

RESEARCH ARTICLE

Open Access

# Genome analysis of *Spiroplasma citri* strains from different host plants and its leafhopper vectors



Rachel Rattner<sup>1†</sup>, Shree Prasad Thapa<sup>2†</sup>, Tyler Dang<sup>3</sup>, Fatima Osman<sup>2</sup>, Vijayanandraj Selvaraj<sup>1</sup>, Yogita Maheshwari<sup>1</sup>, Deborah Pagliaccia<sup>3</sup>, Andres S. Espindola<sup>4</sup>, Subhas Hajeri<sup>5</sup>, Jianchi Chen<sup>1</sup>, Gitta Coaker<sup>2</sup>, Georgios Vidalakis<sup>3</sup> and Raymond Yokomi<sup>1\*</sup>

## Abstract

**Background:** *Spiroplasma citri* comprises a bacterial complex that cause diseases in citrus, horseradish, carrot, sesame, and also infects a wide array of ornamental and weed species. *S. citri* is transmitted in a persistent propagative manner by the beet leafhopper, *Neolaliturus tenellus* in North America and *Circulifer haematiceps* in the Mediterranean region. Leafhopper transmission and the pathogen's wide host range serve as drivers of genetic diversity. This diversity was examined in silico by comparing the genome sequences of seven *S. citri* strains from the United States (BR12, CC-2, C5, C189, LB 319, BLH-13, and BLH-MB) collected from different hosts and times with other publicly available spiroplasmas.

**Results:** Phylogenetic analysis using 16S rRNA sequences from 39 spiroplasmas obtained from NCBI database showed that *S. citri* strains, along with *S. kunkelii* and *S. phoeniceum*, two other plant pathogenic spiroplasmas, formed a monophyletic group. To refine genetic relationships among *S. citri* strains, phylogenetic analyses with 863 core orthologous sequences were performed. Strains that clustered together were: CC-2 and C5; C189 and R8-A2; BR12, BLH-MB, BLH-13 and LB 319. Strain GII3-3X remained in a separate branch. Sequence rearrangements were observed among *S. citri* strains, predominantly in the center of the chromosome. One to nine plasmids were identified in the seven *S. citri* strains analyzed in this study. Plasmids were most abundant in strains isolated from the beet leafhopper, followed by strains from carrot, Chinese cabbage, horseradish, and citrus, respectively. All these *S. citri* strains contained one plasmid with high similarity to plasmid pSci6 from *S. citri* strain GII3-3X which is known to confer insect transmissibility. Additionally, 17 to 25 prophage-like elements were identified in these genomes, which may promote rearrangements and contribute to repetitive regions.

(Continued on next page)

\* Correspondence: [ray.yokomi@usda.gov](mailto:ray.yokomi@usda.gov)

†Rachel Rattner and Shree Prasad Thapa contributed equally to this work.

<sup>1</sup>Crop Diseases, Pests, and Genetics Research Unit, San Joaquin Valley Agricultural Sciences Center, USDA Agricultural Research Service, Parlier, CA 93648, USA

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



© The Author(s). 2021 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

(Continued from previous page)

**Conclusions:** The genome of seven *S. citri* strains were found to contain a single circularized chromosome, ranging from 1.58 Mbp to 1.74 Mbp and 1597–2232 protein-coding genes. These strains possessed a plasmid similar to pSci6 from the G13–3X strain associated with leafhopper transmission. Prophage sequences found in the *S. citri* genomes may contribute to the extension of its host range. These findings increase our understanding of *S. citri* genetic diversity.

**Keywords:** Citrus stubborn disease, Beet leafhopper, Spiroplasma, Sequencing, Genome assembly, Prophage

## Background

Spiroplasmas (helical mollicutes: Firmicutes: Mollicutes: Entomoplasmatales: Spiroplasmataceae) are wall-less, gram-positive bacteria with mobile helical cells. The bacteria are fastidious, culturable in cell-free media [1], and have a diverse host range [2]. Spiroplasmas are found in many arthropods including bees, flies, mosquitoes, scorpion flies, beetles, and ticks [2] and have host relationships that range from commensal, mutualistic, and pathogenic [3]. Plant pathogenic spiroplasmas cause economic damage to crops and ornamentals. These pathogens include: *Spiroplasma citri*, causal agent of citrus stubborn disease (CSD) [4], brittle root of horseradish [5], and carrot purple leaf [6]; *S. kunkelli*, the causal agent of corn stunt [7]; and *S. phoeniceum*, isolated from periwinkle showing symptoms typical of mycoplasma-like organisms [8]. Plant pathogenic spiroplasmas are transmitted in a persistent propagative manner by leafhoppers. Vectors of *S. citri* are the beet leafhopper (BLH), *Neotalitrus* (syn. *Circulifer*) *tenellus* (Baker) [9] in North America and *Circulifer haematoceps* (Mulsant et Rey) in the Mediterranean region [10]. *S. kunkelli* is transmitted by *Dalbulus maidis* (DeLong) [7] and *S. phoeniceum* was experimentally transmitted by *Macrostes fascifrons* (Stål) [8].

Characterization of spiroplasmas have been based on morphological and biological properties. However, because growth, metabolism, and DNA-DNA relatedness studies are time consuming, serological deformation tests and enzyme-linked immunosorbent assays have been used for identification of new spiroplasma groups in accordance to the International Subcommittee on the Taxonomy of Mycoplasmatales [2, 11]. Recently, long-read high-throughput sequencing technology and whole genome sequencing of bacteria have become cost-effective and offers a precise method to differentiate bacterial species and strains that have highly repetitive regions in its genome [12, 13].

The pathogen's wide host range and vector transmission serve as bottlenecks and drivers of genetic diversity of *S. citri* populations. Although *S. citri* exist in free living form in insect hemolymph and appropriate culture media, the pathogen in the vector must enter and move through the salivary gland and exit into the salivary duct and be expelled by the vector during probing and/or

feeding in plants where *S. citri* infects host phloem tissue and exists intracellularly and is phloem-limited. Therefore, the objective of this study is to examine the genomes of *S. citri* collected from diverse hosts from different locations and times; and analyze the relationship between the genotype and phenotype of *S. citri* from citrus and horseradish (perennial crops); carrot and Chinese cabbage (annual crops); and from the BLH vector. The analysis was performed on whole-genome sequences of seven newly sequenced strains of *S. citri* and compared amongst each other, other *S. citri* strains, and spiroplasmas from other hosts. New insights in the evolution and diversity of *S. citri* is presented herein.

## Results

### Genome assembly and annotation

Cultures of six strains of *Spiroplasma citri* were established and sequences reported previously (Table 1) [14, 15] and a new strain, C5, is reported here. Briefly, *S. citri* strains C189 and LB 319 were isolated from the woody crop, citrus. BR12, CC-2, and C5 strains were isolated from the seasonal crops such as horseradish, Chinese cabbage, and carrot, respectively. BLH-13 and BLH-MB strains were isolated from the BLH. The complete genomes of the six strains were acquired using the long-read technology, PacBio [14, 15] and C5 was obtained using Nanopore sequencing technology. Sequences from each strain were assembled into single chromosomal contigs. Contigs that did not associate with the chromosome were designated as putative plasmids (Table 2). The chromosome and plasmid status of each contig were further confirmed by BLASTn analyses against the GenBank database Release 236 (Supplementary Table S1). The circular chromosome for all seven strains ranged from 1,576,550 to 1,742,208 bp, with an average G + C content of 25.4%. Total genome size ranged from 1,611,714 to 1,832,173 bp in strains isolated from plants and 1,968,976 to 2,155,613 bp in strains isolated from the BLH. Annotation of each contig was performed by the NCBI Prokaryotic Genome Annotation Pipeline (PGAP), which predicted 32 tRNA genes, three rRNA genes and protein-coding genes which ranged between 1597 and 2232. Extrachromosomal DNAs, characterized as putative plasmids varied in all the strains viz., one or

**Table 1** *Spiroplasma citri* strains analyzed in this study

Strain	Host	Location	Year of collection	Reference
C189	Citrus	Riverside, California	1960	[14]
LB 319	Citrus	Ducor, California	2007	[14]
BR12	Horseradish	Collinsville, Illinois	1984	[14]
CC-2	Chinese cabbage	Fresno, California	2016	[15]
C5	Carrot	Bakersfield, California	2005	This study
BLH-13	Beet leafhopper	Mettler, California	2010	[14]
BLH-MB	Beet leafhopper	Parlier, California	2011	[14]

two plasmids from citrus, two plasmids from horseradish, three plasmids from Chinese cabbage, seven plasmids from carrot, and eight or nine plasmids from the BLH. Putative plasmid sizes ranged from 2047 bp to 135,023 bp (Supplementary Table S1). Seven of the 32 plasmids identified in these seven strains could not be circularized and further research is needed to determine if they are linear or products of sequencing error or culturing conditions.

#### Phylogenomics

Molecular phylogenetic inference of 39 spiroplasmas was performed using 16S rRNA genes in the NCBI database. Analysis of this gene sequence indicated that *S. citri* strains are closely related, but not identical. The phylogeny inferred from the 16S rRNA gene shows that *S. citri* strains formed a monophyletic group with plant pathogenic *S. kunkelii*, *S. phoeniceum*, and a honeybee pathogen, *S. melliferum* (Fig. 1, Supplementary Table S2). To facilitate a high-resolution comparison of *S. citri*

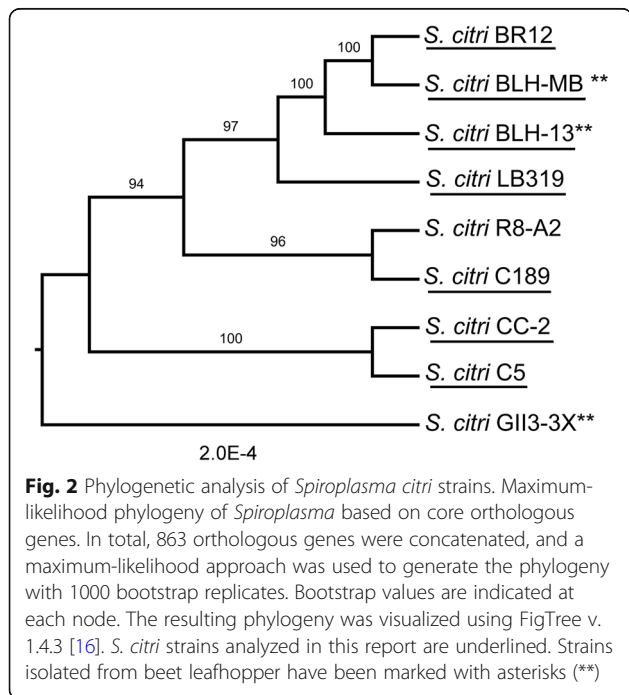
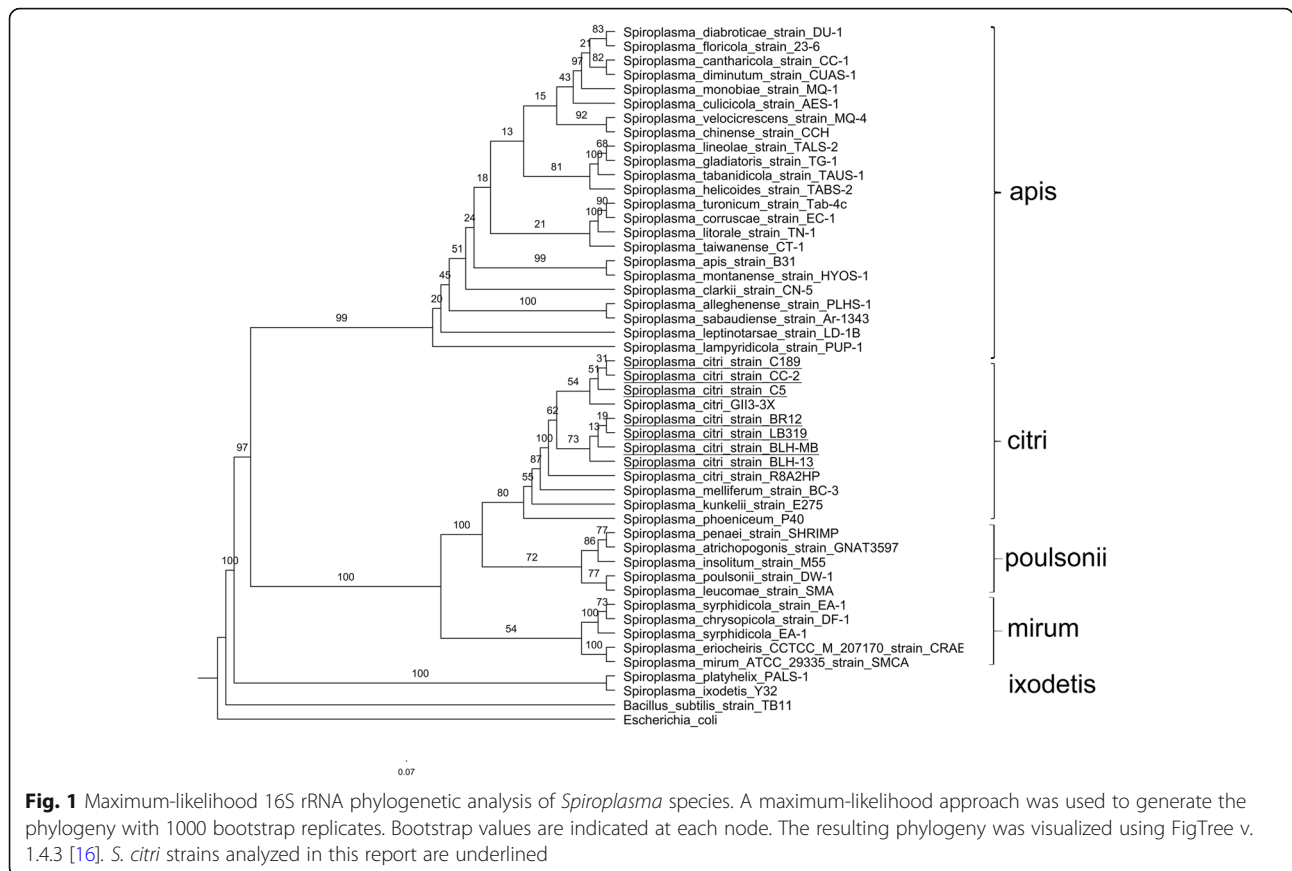
strains, core genomes were analyzed for nine *S. citri* genomes available in NCBI. Using the orthoMCL pipeline, a total of 863 orthologous genes were identified as conserved among the *S. citri* strains. The 863 orthologous genes were concatenated, and a maximum-likelihood approach was employed to generate a *S. citri* phylogeny (Fig. 2). Phylogenetic analyses with the core orthologous sequences among the *S. citri* strains showed citrus strains C189 from southern California and R8-A2 from Morocco clustered together. CC-2, isolated from Chinese cabbage, and C5, isolated from carrot, clustered together. Strains LB 319, BLH-13, BLH-MB, and BR12 clustered together in a separate clade. There was clear separation of *S. citri* from *S. kunkelii* (Supplementary Fig. S1).

#### Comparative genomics

The circular chromosome of the seven *S. citri* strains from the U.S. was compared via BLASTn to the R8-A2 strain from citrus in Morocco as the reference

**Table 2** Genome assembly statistics for *Spiroplasma citri* strains analyzed in this study

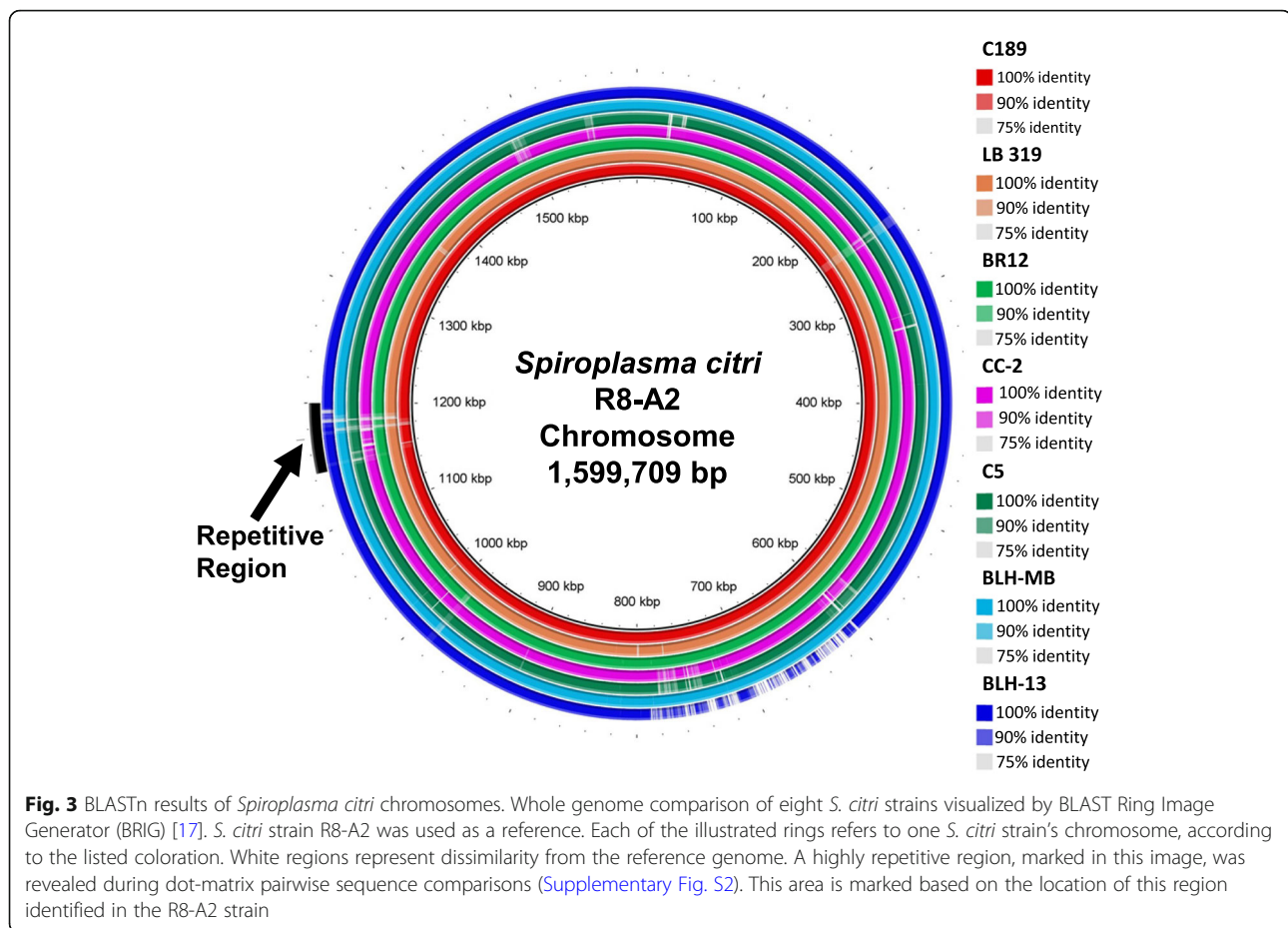
<i>Spiroplasma citri</i> strain	<i>S. citri</i> C189	<i>S. citri</i> LB 319	<i>S. citri</i> BR12	<i>S. citri</i> CC-2	<i>S. citri</i> C5	<i>S. citri</i> BLH-13	<i>S. citri</i> BLH-MB
Chromosome size (bp)	1,577,041	1,734,522	1,731,112	1,709,192	1,618,536	1,576,550	1,742,208
Combined size of plasmids (bp)	34,663	95,488	101,061	82,444	126,436	392,426	413,405
Chromosome + plasmid size (bp)	1,611,704	1,830,010	1,832,173	1,791,636	1,795,359	1,968,976	2,155,613
No. of plasmids	1	2	2	3	7	8	9
Chromosome GC content (%)	25.6	25.4	25.4	25.6	25.6	25.4	25.4
Total no. genes	1946	2207	2200	2068	2064	2594	2411
Total protein-coding genes	1597	1853	1876	1716	1701	2232	2082
Total rRNA genes	3	3	3	3	3	3	3
Total tRNA genes	32	32	32	32	32	32	32
GenBank Accessions	CP047426.1, CP047427.1 [15]	CP046371.1 - CP046373.1 [15]	CP046368.1 - CP046370.1 [15]	CP042472.1 - CP042475.1 [14]	CP053304.1 - CP053311.1 (this study)	CP047428.1 - CP047436.1 [15]	CP047437.1 - CP047446.1 [15]



sequence. Visualization of these results was performed using the BLAST Ring Image Generator (BRIG). This genome level comparison among *S. citri* strains isolated from different sources showed a high level of homogeneity among each other and the reference genome, R8-A2 (Fig. 3). *S. citri* strains C189, LB 319, and BR-12, which were isolated from citrus and horseradish, appear most similar to R8-A2. A large region of dissimilarity near the middle of the chromosome is notable in the BLH-13 strain, ranging from ~ 600 kbp to ~ 800 kbp. These differences were not found in the BLH-MB strain, but some variability in this region can be seen in CC-2 and C5.

Dot-matrix pairwise sequence comparisons revealed a highly repetitive region at ~ 1200 kbp (Supplementary Fig. S2). The repetitive region, marked in Fig. 3, is an area of higher dissimilarity in the chromosome among the *S. citri* strains analyzed in this study.

Pairwise whole genome comparisons were performed with *S. citri* strains BLH-13, LB 319, and CC-2 which were selected to represent the biological diversity in this study. This comparison revealed high genome similarity, higher numbers of shared genes, and limited genome re-arrangements as observed in the center region (Fig. 4a). In contrast,



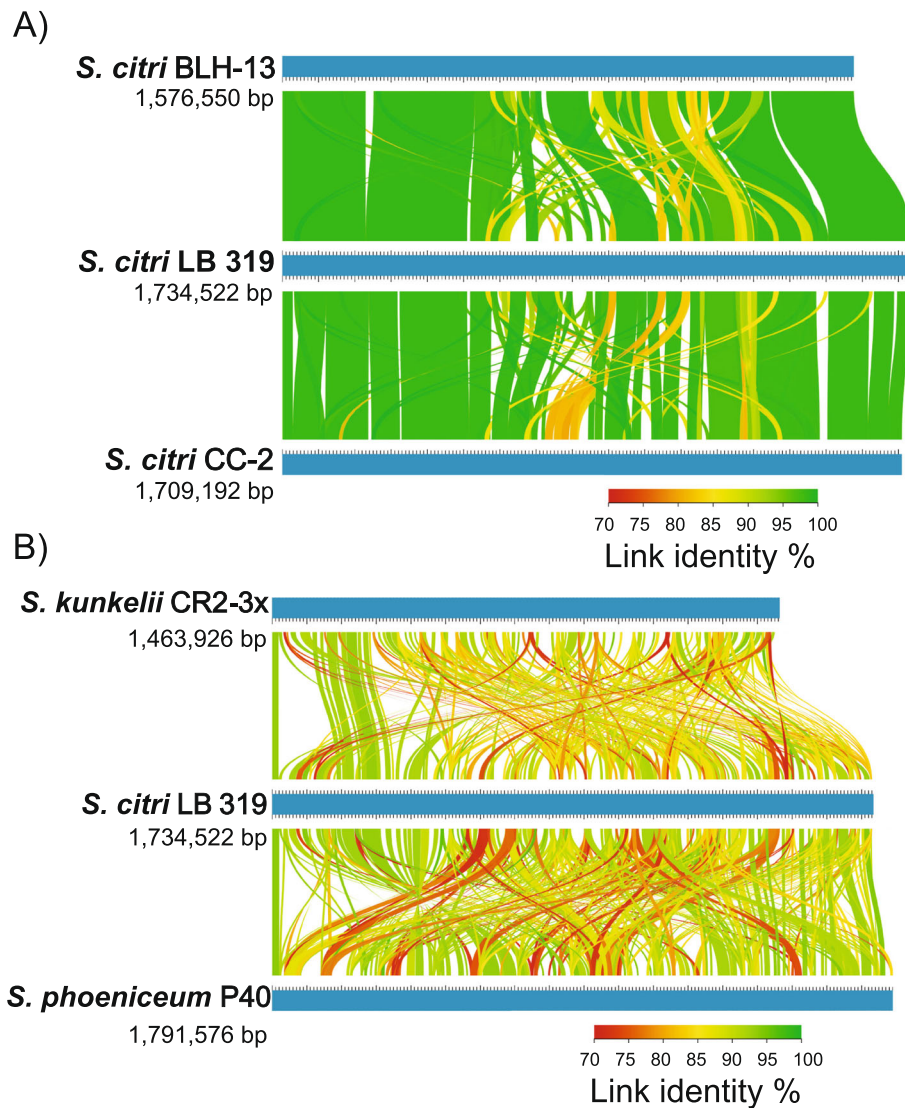
genome comparison among different species of plant pathogenic spiroplasmas were examined using LB 319 from citrus, *S. phoeniceum* P40 from periwinkle, and *S. kunkelii* CR2-3X from corn. Here, *S. citri*, *S. kunkelii*, and *S. phoeniceum* showed significant differences in gene content, low level of genome similarity, and extensive genome rearrangements. *S. kunkelii* and *S. phoeniceum* also exhibited fewer regions of genome similarity and extensive genomic rearrangements (Fig. 4b).

Homologous genes were also identified among LB 319, BLH-13, and CC-2 (Fig. 5a). These *S. citri* strains shared 990 core homologous gene clusters, with 42 and 43 homologous gene clusters specific to each of these strains. Among different plant pathogenic *Spiroplasma* spp., LB 319, *S. phoeniceum* P40, and *S. kunkelii* CR2-3X, shared 755 core homologous gene clusters (Fig. 5b). There were 201 to 424 homologous gene clusters specific to *S. phoeniceum* P40 and *S. kunkelii* CR2-3x, respectively. LB 319 shared 120 homologous gene clusters with *S. phoeniceum*, while sharing only 37 homologous isolated from *S. kunkelii*. Additionally, *S. phoeniceum* and *S. kunkelii* shared 271

homologous gene clusters that were absent from LB 319.

#### Functional assignment of *S. citri* LB 319 protein-coding sequences

Due to similarity of the chromosome of the seven *S. citri* strains studied, LB 319 was selected for further characterization. LB 319 had 1750 annotated protein coding sequences (CDS) and the functional classification of these protein coding sequences assigned only 553 CDS (32%) in different clusters of orthologous groups (COGs). The most abundant functional category was DNA replication, recombination and repair, followed by translation. These categories mainly consist of gDNA polymerases (*dnaE*, *dnaN*, *dnaX*, *holA*, *holB*, *polC*), nucleotide excision repair (*uvrA*, *uvrB*, *uvrC*), DNA topoisomerases (*gyrA*, *gyrB*, *parC*, *parE*), ribosomal proteins, and tRNA synthetases genes. Other important functional categories include translation (COG category K), nucleotide metabolism and transport (COG category F), and transcription (COG category O) (Fig. 6). Descriptive functional information of the genes is included in [Supplemental Table S3](#). The low number of assigned COGs



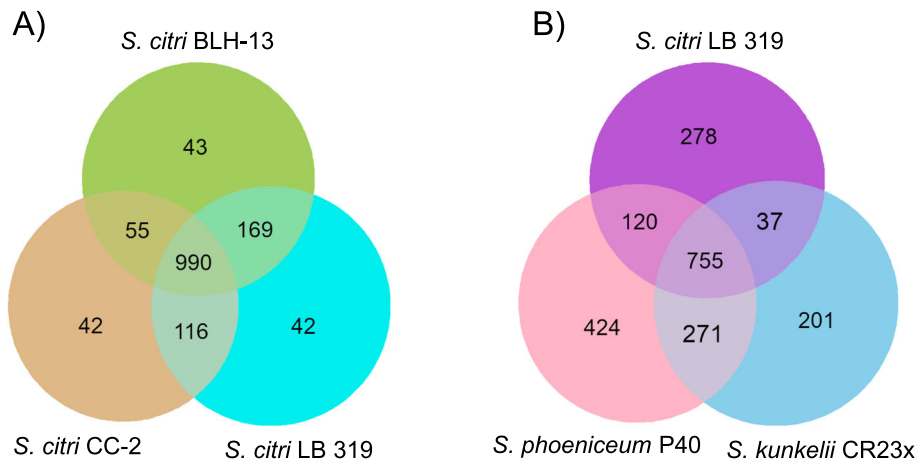
**Fig. 4** Genome-wide comparison of pathogenic *Spiroplasma* species. Linear chromosomal maps were built using AliTV v. 1.0 visualization software [18], based on whole-genome alignments with Lastz v. 1.0.4 aligner [19]. Both panels depict pairwise comparisons, expressed as percentage of nucleotide similarity, that connect different homologous genomic regions. Chromosomes are completely finished and pictured in blue. **a** *S. citri* strains BLH-13 isolated from the beet leafhopper (BLH), LB 319 isolated from citrus and CC-2 isolated from Chinese cabbage. **b** *S. citri* LB 319, *S. phoeniceum* P40 and *S. kunkelii* CR2-3X

suggests that a large proportion of them may be fragments of unrecognized pseudogenes. Genes involved in mismatch repair like *mutS*, *mutI*, *mutH*, *exoI*, *exoX*, *recI* and genes involved in homologous recombination like *recA*, *recB*, *recC* etc. are missing or truncated.

#### Plasmid variability between *S. citri* strains

Plasmids are genetic elements which may encode virulence factors and also play important roles in establishing host range [20–23]. Although a high level of similarity was found in the chromosome of *S. citri* strains, more diversity

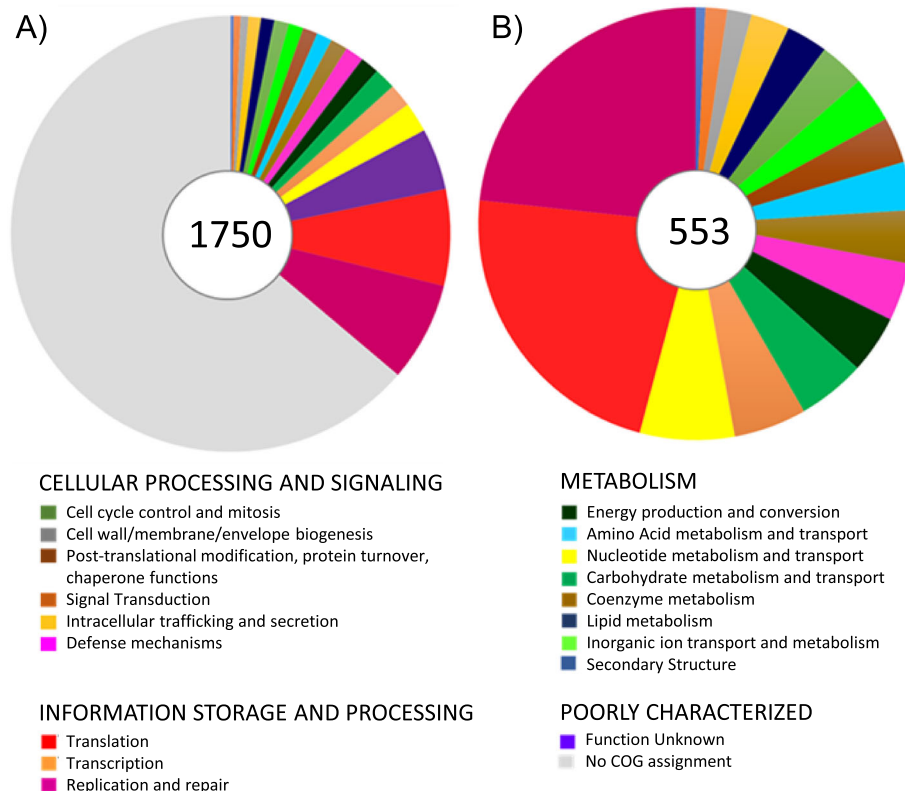
was found in the number of putative plasmids associated with these strains (Fig. 7a). For the sake of this study these putative plasmids are referred to as plasmids since the DNA in 25 of 32 plasmid-like contigs were circularized. Eight and nine plasmids were found in *S. citri* strains isolated from the BLH. Strains isolated from carrot and Chinese cabbage contained seven and three plasmids, respectively. *S. citri* strains isolated from citrus and horseradish possessed one to two plasmids. *S. citri* adhesion-related proteins (ScARPs), which are expected to be involved in *S. citri* adhesion to insect cells [25, 26], were



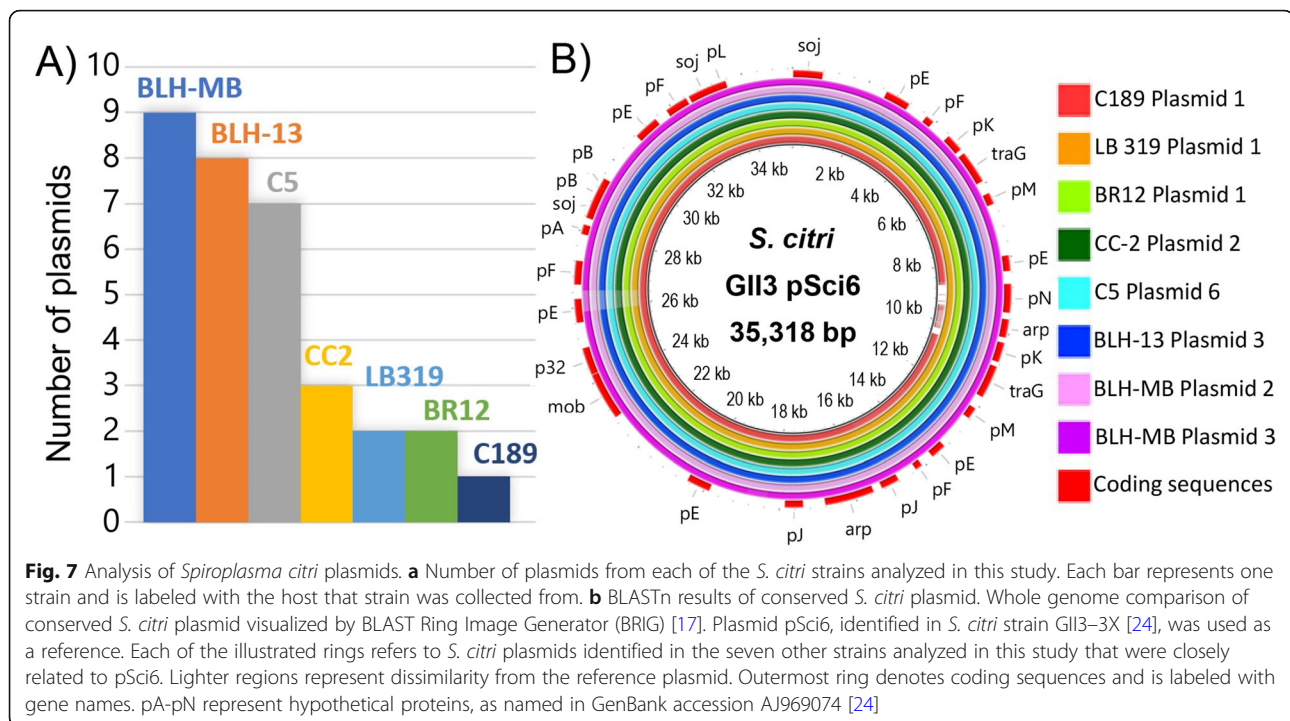
**Fig. 5** Numbers of shared and genome-specific homologous gene clusters. The Venn diagrams show the number of shared and genome-specific homologous gene clusters among the genomes compared. **a** Comparison among *Spiroplasma citri* BLH-13, CC-2, and LB 319. **b** Comparison among *S. citri* LB 319, *S. phoeniceum* P40, and *S. kunkelii* CR2-3X

predicted in several plasmids by NCBI’s Prokaryotic Genome Annotation Pipeline (PGAP). These ScARPs were present in one plasmid in BR12 and CC-2, two plasmids in C5 and BLH-MB, and three plasmids in BLH-13. No full-length ScARPs were predicted in C189 or LB 319

plasmids. *S. citri* strain C189, isolated from citrus in 1960, retained only one plasmid, pScp-C189–1. BLAST results revealed that this plasmid was highly similar to plasmid pSci6, identified in *S. citri* strain GII3–3X [24]. All strains analyzed in this study contained at least one plasmid with



**Fig. 6** Functional classification of protein-coding genes in *S. citri* LB 319. The functional categorization of each protein-coding gene was classified according to the COG assignments. The pie graph indicates the COG distribution statistics. Each color represents a COG functional description. **a** All annotated protein-coding genes in the *S. citri* LB 319 genome. **b** Protein-coding genes that could be assigned functional category



very high similarity to pSci6. (Fig. 7b). BLH-MB possessed two plasmids which resembled pSci6.

### Prophage prevalence in *S. citri* strains

The PHASTER (PHAge Search Tool – Enhanced Release) web server was used to identify and annotate putative prophage regions within the *S. citri* genomes. Seventeen to 25 prophage-like elements were identified in the *S. citri* genomes (Supplementary Fig. S3). Plasmid pScpLB319–2 in LB 319 and pScpC5–3 in C5 contained prophage sequences (Supplementary Fig. S3 B,E). Two plasmids in BLH-13 (pSciBLH13–1 and pSciBLH13–6) and three plasmids in BLH-MB (pSciBLHMB-1, pSciBLHMB-7, and pSciBLHMB-8) possessed prophage sequences (Supplementary Fig. S3 F,G). No prophage sequences were predicted in plasmids from C189, BR12, or CC-2. A large proportion of these prophage sequences were homologous to plectrovirus SpV1 [27] and SVTS2 [28] (Supplementary Table S4). While multiple plectrovirus SpV1 sequences were found to be distributed throughout the chromosome, a concentrated area of SVTS2 sequences was found at approximately 1.2 Mbp. This region of the chromosome was found to have highly repetitive sequences in all *S. citri* genomes in this study as shown by dot-matrix pairwise sequence comparisons (Fig. 3, Supplementary Fig. S2). Further characterization of the prophage region in the

chromosome and plasmids will be examined in the future.

### Putative virulence components in *S. citri*

*S. citri* does not have specialized protein secretion apparatus, such as the type II and type III secretion systems. *S. citri* utilizes a sec-dependent protein export pathway. In *S. citri* LB 319, *secY* (GMI18\_RS01645), *secA* (GMI18\_RS01115), *ftsY* (GMI18\_RS00765), *ffh* (GMI18\_RS02240), *secE* (GMI18\_RS01245), and *yidC* (GMI18\_RS10485) are the genes involved in the sec-dependent pathway and are conserved across *S. citri* strains. Several sequences in plasmid pScp-C189–1 were homologous to genes associated with type IV secretion systems, including Mob-like transmembrane proteins, TraG proteins, and proteins containing TraM recognition domains and type IV secretion system DNA-binding domains. This plasmid is conserved across *S. citri* strains. Fructose operon genes are major pathogenicity determinants in *S. citri* [29, 30]. *fruR* (GMI18\_RS08785), *fruA* (GMI18\_RS08780) and *fruK* (GMI18\_RS08775) are genes present in the fructose operon and were present in all the *S. citri* strains analyzed in this study. *fruR* codes for bacterial transcriptional regulators of carbohydrate catabolic operons. *fruA* codes for the permease of the phosphoenolpyruvate:fructose phosphotransferase system. This fructose permease allows uptake and concomitant phosphorylation of fructose into fructose-1-phosphate. *fruK* codes for a 1-



phosphofructokinase, which uses ATP to phosphorylate fructose-1-phosphate into fructose-1,6-bisphosphate [30].

## Discussion

Prior to advances in long-read sequencing, the best assembly of *S. citri* was strain GII3–3X, which contained 39 chromosomal contigs [24]. The first sequence of *S. citri* containing a single chromosomal contig was reported in 2017 for Moroccan *S. citri* strain R8-A2<sup>T</sup> [31]. Complete genomes, with single chromosomal contigs, of six more strains of *S. citri* isolated in the United States were reported in 2020 [14, 15] and a seventh strain, C5, isolated from carrot, is reported in this study. Chromosome sizes of these genomes were similar in size to strain R8-A2<sup>T</sup> (~1.6 Mbp), although most of the U.S. strains were slightly larger in size. Additionally, the number of predicted protein coding regions was higher in U.S. strains compared to the R8-A2<sup>T</sup> genome. One set of rRNA genes and 32 tRNA genes were predicted in U.S. strains, which is consistent with the R8-A2<sup>T</sup> strain [31].

Phylogenetic analyses between *S. citri* strains showed citrus strains C189 from southern California and R8-A2 from Morocco clustered together. C189 was originally obtained in 1957 from Washington Navel in Riverside, California by graft transmission to sweet orange seedling [4] and has been continuously maintained in the greenhouse. R8-A2 was originally obtained from Washington Navel in Morocco by graft transmission to sour orange [32]. Therefore, both strains were isolated and cultured from greenhouse citrus hosts many years after initially acquired and sequenced. Strains LB 319, BLH-13, BLH-MB, and BR12 clustered together. In these cases, *S. citri* was isolated and cultured directly from these field sources and sequenced. CC-2 and C5, which originated from annual hosts with similar row crop ecologies, clustered together in a separate clade. Moreover, the analysis of core orthologous genes suggested that strains from citrus were very closely related and BLH strains were more closely related to citrus strains than those from Chinese cabbage or carrot. *S. citri* strain GII3–3X, isolated from *C. haematopceps*, did not group with strains isolated from BLHs from California, but this may be due to the incomplete genome sequence of GII3–3X.

A high level of homology and synteny was observed between the *S. citri* strains in this study, with some dissimilar regions and genomic rearrangements appearing in the center region of the chromosome. However, comparison with the most closely related species, *S. phoeniceum* and *S. kunkelii*, showed that the chromosomal organization is largely rearranged and exhibits much lower levels of sequence similarity. *S. phoeniceum* is associated with lethal yellows in periwinkle [33], while *S. kunkelii* causes corn stunt disease of *Zea mays* L. [34].

Large rearrangements were also observed when comparing *S. citri* to *S. melliferum*, a honeybee pathogen [35]. When comparing gene content, *S. citri* strains shared approximately 80% of the homologous gene clusters observed, with about 3% of gene clusters being unique to each strain. However, *S. citri* strain LB 319 shared approximately 64% of gene clusters with *S. phoeniceum* and *S. kunkelii*, while about 23% of gene clusters were unique to *S. citri*. These differences may be caused by differential gene loss, phage-mediated horizontal gene acquisition, and by ecological and biological diversification [35, 36].

Moreover, some of this variation may be due to variation in prophage sequences [37], which are viral or phage genomic DNA sequences integrated into a bacterial genome. *S. citri* is highly susceptible to viral invasion, due to its lack of a cell wall [38]. Between seventeen and twenty-five areas of the genome were predicted to contain prophage insertions in the *S. citri* genomes studied which contributed to the variations in chromosome size. Most of these sequences observed in *S. citri* are homologous to SpV1, a plectrovirus, and were dispersed throughout the genome [37]. PHASTER analysis demonstrated that SpV1 was prevalent throughout the chromosome of all seven *S. citri* strains, but the number and positions were inconsistent (Supplemental Fig. S3). SpV1 viral sequences have been associated with major variations of the *S. citri* genome [39]. These viral sequences, integrated into the *Spiroplasma* chromosome, can have a large effect on genome stability. A model for the evolution of the *Spiroplasma* genome has been linked to viral invasion, which could account for intraspecific genome size variation, low conservation in chromosomal organization, and a gain of lineage-specific genes [36]. The rearrangements and genome instability are apparent in pairwise comparisons of *S. citri* strains, and even more so when compared to their closest relatives, *S. phoeniceum* and *S. kunkelii*. Viral invasions likely promoted these rearrangements in plant pathogenic bacteria and could attribute to their adaptation to different hosts. In contrast to SpV1, multiple copies of phage sequences homologous to SVTS2 were present in one specific, repetitive region in all the *S. citri* strains studied (Supplementary Fig. S3). Integration of SVTS2 viral sequences into the chromosome of *S. citri* has been associated with the resistance of *S. citri* to spiroplasma virus SVTS2 [40]. This may be why SVTS2 sequences are not prevalent throughout the genome and repeating elements containing SVTS2 sequences are conserved across strains.

*S. citri* does not have specialized protein secretion apparatus like the Type II and Type III secretion systems, but instead utilizes the sec-dependent protein export pathways. *S. citri* does not have T4SS. Components of

the T4SS are present and highly conserved in *S. citri* strains analyzed in this study. However, their importance in virulence has yet to be verified. T4SSs are related to bacterial conjugation systems and secrete substrates into a wide range of target cells, including other bacteria and eukaryotic cells. T4SSs are known to play a role in pathogenesis in a wide range of bacteria by genetic exchange and delivery of effector molecules to target cells [41].

Fructose operon genes are major virulence components in *S. citri*, which utilizes fructose for pathogenicity and growth in plants [29]. The fructose operon consists of three genes: *fruR*, *fruA*, and *fruK*. *S. citri* mutants of *fruR*, which likely encodes for the putative regulator protein of the fructose operon, abolished expression of all three genes of the operon. This mutant exhibited severely impaired pathogenicity, which could be restored when the mutant was complemented with functional fructose operon genes [30]. This work proposed an explanation of the role of fructose utilization in the pathogenicity of *S. citri*, suggesting that companion cells in the plant host compete for fructose. *S. citri* uses fructose as a carbon and energy source, resulting in a reduced fructose concentration in the plant companion cells, modifying the distribution of photoassimilates, leading to disease symptoms [30].

Every *S. citri* strain studied had a unique pattern of extrachromosomal DNA and the presence or absence, position (free or integrated), and number of these sequences is a significant source of variation among strains [42]. After several years of maintenance in plants, *S. citri* strain BR3-G showed chromosomal rearrangements compared to strain BR3-T, which was transmitted from plant to plant by the BLH [39]. Strain BR3-G was found to be non-transmissible by the BLH which was correlated to a large deletion of a SpV1-related transposase gene [39]. Prolonged cultivation of bacteria has been reported to cause free plasmid DNA to be integrated into the chromosome through recombination events [43–45]. Strains isolated from BLH examined in this study contained two or three plasmids with predicted prophage sequences. The strains isolated from plants exhibited one or no plasmids with predicted prophage sequences. The plasmids that contain viral sequences were homologous to those in the chromosome of *S. citri* strain R8-A2. It is plausible that after transmission by the BLH to plants, plasmids containing viral sequences could be incorporated by recombination with chromosomal plectroviral sequences. This would result in fewer plasmids in perennial plants compared to annual plants, which was observed in this study.

Plasmids of phytoplasmas and spiroplasmas are known to be involved in insect transmissibility [46–49]. All the newly sequenced *S. citri* strains contained at least one plasmid with high homology to plasmid pSci6, identified

in *S. citri* strain GII3–3X [24]. pSci6 plasmid encodes P32 protein, associated with insect transmissibility, and this plasmid confers insect transmissibility into non-transmissible strains of *S. citri* [50]. P32 has been suggested to interact with surface membrane proteins and may be necessary, but not sufficient for spiroplasma adhesion and invasion of insect cells [51]. In strain GII-3, eight proteins belonging to the ScARP protein family, which are expected to be involved in *S. citri* adhesion to insect cells, were encoded by five plasmids [24–26]. In previous studies, a loss of the high-molecular-mass plasmids carrying ScARP genes was correlated with a non-transmissible phenotype [51]. ScARP genes were found to be present in plasmids isolated from perennial crops and the leafhopper vector in this study. These plasmids are highly similar to pSci2 and pSci5 from *S. citri* strain GII3–3X and pBJS-O from *S. citri* strain BR3–3X [24, 52] (Supplementary Table S1). Interestingly, no full-length ScARPs were predicted in C189 or LB 319 plasmids. The lack of additional plasmids containing ScARP genes, seen in strains obtained from perennial crops, suggest these strains could have lost their ability to be transmitted by the BLH.

Several other genes have been assigned putative functions in plasmid pSci6. This included *soj-parA*, which is involved in DNA partitioning, and *traG* and *mob*, which are associated with DNA transfer and are suggested to be involved in a Type IV secretory pathway [24]. A coding region was also identified in plasmid pScp-C189–1, which has homology to a plasmid replication-relaxation (relaxase) family protein. Plasmids in *S. citri* have been suggested to be horizontally transferred between cells by conjugation [24]. Several important genes are at the core of plasmid conjugation, including Type IV coupling proteins and relaxases [53]. The *traG* and *mob* genes found in pSci6 correspond to VirB4/D4 components of the type IV secretion pathway, which allows for the translocation of DNA through cytoplasmic membranes [54]. Walled bacteria require many components in their conjugation system; however, these components may not be necessary in *S. citri*, a wall-less bacterium. In addition to *traG* and *mob*, Saillard et al. suggested that pSci6 should contain a relaxase, but this family of proteins was not reported [24]. The replication-relaxation (relaxase) family protein is essential for plasmid replication and plasmid DNA relaxation, part of conjugative DNA transfer in bacteria [55–57]. A BLAST search of the coding sequences in pScp-C189–1 revealed homology to a replication-relaxation family protein. This further supports previous studies that reported genetic exchanges by a conjugation-like process in *S. citri* [58]. The genes identified in this plasmid encode for virulence-associated proteins

involved in adhesion and conjugal DNA transfer. The occurrence of this persistent, conserved plasmid suggests it plays an important role in this pathogen.

## Conclusion

Six *Spiroplasma citri* genomes were published recently, but were not fully analyzed [14, 15]. Those sequences, along with C5, a strain de novo assembled in this study, greatly expanded the availability of *S. citri* genomes and allowed performance of extensive in silico comparative genomic studies that provide insights into this organism's genetic diversity and evolution. An extremely high level of homogeneity was observed in the chromosomal contigs across *S. citri* strains. Variation in plasmid number may play an important role in insect transmission and virulence. Moreover, differences in genome size and stability appear to result from variations in number and site of plectroviral sequences inserted into the genome. These features likely contribute to *S. citri* adaptation to different hosts and transmissibility by leafhopper vectors. Further studies will be necessary to validate the roles of plasmids and viral insertion sequences in *S. citri* strains isolated from various hosts.

## Methods

### Strain isolation and DNA preparation

Cultures of *S. citri* strains CC-2, C189, BR-12, LB 319, BLH-13, and BLH-MB were isolated, and DNA extracted in a previous study [14, 15]. *S. citri* strain C5 was collected in 2005 from carrot growing in NW Bakersfield, California. Briefly, *S. citri* was isolated and grown in LD8 medium [59]. Later the cultures were triple cloned and stored at  $-80^{\circ}\text{C}$  until further use. Cultures were re-established and total genomic DNA was extracted by CTAB [60] or by DNeasy Blood and Tissue Extraction kit (Qiagen). Collection details of *S. citri* in this study are listed in Table 1.

### Whole-genome shotgun sequencing

*S. citri* strain C5 was sequenced on the Oxford Nanopore (Oxford, United Kingdom) MinION platform [61]. The library was prepared with Oxford Nanopore (Oxford, United Kingdom) Rapid Barcoding Kit-SQK-RBK004 according to the manufacturer's specifications. Data was collected using MinION Release 19.10.1. Bases were called using Guppy v.3.4.5 and the adapter screening and quality filtering of raw sequencing data were performed using Fastp v 0.20.0 [62]. Remaining *S. citri* strains' genomes were sequenced previously on PacBio (Menlo Park, CA, USA) RS II platform [14, 15].

### De novo sequencing assembly

Sequences from *S. citri* strains CC-2, C189, BR-12, LB 319, BLH-13, and BLH-MB were reported previously [14, 15]. Briefly, raw reads were filtered, subreads were established

by PacBio, and assembled into contigs using Canu 1.8 [63]. For *S. citri* strain C5, contigs were assembled using Canu 1.8. An additional polishing step was performed by medaka v 1.0.3 (Oxford Nanopore Technologies) and frame-shift-corrected by DIAMOND v 0.9.28 [64] and MEGAN v 6.18.4 [65], following the pipeline described by Arumugam et al. [66]. Approximately 500 bp segments from each end of a contig were used for BLASTn search to check the contig singularity. Appropriate reads connecting both ends were used for enclosure. The chromosome and plasmid status of each contig were further confirmed by BLASTn analyses against the GenBank database. The genome sequence data was deposited in the NCBI database (accession numbers shown in Table 2). Annotation of each contig was performed by the NCBI Prokaryotic Genome Annotation Pipeline (PGAP) [67].

### Phylogenetic analyses

The 16S rRNA sequences of thirty-nine *Spiroplasma* species were obtained from the NCBI database. Sequence alignments were carried out with the PRANK alignment tool [68]. Maximum-likelihood approach was used to reconstruct the phylogenetic tree using RAxML software [69]. Bootstrapping was performed with 1000 replicates. The resulting phylogeny was visualized with FigTree [16].

Orthologous genes of *S. citri* isolates were predicted using the OrthoMCL v. 2.0 pipeline [70]. All-versus-all BLASTN ( $E$  value  $< 10^{-5}$ , alignment coverage  $> 50\%$ ) comparison of all gene sequences for each species was performed and orthologous genes were clustered by OrthoMCL v. 2.0. Multiple sequence alignment was done with PRANK v. 1.70.427 [71]. The sequence alignments were concatenated by FASconCAT v. 1.1, yielding a gene super-matrix [72]. Maximum-likelihood approach was used to reconstruct the phylogenetic tree using RAxML v. 8.2 software with 1000 bootstrap replicates [69]. The resulting phylogeny was visualized using FigTree v. 1.4.3 [16].

### Bioinformatics analysis

Large genome comparison of eight *S. citri* sequences was computed and visualized with the use of BLAST Ring Image Generator (BRIG) v 0.95 [17]. Pairwise genome alignment was achieved by the Lastz v. 1.04 program [19]. The results were visualized using AliTV v. 1.0 [18]. Shared and genome-specific genes were identified between the *S. citri* strains isolated from different sources and among *S. citri* LB 319, *S. kunkelii* CR2-3X, and *S. phoeniceum* P40. The sequence similarity search step in the OrthoMCL analysis was conducted at the nucleotide level [70]. Functional annotation of COG was done using eggNOG-mapper [73]. Prophage sequences were predicted using PHASTER online server [74].

## Abbreviations

PGAP: Prokaryotic Genome Annotation Pipeline; CSD: citrus stubborn disease; CC: Chinese cabbage; LB: Liberty Bell; BR: brittle root; BLH: beet leafhopper; BRIG: BLAST Ring Image Generator; COG: cluster of orthologs groups; ScARPs: *S. citri* adhesion-related proteins; PHASTER: PHAge Search Tool – Enhanced Release

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12864-021-07637-8>.

**Additional file 1: Supplementary Table S1.** BLASTn similarity search of *Spiroplasma citri* putative plasmid (contigs) against available *Spiroplasma* sequences deposited in NCBI database.

**Additional file 2: Supplementary Table S2.** GenBank accession numbers for the 16S rRNA *Spiroplasma* sequences used for phylogenetic analyses in Fig. 1.

**Additional file 3: Supplementary Table S3.** COG functional categories of *Spiroplasma citri* LB319

**Additional file 4: Supplementary Table S4.** Summary of PHAge Search Tool Enhanced Release (PHASTER) analysis of *Spiroplasma citri* strains

**Additional file 5: Supplementary Fig. S1.** Phylogenetic analysis of *Spiroplasma* strains. Maximum-likelihood phylogeny of *Spiroplasma* based on core orthologous genes. In total, 601 orthologous genes were concatenated, and a maximum-likelihood approach was used to generate the phylogeny with 1000 bootstrap replicates. Bootstrap values are indicated at each node. The resulting phylogeny was visualized using FigTree v.1.4.3 [70]. *S. citri* strains analyzed in this report are underlined.

**Additional file 6: Supplementary Fig. S2.** Dot-Matrix representation of a BLASTn comparison of complete *Spiroplasma citri* genome sequences. The chromosome of each newly assembled strain (X-axis) was compared to the reference sequence, strain R8-A2 (Y-axis). A) C189; B) LB 319; C) BR12; D) CC-2; E) C5; F) BLH-13; G) BLH-MB.

**Additional file 7: Supplementary Fig. S3.** Predicted prophage sequences in *Spiroplasma citri* genomes. Prophage annotation was performed by PHASTER (PHAge Search Tool – Enhanced Release). The program predicts the completeness of the predicted prophages, based on the proportion of phage genes in the identified region. Green bars are scored as intact (score > 90); blue bars are questionable (score 70–90); red bars are incomplete (score < 70). Strains with plasmids predicted to contain phage genes are included (Fig. S3B, E, F, and G). Approximate location of repetitive region identified by dot-matrix pairwise sequence comparisons (Supplementary Fig. S2) is highlighted in yellow.

## Acknowledgements

We thank Robert DeBorde of the United States Department of Agriculture-Agricultural Research Service, San Joaquin Valley Agricultural Sciences Center, Parlier, CA for technical assistance. Jacqueline Fletcher and Kitty Cardwell, Entomology & Plant Pathology Department, Institute of Biosecurity and Microbial Forensics, Oklahoma State University, Stillwater, Oklahoma, are gratefully acknowledged for providing cultures of *S. citri* strain BR12 and C5. We acknowledge the dedication of the past and current Citrus Clonal Protection Program (CCPP) personnel at the University of California, Riverside for maintaining *in planta* the originally cultured *S. citri* isolate C189 since 1957 [4]. Mention of trade names or commercial products in this publication is solely for providing specific information and does not imply recommendation or endorsement by the USDA. USDA is an equal opportunity provider and employer.

## Authors' contributions

RKY, GV, FO and DP conceptualized the study and acquired the *S. citri* cultures. Funds for this research were from grants to RKY and GV. DP, YM, and VS cultured the pathogens and purified the DNA. FO supervised the PacBio submissions. AE and TD cultured, purified DNA, and provided raw sequence data for *S. citri* strain C5, JC provided technical assistance for genome assembly and bioinformatics, SH provided some technical

assistance. RR and ST conducted data curation and analysis. RR, ST and RY wrote the manuscript and DP, SH, JC, GC, and GV edited the manuscript. All author(s) read and approved the final manuscript.

## Funding

This work was supported by USDA 2034–22000-013–10R and two grants from the Citrus Research Board: Project 5300–191 to R. Yokomi; and Project 5100–153 to G. Vidalakis. G. Coaker and S. Thapa were supported by grants from the USDA (2019–70016-29796, 2016–70016-24833). Additional support for this work was provided in part by the Citrus Research Board (Project 6100 and 5300–179), the USDA National Institute of Food and Agriculture, Hatch Project 1020106 and the National Clean Plant Network (AP17PPQS&T00C118 and AP18PPQS&T00C107). The sequencing was carried out at the DNA Technologies and Expression Analysis Cores at the UC Davis Genome Center, supported by NIH Shared Instrumentation Grant 1S10OD010786–01. The funding bodies played no role in the design of the study and collection, analysis or interpretation of data and writing of the manuscript.

## Availability of data and materials

The data described in this study can be freely and openly accessed from the NCBI database. All sequence data has been deposited in GenBank under the accession numbers CP042472–CP042475, CP0426368–CP046373, CP047426–CP047446, and CP053304–CP053311. The versions described in this paper are the first versions. PacBio sequencing reads for BioProject PRJNA558054 have been deposited in the NCBI Sequence Read Archive (SRA) under accession number SRR9903453 for CC-2. PacBio sequencing reads for BioProject PRJNA591027 have been deposited in the NCBI SRA under accession numbers SRR10843927 for C189; SRR10507068 for LB 319; SRR10507067 for BR12; SRR10507065 for BLH-13; and SRR10507066 for BLH-MB. Nanopore sequencing reads for BioProject PRJNA625113 have been deposited in the NCBI SRA under accession number SRR11536473 for C5. *S. citri* cultures for BLH-13, BLH-MB, BR12, LB 319, and CC-2 have been deposited in ATCC no. SD-7532-7535 and SD-7278, respectively. However, ATCC Biorepository Service Agreement 2019-BRS-00049 maintains confidentiality of information regarding these cultures. C189 and C5 cultures are available from the corresponding author.

## Declarations

### Ethics approval and consent to participate

Not applicable.

### Consent for publication

Not applicable.

### Competing interests

The authors declare that they have no competing interests.

## Author details

<sup>1</sup>Crop Diseases, Pests, and Genetics Research Unit, San Joaquin Valley Agricultural Sciences Center, USDA Agricultural Research Service, Parlier, CA 93648, USA. <sup>2</sup>Department of Plant Pathology, University of California, Davis, CA 95616, USA. <sup>3</sup>Department of Microbiology and Plant Pathology, University of California, Riverside, CA 92521, USA. <sup>4</sup>Department of Entomology & Plant Pathology and Institute of Biosecurity and Microbial Forensics, Oklahoma State University, Stillwater, OK 74078, USA. <sup>5</sup>Citrus Pest Detection Program, Central California Tristeza Eradication Agency, Tulare, CA 93274, USA.

Received: 2 September 2020 Accepted: 21 April 2021

Published online: 22 May 2021

## References

- Whitcomb RF. The mycoplasmas v3: plant and insects mycoplasmas. Amsterdam: Elsevier; 2012.
- Gasparich GE. Spiroplasmas: evolution, adaptation and diversity. Front Biosci. 2002;7:d619–40.
- Regassa LB, Gasparich GE. Spiroplasmas: evolutionary relationships and biodiversity. Front Biosci. 2006;11(1):2983–3002. <https://doi.org/10.2741/2027>.
- Fudl-Allah AESA, Calavan EC, Igwegbe ECK. Culture of a mycoplasma-like organism associated with stubborn disease of citrus. Phytopathology. 1972; 62(7):729–31. <https://doi.org/10.1094/Phyto-62-729>.

5. Fletcher J. Brittle root disease of horseradish: evidence for an etiological role of *Spiroplasma citri*. *Phytopathology*. 1981;71(10):1073–80. <https://doi.org/10.1094/Phyto-71-1073>.
6. Lee I-M, Bottner KD, Munyaneza JE, Davis RE, Crosslin JM, du Toit LJ, et al. Carrot purple leaf: a new spiroplasma disease associated with carrots in Washington state. *Plant Dis*. 2006;90(8):989–93. <https://doi.org/10.1094/PD-90-0989>.
7. Chen TA, Liao CH. Corn stunt spiroplasma: isolation, cultivation, and proof of pathogenicity. *Science*. 1975;188(4192):1015–7. <https://doi.org/10.1126/science.188.4192.1015>.
8. Saillard C, Vignault JC, Bove JM, Raie A, Tully JG, Williamson DL, et al. *Spiroplasma phoeniceum* sp. nov., a new plant-pathogenic species from Syria. *Int J Syst Bacteriol*. 1987;37(2):106–15. <https://doi.org/10.1099/00207713-37-2-106>.
9. Liu HY, Gumpf DJ, Oldfield GN, Calavan EC. The relationship of *Spiroplasma citri* and *Circulifer tenellus*. *Phytopathology*. 1983;73(4):585–90. <https://doi.org/10.1094/Phyto-73-585>.
10. Fos A, Bové JM, Lallemand J, Saillard C, Vignault JC, Ali Y, et al. The leafhopper *Neolittorae haematoceps* (Mulsant & Rey) is a vector of *Spiroplasma citri* in the Mediterranean. *Ann Inst Pasteur Microbiol*. 1986; 137A(1):97–107. [https://doi.org/10.1016/s0769-2609\(86\)80009-6](https://doi.org/10.1016/s0769-2609(86)80009-6).
11. Brown DR, Whitcomb RF, Bradbury JM. Revised minimal standards for description of new species of the class Mollicutes (division Tenericutes). *Int J Syst Evol Microbiol*. 2007;57(11):2703–19. <https://doi.org/10.1099/ijs.0.64722-0>.
12. Didelot X, Bowden R, Wilson DJ, Peto TEA, Crook DW. Transforming clinical microbiology with bacterial genome sequencing. *Nat Rev Genet*. 2012;13(9): 601–12. <https://doi.org/10.1038/nrg3226>.
13. Amarasinghe SL, Su S, Dong X, Zappia L, Ritchie ME, Gouil Q. Opportunities and challenges in long-read sequencing data analysis. *Genome Biol*. 2020; 21:1–16.
14. Yokomi R, Rattner R, Osman F, Maheshwari Y, Selvaraj V, Pagliaccia D, et al. Whole genome sequence of five strains of *Spiroplasma citri* isolated from different host plants and its leafhopper vector. *BMC Res Notes*. 2020;13:1–4.
15. Yokomi R, Chen J, Rattner R, Selvaraj V, Maheshwari Y, Osman F, et al. Genome sequence resource for *Spiroplasma citri*, strain CC-2, associated with citrus stubborn disease in California. *Phytopathology*. 2020;110(2):254–6. <https://doi.org/10.1094/PHYTO-08-19-0304-A>.
16. Rambaut A. FigTree v1. 4. 2012. <http://tree.bio.ed.ac.uk/software/figtree/>.
17. Alikhan N-F, Petty NK, Ben Zakour NL, Beatson SA. BLAST ring image generator (BRIG): simple prokaryote genome comparisons. *BMC Genomics*. 2011;12(1):402. <https://doi.org/10.1186/1471-2164-12-402>.
18. Ankenbrand MJ, Hohlfeld S, Hackl T, Förster F. AliTV—interactive visualization of whole genome comparisons. *PeerJ Comput Sci*. 2017;3:e116. <https://doi.org/10.7717/peerj-cs.116>.
19. Harris RS. Improved pairwise alignment of genomic DNA. Ph.D. Dissertation: Pennsylvania State University; 2007.
20. Elwell LP, Shipley PL. Plasmid-mediated factors associated with virulence of bacteria to animals. *Annu Rev Microbiol*. 1980;34(1):465–96. <https://doi.org/10.1146/annurev.mi.34.100180.002341>.
21. Diard M, Hardt W-D. Evolution of bacterial virulence. *FEMS Microbiol Rev*. 2017;41(5):679–97. <https://doi.org/10.1093/femsre/fux023>.
22. Klümper U, Riber L, Dechesne A, Sannazzarro A, Hansen LH, Sørensen SJ, et al. Broad host range plasmids can invade an unexpectedly diverse fraction of a soil bacterial community. *ISME J*. 2015;9(4):934–45. <https://doi.org/10.1038/ismej.2014.191>.
23. Jain A, Srivastava P. Broad host range plasmids. *FEMS Microbiol Lett*. 2013; 348(2):87–96. <https://doi.org/10.1111/1574-6968.12241>.
24. Saillard C, Carle P, Duret-Nurbel S, Henri R, Killiny N, Carrère S, et al. The abundant extrachromosomal DNA content of the *Spiroplasma citri* G13-3X genome. *BMC Genomics*. 2008;9:1–13.
25. Yu J, Wayadande AC, Fletcher J. *Spiroplasma citri* surface protein P89 implicated in adhesion to cells of the vector *Circulifer tenellus*. *Phytopathology*. 2000;90(7):716–22. <https://doi.org/10.1094/PHYTO.2000.90.7.716>.
26. Berg M, Melcher U, Fletcher J. Characterization of *Spiroplasma citri* adhesion related protein SARP1, which contains a domain of a novel family designated sarpin. *Gene*. 2001;275(1):57–64. [https://doi.org/10.1016/S0378-1119\(01\)00655-2](https://doi.org/10.1016/S0378-1119(01)00655-2).
27. Renaudin J, Bové JM. SpV1 and SpV4, spiroplasma viruses with circular, single-stranded DNA genomes, and their contribution to the molecular biology of spiroplasmas. *Adv Virus Res*. 1994;44:429–63. [https://doi.org/10.1016/S0065-3527\(08\)60335-8](https://doi.org/10.1016/S0065-3527(08)60335-8).
28. Sha Y, Melcher U, Davis RE, Fletcher J. Common elements of spiroplasma plectroviruses revealed by nucleotide sequence of SVTS2. *Virus Genes*. 2000; 20(1):47–56. <https://doi.org/10.1023/A:1008108106916>.
29. Dubrana MP, Béven L, Arricau-Bouvery N, Duret S, Claverol S, Renaudin J, et al. Differential expression of *Spiroplasma citri* surface protein genes in the plant and insect hosts. *BMC Microbiol*. 2016;16(1):53. <https://doi.org/10.1186/s12866-016-0666-y>.
30. Gaurivaud P, Danet JL, Laigret F, Garnier M, Bové JM. Fructose utilization and virulence of *Spiroplasma citri*. *Mol Plant Microbe In*. 2000; 13(10):1145–55. <https://doi.org/10.1094/MPMI.2000.13.10.1145>.
31. Davis RE, Shao J, Zhao Y, Gasparich GE, Gaynor BJ, Donofrio N. Complete genome sequence of *Spiroplasma citri* strain R8-A2T, causal agent of stubborn disease in Citrus species. *Genome Announc*. 2017;5:e00206–17.
32. Saglio P, L'Hospital M, Lalleche D, et al. *Spiroplasma citri* gen. and sp. n.: a mycoplasma-like organism associated with “stubborn” disease of citrus. *Int J Syst Evol Microbiol*. 1972;23(3):191–204.
33. Nejat N, Vadamalai G, Sijam K, Dickinson M. First report of *Spiroplasma citri* (–induced) associated with periwinkle lethal yellows in Southeast Asia. *Plant Dis*. 2011;95(10):1312. <https://doi.org/10.1094/PDIS-03-11-0251>.
34. Whitcomb RF, Chen TA, Williamson DL, Liao C, Tully JG, Bove JM, et al. *Spiroplasma kunkelii* sp. nov.: characterization of the etiological agent of corn stunt disease. *Int J Syst Bacteriol*. 1986;36(2):170–8. <https://doi.org/10.1099/00207713-36-2-170>.
35. Lo W-S, Chen L-L, Chung W-C, Gasparich GE, Kuo C-H. Comparative genome analysis of *Spiroplasma melliferum* IPMB4A, a honeybee-associated bacterium. *BMC Genomics*. 2013;14(1):22. <https://doi.org/10.1186/1471-2164-14-22>.
36. Ku C, Lo W-S, Chen L-L, Kuo C-H. Complete genomes of two dipteran-associated spiroplasmas provided insights into the origin, dynamics, and impacts of viral invasion in spiroplasma. *Genome Biol Evol*. 2013;5(6):1151–64. <https://doi.org/10.1093/gbe/evt084>.
37. Melcher U, Sha Y, Ye F, Fletcher J. Mechanisms of *Spiroplasma* genome variation associated with SpVI-like viral DNA inferred from sequence comparisons. *Microb Comp Genomics*. 1999;4(1):29–46. <https://doi.org/10.1089/omi.1.1999.4.29>.
38. Rakonjac J, Bennett NJ, Spagnuolo J, Gagic D, Russel M. Filamentous bacteriophage: biology, phage display and nanotechnology applications. *Curr Issues Mol Biol*. 2011;13(2):51–76.
39. Ye F, Melcher U, Rascoe JE, Fletcher J. Extensive chromosome aberrations in *Spiroplasma citri* strain BR3. *Biochem Genet*. 1996;34(7-8):269–86. <https://doi.org/10.1007/BF02399947>.
40. Sha Y, Melcher U, Davis RE, Fletcher J. Resistance of *Spiroplasma citri* lines to the virus SVTS2 is associated with integration of viral DNA sequel into host chromosomal and extrachromosomal DNA. *Appl Environ Microbiol*. 1995; 61(11):3950–9. <https://doi.org/10.1128/AEM.61.11.3950-3959.1995>.
41. Cascales E, Christie PJ. The versatile bacterial type IV secretion systems. *Nat Rev Microbiol*. 2003;1(2):137–49. <https://doi.org/10.1038/nrmicro753>.
42. Melcher U, Fletcher J. Genetic variation in *Spiroplasma citri*. *Eur J Plant Pathol*. 1999;105(6):519–33. <https://doi.org/10.1023/A:1008757716803>.
43. Nur I, LeBlanc DJ, Tully JG. Short, interspersed, and repetitive DNA sequences in *Spiroplasma* species. *Plasmid*. 1987;17(2):110–6. [https://doi.org/10.1016/0147-619X\(87\)90017-5](https://doi.org/10.1016/0147-619X(87)90017-5).
44. Biškup UG, Strle F, Ružič-Sabljić E. Loss of plasmids of *Borrelia burgdorferi* sensu lato during prolonged in vitro cultivation. *Plasmid*. 2011;66(1):1–6. <https://doi.org/10.1016/j.plasmid.2011.02.006>.
45. Campbell AM. Chromosomal insertion sites for phages and plasmids. *J Bacteriol*. 1992;174(23):7495–9. <https://doi.org/10.1128/JB.174.23.7495-7499.1992>.
46. Nishigawa H, Oshima K, Kakizawa S, Jung H-Y, Kuboyama T, Miyata S, et al. A plasmid from a non-insect-transmissible line of a phytoplasma lacks two open reading frames that exist in the plasmid from the wild-type line. *Gene*. 2002;298(2):195–201. [https://doi.org/10.1016/S0378-1119\(02\)00975-7](https://doi.org/10.1016/S0378-1119(02)00975-7).
47. Oshima K, Shiomi T, Kuboyama T, Sawayanagi T, Nishigawa H, Kakizawa S, et al. Isolation and characterization of derivative lines of the onion yellows phytoplasma that do not cause stunting or phloem hyperplasia. *Phytopathology*. 2001;91(11):1024–9. <https://doi.org/10.1094/PHYTO.2001.91.11.1024>.
48. Bai X, Fazzolari T, Hogenhout SA. Identification and characterization of traE genes of *Spiroplasma kunkelii*. *Gene*. 2004;336(1):81–91. <https://doi.org/10.1016/j.gene.2004.03.022>.
49. Davis RE, Dally EL, Jomantiene R, Zhao Y, Roe B, Lin S, et al. Cryptic plasmid pSKU146 from the wall-less plant pathogen *Spiroplasma kunkelii* encodes an

- adhesin and components of a type IV translocation-related conjugation system. *Plasmid*. 2005;53(2):179–90. <https://doi.org/10.1016/j.plasmid.2004.09.002>.
50. Berho N, Duret S, Renaudin J. Absence of plasmids encoding adhesion-related proteins in non-insect-transmissible strains of *Spiroplasma citri*. *Microbiology*. 2006;152(3):873–86. <https://doi.org/10.1099/mic.0.28541-0>.
  51. Killiny N, Batailler B, Foissac X, Saillard C. Identification of a *Spiroplasma citri* hydrophilic protein associated with insect transmissibility. *Microbiology*. 2006;152(4):1221–30. <https://doi.org/10.1099/mic.0.28602-0>.
  52. Joshi BD, Berg M, Rogers J, Fletcher J, Melcher U. Sequence comparisons of plasmids pBJS-O of *Spiroplasma citri* and pSKU146 of *S. kunkelii*: implications for plasmid evolution. *BMC Genomics*. 2005;6(1):175. <https://doi.org/10.1186/1471-2164-6-175>.
  53. Smillie C, Garcillán-Barcia MP, Francia MV, Rocha EPC, de la Cruz F. Mobility of plasmids. *Microbiol Mol Biol Rev*. 2010;74(3):434–52. <https://doi.org/10.1128/MMBR.00020-10>.
  54. Chen I, Christie PJ, Dubnau D. The ins and outs of DNA transfer in bacteria. *Science*. 2005;310(5753):1456–60. <https://doi.org/10.1126/science.1114021>.
  55. Zechner EL, de la Cruz F, Eisenbrandt R, Grahn AM, Koraimann G, Lanka E, et al. Conjugative-DNA transfer processes. In: Thomas CM, editor. *Horizontal gene pool*. Amsterdam: Harwood Academic Publishers; 2000. p. 93–180.
  56. Francia MV, Varsaki A, Garcillán-Barcia MP, Latorre A, Drainas C, de la Cruz F. A classification scheme for mobilization regions of bacterial plasmids. *FEMS Microbiol Rev*. 2004;28(1):79–100. <https://doi.org/10.1016/j.femsre.2003.09.001>.
  57. Garcillán-Barcia MP, Francia MV, de la Cruz F. The diversity of conjugative relaxases and its application in plasmid classification. *FEMS Microbiol Rev*. 2009;33(3):657–87. <https://doi.org/10.1111/j.1574-6976.2009.00168.x>.
  58. Barroso G, Labarere J. Chromosomal gene transfer in *Spiroplasma citri*. *Science*. 1988;241(4868):959–61. <https://doi.org/10.1126/science.3261453>.
  59. Lee I-M. New media for rapid growth of *Spiroplasma citri* and corn stunt spiroplasma. *Phytopathology*. 1984;74(1):84–9. <https://doi.org/10.1094/Phyto-74-84>.
  60. Doyle J. DNA protocols for plants. In: Hewitt GM, Johnston AWB, Young JPW, editors. *Molecular techniques in taxonomy*. Berlin, Heidelberg: Springer; 1991. p. 283–93. [https://doi.org/10.1007/978-3-642-83962-7\\_18](https://doi.org/10.1007/978-3-642-83962-7_18).
  61. Mikheyev AS, Tin MMY. A first look at the Oxford Nanopore MinION sequencer. *Mol Ecol Resour*. 2014;14(6):1097–102. <https://doi.org/10.1111/1755-0998.12324>.
  62. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*. 2018;34(17):i884–90. <https://doi.org/10.1093/bioinformatics/bty560>.
  63. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res*. 2017;27(5):722–36. <https://doi.org/10.1101/gr.215087.116>.
  64. Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. *Nat Methods*. 2015;12(1):59–60. <https://doi.org/10.1038/nmeth.3176>.
  65. Huson DH, Auch AF, Qi J, Schuster SC. MEGAN analysis of metagenomic data. *Genome Res*. 2007;17(3):377–86. <https://doi.org/10.1101/gr.5969107>.
  66. Arumugam K, Bağcı C, Bessarab I, Beier S, Buchfink B, Górská A, et al. Annotated bacterial chromosomes from frame-shift-corrected long-read metagenomic data. *Microbiome*. 2019;7:1–13.
  67. Tatusova T, DiCuccio M, Badretdin A, Chetvermin V, Nawrocki EP, Zaslavsky L, et al. NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Res*. 2016;44(14):6614–24. <https://doi.org/10.1093/nar/gkw569>.
  68. Löytynoja A, Goldman N. An algorithm for progressive multiple alignment of sequences with insertions. *PNAS*. 2005;102(30):10557–62. <https://doi.org/10.1073/pnas.0409137102>.
  69. Stamatakis A. RAXML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics*. 2006;22(21):2688–90. <https://doi.org/10.1093/bioinformatics/btl446>.
  70. Li L, Stoeckert CJ, Roos DS. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res*. 2003;13(9):2178–89. <https://doi.org/10.1101/gr.1224503>.
  71. Löytynoja A. Phylogeny-aware alignment with PRANK. *Methods Mol Biol*. 2014;1079:155–70. [https://doi.org/10.1007/978-1-62703-646-7\\_10](https://doi.org/10.1007/978-1-62703-646-7_10).
  72. Kück P, Meusemann K. FASconCAT: convenient handling of data matrices. *Mol Phylogenet Evol*. 2010;56(3):1115–8. <https://doi.org/10.1016/j.ympev.2010.04.024>.
  73. Huerta-Cepas J, Forslund K, Coelho LP, Szklarczyk D, Jensen LJ, von Mering C, et al. Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper. *Mol Biol Evol*. 2017;34(8):2115–22. <https://doi.org/10.1093/molbev/msx148>.
  74. Arndt D, Grant JR, Marcu A, Sajed T, Pon A, Liang Y, et al. PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Res*. 2016;44(W1):W16–21. <https://doi.org/10.1093/nar/gkw387>.

## Publisher's Note

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

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

