights into the origin, dynamics, and impacts of viral invasion in spiroplasma. *Genome Biol. Evol.* 5, 1151–1164. doi: 10.1093/gbe/evt084

Labroussaa, F., Lebaudy, A., Baby, V., Gourgues, G., Matteau, D., Vashee, S., et al. (2016). Impact of donor-recipient phylogenetic distance on bacterial genome transplantation. *Nucleic Acids Res.* 44, 8501–8511. doi: 10.1093/nar/gkw688

Labroussaa, F., Torres-Puig, S., and Jores, J. (2023). "Chapter 1 - Genome transplantation in *Mollicutes*" in *Methods in Microbiology,* eds. V. Gurtler and M. Calcutt (New York, NY: Academic Press), 3–32.

Lartigue, C., Glass, J. I., Alperovich, N., Pieper, R., Parmar, P. P., Hutchison, C. A. 3rd, et al. (2007). Genome transplantation in bacteria: changing one species to another. *Science* 317, 632–638. doi: 10.1126/science.1144622

Lartigue, C., Lambert, B., Rideau, F., Dahan, Y., Decossas, M., Hillion, M., et al. (2022). Cytoskeletal components can turn wall-less spherical bacteria into kinking helices. *Nat. Comm.* 13:6930. doi: 10.1038/s41467-022-34478-0

Lartigue, C., Vashee, S., Algire, M. A., Chuang, R. Y., Benders, G. A., Ma, L., et al. (2009). Creating bacterial strains from genomes that have been cloned and engineered in yeast. *Science* 325, 1693–1696. doi: 10.1126/science.1173759

Lo, W. S., Chen, L. L., Chung, W. C., Gasparich, G. E., and Kuo, C.-H. (2013). Comparative genome analysis of *Spiroplasma melliferum* IPMB4A, a honeybee-associated bacterium. *BMC Genomics* 14:22. doi: 10.1186/1471-2164-14-22

Noskov, V. N., Karas, B. J., Young, L., Chuang, R. Y., Gibson, D. G., Lin, Y. C., et al. (2012). Assembly of large, high G+C bacterial DNA fragments in yeast. *ACS Synth. Biol.* 1, 267–273. doi: 10.1021/sb3000194

Noskov, V. N., Segall-Shapiro, T. H., and Chuang, R. Y. (2010). Tandem repeat coupled with endonuclease cleavage (TREC): a seamless modification tool for genome engineering in yeast. *Nucleic Acids Res.* 38, 2570–2576. doi: 10.1093/nar/gkq099

Regassa, L. B., and Gasparich, G. E. (2006). Spiroplasmas: evolutionary relationships and biodiversity. *Front. Biosci.* 11, 2983–3002. doi: 10.2741/2027

Rideau, F., Le Roy, C., Descamps, E. C. T., Renaudin, H., Lartigue, C., and Bébéar, C. (2017). Cloning, stability, and modification of *Mycoplasma hominis* genome in yeast. *ACS Synth. Biol.* 6, 891–901. doi: 10.1021/acssynbio.6b00379

Ruiz, E., Talenton, V., Dubrana, M. P., Guesdon, G., Lluch-Senar, M., Salin, F., et al. (2019). CReasPy-cloning: a method for simultaneous cloning and engineering of megabase-sized genomes in yeast using the CRISPR-Cas9 system. *ACS Synth. Biol.* 8, 2547–2557. doi: 10.1021/acssynbio.9b00224

Tagwerker, C., Dupont, C. L., Karas, B. J., Ma, L., Chuang, R. Y., Benders, G. A., et al. (2012). Sequence analysis of a complete 1.66 Mb *Prochlorococcus marinus* MED4 genome cloned in yeast. *Nucleic Acids Res.* 40, 10375–10383. doi: 10.1093/nar/gks823

Tsarmpopoulos, I., Gourgues, G., Blanchard, A., Vashee, S., Jores, J., Lartigue, C., et al. (2016). In-yeast engineering of a bacterial genome using CRISPR/Cas9. *ACS Synth. Biol.* 5, 104–109. doi: 10.1021/acssynbio.5b00196

Tully, J. G., Rose, D. L., Whitcomb, R. F., and Wenzel, R. P. (1979). Enhanced isolation of *Mycoplasma pneumoniae* from throat washings with a newly-modified culture medium. *J. Infect. Dis.* 139, 478–482. doi: 10.1093/infdis/139.4.478

Tully, J. G., Whitcomb, R. F., Rose, D. L., and Bové, J. M. (1982). *Spiroplasma mirum*, a new species from the rabbit Tick (*Haemaphysalis leporispalustris*). *Int. J. Syst. Evol. Microbiol.* 32, 92–100. doi: 10.1099/00207713-32-1-92

Venter, J. C., Glass, J. I., Hutchison, C. A. 3rd., and Vashee, S. (2022). Synthetic chromosomes, genomes, viruses, and cells. *Cell* 185, 2708–2724. doi: 10.1016/j.cell.2022.06.046

Whitcomb, R. F. (1980). The Genus Spiroplasma. *Annu. Rev. Microbiol.* 34, 677–704. doi: 10.1146/annurev.mi.34.100180.003333

Whitcomb, R. F., Chen, T. A., Williamson, D. L., Liao, C., Tully, J. G., Bové, J. M., et al. (1986). *Spiroplasma kunkelii* sp. nov.: characterization of the etiological agent of corn stunt disease. *Int. J. Syst. Evol. Microbiol.* 36, 170–178. doi: 10.1099/00207713-36-2-170

Whitcomb, R. F., French, F. E., Tully, J. G., Gasparich, G. E., Rose, D. L., Carle, P., et al. (1997). *Spiroplasma chrysopicola* sp. nov., *Spiroplasma gladiatoris* sp. nov., *Spiroplasma helicoides* sp. nov., and *Spiroplasma tabanidicola* sp. nov., from Tabanid (Diptera: Tabanidae) Flies. *Int. J. Syst. Evol. Microbiol.* 47, 713–719. doi: 10.1099/00207713-47-3-713

Zhao, G., Lu, D., Li, M., and Wang, Y. (2023). Gene editing tools for mycoplasmas: references and future directions for efficient genome manipulation. *Front. Microbiol.* 14:1191812. doi: 10.3389/fmicb.2023.1191812

Frontiers in Microbiology

Explores the habitable world and the potential of microbial life

The largest and most cited microbiology journal which advances our understanding of the role microbes play in addressing global challenges such as healthcare, food security, and climate change.

Discover the latest Research Topics

See more →

Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact