frontiers | Frontiers in Microbiology

TYPE Original Research
PUBLISHED 02 October 2023
DOI 10.3389/fmicb.2023.1252709

Check for updates

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

COPYRIGHT
© 2023 Zhang, Wang, Li, Shan, Li, Huang and He. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

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

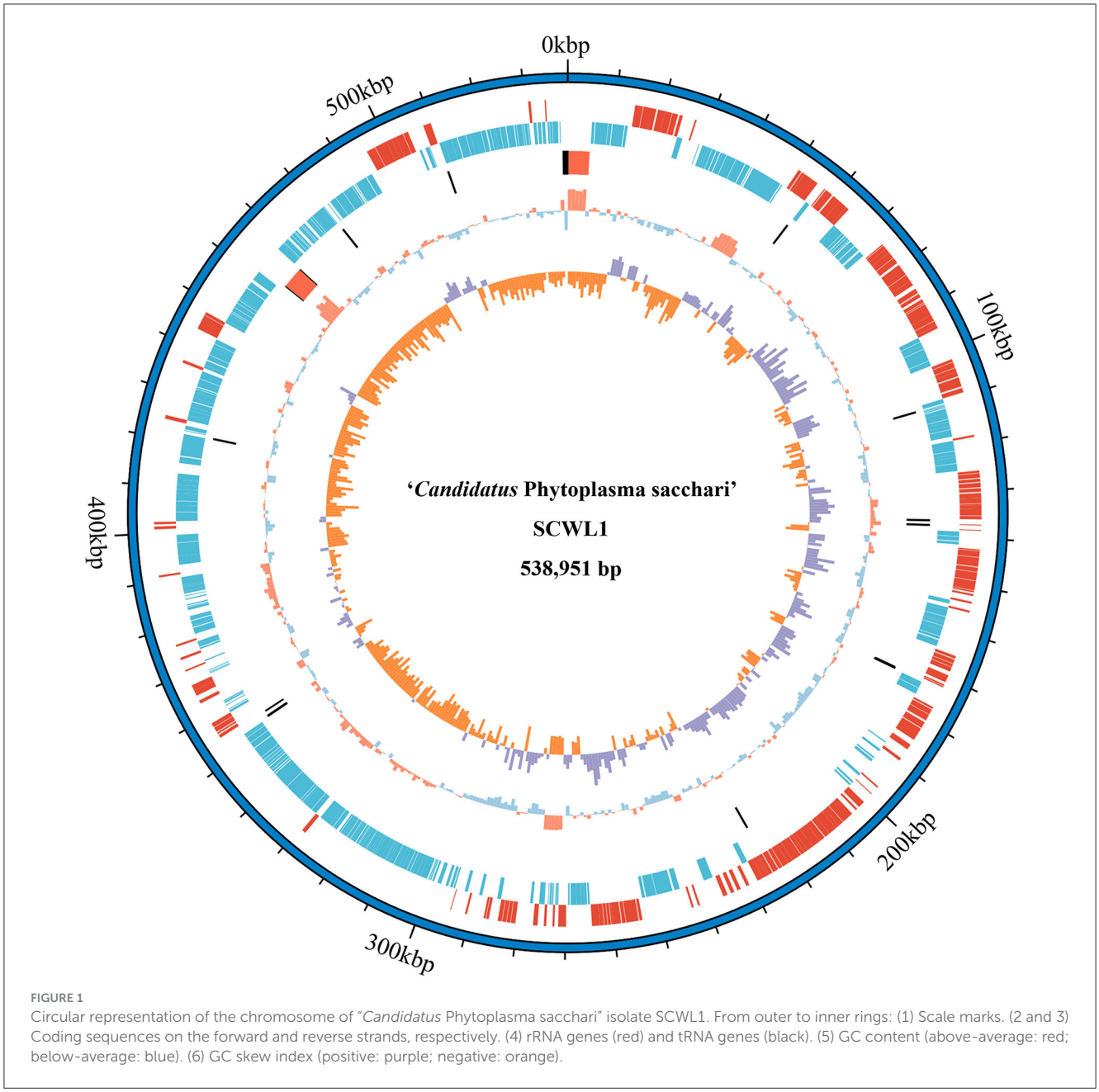

FIGURE 1
Circular representation of the chromosome of "*Candidatus* Phytoplasma sacchari" isolate SCWL1. From outer to inner rings: (1) Scale marks. (2 and 3) Coding sequences on the forward and reverse strands, respectively. (4) rRNA genes (red) and tRNA genes (black). (5) GC content (above-average: 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.

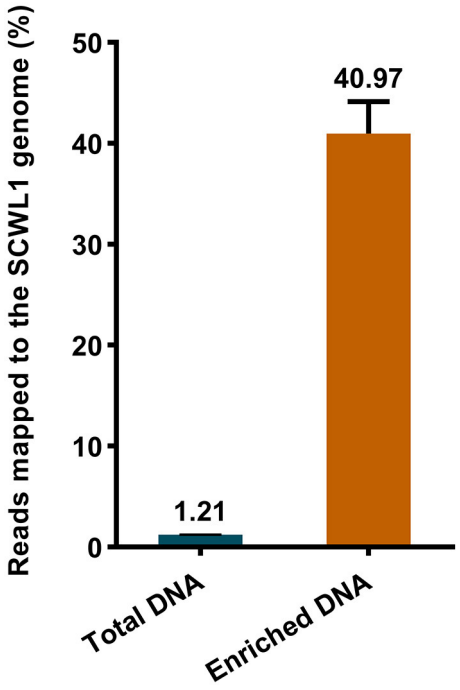

FIGURE 2
Percentage of sequence reads mapped to the genome of "*Candidatus* Phytoplasma sacchari" isolate SCWL1.

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

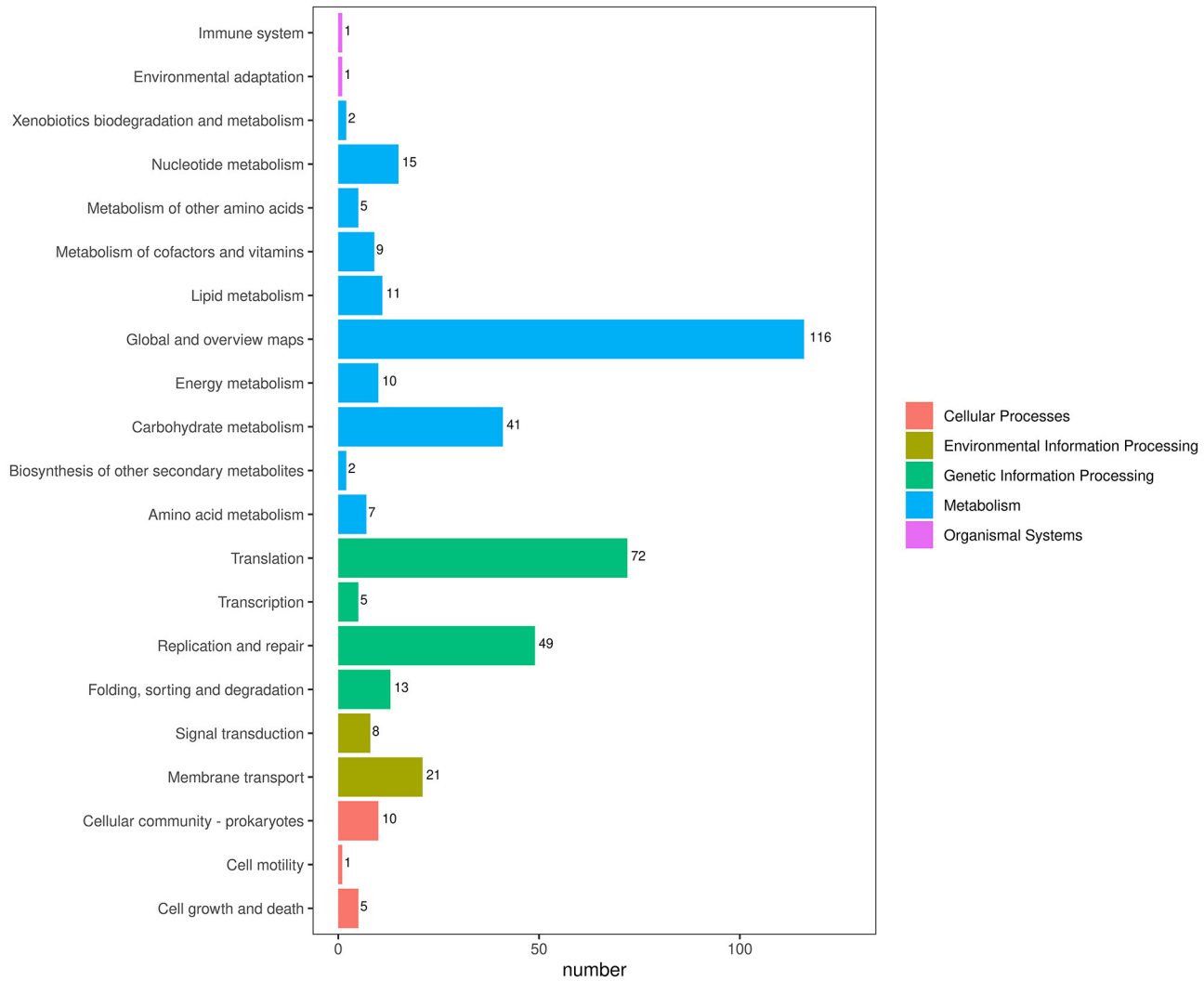

FIGURE 3
Classification map of Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway annotation analysis of the genome of "*Candidatus* Phytoplasma sacchari" isolate SCWL1.

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

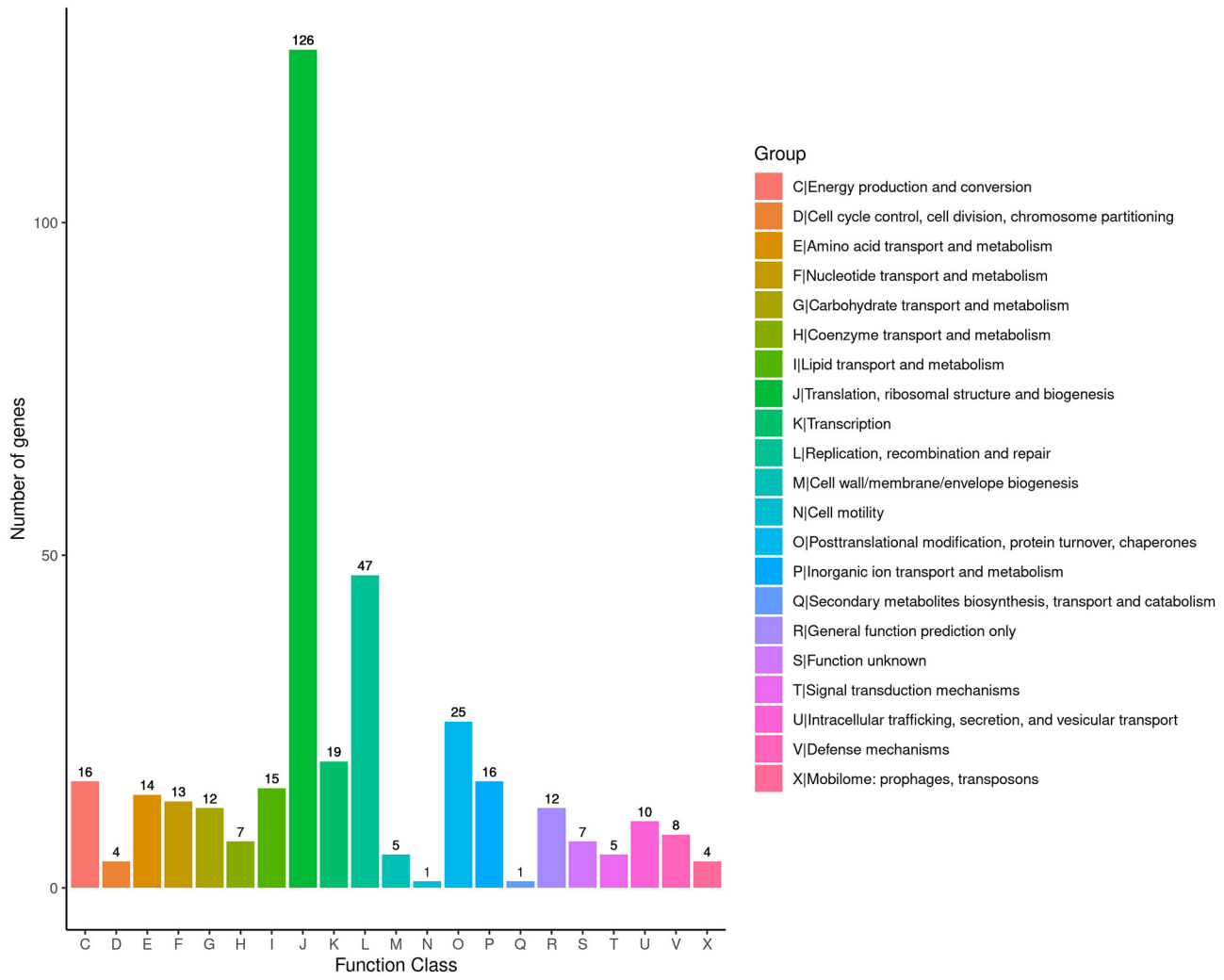

FIGURE 4
Classification map of Cluster of Orthologous Groups of proteins (COG) annotation analysis of the genome of "*Candidatus* Phytoplasma sacchari" isolate SCWL1.

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

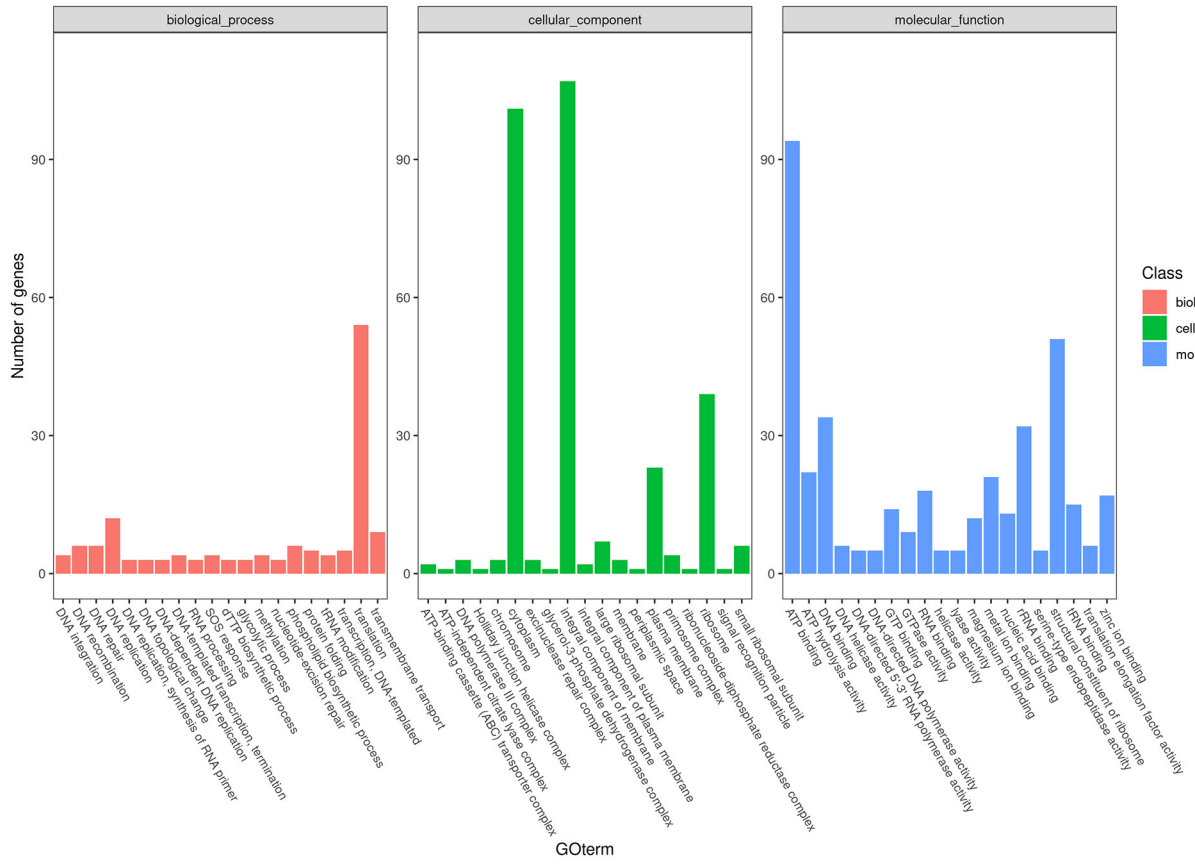

FIGURE 5
Classification map of Gene Ontology (GO) annotation analysis of the genome of "*Candidatus* Phytoplasma sacchari" isolate SCWL1.

10.3389/fmicb.2023.1252709

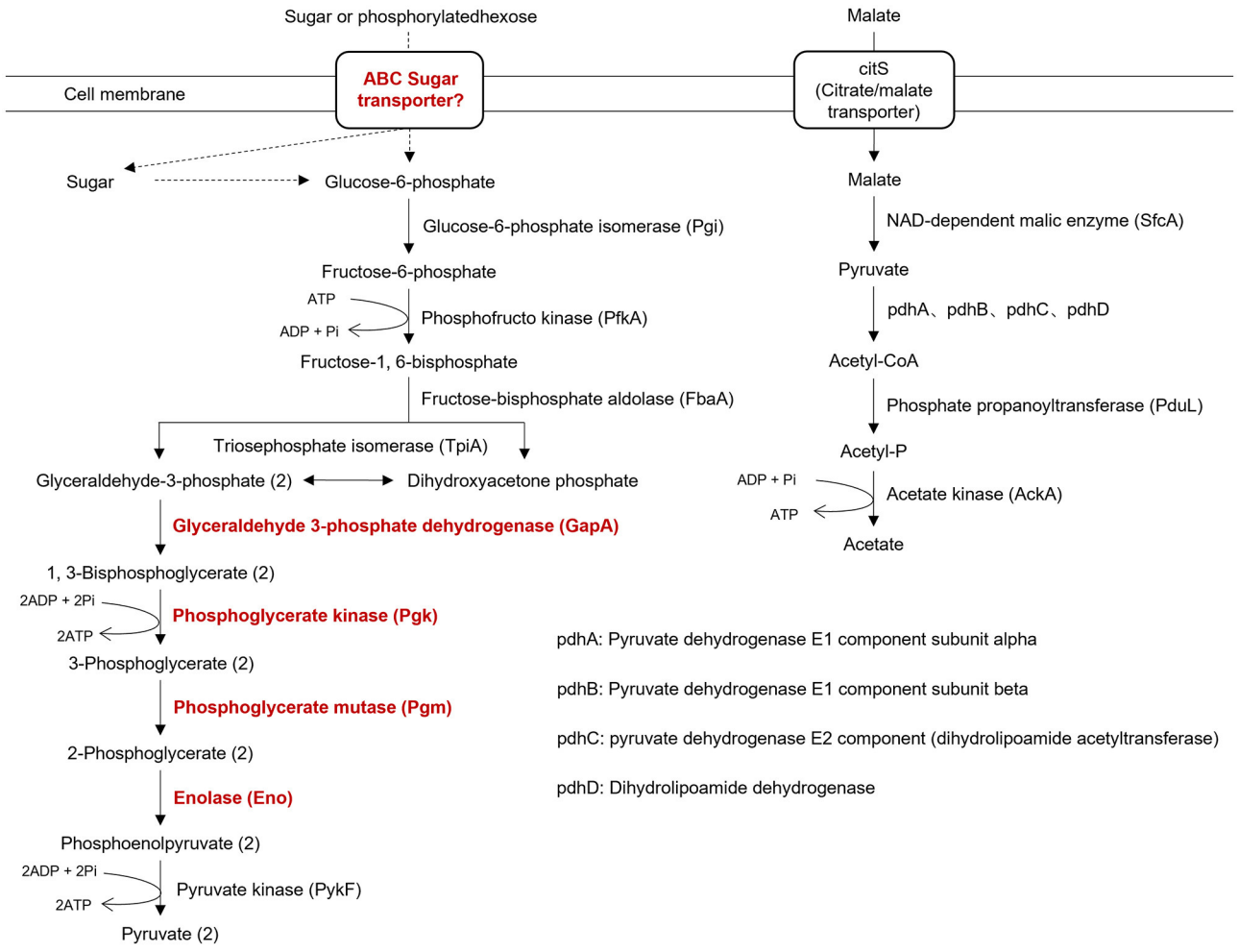

FIGURE 6
Annotation of genes related to energy-yielding pathways in the genome of isolate SCWL1. The proteins marked in red indicate that they are not absent in the genome of isolate SCWL1 genome; the dotted line represents the absence of the enzyme involved in the reaction.

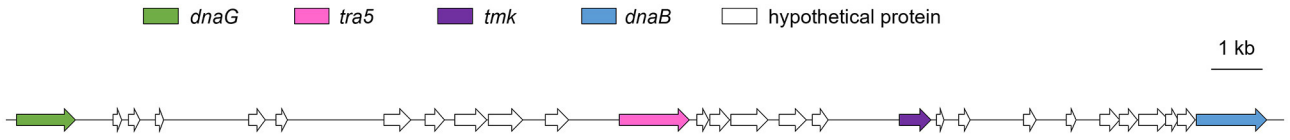

FIGURE 7
Potential mobile units (PMUs) in the genome of "*Candidatus* Phytoplasma sacchari" isolate SCWL1. dnaG, DNA primase; tra5, IS3 family transposase; hflB, ATP-dependent Zn protease; tmk, thymidylate kinase; dnaB, replicative DNA helicase.

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.

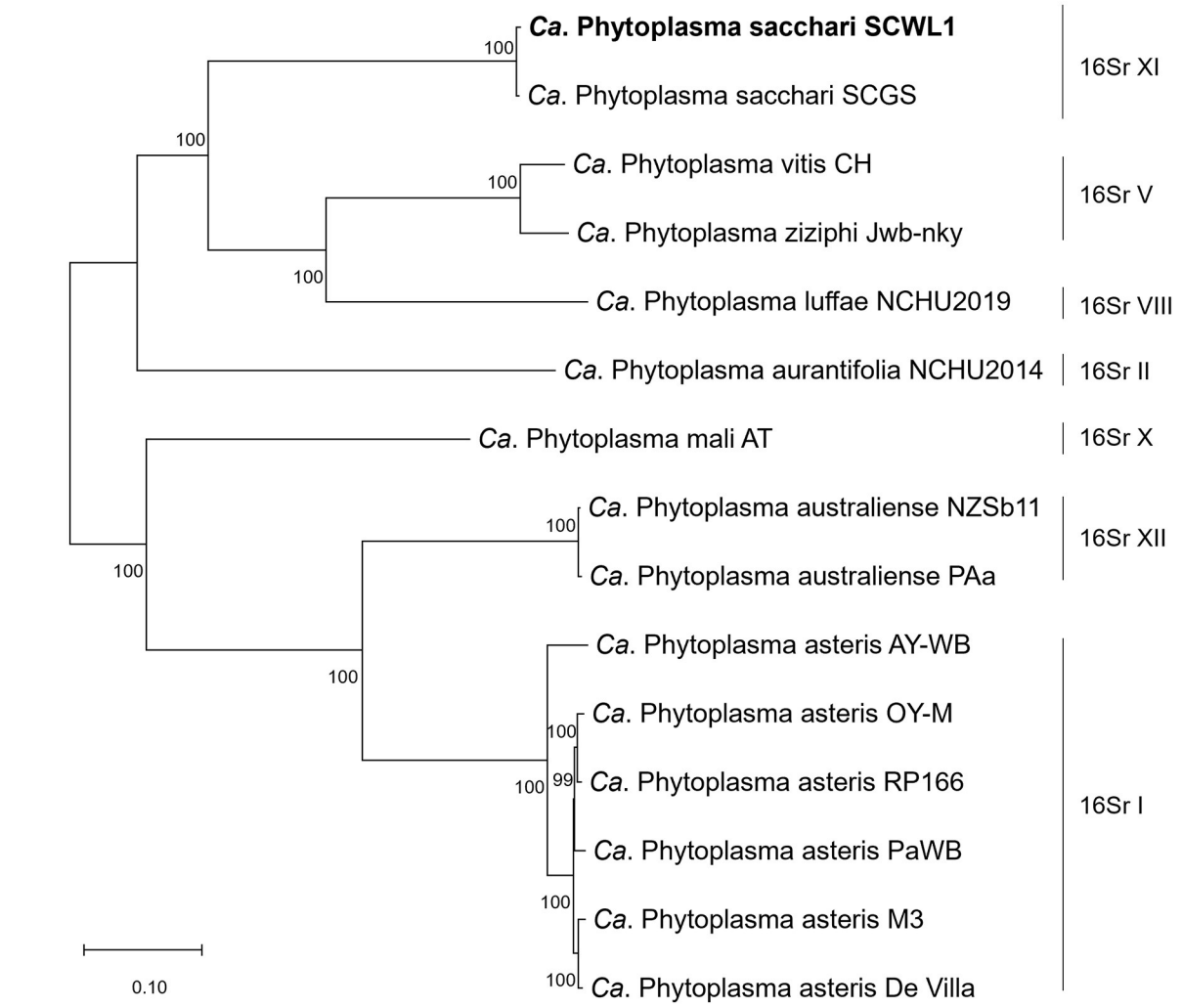

FIGURE 8
Maximum likelihood phylogeny of phytoplasmas inferred based on a concatenated alignment of 191 single-copy orthologous proteins.

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

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.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