Phytopathologia Mediterranea
The international journal of the
Mediterranean Phytopathological Union

Firenze University Press
www.fupress.com/pm

Research Papers

Ribosomal protein coding genes *SSU12p* and *LSU36p* as molecular markers for phytoplasma detection and differentiation

Weier CUI[1], Alan ZAMORANO[1], Nicolás QUIROGA[1], Assunta BERTACCINI[2], Nicola FIORE[1,*]

[1] *University of Chile, Faculty of Agronomic Sciences, Department of Plant Health, Santiago, Chile. Av. Santa Rosa 11315, Santiago, Chile*
[2] *Alma Mater Studiorum - University of Bologna, Department of Agricultural and Food Sciences, viale G. Fanin 40, 40127, Bologna, Italy*
*Corresponding author. E-mail: nfiore@uchile.cl
Shared co-first authorship: Weier Cui and Alan Zamorano

Summary. Detection and classification of phytoplasmas mainly rely on amplification of the 16S rRNA gene followed by RFLP analysis and/or sequencing, because these organisms lack complete phenotypic characterization. Other conserved genomic *loci* have been exploited as additional molecular markers for phytoplasma differentiation. Two *loci*, *SSU12p* and *LSU36p*, selected by whole-genome comparison of 12 phytoplasma strains, were used for primer design, and were successfully tested on DNA samples from plants infected by phytoplasmas belonging to ten 16S ribosomal groups. The phylogenetic trees inferred from *SSU12p* and *LSU36p loci* were highly congruent to the trees derived from 16S rRNA and *tuf* genes of the same phytoplasma strains. Virtual RFLP analysis of the amplified *SSU12p* gene showed distinct patterns for most of the phytoplasma ribosomal subgroups tested. These results show that *SSU12p* and *LSU36p* genes are reliable additional markers for phytoplasma detection and differentiation.

Keywords. PCR, 16S rRNA gene, *tuf* gene, RFLP.

OPEN ACCESS

Citation: W. Cui, A. Zamorano, N. Quiroga, A. Bertaccini, N. Fiore (2021) Ribosomal protein coding genes *SSU12p* and *LSU36p* as molecular markers for phytoplasma detection and differentiation. *Phytopathologia Mediterranea* 60(2): 281-292. doi: 10.36253/phyto-11993

Accepted: July 7, 2021

Published: September 13, 2021

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Competing Interests: The Author(s) declare(s) no conflict of interest.

Editor: Nihal Buzkan, Kahramanmaraş Sütçü Imam University, Turkey.

INTRODUCTION

Phytoplasmas are obligate intracellular pathogens that reside and multiply in the phloem tissues of plants and in insect hosts. They are associated with severe diseases of economically important plants, including aster yellows, coconut lethal yellowing, apple proliferation, pear decline, peach X disease and ash yellows. Australian grapevine yellows, which is associated with three phytoplasmas, causes up to 54% yield losses (Glenn, 2000). In Brazil, yield losses caused by maize bushy stunt are estimated to be worth $US 16.5 million (Oliveira *et al.*, 2003). Due to the difficulty to culture phytoplasmas (Contaldo and Bertaccini, 2019) and the lack of a complete phenotypic characterization of these organisms, phytoplasmas classification is based on their 16S rRNA gene sequences, that are conserved and widely used for prokary-

ote identification (Lee *et al*., 1993; Ludwig and Schleifer, 1994; Seemüller *et al*., 1994; Schneider *et al*, 1995; Jenkins *et al*., 2012). A provisional naming system (IRPCM, 2004) assigned '*Candidatus* Phytoplasma' species to strains whose 16S rRNA gene sequence has less than 97.5% identity to any previously described '*Ca*. Phytoplasma' species. A set of 17 restriction enzymes was selected to generate the restriction fragment length polymorphism (RFLP) profile of the R16F2n/R2 fragment of the 16S rRNA gene. By this approach, 16Sr groups and subgroups have been identified (Lee *et al*., 1998; Wei *et al*., 2008). The '*Ca*. Phytoplasma' species and the RFLP-generated ribosomal groups and subgroups are therefore the two approaches used to classify these prokaryotes. However, one 16S ribosomal group may contain one or more '*Ca*. Phytoplasma' species, whereas all the strains within one ribosomal subgroup belong to the same '*Ca*. Phytoplasma' species (Bertaccini and Lee, 2018).

Considering the stringency of the 16S rRNA gene in assigning '*Ca*. Phytoplasma' species, there are limitations in differentiating closely related strains, so other *loci* have been described and utilized as additional molecular markers for phytoplasma strain differentiation. Other markers have been used in phytoplasma phylogenetic studies, including the 16S-23S intergenic spacer, the 23S rRNA gene, the ribosomal protein operon (*rp19-rpl22-rps3*), the elongation factor Tu (*tuf*), protein translocase units (*secA* and *secY*), the chaperonin 60 (*cpn60*), and the subunit β of RNA polymerase (*rpoB*) (Marcone *et al*., 2000; Martini *et al*., 2002, 2007; Hodgetts *et al*., 2008; Lee *et al*., 2010; Makarova *et al*., 2012; Valiunas *et al*., 2013). The methionine aminopeptidase gene (*map*)-*uvrB-degV*, *nusA* and *vmp1* was also used for differentiation of strains within, respectively, the 16SrV, 16SrI and 16SrX-II-A groups and subgroups (Shao *et al*., 2006; Arnaud *et al*., 2007; Cimerman *et al*., 2009). All these markers except *vmp1* have also been used for differentiation of other bacteria (Pérez-López *et al*., 2016), confirming the suitability of a gene-based strategy also for phytoplasma strain differentiation.

Besides providing classification, the 16S rRNA gene also serves as the most important detection marker for phytoplasmas. Several sets of primers have been designed to amplify different fragments from this gene. The combination of P1/P7 and R16F2n/R2 is the most employed for phytoplasma detection, but other primer sets as well as ribosomal group-specific primers are useful for detection of multiple phytoplasma infections and/or heterogeneous phytoplasma populations (Duduk *et al*., 2013). Other *loci* are also used as detection markers, including *tuf*, *rpoB*, *cpn60*, *nusA* and *vmp1* (Shao *et al*., 2006; Cimerman *et al*., 2009; Makarova *et al*., 2012; Valiunas *et al*., 2013; Dumonceaux *et al*., 2014). However, the lack of universal primers (*rpoB*, *secY*, *rp*), the narrow detection range (*nusA*, *vmp1*, *map-uvrB-degV*) and the high rate of false positives (*cpn60*) severely reduce their detection efficiency.

In the present study, new molecular markers for phytoplasma detection and differentiation were designed and tested. Using whole genome comparisons, the phytoplasma genome conserved regions *SSU12p* and *LSU36p* were selected for primer design, and tested to verify their usefulness as molecular markers for a range of phytoplasma strains.

MATERIALS AND METHODS

Phytoplasma strains and nucleic acid preparation

Thirty-three phytoplasma strains collected from various host plant species from different geographic regions worldwide were used. The strains were identified on their 16S ribosomal DNA (Cui *et al*., 2019; EPPO-Q-bank, 2020), and belong to the 16Sr groups: -I, -II, -III, -V, -VI, -VII, -IX, -X, -XII and -XIII. The strain names, acronyms, 16Sr groups/subgroups, and providers are listed in Table 1. The DNAs from strawberry plants infected by the StrPh-CL strains were extracted as described by Cui *et al.* (2019). DNA samples provided by EPPO-Q-bank were extracted as described by Makarova *et al.* (2012).

Primer design

Direct and nested PCR primers were designed by comparing the conserved genomic regions of 12 phytoplasma strains available from the GenBank, including those associated with the diseases aster yellows witches' broom (AYWB) (CP000061), onion yellows mild strain (OYM) (NC_005303), peanut witches' broom (PnWB) strain NTU2011 (AMWZ00000000), *Echinacea purpurea* witches' broom (*E. purpurea* WB) strain NCHU2014 (LKAC00000000), Italian clover phyllody (ItClPh) strain MA1 (AKIM00000000), *Vaccinium* witches' broom (VacWB) strain VAC (AKIN00000000), milkweed yellows (MWY) strain MW1 (AKIL00000000), poinsettia branch-inducing phytoplasma (PoiBI) strain JR1 (AKIK00000000), '*Ca*. P. mali' strain AT (CU469464), '*Ca*. P. australiense' (AUSGY) (AM422018), strawberry lethal yellows phytoplasma (CPA) strain NZSb11 (CP002548), and phytoplasma Vc33 (LLKK00000000). Whole-genome comparison was performed with the "Sequence-based comparison" tool on the Rapid Annotation using the Subsystem Technology (RAST) server

Table 1. Phytoplasma strains used in this study, and their corresponding *SSU12p* and *LSU36p* amplicons.

16Sr group	Subgroup	Associated disease	Acronym	Provided by	Amplicon size[a]		GenBank Accession Number	
					SSU12p	LSU36p	SSU12p	LSU36p
16SrI	A	Aster yellows witches' broom	AYWB	GenBank	748	465	CP000061	
	B	*Catharanthus* virescence	CTV	EPPO-Q-bank	766	468	MT161513	MT161546
	B	Onion yellows mild strain	OY-M	GenBank	766	461	NC_005303	
	B	Primula yellows	PRIVA	EPPO-Q-bank	766	461	MT161512	MT161545
	B	Maize bushy stunt	MBS	GenBank	766	461	CP015149	
	C	Clover phyllody	KVE	EPPO-Q-bank	758	462	MT161514	MT161547
	F	Aster yellows from apricot	A-AY	EPPO-Q-bank	735	462	MT161515	MT161548
16SrII	A	Peanut witches' broom	PnWB	GenBank	703	411	AMWZ00000000	
	A	*Echinacea purpurea* witches' broom	*E. purpurea* WB	GenBank	703	411	LKAC00000000	
	C	Faba bean phyllody	FBP	EPPO-Q-bank	703	409	MT161516	MT161549
	D	Tomato big bud	TBB	EPPO-Q-bank	703	418	MT161517	MT161550
16SrIII	A	Peach X disease	WX	GenBank	738	417	LHCF00000000	
	A	Peach X disease	CX	EPPO-Q-bank	739	417	MT161518	MT161551
	A	Peach yellow leaf roll	PYLR	EPPO-Q-bank	739	417	MT161519	MT161552
	B	Italian clover phyllody	ItClPh	GenBank	740	417	AKIM00000000	
	B	*Taraxacum* leaf reddening	TA	EPPO-Q-bank	740	417	MT161520	MT161553
	B	Plum leptonecrosis	LNI	EPPO-Q-bank	740	417	MT161521	MT161554
	D	Goldenrod yellows	GR	EPPO-Q-bank	740	417	MT161522	MT161555
	E	Spirea stunt	SPI	EPPO-Q-bank	739	420	MT161523	MT161556
	F	*Vaccinium* witches' broom	VacWB	GenBank	740	417	AKIN00000000	
	F	*Solanum* big bud	SBB	EPPO-Q-bank	741	417	MT161524	MT161557
	F	Milkweed yellows	MW1	EPPO-Q-bank	740	417	MT161525	MT161558
	F	Milkweed yellows	MWY	GenBank	740	417	AKIL00000000	
	H	Poinsettia branch inducing	PoiBI	GenBank	739	417	AKIK00000000	
	H	Poinsettia branch inducing	JR	EPPO-Q-bank	739	417	MT161526	MT161559
	J	Phytoplasma Vc33	Vc33	GenBank	740	417	LLKK00000000	
16SrV	A	Elm yellows	EY	EPPO-Q-bank	724	419	MT161527	MT161560
16SrVI	A	Clover proliferation	CP1	EPPO-Q-bank	722	381	MT161528	MT161561
16SrVII	A	Ash yellows	ASHY	EPPO-Q-bank	719	398	MT161529	MT161562
16SrIX	B	Almond witches' broom	SA213	GenBank	723	489	JPSQ00000000	
	C	*Picris echioides* yellows	PEY	EPPO-Q-bank	737	677	MT161530	MT161563
	C	Naxos periwinkle yellows	NAXOS	EPPO-Q-bank	737	677	MT161531	MT161564
16SrX	A	Apple proliferation	AP	EPPO-Q-bank	723	443	MT161532	MT161565
	A	Apple proliferation	AT	GenBank	723	443	CU469464	
	B	European stone fruit yellows	ESFY	EPPO-Q-bank	724	489	MT161533	MT161566

(Continued)

Table 1. (Continued).

16Sr group	Subgroup	Associated disease	Acronym	Provided by	Amplicon size[a]		GenBank Accession Number	
					SSU12p	LSU36p	SSU12p	LSU36p
	B	Plum leptonecrosis	LNp	EPPO-Q-bank	724	487	MT161534	MT161567
	C	Pear decline	PD	EPPO-Q-bank	724	442	MT161535	MT161568
16SrXII	A	Moliére disease	MOL	EPPO-Q-bank	723	406	MT161536	MT161569
	A	Grapevine yellows	SA-1	GenBank	723	406	MPBG00000000	
	A	Grapevine yellows	CH1	EPPO-Q-bank	723	406	MT161537	MT161570
	B	Australian grapevine yellows	AUSGY	GenBank	745	401	AM422018	
	C	Strawberry lethal yellows	CPA	GenBank	745	401	CP002548	
16SrXIII	F	Strawberry phyllody	StrPh-CL1	University of Chile	739	406	MT161538	MT161571
			StrPh-CL3		739	406	MT161540	MT161573
			StrPh-CL5		739	406	MT161542	MT161575
			StrPh-CL6		739	406	MT161543	MT161576
			StrPh-CL7		739	406	MT161544	MT161577
	K	Strawberry phyllody	StrPh-CL2	University of Chile	737	406	MT161539	MT161572
			StrPh-CL4		737	406	MT161541	MT161574

[a]bp, the length of the primers is not included.

(http://rast.nmpdr.org/rast.cgi). The annotated genes were sorted based on similarity, and four genes with the greatest similarities, except for the 16S rRNA, were selected for primer design. These were: the small subunit ribosomal protein S12p (*SSU12p*), and the large subunit ribosomal proteins L2p, L27p and L36p (*LSU2p*, *LSU27p* and *LSU36p*). For each gene, a region containing the gene and 500 bp flanking the region upstream and downstream was used for the alignment. Direct and nested primers were selected within the most conserved regions (Table 2).

Cloning and sequencing

PCRs were carried out using the Invitrogen™ Platinum™ Taq DNA Polymerase system. Each reaction was performed in a 30 µL volume containing 1× reaction buffer, 2.5 mM $MgCl_2$ and 2 U Taq polymerase, supplied with 0.3 mM dNTP, 0.8 µM of each primer and sterile double distilled water. One µL (20 ng) of nucleic acid was used as template, and 0.2 µL of the amplicon was used as template for the nested assays. A sample devoid of DNA template was enclosed as negative control. PCR was initiated by a 5 min denaturation at 94°C, followed by 35 cycles of 30 s denaturation at 94°C, 45 s annealing at respective temperatures (Table 2) and 1 min extension at 72°C, and a final extension at 72°C for 7 min. A three step annealing strategy was used, with each step of 15 s, for all the primer sets (Table 2).

The PCR products were resolved in 1.2% agarose gels with ethidium bromide. Amplicons, corresponding to approx. 820 bp for the *SSU12p* gene and 420 to 530 bp for the *LSU36p* gene from nested PCR, were recovered and cloned into the vector pGEM®-T Easy (Promega). The plasmids were transformed into *E. coli* TOP10 chemically competent cells (Life Technologies), and the clones were sequenced using the T7/SP6 primers in both directions. Each pair of sequences was aligned and assembled using BioEdit. Three individual clones for each amplicon from each sample were analyzed. Each PCR, cloning and sequencing was repeated at least three times.

Phylogenetic analyses

The consensus sequence of each amplicon was submitted to the NCBI GenBank database (Table 1). Sequence information of the 12 phytoplasma strains used for the primer design was obtained from the same database. Sequence information of four other strains, including maize bushy stunt phytoplasma (MBS) strain M3 (CP015149), '*Ca*. P. pruni' strain CX

Table 2. Universal primers designed for amplification of *SSU12p* and *LSU36p loci*.

Target	PCR	Primer	Sequence[a]	T°C[b]		
				1	2	3
SSU12p	Direct	ItSSU12pF	ATGCCTACTRTTTCWCAATTAATTA	51.8	48.3	44.0
		ItSSU12pR	ATCTTAAACCTAAAGATTGRCGTC			
	Nested	ItSSu12pFn	AAAACCTAACTCCGCTTT	49.7	45.8	44.1
		ItSSu12pR1n	TTATGAAAAGTGGTAAAAAAG			
		ItSSu12pR2n	TTATGAAAGATGGMAAAAAGG			
LSU2p	Direct	ItLSu2pF	CTCATGYAAGTGTTTATCA	55.0	47.0	42.0
		ItLSu2pR	CTAAACGTGYTTTTCKAGG			
	Nested	ItLSu2pFn	YACTAGCAAYGTTTTRCC	56.0	47.0	42.0
		ItLSu2pRn	CCTAATTTATGWCCCACCAT			
LSU27p	Direct	ItLSu27pF1	AAAAATATCGTTTAAAACAAGG	55.0	47.0	42.0
		ItLSu27pF2	AAAAATATCGTTGTAAACAAGG			
		ItLSu27pR	GATATAGTTTGTGCTTCBGTTTC			
	Nested	ItLSu27pFn	GTTCCTCTTTGGCGRTAA	56.0	53.0	48.0
		ItLSu27pRn	TTAGAATGAGAATCACGACC			
LSU36p	Direct	ItLSU36pF1	GACTTTTTGCATTGAACC	51.6	47.2	42.2
		ItLSU36pF2	GACTTTTTGTGTTGAACC			
		ItLSU36pR	CGTTGTTTCTAGTTTTTTGHCC			
	Nested	ItLSU36pFn	AAGTGCTCATTTTGAACAYAC	50.0	47.7	43.1
		ItLSU36pRn	TTAYCCTTGTCTTTGATTRT			

[a]Degenerate nucleotides: R = A or G, W = A or T, H = A or C or T, Y = C or T.
[b]Annealing temperatures: a three-step annealing was applied: 15 s of T1 followed by 15 s of T2, and then 15 s of T3.

(LHCF00000000), '*Ca*. P. *phoenicium*' strain SA213 (JPSQ00000000), and '*Ca*. P. solani' (STOL) strain SA-1 (MPBG00000000), was also retrieved from the database for constructing phylogenetic trees. The sequence alignment was performed using ClustalW. *Acholeplasma laidlawii* (CP000896) served as outgroup. The phylogenetic trees were constructed using the Molecular Evolutionary Genetics Analysis program (MEGA7) (Kumar *et al*., 2016). Diversity indices, represented by the distances within and between groups, were also calculated using MEGA7. Virtual RFLP was performed using Vector NTI.

RESULTS

PCR amplification

In all the PCR reactions, several annealing temperatures were tested to select the best combinations. Since one reaction may include more than two primers and each primer may contain several degenerate nucleotides, a range of melting temperatures was calculated using the online tool (IDT OligoAnalyzer). This range could exceed the suggested melting temperature difference (5°C) for primer design, and when a single annealing temperature was utilized, not all the primers would anneal as efficiently, and nonspecific amplicons might be produced. Therefore, a three-step annealing strategy was used to optimize the reactions. Three annealing temperatures were selected at the maximum, mean and minimum points in the melting temperature range, each step lasting 15 sec. For each primer set, several adjustments were made before establishing the optimal combination (Table 2).

For all the samples used in this study, the PCR assay using the ItSSU12pF/ItSSU12pR primer pair produced clear bands approximately ranging from 750 bp to 820 bp (Figure 1A, Table 1). Subsequent nested PCR also generated clear bands (data not shown). These PCR amplicons were cloned and sequenced.

The PCR assay using the ItLSU36pF1/2/ItLSU36pR primers generated multiple bands or smears, and, in several cases, the expected products were not visible (Figure 1B). However, the subsequent nested PCR using the ItLSU36pFn/ItLSU36pRn primers always generated a strong and clear amplicon, and, in some cases, longer but significantly weaker bands (Figure 1C, Table 1). The strongest bands from each sample were recovered from the gels, cloned and sequenced.

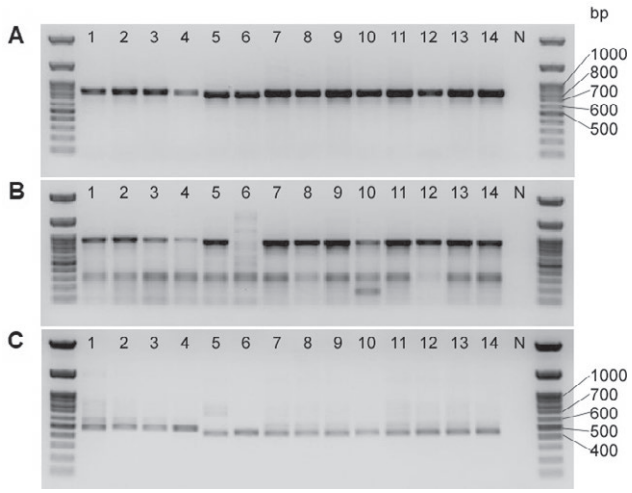

Figure 1. Agarose electrophoresis of the PCR results using the primer sets ItSSU12pF/ItSSU12pR (A), ItLSU36pF1/2/ItLSU36pR (B) and ItLSU36pFn/ItLSU36pRn (C). Lanes: 1. PRIVA, 2. CVT, 3. KVE, 4. A-AY, 5. FBP, 6. TBB, 7. CX, 8. PYLR, 9. TA, 10. LNII, 11. GR, 12. SP1, 13. MW1, 14. JR, N = negative control. Ladder: Maestrogen AccuRuler 100 bp Plus (ThermoFisher).

The direct and nested PCR assays targeting *LSU2p* and *LSU27p* genes failed to produce satisfactory results. In each trial, less than half of the tested samples showed amplification, and the results were not repeatable (data not shown). Optimization of the annealing temperatures failed to achieve consistent results and these primers were therefore discarded.

The specificity of the remaining primers was tested for detection of *Xylella fastidiosa*, *Agrobacterium tumefaciens*, *Pantoea agglomerans*, *Clavibacter michiganensis* subsp. *michiganesis*, *Pseudomonas syringae* pv. *tomato*, *P. syringae* pv. syringae, *Ralstonia solanacearum*, *Curtobacterium flaccumfaciens* pv. *flaccumfaciens*, *Xanthomonas arboricola* pv. *juglandis*, '*Candidatus* Liberibacter solanacearum', and '*Ca*. L. asiaticus', with no amplification (data not shown).

Loci structures

The amplicons generated by ItSSU12pF/ItSSU12pR covered the full length of the *SSU12p* gene, the partial sequence of *SSU7p* gene, and the intergenic region between the two genes. In the following text, this amplicon and its corresponding genomic *locus* are referred to as *SSU12p*. Sequence alignment of all the amplified samples and selected strains retrieved from the GenBank showed that *SSU12p* presented greater variation among ribosomal groups and subgroups compared with 16S rRNA (Supplementary Figure S1). The most relevant variation lay between 396 to 431 nt in strain AYWB (16SrI-A) corresponding to the intergenic region, which was less conserved than the coding genes, where the phytoplasma strains of the same ribosomal group and/or subgroup were featured by specific insertions and deletions.

The amplicon generated by ItLSU36pFn/ItLSU36pRn covered approximately 80% of the *LSU36p* gene, the full length of the gene encoding bacterial protein translation initiation factor 1 (IF-1), approx. 5% of the *map* gene, and the two intergenic regions. This amplicon and its corresponding genomic *locus* is referred to as *LSU36p* in the following text. Sequence alignment showed that the most conserved region was that encompassing the first 102 nt in the sequence of the strain AYWB, corresponding to the *LSU36p* gene (Supplementary Figure S2). The rest of the amplicon was highly variable among ribosomal groups and/or subgroups, with especially low similarity between the three tested strains in the 16SrIX group and the rest of the strains. However, by comparing this region in AYWB and '*Ca*. P. phoenicium' strain SA213 (16SrIX-B) it was observed that the low similarity was mainly located in the two intergenic regions, which also resulted in differences in length among the tested strains (Supplementary Figures S2 and S3).

Phylogenetic analyses

Phylogenetic trees were constructed with the *SSU12p* and *LSU36p* sequences separately as well as with the concatenated sequences (Figures 2A, 2B and 2C). The three trees showed clear separation of the phytoplasmas classified in the different 16Sr groups, the only exception being the 16SrXII-A subgroup ('*Ca*. P. solani'), which was more closely related to the 16SrI group than to the other 16SrXII subgroups in the *SSU12p* tree (Figure 2A). A number of subgroups and their corresponding '*Ca*. Phytoplasma' species were also clearly separated, e.g. 16SrI-B ('*Ca*. P. asteris'), 16SrXII-A ('*Ca*. P. solani'), 16SrXIII-F, 16SrXIII-K, 16SrX-A ('*Ca*. P. mali'), 16SrX-B ('*Ca*. P. prunorum') and 16SrX-C ('*Ca*. P. pyri'). The three trees showed significant consistency with those inferred from the 16S rRNA and *tuf* genes (Figure 2D and 2E).

To further evaluate the efficiency of the *SSU12p* and *LSU36p* for phytoplasma strain differentiation, the diversity indices within each 16Sr group and between any two groups were calculated, and paired t-Tests were performed to compare the set of "between group mean distance" indices from each marker (Supplementary Table S1). Both sets of indices from *SSU12p* and *LSU36p* were significantly higher than that of 16S rRNA ($P < 0.01$), suggesting that these two markers could efficiently separate the strains in different ribosomal groups. The indi-

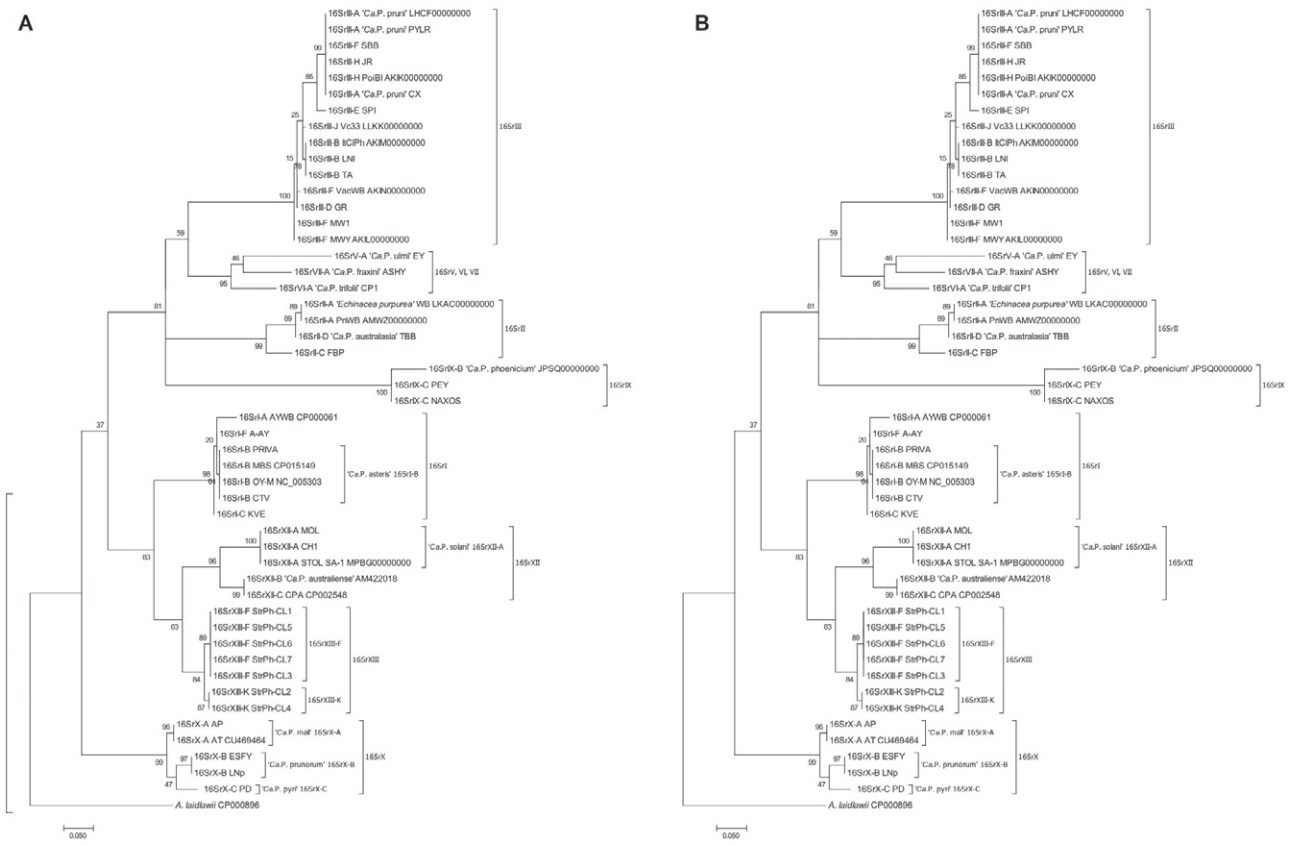

Figure 2. Phylogenetic trees inferred from *SSU12p* (A), *LSU36p* (B), concatenated sequences of *SSU12p* and *LSU36p* (C), 16S rRNA (D) and *tuf* (E). The trees were inferred using the Maximum Likelihood method based on the Tamura-Nei model (Tamura and Nei, 1993). Numbers at the nodes indicate the bootstrap values. The bars indicate substitutions per nucleotide position. *Acholeplasma laidlawii* was used as outgroup. (*continued*)

ces from *LSU36p* are also significantly higher than that of the *tuf* gene, suggesting that using *LSU36p* would improve the differentiation of phytoplasma strains.

RFLP analyses

The *SSU12p* sequences were further examined using by virtual RFLP with 18 restriction enzymes including: *Alu*I, *Bam*HI, *Bfa*I, *Bat*UI, *Dra*I, *Eco*RI, *Hae*III, *Hha*I, *Hinf*I, *Hpa*I, *Hpa*II, *Kpn*I, *Mse*I, *Rsa*I, *Sau*3AI, *Ssp*I, *Taq*I and *Tha*I (Supplementary Figure S4). *Eco*RI had no restriction site in any of the 49 samples examined, *Bam*HI, *Hpa*I and *Kpn*I each recognized only a single restriction site in one subgroup, *Hae*III recognized only one restriction site in two subgroups, and *Mse*I recognized up to 17 restriction sites in some subgroups. These enzymes were therefore not suitable for RFLP analyses. Another seven enzymes, *Alu*I, *Bst*UI, *Dra*I, *Hha*I, *Rsa*I, *Taq*I and *Tha*I generated two patterns within only one subgroup, and were therefore not suitable for the general phytoplasma differentiation. A set of five enzymes, *Bfa*I, *Hinf*I, *Hpa*II, *Sau*3AI and *Ssp*I clearly separated all the 16Sr groups (Figure 3). However, subgroups 16SrII-A and-D, 16SrIII-A and -E, 16SrIII-D and -F, 16SrXII-B and-C, and 16SrXIII-F and-K still showed the same restriction patterns. Some of these pairs could be distinguished by additional enzymes, including: 16SrIII-D and -F distinguished by *Alu*I, 16SrXII-B and -C by *Hha*I, and 16SrXIII-F and -K by *Alu*I, *Bst*UI, *Hha*I, *Rsa*I and *Tha*I. The 16SrII-A and -D and 16SrIII-A and -E remained unresolved.

Several 16S ribosomal groups and subgroups were featured with specific SNP sites, some of which contributed to their distinct RFLP patterns. For example, all the sample strains from the 16SrI group shared the specific sites of 11T, 183C, 362G, 375T, 376T, 471C, 519C, 630C and 633C, whereas those belonging to the 16SrIII group were marked with the sites of 30T, 48G, 58A, 275G and 320G (Supplementary Figure S1). The 578A site of the

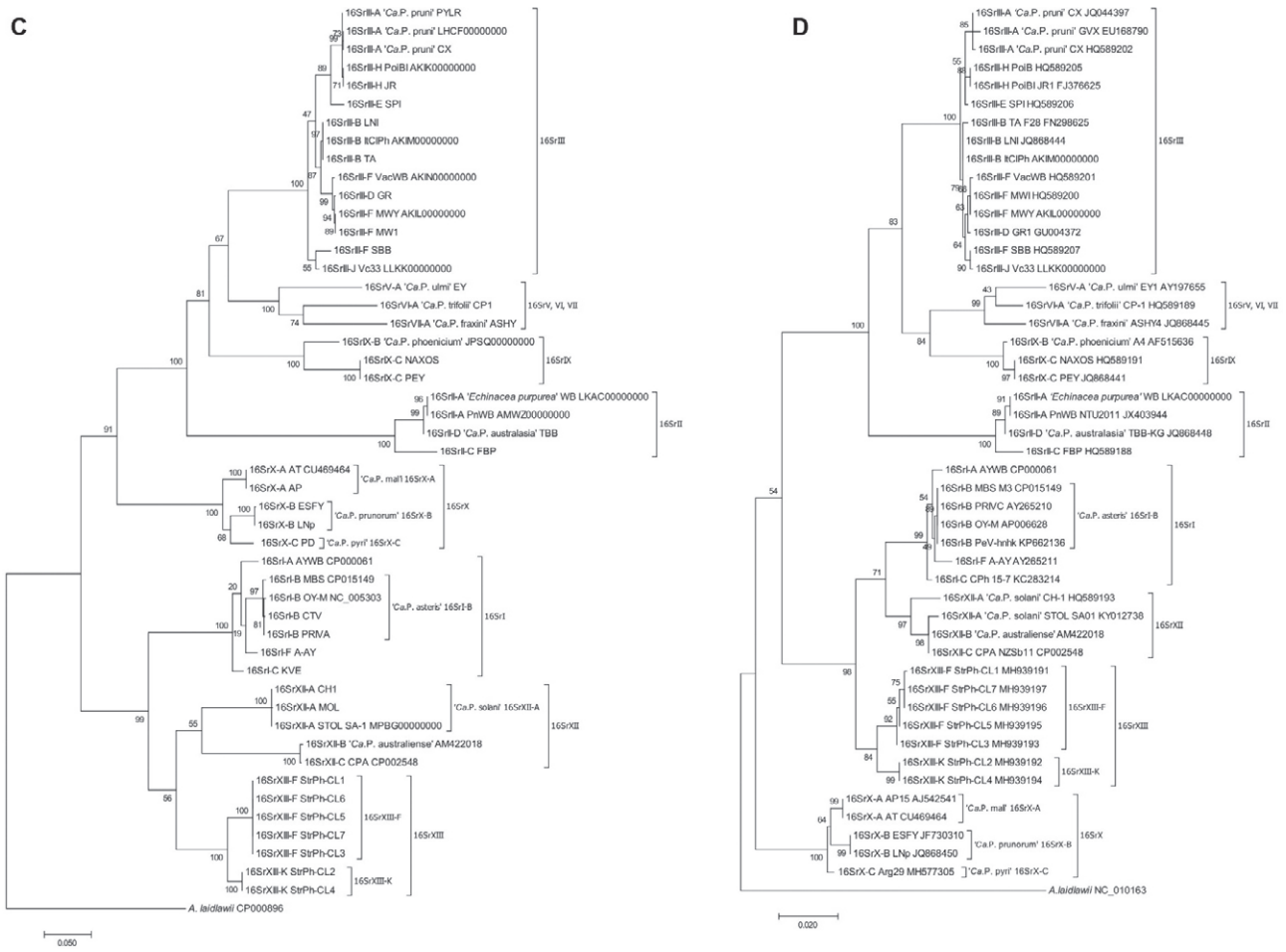

Figure 2. (*continued*)

strains from the 16SrXIII-F subgroup, and the 576A site of the strains from the 16SrXIII-K subgroup, also resulted in a specific *Ssp*I restriction site, producing a triple-band pattern for these two subgroups on the virtual RFLP. The 218G site unique to the strains from the 16SrIX-C subgroup resulted in a specific *Hpa*II restriction site, generating double bands on the virtual RFLP for this subgroup (Figure 3, Supplementary Figures S1 and S4).

DISCUSSION

Using 33 DNA samples and 16 sequences retrieved from the GenBank belonging to ten 16Sr groups and 27 subgroups, this study has shown that both *SSU12p* and *LSU36p* are suitable *loci* for phytoplasma detection and differentiation. In the RFLP analyses using amplicons generated by *SSU12p*, a set of seven enzymes, including *Bfa*I, *Hinf*I, *Hpa*II, *Sau*3AI, *Ssp*I, *Alu*I and *Hha*I, were able to identify all the phytoplasmas in the 16Sr groups, and in all but four subgroups (16SrII-A/-D and 16SrIII-A/-E) examined.

The primers for *SSU12p* and *LSU36p* amplified phytoplasma sequences from all the samples tested, proving that they are amplifying conserved regions in a robust manner. The *SSU12p* primers generated in direct PCR clear, single-band products. According to the literature, *SSU12p* is to date peerless for phytoplasma PCR detection, considering its ability to generate a unique specific band in direct PCR using a single pair of primers from a wide range of phytoplasmas. The high consistency of *SSU12p* for phytoplasma identification with 16S rRNA and *tuf* genes confirms its reliability, suggesting that the application of this pair of primers is appropriate for rapid and efficient phytoplasma detection and identification. The *LSU36p* primers, on the other hand, requires nested amplification, and the resulting products may vary

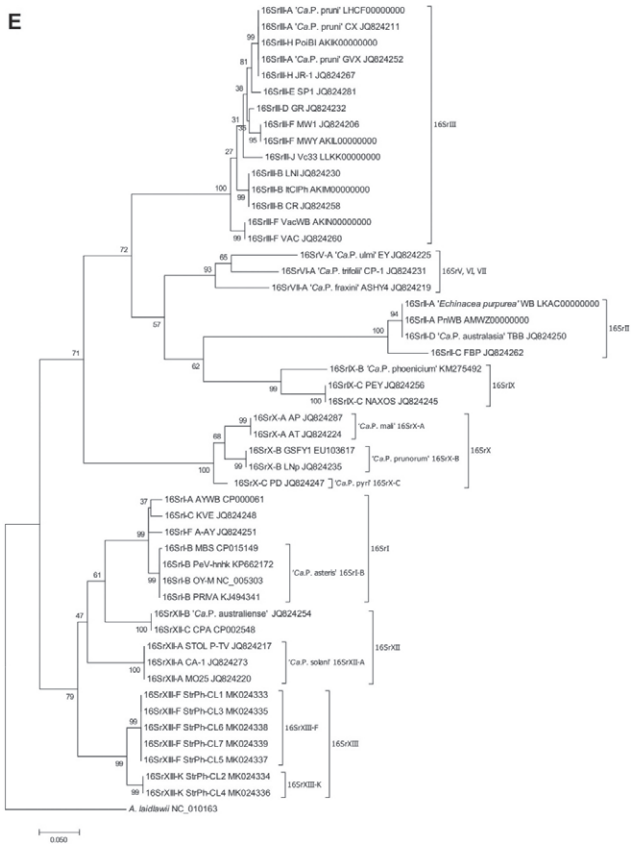

Figure 2. (*continued*)

significantly in size. However, the relatively high value of between-group mean distance indices suggests that *LSU36p* has potential for resolving closely related strains. Further study focused on other strains belonging to different subgroups from the same ribosomal group is required for confirmation.

Due to different evolutionary processes, phylogenetic trees derived from different genome *loci* may show conflicting structures. One way to interpret the conflicting information is to concatenate the *loci* for phylogenetic analyses. Although concatenation is a controversial method because of potential misspecification of models, it provides longer sequences to overcome sampling errors (Holland *et al.,* 2004). In the present study, the phylogenetic trees inferred from *SSU12p*, *LSU36p* and *SSU12p* plus *LSU36p* showed clear and unambiguous consistency of ramification of phytoplasma subgroups within most of the 16Sr groups, confirming the robustness of the concatenation methods.

The only exception was the 16SrIII group, which showed unclear relationships among several subgroups. For example, in the *SSU12p* tree, the strain SBB from the 16SrIII-F subgroup formed a clade with the Vc33 from the 16SrIII-J subgroup, while the other two strains from the 16SrIII-F group, MWY and MW1, were grouped with strains from the 16SrIII-B and 16SrIII-D subgroups. This was probably due to the intrinsic structure of the 16SrIII group, since the trees from both *tuf* and 16S rRNA also showed unclear structures within this phytoplasma group. A similar conflict occurred within the 16SrIII group in independent studies analyzing 16S rRNA and *secY* phylogenies (Lee *et al*., 2010; Fernández *et al*., 2017). The two copies of the 16S rRNA gene of phytoplasmas in this group very often present interoperon heterogeneity. Data from *secY* and *tuf* genes, both present in the genome in single copy, indicated that the confusing tree structures were not incidental. These results suggest that the subgroup classification within the 16SrIII group may not reflect phylogenetic interrelationship and the RFLP-based classification may be biased, because this classification solely depends on the restriction sites of a selected set of enzymes while the SNPs in sequences other than these sites are neglected.

The reliability of *SSU12p* and *LSU36p* as phytoplasma markers confirms that genome comparison is an approach that could also be used for selecting genes to differentiate these bacteria. A larger number of samples than used in the present study, containing strains from untested groups and subgroups, will help to confirm the wide reliability of this detection system. The development of next-generation sequencing and long-read sequencing has built an expanding genomic database of microbial pathogens. Comparative genomics has been used to study the mechanisms of pathogenicity, molecular epidemiology, molecular diagnostics, multi-locus sequence typing, and transmission prediction (Avarre *et al*., 2011; Bastardo *et al*., 2012; Walker *et al*., 2014; Bayliss *et al*., 2017; Aly *et al*., 2019). As more phytoplasma genomes are being sequenced, comparative genomics has also become the trend for analyses in genome reports (Sparks *et al*., 2018; Wang *et al*., 2018; Music *et al*., 2019; Cho *et al*., 2019). Approaches on the genome level will likely be increasingly applied to phytoplasmas for understanding their adaptations to diverse host species. However, the identification of new markers for detection and differentiation of phytoplasmas strains is still a necessary tool for developing knowledge of epidemiology and management of phytoplasma-associated diseases that aim to avoid their pandemic distribution.

ACKNOWLEDGEMENTS

This study was supported by the National Fund for Scientific and Technological Development (FONDECYT) of Chile, project Nos. 1140883 and 11160719, and Postdoctoral Project 2017, No. 3170120.

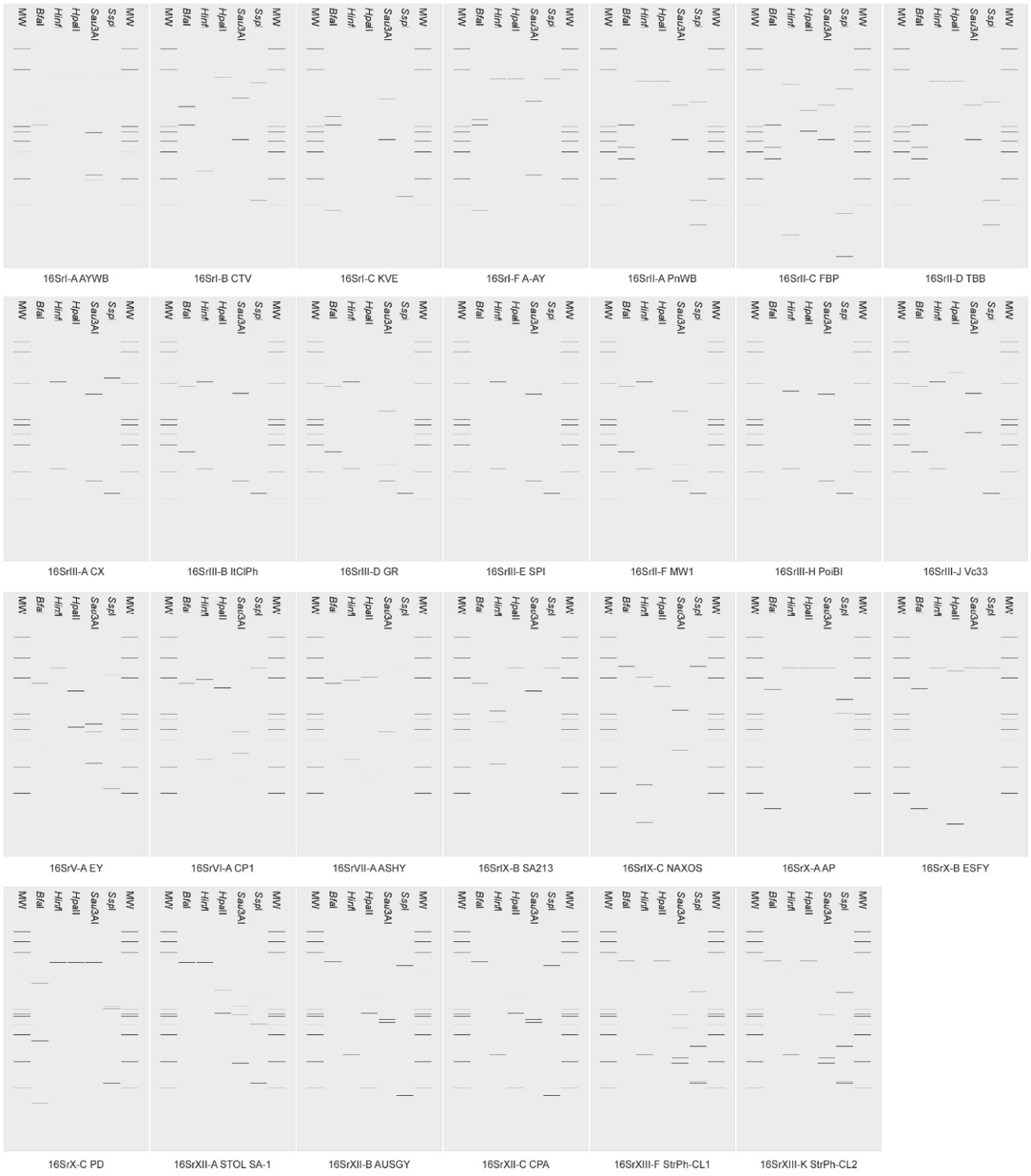

Figure 3. *In silico* RFLP patterns of *SSU12p* sequences from 27 phytoplasma strains, representing all the ribosomal groups and subgroups used. Restriction enzymes: *Bfa*I, *Hinf*I, *Hpa*II, *Sau*3AI, *Ssp*I. Size marker: phiX174 digested by *Bsu*RI and *Hae*III.

LITERATURE CITED

Aly M.A., Domig K.J., Kneifel W., Reimhult E., 2019. Whole genome sequencing based comparison of food isolates of *Cronobacter sakazakii*. *Frontiers in Microbiology* 10: 1464.

Arnaud G., Malembic-Maher S., Salar P., Bonnet P., Maixner M., … Foissac X., 2007. Multilocus sequence typing confirms the close genetic interrelatedness of three distinct "flavescence dorée" phytoplasma strain clusters and group 16SrV phytoplasmas infecting grapevine and alder in Europe. *Applied and Environmental Microbiology* 73: 4001–4010.

Avarre J.C., Madeira J.P., Santika A., Zainun Z., Baud M., … Maskur M., 2011. Investigation of Cyprinid herpesvirus-3 genetic diversity by a multi-locus variable number of tandem repeats analysis. *Journal of Virological Methods* 173: 320–327.

Bastardo A., Ravelo C., Romalde J.L., 2012. Multilocus sequence typing reveals high genetic diversity and epidemic population structure for the fish pathogen *Yersinia ruckeri*. *Environmental Microbiology* 14: 1888–1897.

Bayliss S.C., Verner-Jeffreys D.W., Bartie K.L., Aanensen D.M., Sheppard S.K., … Feil E.J., 2017. The promise of whole genome pathogen sequencing for the molecular epidemiology of emerging aquaculture pathogens. *Frontiers in Microbiology* 8: 1–18.

Bertaccini A., Lee I.-M., 2018. Phytoplasmas: an update. In: *Phytoplasmas: Plant Pathogenic Bacteria – I: Characterization and Epidemiology of Phytoplasma-Associated Disease* (G.P. Rao, A. Bertaccini, N. Fiore, L.W. Liefting, eds.), Springer Nature, Singapore, 1-29.

Cho S.-T., Lin C.-P., Kuo C.-H., 2019. Genomic characterization of the periwinkle leaf yellowing (PLY) phytoplasmas in Taiwan. *Frontiers in Microbiology* 10: 2194.

Cimerman A., Pacifico D., Salar P., Marzachi C., Foissac X., 2009. Striking diversity of *vmp1*, a variable gene encoding a putative membrane protein of the "stolbur" phytoplasma. *Applied and Environmental Microbiology* 75: 2951–2957.

Contaldo N., Bertaccini A., 2019. Phytoplasma cultivation. In: *Phytoplasmas: Plant Pathogenic Bacteria - III: Genomics, Host Pathogen Interactions and Diagnosis* (A. Bertaccini, K. Oshima, M. Kube, G.P. Rao, ed.s), Springer Nature, Singapore, 89-104.

Contaldo N., Canel A., Makarova O., Paltrinieri S., Bertaccini A., Nicolaisen M. 2011. Use of a fragment of the *tuf* gene for phytoplasma 16Sr group/subgroup differentiation. *Bulletin of Insectology* 64 (Supplement): S45–S46.

Cui W., Quiroga N., Curkovic S.T., Zamorano A., Fiore N., 2019. Detection and identification of 16SrXIII-F and a novel 16SrXIII phytoplasma subgroups associated with strawberry phyllody in Chile. *European Journal of Plant Pathology* 155: 1039–1046.

Duduk B., Paltrinieri S., Lee I-M., Bertaccini A., 2013. Nested PCR and RFLP analysis based on the 16S rRNA gene. In: *Phytoplasma: Methods and Protocols* (M. Dickinson, J. Hodgetts, ed.), Humana Press, Totowa, New Jersey, United States of America, 159–171.

Dumonceaux T.J., Green M., Hammond C., Perez E., Olivier C., 2014. Molecular diagnostic tools for detection and differentiation of phytoplasmas based on chaperonin-60 reveal differences in host plant infection patterns. *PLoS One* 9: e116039.

EPPO-Q-bank, 2020. https://qbank.eppo.int/

Fernández F.D., Meneguzzi N.G., Conci L.R., 2017. Identification of three novel subgroups within the X-disease group phytoplasma associated with strawberry redness disease. *International Journal of Systematic and Evolutionary Microbiology* 67: 753–758.

Glenn D., 2000. National program for the management of phytoplasmas in Australian grapevines. Final report CRCV 95/2 for Grape and Wine Research and Development Corporation. Loxton, Australia, 174 pp.

Hodgetts J., Boonham N., Mumford R., Harrison N., Dickinson M., 2008. Phytoplasma phylogenetics based on analysis of *secA* and 23S rRNA gene sequences for improved resolution of candidate species of *'Candidatus* Phytoplasma'. *International Journal of Systematic and Evolutionary Microbiology* 58: 1826–1837.

Holland B.R., Huber K.T., Moulton V., Lockhart P.J., 2004. Using consensus networks to visualize contradictory evidence for species phylogeny. *Molecular Biology and Evolution* 21: 1459-1461.

IRPCM 2004. '*Candidatus* Phytoplasma', a taxon for the wall-less, non-helical prokaryotes that colonize plant phloem and insects. *International Journal of Systematic and Evolutionary Microbiology* 54: 1243–1255.

Jenkins C., Ling C.L., Ciesielczuk H.L., Lockwood J., Hopkins S., … Kibbler C.C., 2012. Detection and identification of bacteria in clinical samples by 16S rRNA gene sequencing: comparison of two different approaches in clinical practice. *Journal of Medical Microbiology* 61: 483–488.

Kumar S., Stecher G., Tamura K., 2016. MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution* 33: 1870–1874.

Lee I-M., Hammond R.W., Davis R.E., Gundersen D.E., 1993. Universal amplification and analysis of patho-

gen 16S rDNA for classification and identification of mycoplasmalike organisms. *Phytopathology* 83: 834–842.

Lee I-M., Gundersen-Rindal D.E., Davis R.E., Bartoszyk I.M., 1998. Revised classification scheme of phytoplasmas based on RFLP analyses of 16S rRNA and ribosomal protein gene sequences. *International Journal of Systematic Bacteriology* 48: 1153–1169.

Lee I-M., Bottner-Parker K.D., Zhao Y., Davis R.E., Harrison N.A., 2010. Phylogenetic analysis and delineation of phytoplasmas based on *secY* gene sequences. *International Journal of Systematic and Evolutionary Microbiology* 60: 2887–2897.

Ludwig W., Schleifer K.H., 1994. Bacterial phylogeny based on 16S and 23S rRNA sequence analysis. *FEMS Microbiology Reviews* 15: 155–173.

Makarova O., Contaldo N., Paltrinieri S., Kawube G., Bertaccini A., Nicolaisen M., 2012. DNA barcoding for identification of *'Candidatus* Phytoplasmas' using a fragment of the elongation factor *Tu* gene. *PLoS One* 7: e52092.

Marcone C., Lee I.-M., Davis R.E., Ragozzino A., Seemüller E., 2000. Classification of aster yellows-group phytoplasmas based on combined analyses of rRNA and *tuf* gene sequences. *International Journal of Systematic and Evolutionary Microbiology* 50: 1703–1713.

Martini M., Botti S., Marcone C., Marzachì C., Casati P., … Bertaccini A., 2002. Genetic variability among "flavescence dorée" phytoplasmas from different origins in Italy and France. *Molecular and Cellular Probes* 16: 197–208.

Martini M., Lee I.-M., Bottner K.D., Zhao Y., Botti S., … Osler R., 2007. Ribosomal protein gene-based phylogeny for finer differentiation and classification of phytoplasmas. *International Journal of Systematic and Evolutionary Microbiology* 57: 2037–2051.

Music M.S., Samarzija I.S., Hogenhout S.A., Haryono M., Cho S-T., Kuo C-H., 2019. The genome of '*Candidatus* Phytoplasma solani' strain SA-1 is highly dynamic and prone to adopting foreign sequences. *Systematic and Applied Microbiology* 42: 117-127.

Oliveira E. de, Resende R. de O., Pecci M. de la P.G., Laguna I.G., Herrera P., Cruz I., 2003. Occurrence of viruses and stunting diseases and estimative of yield losses by mollicutes in corn in Paraná State, Brazil. *Pesquisa Agropecuária Brasileira* 38: 19–25.

Pérez-López E., Olivier C.Y., Luna-Rodríguez M., Dumonceaux T.J., 2016. Phytoplasma classification and phylogeny based on *in silico* and *in vitro* RFLP analysis of *cpn60* universal target sequences. *International Journal of Systematic and Evolutionary Microbiology* 66: 5600–5613.

Schneider B., Seemüller E., Smart C.D., Kirkpatrick B.C., 1995. Phylogenetic classification of plant pathogenic mycoplasmalike organisms or phytoplasmas. In: *Molecular and Diagnostic Procedures in Mycoplasmology* (S. Razin, J.G. Tully, eds.), Academic Press, San Diego, California, United States of America, 369-380.

Seemüller E., Schneider B., Maurer R., Ahrens U., Daire X., … Stackebrandt E., 1994. Phylogenetic classification of phytopathogenic mollicutes by sequence analysis of 16S ribosomal DNA. *International Journal of Systematic Bacteriology* 44: 440–446.

Shao J., Jomantiene R., Dally E.L., Zhao Y., Lee I.-M., … Davis R.E., 2006. Phylogeny and characterization of phytoplasmal *nusA* and use of the *nusA* gene in detection of group 16SrI strains. *Journal of Plant Pathology* 88: 193–201.

Sparks M.E., Bottner-Parker K.D., Gundersen-Rindal D.E., Lee I.-M., 2018. Draft genome sequence of the New Jersey aster yellows strain of '*Candidatus* Phytoplasma asteris'. *PLoS One* 13: e0192379.

Tamura K., Nei M., 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Molecular Biology and Evolution* 10: 512–526.

Valiunas D., Jomantiene R., Davis R.E., 2013. Evaluation of the DNA-dependent RNA polymerase β-subunit gene (*rpoB*) for phytoplasma classification and phylogeny. *International Journal of Systematic and Evolutionary Microbiology* 63: 3904–3914.

Walker T.M., Lalor M.K., Broda A., Ortega L.S., Morgan M., … Conlon C.P., 2014. Assessment of Mycobacterium tuberculosis transmission in Oxfordshire, UK, 2007–12, with whole pathogen genome sequences: an observational study. *The Lancet Respiratory Medicine* 2: 285–292.

Wang J., Song L., Jiao Q., Yang S., Gao R., … Zhou G., 2018. Comparative genome analysis of jujube witches'-broom phytoplasma, an obligate pathogen that causes jujube witches'-broom disease. *BMC Genomics* 19: 689.

Wei W., Lee I.-M., Davis R.E., Suo X., Zhao Y., 2008. Automated RFLP pattern comparison and similarity coefficient calculation for rapid delineation of new and distinct phytoplasma 16Sr subgroup lineages. *International Journal of Systematic and Evolutionary Microbiology* 58: 2368–2377.