Article

The Complete Genome of the "Flavescence Dorée" Phytoplasma Reveals Characteristics of Low Genome Plasticity

Christophe Debonneville [1,*], Léa Mandelli [1], Justine Brodard [1], Raphaël Groux [1], David Roquis [2] and Olivier Schumpp [1]

[1] Virology, Bacteriology and Phytoplasmology, Agroscope, CH-1260 Nyon, Switzerland; lea.mandelli@agroscope.admin.ch (L.M.); justine.brodard@agroscope.admin.ch (J.B.); raphael.groux@agroscope.admin.ch (R.G.); olivier.schumpp@agroscope.admin.ch (O.S.)

[2] Plant Genome Dynamics, Agroscope, CH-1260 Nyon, Switzerland; david.roquis@agroscope.admin.ch

[*] Correspondence: christophe.debonneville@agroscope.admin.ch

Simple Summary: Phytoplasmas are non-cultivable organisms for which biological and molecular characterization is a challenge. Despite the development of new sequencing technologies, few high-quality phytoplasma genomes are available. This research describes the genome sequencing of the phytoplasma involved in grapevine Flavescence dorée, a devastating disease in European vineyards. Analysis of the gene content and comparison with other representative phytoplasmas revealed specific characteristics of the "Flavescence dorée" phytoplasma chromosome. Their implications for the development of the disease are presented and discussed.

Abstract: Members of the genus '*Candidatus* Phytoplasma' are obligate intracellular bacteria restricted to phloem sieve elements and are able to colonize several tissues and the hemolymph in their insect vectors. The current unfeasibility of axenic culture and the low complexity of genomic sequences are obstacles in assembling complete chromosomes. Here, a method combining pathogen DNA enrichment from infected insects and dual deep-sequencing technologies was used to obtain the complete genome of a phytoplasma causing Grapevine Flavescence dorée. The de novo assembly generated a circular chromosome of 654,223 bp containing 506 protein-coding genes. Quality assessment of the draft showed a high degree of completeness. Comparative analysis with other phytoplasmas revealed the absence of potential mobile units and a reduced amount of putative phage-derived segments, suggesting a low genome plasticity. Phylogenetic analyses identified *Candidatus* Phytoplasma ziziphi as the closest fully sequenced relative. The "Flavescence dorée" phytoplasma strain CH genome also encoded for several putative effector proteins potentially playing a role in pathogen virulence. The availability of this genome provides the basis for the study of the pathogenicity mechanisms and evolution of the Flavescence dorée phytoplasma.

Keywords: genome sequencing; effectors; "Flavescence dorée" phytoplasma

1. Introduction

Phytoplasma diseases affect hundreds of plant species through highly specialized interactions whereby each phytoplasma species colonizes a relatively narrow host range [1–4]. As intracellular parasites exclusively restricted to the phloem of their host plant, they are transmitted in a persistent and propagative manner by hemipterans vectors. This mode of transmission involves multiplication in insect tissues and the crossing of several cell barriers [1,5]. Thus, the interaction with the vector is the result of advanced co-evolution [6,7]. This very high level of specialization for a limited number of hosts masks a strong capacity to adapt to changing environments such as plant phloem or insect tissues [8]. The genes required in these diverse environments are different and significant transcriptomic variation accompanies the transition from plant to vector [9,10].

Because of their small size and lack of a cell wall, phytoplasmas have been associated with organisms of the class Mollicutes since their discovery [11]. However, they are distinguished from the bacteria of the other genera of this group by their inability to be grown under axenic conditions. This specification considerably complicates their characterization and identification. Nevertheless, classification based on the 16S rRNA has led to the description of more than 30 different groups [12–14] including 49 officially published species and several putative species. The impossibility to cultivate them and a GC content of around 20–30% explain why, despite their high number and diversity, only 11 complete and annotated phytoplasma genome sequences are available to date.

These genomes are small (570 to 1000 kbp) and are sometimes associated with plasmids. They can have a high number of repeats and mobile units covering a significant part of the chromosome, suggesting a high plasticity of the genome [8,15–17]. The number of genes is low, ranging from 475 to ~1000, and phytoplasmas notably lack several biochemical pathways normally found in bacteria (reviewed in [2]). Instead, they rely on the import of essential nutrients from their environment, as evidenced by the presence of many genes encoding transport proteins [8]. The relatively small number of metabolic genes contrasts with the presence of numerous secreted effector proteins [15,18–20]. Effectors decrease the plant's defenses [21] or alter its development [20,22–25] and promote the attractiveness of the plant to insect vectors [26–28]. These secreted effectors possess a signal peptide on the N-terminal side that may be confused with a transmembrane domain [29]. Phytoplasma genomes also contain genuine transmembrane proteins acting as effectors, as shown by Imp, Amp or VmpA, that are involved in interactions with their plant hosts or the insect vector [30–33].

In grapevine (*Vitis* sp.), phytoplasmas from various phylogenetic groups are associated with Grapevine Yellows (GY). GY is spread worldwide in wine-growing areas, inducing considerable economic losses. Flavescence dorée, one of the most devastating GY, is caused by Flavescence dorée phytoplasma (FDp), belonging to the 16SrV taxonomic group. Although it can be transmitted by several insects, the propagation of the disease is epidemic only in the presence of the leafhopper *Scaphoideus titanus* Ball [34]. FD is classified as a quarantine disease in the EU [35]. Switzerland is an interesting area for the analysis of this disease because the application of strict measures did not permit the control of the disease south of the Alps, whereas similar measures allowed the eradication of several recent outbreaks north of the Alps.

In this study, we aimed to reconstruct a high-resolution sequence of the FDp genome. A crucial step of enrichment of bacterial DNA from infected insects allowed us to obtain a high read coverage from the deep sequencing [36]. A dual sequencing approach combining Illumina and Oxford Nanopore Technology (ONT) was used to generate a complete circular and annotated chromosome of the "Flavescence dorée" phytoplasma. An analysis of key features was performed as well as a comparison with other existing phytoplasma genomes.

2. Materials and Methods

2.1. Biological Material Source

In November 2020, 5 vines of cv. Chasselas from the outbreak in Chardonne (Vaud, Switzerland) were carefully uprooted and brought to the quarantine facilities at Agroscope. These plants previously tested positive for FDp during the summer season. After overwintering, surviving plants were tested for the presence of FDp by quantitative PCR, as previously described [37]. About 30 *Scaphoideus titanus* reared according to Ripamonti et al. [38] were encaged on a symptomatic shoot for two weeks and then transferred to the healthy *Vicia faba* variety Aguadulce for transmission during two weeks. Two weeks later, the *Vicia faba* plants were tested for the presence of FDp by quantitative PCR. One positive plant was used as feeding source for about 50 *Euscelidius variegatus* 4–5th stage nymph instars for two weeks. After a latency period of two weeks, twenty *E. variegatus* were collected for DNA extraction.

2.2. DNA Extraction and Enrichment

Twenty insects were homogenized in 900 µL of extraction buffer (3% Cetyltrimethylammonium bromide CTAB, 1.4 M NaCl, 25 mM EDTA, 1 M Tris-HCl, 2% β-Mercaptoethanol, pH 8.0) and shaken for 30 min at 600 rpm and 65 °C. 900 µL of Chloroform/Isoamylalcohol was added, homogenized by vortexing for 5 s and centrifuged for 5 min at 3000× g. The aqueous layer was carefully transferred to a new tube, mixed with an equal volume of cold Isopropanol, and incubated 60 min at −20 °C for DNA precipitation. Precipitated material was recovered by 2 min of centrifugation at 10,000× g and washed with 1 mL of 70% Ethanol. A DNA pellet was dried overnight at room temperature and resuspended into 100 µL of PCR-grade water. The extract was treated with 1 mg/mL RNase A for 30 min at 37 °C. A total of 2 µg of DNA was used with the NEBNext Microbiome DNA Enrichment kit (New England Biolabs Inc., Ipswich, MA, USA #E2612S), following manufacturer's instructions. The option "Ethanol Precipitation" was applied to clean up the DNA at the end of the procedure.

2.3. MinION and Illumina Sequencing

DNA was fragmented using a g-tube (Covaris, Woburn, MA, USA #520079) for 1 min at 2000× g. A total of 1 µg of fragmented DNA was used with the Ligation Sequencing kit (Oxford Nanopore Technologies, Oxford, UK #SQK-LSK110), following manufacturer's instructions. The final mixture containing the library was sequenced on a MinION Mk1C using a R9.4.1 flow cell. Sequences from the MinION were basecalled and adapter sequences removed with Guppy v3.4.5 using built-in configuration "dna_r9.4.1_450bps_fast.cfg". Reads were then processed through Filtlong v0.2.0 with –min_mean_q 80 and –min_length 200 to remove low-quality sequences. The long reads were obtained from two sequencing runs of two independent libraries on the same flow cell. They had a N50 of 15.5 kb, as assessed by NanoPlot. Illumina sequencing was performed by Macrogen Europe B.V (Amsterdam, The Netherlands) using TruSeq DNA PCR-Free library preparation, followed by a NovaSeq 2 × 150 bp sequencing run and a depth of 100 M reads. Reads were trimmed through trimmomatic v0.39 with PE –phred33 ILLUMINACLIP LEADING:20 TRAILING:20 SLIDINGWINDOW:4:20 MINLEN:50.

2.4. De Novo Genome Assembly and Quality Assessment

Long reads were assembled with Flye v2.8.3-b1695. All resulting contigs were blasted against a local database (built with makeblastdb application) of phytoplasma complete genomes with Blastn v2.5.0+. Long and short reads were mapped using Minimap2 v2.21-r1071 and Bowtie2 v2.3.4.1, respectively. Mapped reads were processed together for hybrid assembly using Unicycler v0.4.9b. Additionally, mapped reads were visually inspected using Geneious v2022.0.2 for polishing of possible assembly errors based on MinION long reads stackings, similar to Huang et al., 2022 [16]. Samtools depth function was used to assess the sequencing coverage of both long and short reads. Completeness of the draft genome was estimated using BUSCO v5.3.0 [39], with the lineage dataset "mollicutes_odb10" created on 6 March 2020. Additionally, the webserver MiGa (http://microbial-genomes.org, accessed on 20 June 2022) [40] was used for taxonomic evaluation. Sequences were deposited in GenBank database under accession number CP097583 for the genome and PRJNA838420 for the Sequence Read Archive (SRA).

2.5. Functional Annotation

The final draft was annotated with Prokka v1.13 [41]. Functional characterization was assessed with GO FEAT (http://computationalbiology.ufpa.br/gofeat/, accessed on 20 June 2022) [42] and BlastKOALA (https://www.kegg.jp/blastkoala/, accessed on 20 June 2022) [43]. Additionally, manual curation was performed by BlastP searches against the NCBI non-redundant database. Putative effector proteins, including those with transmembrane domain, were predicted with SignalP v5.0 [44] and Phobius (https://phobius.

sbc.su.se/, accessed on 20 June 2022) [45]. The circular map was generated using Circlize v0.4.10 [46]. The first base of the start codon of the dnaA gene was set as base pair 1.

2.6. Genome Comparison and Phylogenetic Analyses

For comparative analysis with representative phytoplasma genomes, OrthoVenn2 (https://orthovenn2.bioinfotoolkits.net/home, accessed on 20 June 2022) [47] was used to identify orthologous clusters. Multiple sequence alignments of single-copy orthologue groups were prepared with MUSCLE v3.8.1551 and concatenated to produce one super-matrix for each phytoplasma. Maximum likelihood phylogeny was inferred with IQtree (http://iqtree.cibiv.univie.ac.at/, accessed on 20 June 2022) [48] with substitution model set on "auto" and bootstrapping with 1000 replicates. For groEL and smpB, trees were inferred by maximum likelihood method in MEGA using the General Time Reversible model and bootstrapping with 500 replicates. All trees were visualized with iTOL (https://itol.embl.de/, accessed on 20 June 2022) [49].

3. Results

3.1. Genome Assembly and Quality Assessment

Twenty *Euscelidius variegatus* fed on FDp-infected broad beans were used as starting material. DNA extract was enriched in phytoplasma genomic DNA through a methyl-CpG-binding, domain-mediated method [36]. Quantitative PCR analysis based on map locus produced a reduced Cq value of roughly five cycles (data not shown) corresponding to a 30-times enrichment. To improve sequencing quality, we used both long-read sequencing from MinION nanopore and short-read sequencing from Illumina. MinION nanopore sequencing generated a total of 1,158,896 long reads (representing 7.3×10^9 bp), and Illumina sequencing generated 128,906,832 short reads (representing 19.5×10^9 bp). In the first step of de novo assembly, long reads were assembled with "Flye", resulting in 10,820 contigs. A comparison with a phytoplasma genome database by MegaBlast identified 65 phytoplasma-related contigs. Among them, a circular contig of 651,724 bp (Figure 1A) showed a bitscore of 4.3×10^4 with the genome of *Candidatus* Phytoplasma ziziphi, a phytoplasma belonging to 16Sr V group [50]. The remaining contigs showed only a poor match with phytoplasma sequences (Supplementary Table S1), and no plasmids were detected. In the second step, all long and short reads were mapped to this contig and only mapped reads were used as inputs for a hybrid assembly with "Unicycler", resulting in a circular chromosome of 654,052 bp. Finally, the reads mapped to chromosome were visually inspected and errors introduced by the long reads corrected at base-pair level to obtain the final assembly of 654,223 bp. Mapping results to this draft found a 15× and 46× mean coverage for long and short reads, respectively.

BUSCO analysis evaluated 151 markers from the dataset "Mollicutes", resulting in 143 genes considered as complete (either single or duplicated), one fragmented gene, and seven missing genes (94.7%, 0.7%, and 4.6% respectively; Figure 1B). In comparison, the assessment of the published genome of *Candidatus* Phytoplasma ziziphi showed a lower BUSCO score (data not shown). Analysis with MiGa [40], an online tool to assess prokaryotic genome assemblies, rated the completeness as "Very high", further confirming the quality of the circular chromosome.
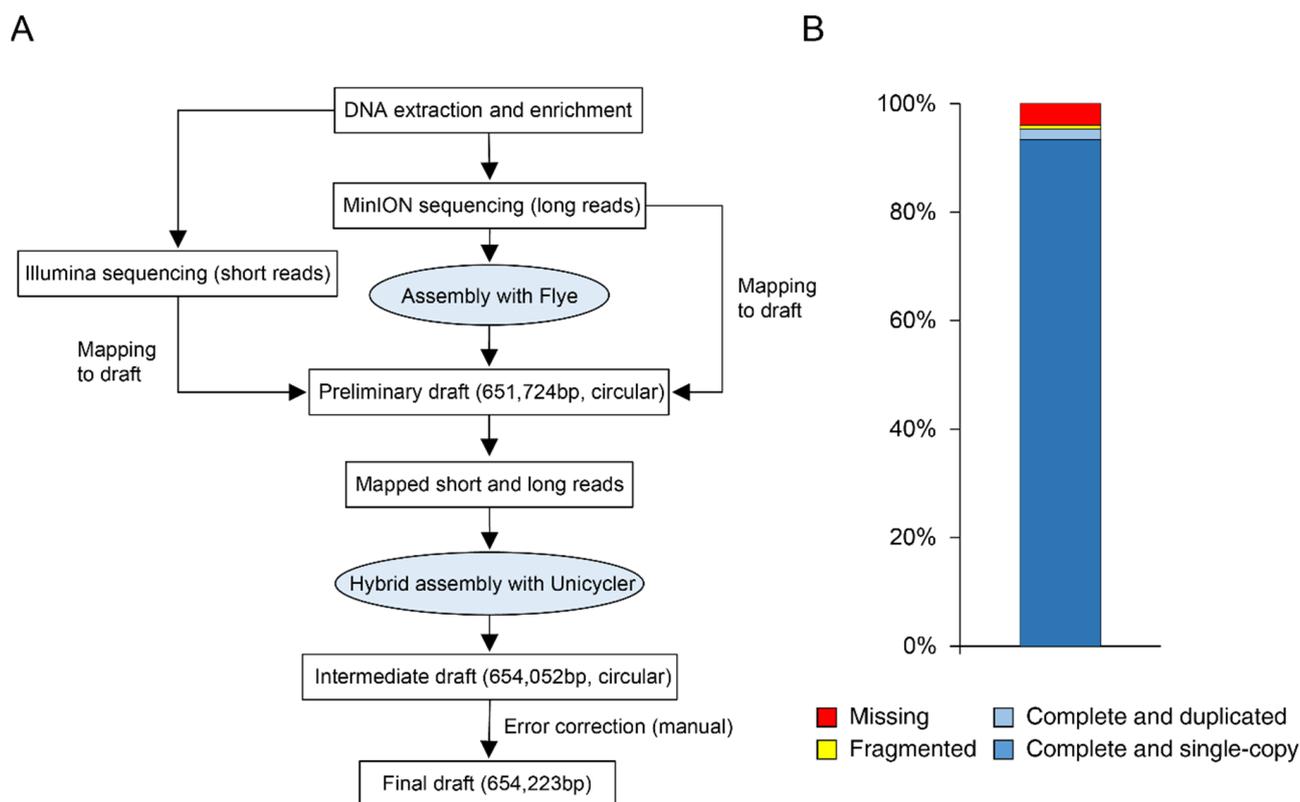
A

B

Figure 1. Genome assembly and quality assessment. (**A**) Workflow proposed as the successful approach for de novo genome assembly. (**B**) BUSCO analysis to assess completeness of the assembled chromosome.

3.2. Genome Characteristics and Annotation

3.2.1. General Features

The genome of Flavescence dorée phytoplasma consists in a circular chromosome of 654,223 bp with a G/C content of 21.7%, which places it among the species with the lowest GC rate [51]. The sequence shows a regular GC-skew pattern comparable to 'Ca. P. mali' chromosome (Figure 2) [52]. The annotation revealed 506 protein-coding genes (CDS), 32 tRNA genes, and 2 rRNA operons (Table 1, Figure 2). CDSs have an average length of 987 bp and cover 77% of the total size of the chromosome. Among the CDSs, 350 proteins had an assigned function, whereas 156 were categorized as hypothetical products. A screen of protein databases searching for homology revealed a match for 470 of the 506 annotated proteins (with an e-value $< 1 \times 10^{-10}$). Proteins with a homologous peer in the genus *Candidatus* Phytoplasma represented 90.9% of the total.

Table 1. General characteristics of FDp strain CH and comparison to representative complete phytoplasma genomes.

| Strain (Accession)	16S rRNA Group	Genome Size (bp)	G/C Content (%)	No. of CDS	Protein-Coding Regions (%)	No. of tRNA Genes	No. of rRNA Operons
FDp strain CH (CP097583)	V	654,223	21.7	506	77	32	2
'Ca. P. ziziphi' jwb-nky (CP02521)	V	750,803	23.2	671	78	32	2
'Ca. P. asteris' AYWB (CP000061)	I	706,569	26.9	636	73.9	32	2
'Ca. P. asteris' OY-M (AP006628)	I	853,092	27.8	752	72.8	32	2
'Ca. P. luffae' (CP054393)	VIII	769,143	23.3	725	80.1	31	2
'Ca. P. australiense' (CP002548)	XII	959,779	27.2	928	75.5	35	2
'Ca. P. mali' (CU469464)	X	601,943	21.4	500	78.2	32	2
'Ca. P. aurantifolia' (CP040925)	II	635,584	24.5	471	66.1	24	2
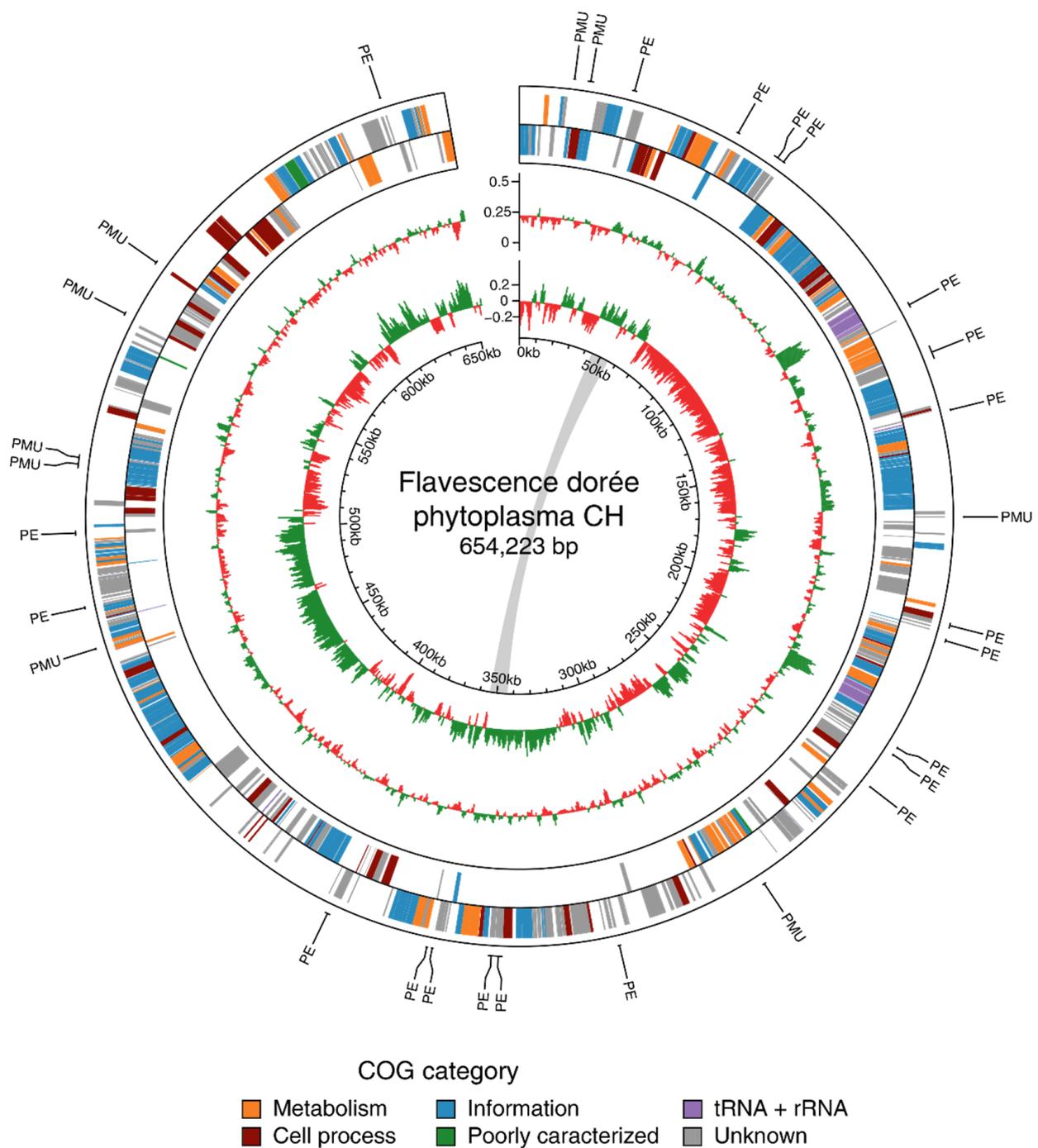
Figure 2. Genome map of the circular chromosome of FDp strain CH. Rings from outside in: (1 and 2) coding sequences on the forward and reverse strand, respectively (color-coded by functional categories); (3) G/C content (above average: green; below average: red); (4) G/C skew (above average: green; below average: red); (5) scale marks. (PE): putative effectors; (PMU): genes belonging to the potential mobile unit core set. The grey line indicates the positions of the 10 kb duplicated region. The G/C content and G/C skew are displayed with a sliding window of 1 kb and a step size of 0.5 kb.

3.2.2. Functional Annotation

Proteins associated to a Gene Ontology (GO) term were classified into the three general categories: "Biological Process" (312 hits), "Cellular Component" (348 hits), and "Molecular Function" (691 hits) (Figure 3A). The most abundant biological process was "Translation" (54 hits), followed by "Methylation" (11 hits) and "Transmembrane transport"

(10 hits). In the cellular component category, the major part of the proteins belonged to "Integral component of membrane" (125 hits), "Cytoplasm" (109 hits), and "Ribosome" (40 hits). Finally, the most abundant molecular functions were "ATP binding" (105 hits), "Structural constituent of ribosome" (52 hits), and "DNA binding" (46 hits) (Figure 3B and Supplementary Table S2). Functional annotation with BlastKOALA (with reference taxonomy group NCBI:txid33926 = Tenericutes) resulted in the assignment of 320 of the 506 CDSs (63.2%) to orthologues in the KEGG database (prokaryotes) [53]. The three main protein families contained 184 proteins classified in "Metabolism", 241 in "Genetic information processing", and 48 in "Signaling and cellular processes" (Figure 4).
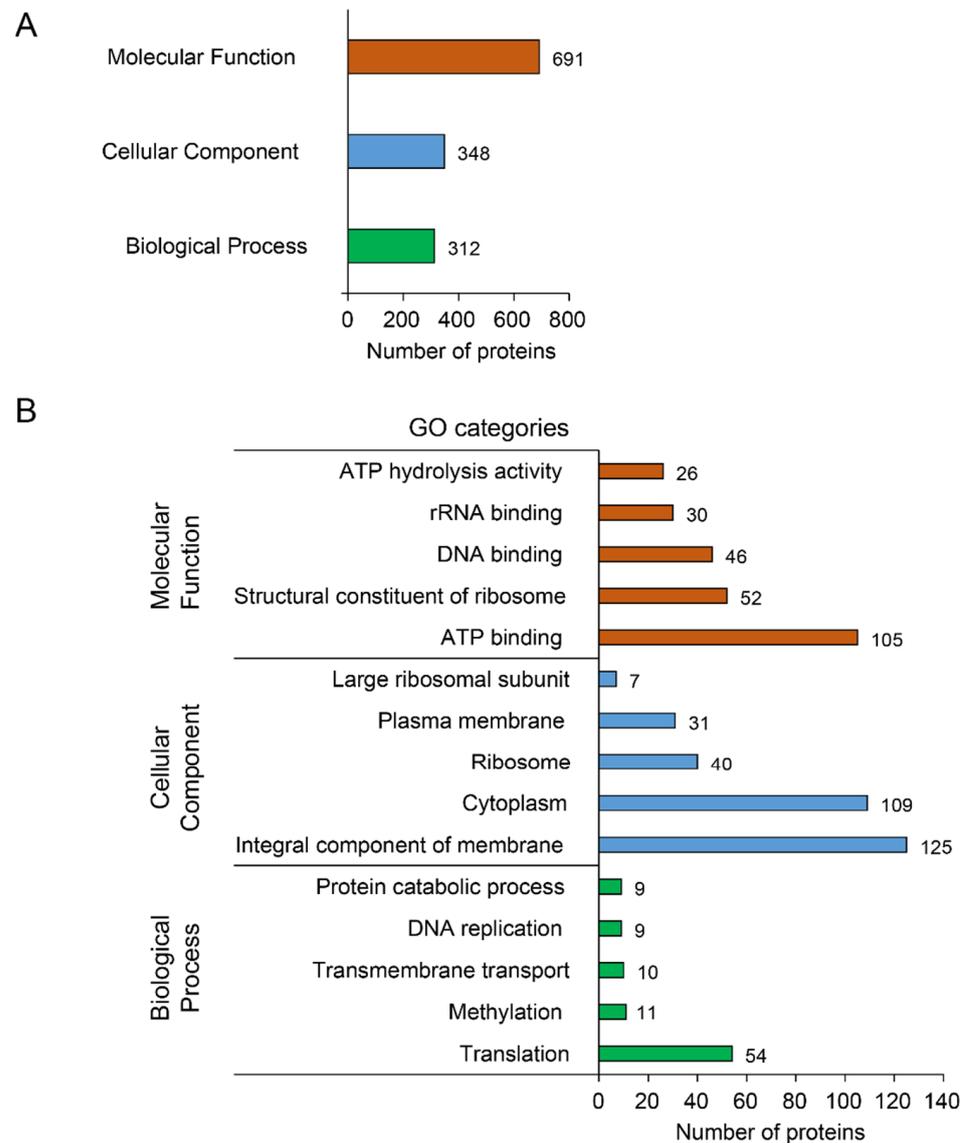

Figure 3. Gene Ontology categories associated with the CDSs from FDp strain CH genome. (**A**) Bar chart showing the distribution among the three main GO domains. One protein can have several molecular functions. (**B**) Bar chart showing the five most represented GO categories in each domain.
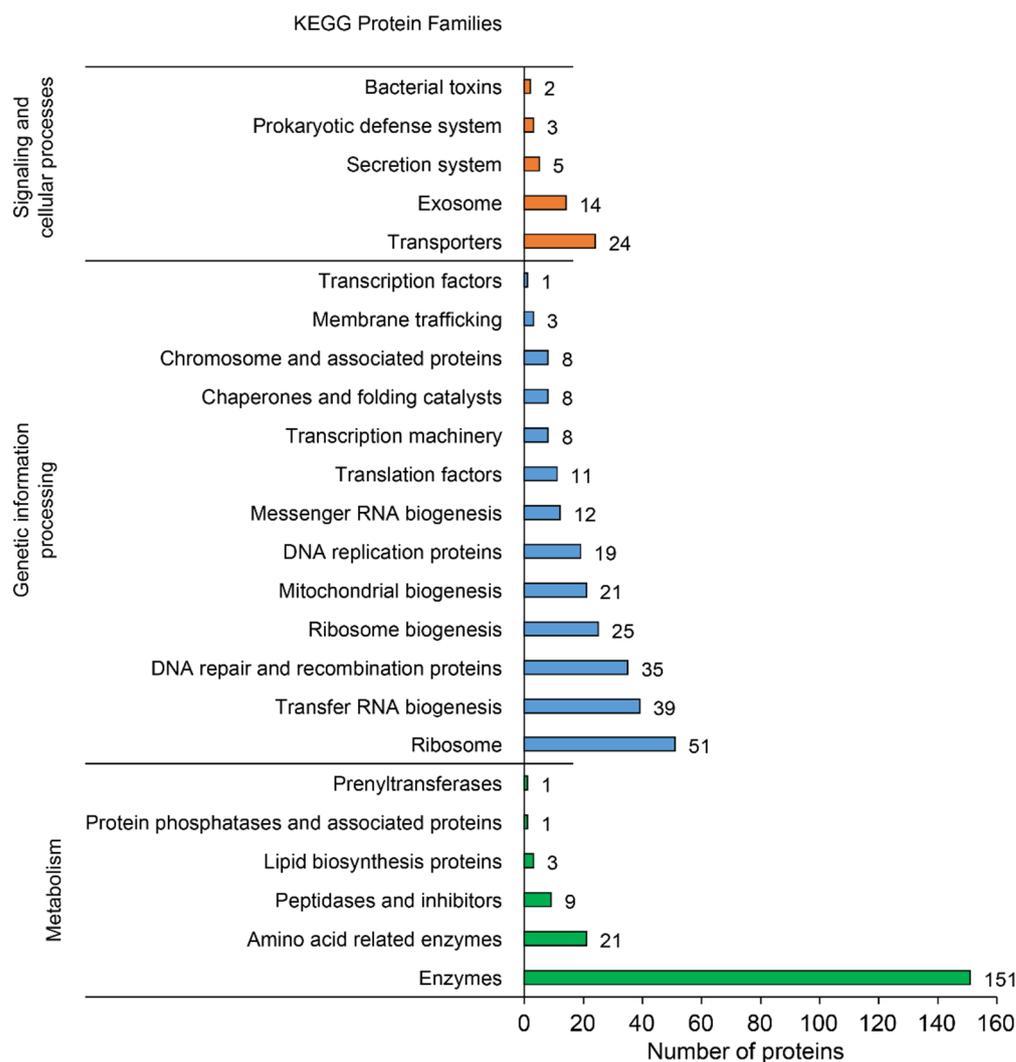
Figure 4. Bar chart representing the KEGG protein families associated with the CDSs from FDp strain CH genome.

3.3. Key Features

3.3.1. Transporters and Metabolic Genes

As shown for other phytoplasmas, FDp has reduced metabolic functions with several incomplete pathways, such as, for example, those for fatty and amino acids biosynthesis. Instead, phytoplasmas depend on the import of essential elements, and the present genome possessed 40 genes involved in transport, most of them being ABC transporters (25 genes; Supplementary Table S3). The pathogen had a complete protein secretion system (sec) with 10 genes affiliated with this function (secA, secE, secY, yidC, ffh, ftsY, dnaJ, dnaK, grpE, and groL) [54]. A total of 25 genes, including 16 coding for ATP-dependent zinc metalloproteases (FtsH proteins), were associated with a peptidase activity. As for the carbohydrate metabolism, the chromosome encoded for complete glycolysis and pyruvate oxidation pathways (Supplementary Table S2). Notably, the genome contained six genes with a restriction endonuclease function, a system that does not belong to the shared core set of phytoplasmas.

3.3.2. Effector Genes and Potential Mobile Units

SignalP v5.0 [44] and Phobius [45] were used to predict the presence of signal peptide (SP) in the set of proteins. A total of 17 putative secreted proteins were identified (Table 2). Phobius was designed to better separate transmembrane domains and signal peptides [29].

It predicted a transmembrane domain instead of a signal peptide for three of these proteins (FlDop_00101, FlDop_00265 and FlDop_00298) but also suggested an SP in FlDop_00246 and FlDop_00266, whereas SignalP v5.0 did not detect them (Supplementary Figure S1). Interestingly, Phobius predicted both a transmembrane domain and an SP for three proteins (FlDop_00023, FlDop_00048, FlDop_00049). These proteins may act as pathogenicity factors, with different mechanisms either attached to the membrane or released in the host's cells. Only one homologue of a member of the SAP repertoire describing putative effector proteins identified in '*Ca*. P. asteris' [18] was present (FlDop_00531), and no homologues to the experimentally characterized effectors SAP05, SAP11, SAP54 and TENGU were found.

Table 2. Putative effectors encoded by the FDp genome.

Protein ID	Function	SignalP5.0	Phobius	Accession # of Orthologues in Other Phytoplasmas (Max. 2)
FlDop_00023	Variable membrane protein B	SP	SP and 1 TM	VIO49504 (Alder yellows P.)
FlDop_00041	Hypothetical protein (htmp2)	no prediction	TM	WP_225696264 (*Ca*. P. sp. AldY-WA1) WP_121463947 (*Ca*. P. ziziphi)
FlDop_00048 *	Hypothetical protein	SP	SP and 1 TM	none
FlDop_00049 *	Hypothetical protein	SP	SP and 1 TM	none
FlDop_00090	Hypothetical protein	SP	SP	WP_225696128 (*Ca*. P. sp. AldY-WA1) WP_121464024 (*Ca*. P. ziziphi)
FlDop_00101	Hypothetical protein	SP	TM	WP_238055118 (*Ca*. P. ziziphi) WP_121464035 (*Ca*. P. ziziphi)
FlDop_00112	Hypothetical protein	SP	SP	AYJ01330 (*Ca*. P. ziziphi)
FlDop_00153	Hypothetical protein (htmp5)	no prediction	TM	WP_121464226 (*Ca*. P. ziziphi)
FlDop_00158 *	Hypothetical protein	SP	SP	none
FlDop_00183	Hypothetical protein	SP	SP	WP_026072021 (Poinsettia branch-inducing P.) WP_152411650 (Milkweed yellows P.)
FlDop_00185	Hypothetical protein (SVM family)	SP	SP	WP_034172411 (Chrysanthemum yellows P.) WP_024563506 (*Ca*. P. tritici)
FlDop_00190	Hypothetical protein	SP	SP	PQP79517 (*Ca*. P. phoenicium) WP_078123062 (Ca. P. aurantifolia)
FlDop_00246 *	Hypothetical protein	no prediction	SP	none
FlDop_00265	Hypothetical protein	SP	TM	WP_121464113 (*Ca*. P. ziziphi)
FlDop_00266	Hypothetical protein	no prediction	SP	WP_121464114 (*Ca*. P. ziziphi)
FlDop_00278 *	Hypothetical protein	SP	SP	none
FlDop_00279 *	Hypothetical protein (htmp1)	no prediction	TM	none
FlDop_00298	Hypothetical protein	SP	TM	WP_012504569 (*Ca*. P. mali) WP_227807101 (Mulberry dwarf P.)
FlDop_00404	Hypothetical protein (htmp3)	no prediction	TM	WP_225696004 (*Ca*. P. sp. AldY-WA1) WP_121463722 (*Ca*. P. ziziphi)
FlDop_00420 *	Hypothetical protein (htmp4)	SP	SP	none
FlDop_00531	SAP21-like protein	SP	SP	QKX95099 (Rapeseed phyllody P.) WP_122225587 (*Ca*. P. solani)

SP: signal peptide; TM: transmembrane domain. * singletons (proteins with no homologue in other analyzed phytoplasma genomes).

Five proteins defined as "hypothetical transmembrane protein (htmp)" were also considered as potential effectors. In addition, FtsH proteins mentioned above can also act as potential virulence agents, as shown for '*Ca*. P. mali' [55] and hypothesized for FDp [10].

Phytoplasma genomes usually contain regions called Potential Mobile Units (PMUs) derived from ancient phage attacks [56,57]. PMUs identified in various phytoplasma species present a set of conserved genes involved in different functions, such as transposition and gene duplication [8,16]. Most described PMUs also harbor putative effector proteins. Interestingly, no PMU-like region had been identified in the genome of FDp. Although genes belonging to the core set of PMUs were present (Supplementary Table S2), they were scattered over the circular genome and were only single copy (Figure 2). A raw estimation of prophage loci with PHASTER [58] identified one incomplete region of about 16 kb (2.5% of the total chromosome size) of putative viral origin.

Another interesting finding was the presence of a duplicated region with no detectable phage-derived characteristics of about 10 kb at positions 43,623 to 54,237 and 344,306 to 354,721 containing eight genes. Most of them were involved in sugar transport and

translation processes. The region showed a good synteny with several phytoplasmas (data not shown).

3.4. Phylogeny and Taxonomy

According to the current classification system based on the 16S rRNA gene [59,60], FDp belongs to group 16SrV. Both 16S rRNA gene copies found in the present genome had 100% identity with the gene from reference isolate FD92, placing the strain in the subgroup V-D and cluster map-FD2 [61]. In agreement, the sequence of the Methionine aminopeptidase (map) gene was 100% identical to the described genotype M54. The phylogenetic tree constructed with DNA sequences from the intracellular chaperone groEL (Figure 5A) grouped the new strain together with *Candidatus* Phytoplasma ulmi, Aster Yellows Phytoplasma, and *Candidatus* Phytoplasma ziziphi, three members of the 16SrV group. An additional tree based on smpB, a gene coding for a protein involved in ribosomal rescue recently shown to resolve species differentiation in *Acholeplasmataceae* [62], placed *Candidatus* Phytoplasma ziziphi as the closest relative of the new strain (Figure 5B). At the whole-genome level, these two phytoplasmas have an average nucleotide identity (ANI) of 92.2%, therefore likely representing distinct species [63]. Analysis with MiGa confirmed that the circular chromosome most likely belonged to a species not present in the database (with *p*-value = 0.0025).
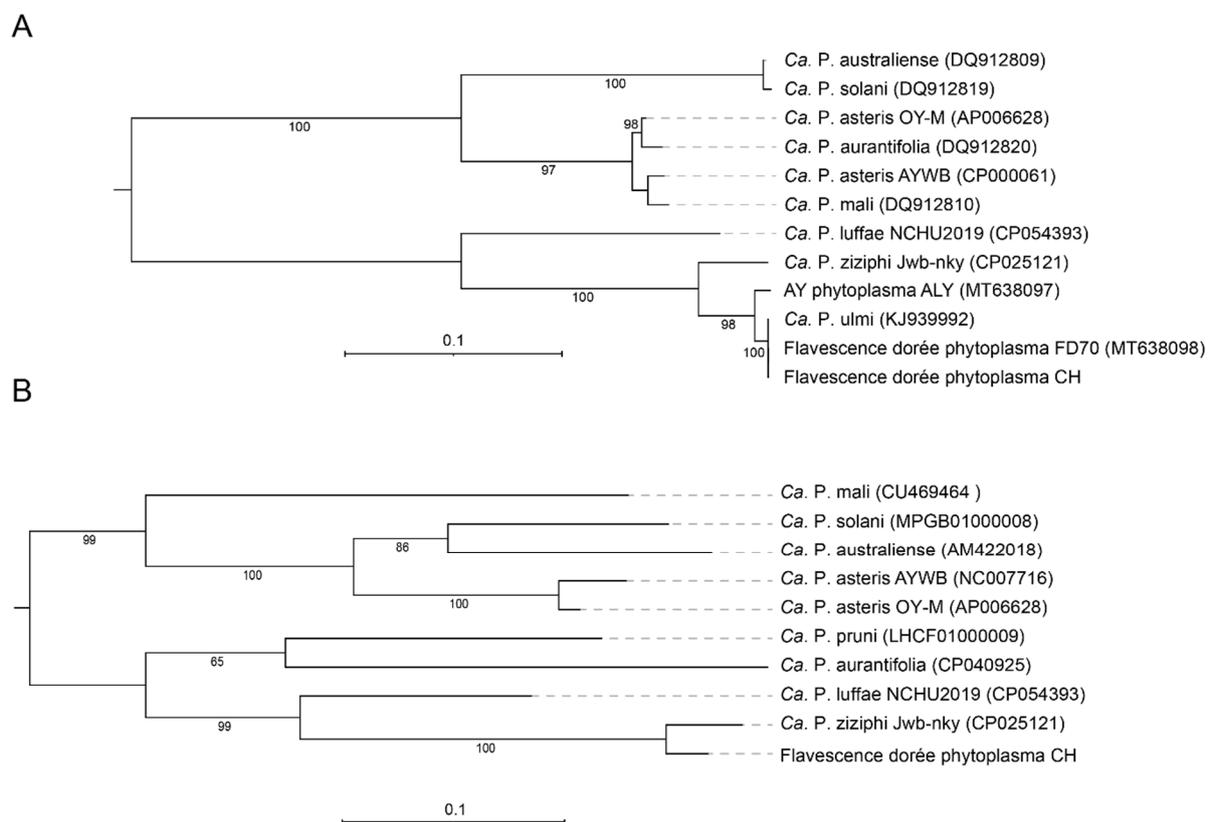

Figure 5. Maximum likelihood phylogeny based on nucleotide sequences of (**A**) groEL and (**B**) smpB genes. The numbers on branches indicate the level of bootstrap support (500 replicates). Support values above 65% are labeled. The scale bar shows the number of substitutions per site.

3.5. Full Genome Comparison

Comparative analysis of gene contents between FDp and eight other phytoplasma genomes (phytoplasmas listed in Table 1 and '*Ca*. P. solani') identified 227 orthologous clusters shared by all phytoplasmas analyzed, including 203 single-copy gene clusters with one protein for each species. The phylogenetic tree based on the concatenated alignment of these proteins confirmed the phylogeny built on smpB, with '*Ca*. P. ziziphi' being the

nearest member and '*Ca*. P. australiense' the most distant one (Figure 6). Moreover, FDp had an average amino acid identity (AAI) of 88.4% and 57.7% when compared to '*Ca*. P. ziziphi' and '*Ca*. P. australiense', respectively.
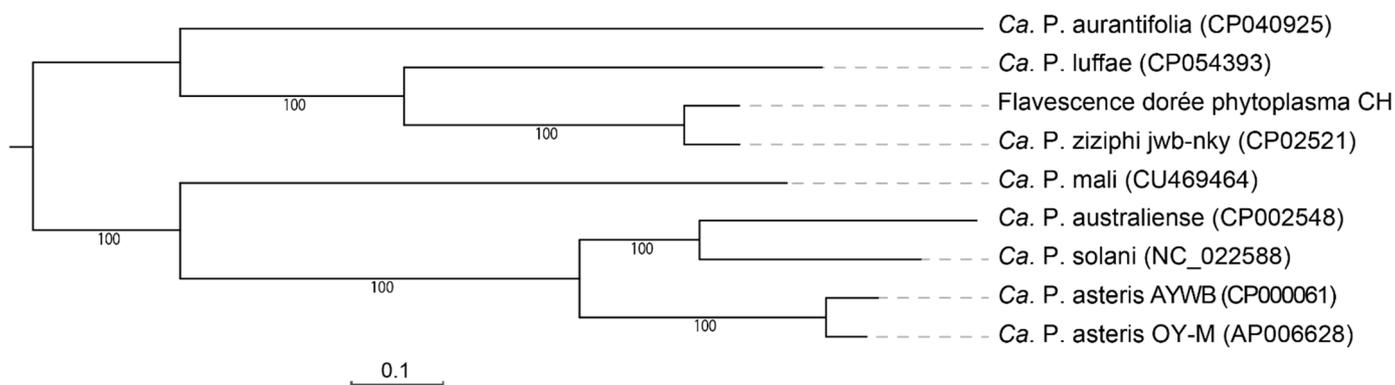

Figure 6. Maximum likelihood phylogeny inferred using 203 single-copy orthologue genes; the concatenated alignment contains 70,619 aligned amino acid sites. The numbers on branches indicate the level of bootstrap support (1000 replicates). Support values above 80% are labeled. The scale bar shows the number of substitutions per site.

Pairwise comparisons showed that FDp and '*Ca*. P. ziziphi' shared the highest number of clusters (Table 3), further corroborating the two phylogenetic analyses. Seven clusters totaling 13 proteins with unknown function were found exclusively in FDp as well as 56 "singletons" (i.e, proteins) with no homologues in other phytoplasma genomes (Supplementary Table S4). The majority of these singletons were hypothetical proteins (46). Notably, seven of them were "putative effectors" and two were associated with a restriction enzyme activity.

Table 3. Pairwise genome comparisons.

	FDp Strain CH	'*Ca*. P. solani'	'*Ca*. P. asteris' OY-M	'*Ca*. P. asteris' AYWB	'*Ca*. P. ziziphi'	'*Ca*. P. mali'	'*Ca*. P. luffae'	'*Ca*. P. aurantifolia'
'*Ca*. P. australiense' (NC_021236)	306	335	332	337	302	301	305	290
'*Ca*. P. aurantifolia' (CP040925)	295	275	282	291	293	295	309	
'*Ca*. P. luffae' (CP054393)	321	300	301	301	322	307		
'*Ca*. P. mali' (NC_011047)	300	299	303	301	302			
'*Ca*. P. ziziphi' jwb-nky (CP02521)	359	296	307	303				
'*Ca*. P. asteris' AYWB (NC_007716)	299	329	384					
'*Ca*. P. asteris' OY-M (AP006628)	302	343						
'*Ca*. P. solani' (GCA_000970375)	304							

The numbers indicate the orthologue clusters between any pair of genomes.

4. Discussion

Despite the small sizes of their genomes, phytoplasmas are challenging organisms to sequence. Because of the low complexity of their chromosomes and the impossibility of axenic culture leading to subsequent excess of host DNA, de novo assemblies often contain unresolved gaps and incomplete genome sequences. The source material for DNA extract is also critical to optimize the abundance of phytoplasma DNA. Thus, various approaches to increase the proportion of pathogen DNA in the extract or the combination of complementary sequencing techniques were recently applied to reconstruct the full genome of several phytoplasmas [36,64]. The "Flavescence dorée" phytoplasma is not an exception

and resisted sequencing efforts for 10 years since the publication of a short communication on a partial assembly in 2011 [65]. Different strategies were tested in our laboratory to achieve whole genome assembly. Sequencing of infected plants with or without Rolling Circle Amplification (RCA) enrichment resulted in numerous gaps and contigs, covering up to a maximum of 80% of the current chromosome. Finally, the successful approach relied on DNA extraction from infected *Euscelidius variegatus* with an enrichment step based on depletion of methylated DNA followed by the combination of both Illumina and ONT sequencing platforms. Using the produced reads, a hybrid strategy with Unicycler led to the assembly of a ~654 kb circular chromosome. Concordantly, the estimated size of the FD92 strain deduced by restriction digests and Pulse-field Gel Electrophoresis (PFGE) is ~670 kb [66].

Quality assessment of the draft genome with BUSCO showed that the sequence encodes for more than 95% of the core genes from the database "Mollicutes" (144/151). Among the seven missing genes, two are actually present (GTPase Era, energy-coupling factor ATPase) but not detected by the software, despite their strong identity with homologues in other phytoplasmas. Interestingly, three genes (Ribose-phosphate pyrophosphokinase, Uracil phosphoribosyltransferase, and Glutamine N5 methyltransferase) are also absent from the phytoplasma genomes listed in Table 1. The gene encoding for Cytidylate kinase is only missing in '*Ca*. P. ziziphi' and '*Ca*. P. luffae', the two phytoplasmas most closely related to FDp. Lastly, Ribonuclease R has been shown to be present in Acholeplasmas but replaced by an inorganic phosphatase in all phytoplasmas [62].

Functional classification of annotated genes revealed that those assigned to KEGG categories "Genetic information processing" (transcription, translation, replication) and "Metabolism" were the most represented. This result is in accordance with previous phytoplasma genome analyses [15,16,67]. A complete set of genes related to glycolysis, either from the upper part (energy demanding) or from the lower part (energy yielding), is present in the genome of FDp, indicating this pathway as the major source of energy. This was already suggested for other phytoplasmas [68,69]. The source of sugars could be sucrose and trehalose, the two major sugar types in plant phloem and insect haemolymph, respectively [70,71]. Maltose ABC-type transporters have been proposed to import these sugars into the cytoplasm [67,68]. An alternative pathway with malate as a carbon source for ATP synthesis is plausible as well. Indeed, three malate/citrate symporters, a malate dehydrogenase, and a phosphate propanoyl transferase were identified in the genome, and malate has been proposed to be a predominant organic acid in the phloem sap [72]. Interestingly, four genes involved in sugar transport are present in the 10 kb duplicated region (2 permeases and 2 substrate-binding proteins). These proteins are members of the Ugp (UgpB1 and UgpC1) and Mal (MalF and MalG) transport systems, which have been shown to be exchangeable in *Escherichia coli* to form a functional periplasmic permease [73]. In addition, a recent study highlighted that the ugp-based system was essential for *Mycobacterium tuberculosis* pathogenicity in humans [74], suggesting that the duplication of this locus in FDp may have a relevance in pathogenicity and virulence of the phytoplasma.

Like other phytoplasmas, FDp harbors a complete sec secretion system to export effectors into host cells [75]. Based on the prediction of signal peptide by SignalP v5.0 and Phobius, 17 putative secreted proteins were identified, among which 14 are devoid of a transmembrane domain and 7 are unique to FDp. This number of predicted effectors is surprisingly low. In comparison, other phytoplasmas (e.g., '*Ca*. P. asteris', '*Ca*. P. ziziphi') usually contain up to five times more candidates in this category [18,50]. Fourteen of them encode a hypothetical protein and only one is a SAP-like protein [18]. Notably, no homologues of SAP05, SAP11, SAP54, and TENGU were found. These proteins produce important morphology changes such as phyllody- or witches' broom-like symptoms in plants infected with various phytoplasmas [23,25,76–79]. Accordingly, vines infected with FDp do not present these types of symptoms. Of interest are also the five genes annotated as hypothetical transmembrane proteins (Htmps) identified in the "Flavescence dorée" chromosome that may mediate interactions with their host environment. Indeed, VmpA

mediates interaction within the midgut and salivary glands of the insect vector [30,80] and Imp interacts with plant actin [31] as well as protein extracts from several insects [81]. Future work is needed to characterize these potential pathogenic agents and their roles in the development of the disease cycle in plant and/or insect hosts.

Putative transposon-like structures named PMUs are unique to phytoplasmas. It has been suggested that they contribute to the genome plasticity of these bacteria [8,82] and are found in the 19 genome assemblies analyzed by Huang and coworkers [16]. In some species, these PMUs account for up to 25% of the genome [16,83]. These regions play a key role in the variance of genome sizes among phytoplasmas. Strikingly, no PMUs were identified on the chromosome of FDp, confirming preliminary results from Carle and colleagues [65]. Although some core PMU genes are present in the sequence, they are scattered throughout the chromosome. Genome instability can also be the consequence of ancient phage integration. Wei et al. [57] determined that large portions of the genome of '*Ca*. P. asteris' strains OY-M and AYWB represented cryptic phage-derived genes (31% and 22.7% respectively). A Prophage Finder tool identified one partial region of putative phage origin in the chromosome of FDp representing less than 2.5% of the total size. Among the 105 proteins with an ATP binding function, sixteen genes encode for FtsH (or HflB) proteins. In *E. coli*, HflB proteins have been shown to prevent prophage integration into the bacterial chromosome by efficient degradation of the λ CII regulatory protein [84]. Moreover, several genes with restriction enzyme activity are predicted in the chromosome and may have played a role in protecting the genome from an integration event following phage infection. However, restriction enzyme activities do not belong to the core set of phytoplasmas [68]. Taken together, these findings suggest that the Flavescence dorée phytoplasma is well armed against phage attacks and is possibly less prone to gene rearrangement than other genomes of the lineage. A regular GC skew pattern (Figure S2) also supports this hypothesis. Remarkably, the genome of '*Ca*. P. solani', a phytoplasma producing the same symptoms in grapevine, shows a completely different layout with a high degree of plasticity, as evidenced by large genome size, numerous gene duplications, and irregular GC skew [67]. It is hypothesized that the genome plasticity favors the spread by multiple insect vectors and infection of a larger diversity of host plants. Small genome size of phytoplasmas is considered to be a consequence of the reductive evolution favored by high intracellular specialization. However, this process is suspected to be balanced by genome plasticity mostly driven by PMUs, leading to gene duplications. Mechanisms by which FDp has evolved towards a more stable genome structure are yet to be determined. Nevertheless, highly specialized interactions restricted to the *Vitis* sp.-*S. titanus*-FDp pathosystem are thought to be responsible for a rapid evolution of the phytoplasma following the introduction of *Scaphoideus titanus* in Europe, presumably in the early 1900s [6]. The reduced number of effectors identified in FDp and the absence of the most common effectors, which are SAP05, SAP11, SAP54, and TENGU, support this hypothesis.

5. Conclusions

This study provides the assembly of the "Flavescence dorée" phytoplasma complete genome. Comparative analysis demonstrated the existence of conserved features among phytoplasmas as well as specific characteristics of the Flavescence dorée pathogen. A significant finding was that the genome did not contain a PMU-like region and appeared to be poor in rearrangement events. Of particular interest would be the functional study of putative effector genes for a better understanding of the interactions between the pathogen and its hosts. Furthermore, future studies of closely related strains detected on *Alnus* sp., presumed to be the ancestor population of FDp [61], will be essential to reconstruct the lineage of Ca. Phytoplasma sp. causing Grapevine Yellows.

References

1. Hogenhout, S.A.; Oshima, K.; Ammarel, D.; Kakizawa, S.; Kingdom, H.N.; Namba, S. Phytoplasmas: Bacteria that manipulate plants and insects. *Mol. Plant Pathol.* **2008**, *9*, 403–423. [CrossRef] [PubMed]
2. Namba, S. Molecular and biological properties of phytoplasmas. *Proc. Jpn. Acad. Ser. B* **2019**, *95*, 401–418. [CrossRef] [PubMed]
3. Bertaccini, A.; Duduk, B.; Paltrinieri, S.; Contaldo, N. Phytoplasmas and Phytoplasma Diseases: A Severe Threat to Agriculture. *Am. J. Plant Sci.* **2014**, *5*, 1763–1788. [CrossRef]
4. Lee, I.M.; Davis, R.E.; Gundersen-Rindal, D.E. Phytoplasma: Phytopathogenic mollicutes. *Annu. Rev. Microbiol.* **2000**, *54*, 221–255. [CrossRef] [PubMed]
5. Koinuma, H.; Maejima, K.; Tokuda, R.; Kitazawa, Y.; Nijo, T.; Wei, W.; Kumita, K.; Miyazaki, A.; Namba, S.; Yamaji, Y. Spatiotemporal dynamics and quantitative analysis of phytoplasmas in insect vectors. *Sci. Rep.* **2020**, *10*, 4291. [CrossRef] [PubMed]
6. Trivellone, V.; Dietrich, C.H. Evolutionary Diversification in Insect Vector–Phytoplasma–Plant Associations. *Ann. Entomol. Soc. Am.* **2021**, *114*, 137–150. [CrossRef]
7. Thébaud, G.; Yvon, M.; Alary, R.; Sauvion, N.; Labonne, G. Efficient transmission of 'Candidatus Phytoplasma prunorum' is delayed by eight months due to a long latency in its host-alternating vector. *Phytopathology* **2009**, *99*, 265–273. [CrossRef]
8. Bai, X.; Zhang, J.; Ewing, A.; Miller, S.A.; Jancso Radek, A.; Shevchenko, D.V.; Tsukerman, K.; Walunas, T.; Lapidus, A.; Campbell, J.W.; et al. Living with genome instability: The adaptation of phytoplasmas to diverse environments of their insect and plant hosts. *J. Bacteriol.* **2006**, *188*, 3682–3696. [CrossRef]
9. Oshima, K.; Ishii, Y.; Kakizawa, S.; Sugawara, K.; Neriya, Y.; Himeno, M.; Minato, N.; Miura, C.; Shiraishi, T.; Yamaji, Y.; et al. Dramatic Transcriptional Changes in an Intracellular Parasite Enable Host Switching between Plant and Insect. *PLoS ONE* **2011**, *6*, e23242. [CrossRef]
10. Jollard, C.; Foissac, X.; Desqué, D.; Razan, F.; Garcion, C.; Beven, L.; Eveillard, S. Flavescence Dorée Phytoplasma Has Multiple ftsH Genes that Are Differentially Expressed in Plants and Insects. *Int. J. Mol. Sci.* **2019**, *21*, 150. [CrossRef]
11. Doi, Y.; Teranaka, M.; Yora, K.; Asuyama, H. 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. Phytopath. Soc. Jpn.* **1967**, *33*, 259–266. [CrossRef]
12. Wei, W.; Lee, I.-M.; Davis, R.E.; Suo, X.; Zhao, Y. Automated RFLP pattern comparison and similarity coefficient calculation for rapid delineation of new and distinct phytoplasma 16Sr subgroup lineages. *Int. J. Syst. Evol. Microbiol.* **2008**, *58*, 2368–2377. [CrossRef] [PubMed]

13. Zhao, Y.; Davis, R.E. Criteria for phytoplasma 16Sr group/subgroup delineation and the need of a platform for proper registration of new groups and subgroups. *Int. J. Syst. Evol. Microbiol.* **2016**, *66*, 2121–2123. [CrossRef]

14. Zhao, Y.; Wei, W.; Lee, I.-M.; Shao, J.; Suo, X.; Davis, R.E. Construction of an interactive online phytoplasma classification tool, iPhyClassifier, and its application in analysis of the peach X-disease phytoplasma group (16SrIII). *Int. J. Syst. Evol. Microbiol.* **2009**, *59*, 2582–2593. [CrossRef] [PubMed]

15. Chung, W.-C.; Chen, L.-L.; Lo, W.-S.; Lin, C.-P.; Kuo, C.-H. Comparative analysis of the peanut witches'-broom phytoplasma genome reveals horizontal transfer of potential mobile units and effectors. *PLoS ONE* **2013**, *8*, e62770. [CrossRef]

16. Huang, C.T.; Cho, S.T.; Lin, Y.C.; Tan, C.M.; Chiu, Y.C.; Yang, J.Y.; Kuo, C.H. Comparative Genome Analysis of '*Candidatus* Phytoplasma luffae' Reveals the Influential Roles of Potential Mobile Units in Phytoplasma Evolution. *Front. Microbiol.* **2022**, *13*, 773608. [CrossRef]

17. Arashida, R.; Kakizawa, S.; Hoshi, A.; Ishii, Y.; Jung, H.Y.; Kagiwada, S.; Yamaji, Y.; Oshima, K.; Namba, S. Heterogeneic dynamics of the structures of multiple gene clusters in two pathogenetically different lines originating from the same phytoplasma. *DNA Cell Biol.* **2008**, *27*, 209–217. [CrossRef]

18. Bai, X.; Correa, V.R.; Toruño, T.Y.; Ammar, E.D.; Kamoun, S.; Hogenhout, S.A. AY-WB phytoplasma secretes a protein that targets plant cell nuclei. *Mol. Plant-Microbe Interact.* **2009**, *22*, 18–30. [CrossRef]

19. Anabestani, A.; Izadpanah, K.; Abbà, S.; Galetto, L.; Ghorbani, A.; Palmano, S.; Siampour, M.; Veratti, F.; Marzachì, C. Identification of putative effector genes and their transcripts in three strains related to '*Candidatus* Phytoplasma aurantifolia'. *Microbiol. Res.* **2017**, *199*, 57–66. [CrossRef]

20. Deng, M.; Ma, F.; Zhang, X.; Huang, J.; Yang, J.; Chen, M.; Zhou, J.; Sun, Q.; Sun, J. Genome-wide identification of jujube witches' broom phytoplasma effectors revealed the role of SJP3 in inducing phyllody. *Sci. Hortic.* **2021**, *290*, 110548. [CrossRef]

21. Sugio, A.; Kingdom Heather, N.; MacLean Allyson, M.; Grieve Victoria, M.; Hogenhout Saskia, A. Phytoplasma protein effector SAP11 enhances insect vector reproduction by manipulating plant development and defense hormone biosynthesis. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, E1254–E1263. [CrossRef] [PubMed]

22. Zhou, J.; Ma, F.; Yao, Y.; Deng, M.; Chen, M.; Zhang, S.; Li, Y.; Yang, J.; Zhang, N.; Huang, J.; et al. Jujube witches' broom phytoplasma effectors SJP1 and SJP2 induce lateral bud outgrowth by repressing the ZjBRC1-controlled auxin efflux channel. *Plant Cell Environ.* **2021**, *44*, 3257–3272. [CrossRef] [PubMed]

23. Tan, C.M.; Li, C.H.; Tsao, N.W.; Su, L.W.; Lu, Y.T.; Chang, S.H.; Lin, Y.Y.; Liou, J.C.; Hsieh, L.C.; Yu, J.Z.; et al. Phytoplasma SAP11 alters 3-isobutyl-2-methoxypyrazine biosynthesis in Nicotiana benthamiana by suppressing NbOMT1. *J. Exp. Bot.* **2016**, *67*, 4415–4425. [CrossRef] [PubMed]

24. Chang, S.H.; Tan, C.M.; Wu, C.T.; Lin, T.H.; Jiang, S.Y.; Liu, R.C.; Tsai, M.C.; Su, L.W.; Yang, J.Y. Alterations of plant architecture and phase transition by the phytoplasma virulence factor SAP11. *J. Exp. Bot.* **2018**, *69*, 5389–5401. [CrossRef] [PubMed]

25. Huang, W.; MacLean, A.M.; Sugio, A.; Maqbool, A.; Busscher, M.; Cho, S.T.; Kamoun, S.; Kuo, C.H.; Immink, R.G.H.; Hogenhout, S.A. Parasitic modulation of host development by ubiquitin-independent protein degradation. *Cell* **2021**, *184*, 5201–5214.e12. [CrossRef]

26. Orlovskis, Z.; Hogenhout, S.A. A Bacterial Parasite Effector Mediates Insect Vector Attraction in Host Plants Independently of Developmental Changes. *Front. Plant Sci.* **2016**, *7*, 885. [CrossRef]

27. Hoshi, A.; Oshima, K.; Kakizawa, S.; Ishii, Y.; Ozeki, J.; Hashimoto, M.; Komatsu, K.; Kagiwada, S.; Yamaji, Y.; Namba, S. A unique virulence factor for proliferation and dwarfism in plants identified from a phytopathogenic bacterium. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 6416–6421. [CrossRef]

28. Tomkins, M.; Kliot, A.; Marée, A.F.M.; Hogenhout, S.A. A multi-layered mechanistic modelling approach to understand how effector genes extend beyond phytoplasma to modulate plant hosts, insect vectors and the environment. *Curr. Opin. Plant Biol.* **2018**, *44*, 39–48. [CrossRef]

29. Garcion, C.; Béven, L.; Foissac, X. Comparison of Current Methods for Signal Peptide Prediction in Phytoplasmas. *Front. Microbiol.* **2021**, *12*, 661524. [CrossRef]

30. Arricau-Bouvery, N.; Duret, S.; Dubrana, M.P.; Batailler, B.; Desque, D.; Beven, L.; Danet, J.L.; Monticone, M.; Bosco, D.; Malembic-Maher, S.; et al. Variable Membrane Protein A of Flavescence Doree Phytoplasma Binds the Midgut Perimicrovillar Membrane of *Euscelidius variegatus* and Promotes Adhesion to Its Epithelial Cells. *Appl. Environ. Microbiol.* **2018**, *84*, e02487-17. [CrossRef]

31. Boonrod, K.; Munteanu, B.; Jarausch, B.; Jarausch, W.; Krczal, G. An immunodominant membrane protein (Imp) of '*Candidatus* Phytoplasma mali' binds to plant actin. *Mol. Plant Microbe Interact.* **2012**, *25*, 889–895. [CrossRef] [PubMed]

32. Galetto, L.; Bosco, D.; Balestrini, R.; Genre, A.; Fletcher, J.; Marzachi, C. The major antigenic membrane protein of "*Candidatus* Phytoplasma asteris" selectively interacts with ATP synthase and actin of leafhopper vectors. *PLoS ONE* **2011**, *6*, e22571. [CrossRef] [PubMed]

33. Rashidi, M.; Galetto, L.; Bosco, D.; Bulgarelli, A.; Vallino, M.; Veratti, F.; Marzachi, C. Role of the major antigenic membrane protein in phytoplasma transmission by two insect vector species. *BMC Microbiol.* **2015**, *15*, 193. [CrossRef] [PubMed]

34. Chuche, J.; Thiéry, D. Biology and ecology of the Flavescence dorée vector *Scaphoideus titanus*: A review. *Agron. Sustain. Dev.* **2014**, *34*, 381–403. [CrossRef]

35. EU. COUNCIL DIRECTIVE 2000/29/EC of 8 May 2000 on Protective Measures against the Introduction into the Community of Organisms Harmful to Plants or Plant Products and against Their Spread within the Community. *OJI* **2000**, *169*, 2000.

36. Nijo, T.; Iwabuchi, N.; Tokuda, R.; Suzuki, T.; Matsumoto, O.; Miyazaki, A.; Maejima, K.; Oshima, K.; Namba, S.; Yamaji, Y. Enrichment of phytoplasma genome DNA through a methyl-CpG binding domain-mediated method for efficient genome sequencing. *J. Gen. Plant Pathol.* **2021**, *87*, 154–163. [CrossRef]

37. Pelletier, C.; Salar, P.; Gillet, J.; Cloquemin, G.; Very, P.; Foissac, X.; Malembic-Maher, S. Triplex real-time PCR assay for sensitive and simultaneous detection of grapevine phytoplasmas of the 16SrV and 16SrXII-A groups with an endogenous analytical control. *Vitis* **2009**, *48*, 87–95. [CrossRef]

38. Ripamonti, M.; Maron, F.; Cornara, D.; Marzachi, C.; Fereres, A.; Bosco, D. Leafhopper feeding behaviour on three grapevine cultivars with different susceptibilities to Flavescence doree. *J. Insect. Physiol.* **2022**, *137*, 104366. [CrossRef]

39. Manni, M.; Berkeley, M.R.; Seppey, M.; Simao, F.A.; Zdobnov, E.M. BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes. *Mol. Biol. Evol.* **2021**, *38*, 4647–4654. [CrossRef]

40. Rodriguez, R.L.; Gunturu, S.; Harvey, W.T.; Rossello-Mora, R.; Tiedje, J.M.; Cole, J.R.; Konstantinidis, K.T. The Microbial Genomes Atlas (MiGA) webserver: Taxonomic and gene diversity analysis of Archaea and Bacteria at the whole genome level. *Nucleic Acids Res.* **2018**, *46*, W282–W288. [CrossRef]

41. Seemann, T. Prokka: Rapid prokaryotic genome annotation. *Bioinformatics* **2014**, *30*, 2068–2069. [CrossRef] [PubMed]

42. Araujo, F.A.; Barh, D.; Silva, A.; Guimaraes, L.; Ramos, R.T.J. GO FEAT: A rapid web-based functional annotation tool for genomic and transcriptomic data. *Sci. Rep.* **2018**, *8*, 1794. [CrossRef] [PubMed]

43. Kanehisa, M.; Sato, Y.; Morishima, K. BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences. *J. Mol. Biol.* **2016**, *428*, 726–731. [CrossRef] [PubMed]

44. Almagro Armenteros, J.J.; Tsirigos, K.D.; Sonderby, C.K.; Petersen, T.N.; Winther, O.; Brunak, S.; von Heijne, G.; Nielsen, H. SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat. Biotechnol.* **2019**, *37*, 420–423. [CrossRef] [PubMed]

45. Kall, L.; Krogh, A.; Sonnhammer, E.L. Advantages of combined transmembrane topology and signal peptide prediction–the Phobius web server. *Nucleic Acids Res.* **2007**, *35*, W429–W432. [CrossRef]

46. Gu, Z.; Gu, L.; Eils, R.; Schlesner, M.; Brors, B. Circlize Implements and enhances circular visualization in R. *Bioinformatics* **2014**, *30*, 2811–2812. [CrossRef]

47. Xu, L.; Dong, Z.; Fang, L.; Luo, Y.; Wei, Z.; Guo, H.; Zhang, G.; Gu, Y.Q.; Coleman-Derr, D.; Xia, Q.; et al. OrthoVenn2: A web server for whole-genome comparison and annotation of orthologous clusters across multiple species. *Nucleic Acids Res.* **2019**, *47*, W52–W58. [CrossRef]

48. Trifinopoulos, J.; Nguyen, L.T.; von Haeseler, A.; Minh, B.Q. W-IQ-TREE: A fast online phylogenetic tool for maximum likelihood analysis. *Nucleic Acids Res.* **2016**, *44*, W232–W235. [CrossRef]

49. Letunic, I.; Bork, P. Interactive Tree Of Life (iTOL) v5: An online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* **2021**, *49*, W293–W296. [CrossRef]

50. Wang, J.; Song, L.; Jiao, Q.; Yang, S.; Gao, R.; Lu, X.; Zhou, G. Comparative genome analysis of jujube witches'-broom Phytoplasma, an obligate pathogen that causes jujube witches'-broom disease. *BMC Genom.* **2018**, *19*, 689. [CrossRef]

51. Razin, S.; Yogev, D.; Naot, Y. Molecular biology and pathogenicity of mycoplasmas. *Microbiol. Mol. Biol. Rev.* **1998**, *62*, 1094–1156. [CrossRef] [PubMed]

52. Kube, M.; Schneider, B.; Kuhl, H.; Dandekar, T.; Heitmann, K.; Migdoll, A.M.; Reinhardt, R.; Seemuller, E. The linear chromosome of the plant-pathogenic mycoplasma '*Candidatus* Phytoplasma mali'. *BMC Genom.* **2008**, *9*, 306. [CrossRef]

53. Kanehisa, M.; Goto, S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **2000**, *28*, 27–30. [CrossRef] [PubMed]

54. Crane, J.M.; Randall, L.L. The Sec System: Protein Export in *Escherichia coli*. *EcoSal Plus* **2017**, *7*. [CrossRef] [PubMed]

55. Seemuller, E.; Sule, S.; Kube, M.; Jelkmann, W.; Schneider, B. The AAA+ ATPases and HflB/FtsH proteases of '*Candidatus* Phytoplasma mali': Phylogenetic diversity, membrane topology, and relationship to strain virulence. *Mol. Plant Microbe Interact.* **2013**, *26*, 367–376. [CrossRef]

56. Jomantiene, R.; Zhao, Y.; Davis, R.E. Sequence-variable mosaics: Composites of recurrent transposition characterizing the genomes of phylogenetically diverse phytoplasmas. *DNA Cell Biol.* **2007**, *26*, 557–564. [CrossRef]

57. Wei, W.; Davis, R.E.; Jomantiene, R.; Zhao, Y. Ancient, recurrent phage attacks and recombination shaped dynamic sequence-variable mosaics at the root of phytoplasma genome evolution. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 11827–11832. [CrossRef]

58. Arndt, D.; Grant, J.R.; Marcu, A.; Sajed, T.; Pon, A.; Liang, Y.; Wishart, D.S. PHASTER: A better, faster version of the PHAST phage search tool. *Nucleic Acids Res.* **2016**, *44*, W16–W21. [CrossRef]

59. Lee, I.-M.; Gundersen-Rindal, D.; Davis, R.E.; Bartoszyk, I. Revised Classification Scheme of Phytoplasmas based on RFLP Analyses of 16S rRNA and Ribosomal Protein Gene Sequences. *Int. J. Syst. Evol. Microbiol.* **1998**, *48*, 1153–1169. [CrossRef]

60. Lee, I.-M.; Hammond, R.E.; Davis, R.E.; Gundersen, D.E. Universal Amplification and Analysis of Pathogen 16S rDNA for Classification and Identification of Mycoplasmalike Organisms. *Phytopathology* **1993**, *83*, 834–842. [CrossRef]

61. Malembic-Maher, S.; Desque, D.; Khalil, D.; Salar, P.; Bergey, B.; Danet, J.L.; Duret, S.; Dubrana-Ourabah, M.P.; Beven, L.; Ember, I.; et al. When a Palearctic bacterium meets a Nearctic insect vector: Genetic and ecological insights into the emergence of the grapevine Flavescence doree epidemics in Europe. *PLoS Pathog.* **2020**, *16*, 244–248. [CrossRef] [PubMed]

62. Zubert, C.; Ilic, A.M.; Duduk, B.; Kube, M. The Genome Reduction Excludes the Ribosomal Rescue System in *Acholeplasmataceae*. *Microb. Physiol.* **2022**, *32*, 45–56. [CrossRef] [PubMed]

63. Bertaccini, A.; Arocha-Rosete, Y.; Contaldo, N.; Duduk, B.; Fiore, N.; Montano, H.G.; Kube, M.; Kuo, C.H.; Martini, M.; Oshima, K.; et al. Revision of the '*Candidatus* Phytoplasma' species description guidelines. *Int. J. Syst. Evol. Microbiol.* **2022**, *72*, 005353. [CrossRef] [PubMed]

64. Tan, C.M.; Lin, Y.C.; Li, J.R.; Chien, Y.Y.; Wang, C.J.; Chou, L.; Wang, C.W.; Chiu, Y.C.; Kuo, C.H.; Yang, J.Y. Accelerating Complete Phytoplasma Genome Assembly by Immunoprecipitation-Based Enrichment and MinION-Based DNA Sequencing for Comparative Analyses. *Front. Microbiol.* **2021**, *12*, 766221. [CrossRef] [PubMed]

65. Carle, P.; Malembic-Maher, S.; Arricau-Bouvery, N.; Desque, D.; Eveillard, S.; Carrere, S.; Foissac, X. 'Flavescence dorée' phytoplasma genome: A metabolism oriented towards glycolysis and protein degradation. *Bull. Insectol.* **2011**, *64*, S13–S14.

66. Malembic-Maher, S.; Constable, F.; Cimerman, A.; Arnaud, G.; Carle, P.; Foissac, X.; Boudon-Padieu, E. A chromosome map of the Flavescence doree phytoplasma. *Microbiology* **2008**, *154*, 1454–1463. [CrossRef] [PubMed]

67. Music, M.S.; Samarzija, I.; Hogenhout, S.A.; Haryono, M.; Cho, S.T.; Kuo, C.H. The genome of '*Candidatus* Phytoplasma solani' strain SA-1 is highly dynamic and prone to adopting foreign sequences. *Syst. Appl. Microbiol.* **2019**, *42*, 117–127. [CrossRef]

68. Kube, M.; Mitrovic, J.; Duduk, B.; Rabus, R.; Seemuller, E. Current view on phytoplasma genomes and encoded metabolism. *Sci. World J.* **2012**, *2012*, 185942. [CrossRef]

69. Oshima, K.; Kakizawa, S.; Nishigawa, H.; Jung, H.Y.; Wei, W.; Suzuki, S.; Arashida, R.; Nakata, D.; Miyata, S.; Ugaki, M.; et al. Reductive evolution suggested from the complete genome sequence of a plant-pathogenic phytoplasma. *Nat. Genet.* **2004**, *36*, 27–29. [CrossRef]

70. Becker, A.; Schloder, P.; Steele, J.E.; Wegener, G. The regulation of trehalose metabolism in insects. *Experientia* **1996**, *52*, 433–439. [CrossRef]

71. Lemoine, R.; La Camera, S.; Atanassova, R.; Dedaldechamp, F.; Allario, T.; Pourtau, N.; Bonnemain, J.L.; Laloi, M.; Coutos-Thevenot, P.; Maurousset, L.; et al. Source-to-sink transport of sugar and regulation by environmental factors. *Front. Plant Sci.* **2013**, *4*, 272. [CrossRef] [PubMed]

72. Zimmermann, M.; Ziegler, H. *Encyclopedia of Plant Physiology, Transport in Plants 1, Phloem Transport*; Springer: Berlin/Heidelberg, Germany, 1975; Volume 1.

73. Hekstra, D.; Tommassen, J. Functional exchangeability of the ABC proteins of the periplasmic binding protein-dependent transport systems Ugp and Mal of Escherichia coli. *J. Bacteriol.* **1993**, *175*, 6546–6552. [CrossRef] [PubMed]

74. De la Torre, L.I.; Vergara Meza, J.G.; Cabarca, S.; Costa-Martins, A.G.; Balan, A. Comparison of carbohydrate ABC importers from Mycobacterium tuberculosis. *BMC Genom.* **2021**, *22*, 841. [CrossRef] [PubMed]

75. Economou, A. Following the leader: Bacterial protein export through the Sec pathway. *Trends Microbiol.* **1999**, *7*, 315–320. [CrossRef]

76. MacLean, A.M.; Orlovskis, Z.; Kowitwanich, K.; Zdziarska, A.M.; Angenent, G.C.; Immink, R.G.; Hogenhout, S.A. Phytoplasma effector SAP54 hijacks plant reproduction by degrading MADS-box proteins and promotes insect colonization in a RAD23-dependent manner. *PLoS Biol.* **2014**, *12*, e1001835. [CrossRef]

77. MacLean, A.M.; Sugio, A.; Makarova, O.V.; Findlay, K.C.; Grieve, V.M.; Toth, R.; Nicolaisen, M.; Hogenhout, S.A. Phytoplasma effector SAP54 induces indeterminate leaf-like flower development in Arabidopsis plants. *Plant Physiol.* **2011**, *157*, 831–841. [CrossRef]

78. Pecher, P.; Moro, G.; Canale, M.C.; Capdevielle, S.; Singh, A.; MacLean, A.; Sugio, A.; Kuo, C.H.; Lopes, J.R.S.; Hogenhout, S.A. Phytoplasma SAP11 effector destabilization of TCP transcription factors differentially impact development and defence of Arabidopsis versus maize. *PLoS Pathog.* **2019**, *15*, e1008035. [CrossRef]

79. Sugawara, K.; Honma, Y.; Komatsu, K.; Himeno, M.; Oshima, K.; Namba, S. The alteration of plant morphology by small peptides released from the proteolytic processing of the bacterial peptide TENGU. *Plant Physiol.* **2013**, *162*, 2005–2014. [CrossRef]

80. Arricau-Bouvery, N.; Duret, S.; Dubrana, M.P.; Desque, D.; Eveillard, S.; Brocard, L.; Malembic-Maher, S.; Foissac, X. Interactions between the flavescence doree phytoplasma and its insect vector indicate lectin-type adhesion mediated by the adhesin VmpA. *Sci. Rep.* **2021**, *11*, 11222. [CrossRef]

81. Trivellone, V.; Ripamonti, M.; Angelini, E.; Filippin, L.; Rossi, M.; Marzachi, C.; Galetto, L. Evidence suggesting interactions between immunodominant membrane protein Imp of Flavescence doree phytoplasma and protein extracts from distantly related insect species. *J. Appl. Microbiol.* **2019**, *127*, 1801–1813. [CrossRef]

82. Toruno, T.Y.; Music, M.S.; Simi, S.; Nicolaisen, M.; Hogenhout, S.A. Phytoplasma PMU1 exists as linear chromosomal and circular extrachromosomal elements and has enhanced expression in insect vectors compared with plant hosts. *Mol. Microbiol.* **2010**, *77*, 1406–1415. [CrossRef] [PubMed]

83. Andersen, M.T.; Liefting, L.W.; Havukkala, I.; Beever, R.E. Comparison of the complete genome sequence of two closely related isolates of '*Candidatus* Phytoplasma australiense' reveals genome plasticity. *BMC Genom.* **2013**, *14*, 529. [CrossRef] [PubMed]

84. Shotland, Y.; Shifrin, A.; Ziv, T.; Teff, D.; Koby, S.; Kobiler, O.; Oppenheim, A.B. Proteolysis of bacteriophage lambda CII by Escherichia coli FtsH (HflB). *J. Bacteriol.* **2000**, *182*, 3111–3116. [CrossRef] [PubMed]