Annals of Applied Biology
An international journal of the aab

Multilocus sequence typing and phylogenetic analysis revealed two distinct almond witches'-broom phytoplasma subpopulations in Iran

Journal:	*Annals of Applied Biology*
Manuscript ID	AAB-2020-0356.R2
Manuscript Type:	Original Article
Date Submitted by the Author:	24-Dec-2020
Complete List of Authors:	Mosayyebi, Niloofar; Shahrekord University Faculty of Agriculture, Plant protection Mehraban, Zahra; Shahrekord University Faculty of Agriculture, Plant Protection Siampour, Majid; Shahrekord University, Plant protection Babaei, Ghobad; AREEO, Plant Protection Quaglino, Fabio; Università degli Studi di Milano, Di.Pro.Ve.-sez. Patologia Vegetale
Key Words:	Microbiology, Population Dynamics, Bacterial Pathogens, Phytoplasmas

Moltilocus sequence typing

Multilocus sequence typing and phylogenetic analysis revealed two distinct almond witches'-broom phytoplasma subpopulations in Iran

Niloofar Mosayyebi[a,*], Zahra Mehraban[a,*], Majid Siampour[a**], Ghobad Babaei[b], Fabio Quaglino[c]

[a] *Department of Plant Protection, College of Agriculture, Shahrekord University, Shahrekord, Iran*

[b] *Plant Protection Research Department, Chaharmahal and Bakhtiari Agricultural and Natural Resources Research and Education Center, AREEO, Shahrekord, Iran*

[c] *Department of Agricultural and Environmental Sciences, Production, Landscape, Agroenergy, Universita degli Studi di Milano, Milan, Italy*

[*] These authors contributed to this work equally

**** Corresponding author**: E-mail address: siampour@sku.ac.ir (M. Siampour)

Summary

During a field survey conducted in different geographical regions of Iran, phytoplasma-like symptoms were observed in peach, almond, wild almond and GF-677 rootstock. Based on 16S rDNA amplification, followed by nucleotide sequence and phylogenetic analyses, almond witches'-broom (AlmWB) phytoplasma strains were found in association with symptomatic plants. Based on the single nucleotide polymorphisms within 16S rDNA sequence, most of the Iranian AlmWB phytoplasma strains were distinguished from the Lebanese strains. Further molecular typing, based on sequence and phylogenetic analyses of *rplV-rpsC* (ribosomal proteins) and *imp* (immunodominant membrane protein) genes, indicated that AlmWB phytoplasma strains can be differentiated in two subpopulations: Ir, including exclusively Iranian strains, and Lb, comprising strains from Lebanon and Iran. Moreover, the selection pressure analysis of *imp* gene sequences suggested that these two AlmWB phytoplasma subpopulations had some distinct biological or ecological properties. Finally, results of the sequence and phylogenetic analyses performed on *rplV-rpsC* and *imp* genes showed a considerably high genetic distance between AlmWB phytoplasma strains (16SrIX-B and variants) and phytoplasmas belonging to other 16SrIX subgroups. The significance of genetic variation in phytoplasmas of 16SrIX group is discussed in relation to their biological characteristics and geographical distinction.

Keywords: phytoplasma, almond witches'-broom, membrane protein, biology, phylogeny

Introduction

Phytoplasmas are a group of ~~uncultivated~~ plant pathogenic *Mollicutes* ~~causing~~ associated with diseases in more than 1,000~~1000~~ plant species. They are specifically transmitted by sap sucking insect vectors in a persistent-propagative manner (Hogenhout *et al*., 2008). The highly conserved 16S rRNA gene has been used as a marker for identification and classification of phytoplasmas (Bertaccini and Lee, 2018). According to the RFLP analysis of the 16S rRNA gene sequence, phytoplasmas were classified in more than 33 ribosomal groups. Moreover, 43 '*Candidatus* Phytoplasma' species have been so far described (IRPCM, 2004; Bertaccini and Lee, 2018). However, the resolving power of 16S rRNA gene is insufficient for differentiation of biologically or ecologically distinct phytoplasma strains. This was implemented with the use of less conserved markers such as ribosomal protein (rp) or *secY* genes for differentiation of closely related phytoplasmas (Martini *et al*., 2007; Lee *et al*., 2010; Siampour *et al*., 2019b).

Phytoplasmas within the ribosomal group 16SrIX (related to '*Ca*. Phytoplasma phoenicium') are composed of strains associated with diverse diseases in numerous plant host species across the world. Based on RFLP analysis of the 16S rRNA gene sequence, phytoplasmas within the group 16SrIX were classified into several subgroups (Lee *et al*., 2012; Pérez-López *et al*., 2018).

Phytoplasmas in 16SrIX group were reported in association with diseases in more than 10 plant species in Iran. These included strains belonging to the subgroups 16SrIX-B, -C, -D, -J and -I (Siampour *et al*. 2019a; Esmaeilzadeh-Hosseini *et al*., 2018). Phytoplasmas of the subgroup 16SrIX-C had the largest host range and were scattered in wide geographical regions in this country (Siampour *et al*., 2019a). The phytoplasma strains of subgroup 16SrIX-B and its variants, described as subgroups 16SrIX-D, -F and -G, were associated with AlmWB disease (Verdin *et al*.,

2003; Molino Lova *et al*., 2011). The natural *Prunus* hosts of the AlmWB phytoplasma in Iran or Lebanon were almond, peach, nectarine, apricot, wild almonds (*Prunus scoparia* and *P. orientalis*), and the rootstock GF-677 (Molino Lova *et al*., 2011; Salehi *et al*., 2015; Verdin *et al*., 2003; Salehi *et al*., 2011). The main characteristic symptom of the AlmWB on almond, peach and nectarine was the development of the witches'-broom from the crown and the trunk of the infected trees (Verdin *et al*., 2003). Although no overall economic loss has been estimated, many orchards were abandoned or replanted due to the AlmWB disease in Iran.

Assessment of the genetic variability of Iranian phytoplasma strains of group 16SrIX has been based mainly on the sequence variation of the 16S rRNA gene. In the present study, more variable *rplV-rpsC* housekeeping genes and also a putative biologically important gene (*imp*) were used to elucidate how genetic variability could be related to the geographical or biological diversity of these phytoplasmas.

Materials and Methods

Sample collection and DNA extraction

In 2018-19, leaf samples were collected from a total of 40 almond, peach, and wild almond (*Prunus scoparia*) trees, and GF-677 (*Prunus amygdalus* x *Prunus persica*) seedlings exhibiting phytoplasma-like symptoms in orchards located in Chaharmahal-Bakhtiari, Fars and Kermanshah provinces, respectively, in Center, Southwest and Northwest of Iran. Leaf samples were also collected from seven symptomless peach trees from orchards of Chaharmahal-Bakhtiari province. Moreover, sesame phyllody (SEJ2 and WIY2) and lettuce phyllody (LET1 and LET2) phytoplasma strains (belonging to 16SrIX group) maintained in potted periwinkle plants

(*Catharanthus roseus* (L.) G. Don) were examined in this study (Salehi *et al*., 2007; 2017). Total DNA was extracted from 100 mg of the midribs according to Abou-Jawdah *et al*. (2002).

Phytoplasma detection and identification

All DNAs were tested for phytoplasma presence through16S rRNA gene amplification by nested PCR analysis using universal primer pairs P1/P7 (Deng and Hiruki, 1991; Schneider *et al*., 1995) followed by R16F0/R16R1 (Lee *et al*., 1993). Total DNA extracted from ten seed grown peach and almond, and three seed grown periwinkle plants (maintained in insect proof cages) were utilized as negative controls. Total DNA extracted from two periwinkle plants infected by cucumber phyllody phytoplasma (CuPh, 16SrII-D; GenBank accession no. KY412986) were also utilized as positive controls in PCR assays (Siampour *et al*., 2019b).

PCR was carried out in 20 µl reactions using 2x master mix (Ampliqon, Denmark) with 0.4 µM of each primer and 100 ng of total DNA. The amplification of an expected 1.4 kbp phytoplasma DNA fragment was visualized by electrophoresis on a 1.2% agarose gel. The amplicons were purified and sequenced in both directions using the same primers (Codon genetic group, Iran). For ribosomal group/subgroup attribution, virtual RFLP analysis of the 1.2 kbp 16S rRNA gene sequences (delimited by R16F2n/R16R2 primers; Gundersen and Lee, 1996) was performed using the *i*PhyClassifier online tool (Zhao *et al*., 2009). For '*Ca.* phytoplasma' species attribution, 16S rDNA nucleotide sequences were aligned with those of '*Ca*. Phytoplasma phoenicium' strains (strains A4 and 21; Verdin *et al*., 2003) and the pairwise sequence identity values were calculated using SDT software (Muhire *et al*., 2014). Single nucleotide polymorphisms (SNPs) analysis of 16S rDNA fragment delimited by R16F2n/R16R2 primers (Gundersen and Lee, 1996) was used to assign AlmWB phytoplasma strains to different 16SrIX-B (AlmWB) genetic lineages (Salehi *et*

al., 2018; 2020). To detect the SNP sites, the 16S rDNA sequence of AlmWB phytoplasmas obtained in this study was compared with that of other Iranian and Lebanese AlmWB strains present in GenBank (accession numbers listed in Table 2) as detailed by Salehi *et al*. (2018 and 2020).

Molecular and phylogenetic analyses using variable genes

Fragments of the *rplV-rpsC* (rp) and *imp* genes of the phytoplasma strains used in this study were amplified and sequenced. Nested PCR with primer pairs rpL2F3/rp(I)R1A (Martini *et al*., 2007) followed by rplF2/rpsR2 (GGGKAATTTTCRCCAACAAG/CAGCTCTAAAAGTATTTAAAGG), designed in this study, was performed to amplify an *rplV-rpsC* fragment of about 1000 bp; covering the complete *rplV* gene, and 550 bp from the 5' end of *rplC* gene. For the amplification of *imp* gene, two primer pairs impF1/impR1 (5'-AAGCGCATTCTGAAGAAATGG-3'/5'-AGAACATGATGAAAAAACAGA-3') and impF2/impR2 (5'-TCAYCCAGAATTTTTATCAAG-3'/5'-AGGAGAAATAATATTTCATG-3') were designed based on the conserved regions of the flanking genes, *dnaD* and *pyrG*, from the genome draft of the '*Ca*. Phytoplasma phoenicium'-related strains SA213 (subgroup 16SrIX-B) and ChiP (subgroup 16SrIX-C) (accession numbers listed in Table S1). Nested PCR assay using primer pairs impF1/impR1 followed by impF2/impR2 was carried out to amplify the *imp* gene full sequence of 16SrIX phytoplasmas. PCRs were run in an automated thermal cycler (Mastercycler gradient, Eppendorf, Germany). The amplicons were purified and sequenced in both directions. The nucleotide sequence assembly for each PCR amplicon of *rplV-rpsC* and *imp* genes was generated using DNAMAN program (Lynnon Biosoft). The 16S rRNA, *rplV-rpsC* and *imp* gene sequences obtained in this study and those of other phytoplasma strains from the Genbank database

(accession numbers listed in Tables 1 and S1) were each aligned using the ClustalW Program. The multiple alignments were used to reconstruct phylogenetic trees using maximum likelihood method in Mega version 7 program (Kumar *et al*., 2016). The robustness of the branches was assessed using bootstrap resampling with 1,000 replications.

Analysis of selection pressure

To study the impact of selection pressure, a multiple sequence alignment of the *imp* gene was generated based on the deduced amino acid sequence alignment using the program DAMBE (http://dambe.bio.uottawa.ca/DAMBE/dambe.aspx). The overall average of the synonymous substitutions per synonymous sites (*dS*) and non-synonymous substitution per non-synonymous sites (*dN*) was computed in pairwise comparisons using the Nei-Gojobori method with Jukes-Cantor model in the MEGA version 7 (Kumar *et al*., 2016). The significant difference between *dN* and *dS* was tested as described by Messier and Stewart (1997). The statistically higher values of *dN* over *dS* (*dN*>*dS*) supports the evidence for presence of positive selection in the *imp* gene.

Structural prediction of Imp protein

The program TOPOCONS (Bernsel *et al*., 2009), was used to predict the presence and orientation of the transmembrane domains in the deduced amino acid sequence of the *imp* gene. The programs DeepCoil (https://toolkit.tuebingen.mpg.de/tools/deepcoil) and Paircoil2 (http://cb.csail.mit.edu/cb/paircoil2/paircoil2.html) were used to identify putative coiled coil structure along the Imp amino acid sequence.

Results

Disease symptoms on the *Prunus* species

Almond trees in Chaharmahal-Bakhtiari province showed the symptoms of witches'-broom on the trunk, shortening of internodes, late flowering, dieback and decline. Based on the symptom observations, the maximum disease incidence of ~13% (12 out 95 trees) was recorded in an almond orchard in this province. The main symptoms observed on the almond trees in Fars and Kermanshah provinces were severe witches'-broom on the trunk and canopy, small and yellow leaves, decline and death of the trees (Figure 1A, B, D). Based on symptoms, the maximum disease incidence of 34% (30 out of 90 trees) was observed in an orchard in Fars province. Symptomatic almond trees were found only in three almond orchards in two regions of Kermanshah province. Only seven out of about 2000 almond trees in these three orchards were symptomatic.

The main symptoms observed on the peach trees in orchards of Chaharmahal-Bakhtiari province were mild reddening or yellowing of the foliage with the leaf margins curled upward along the midrib and shortening of internodes (Figure 1C). Reddish-purple leaf spots were also frequently observed in the early season giving the leaves shot holed appearance later in the season. These leaf spots could also be resulted from infection by plant pathogens other than phytoplasmas. Dieback and decline were the other symptoms observed in some peach trees in Chahramahal-Bakhtiari province. Based on symptomatology study, the maximum disease incidence observed in one of the peach orchards in this province was ~30% (40 out of 130 trees). No witches'-broom symptoms were observed on peach trees in this region. Four out of seven GF-677 seedlings, sporadically grown around one of the peach orchards in the same region were symptomatic. The main symptom observed on these seedlings was small leaves grown sparsely on the canopy.

The disease symptoms on the wild almond in Fars province was severe witches'-broom formed on the main trunk, leaf yellowing, perpendicular development of many shoots on the main branches and decline of the trees (Figure 1E).

Phytoplasma detection and identification using 16S rRNA gene

A total of 40 symptomatic trees of *Prunus* species were tested (Table S2). A DNA fragment of the expected size (about 1.4 kbp) was amplified from all symptomatic *Prunus* samples, periwinkle seedlings infected by sesame phyllody (SEJ2 and WIY2) and lettuce phyllody (LET1 and LET2) phytoplasmas, and from the positive controls. No 16S rDNA amplicon was obtained from DNA of healthy plants used as negative controls (except for one sample; Table S2).

The 16S rRNA gene sequence from 20 phytoplasma strains including 16 strains detected in *Prunus* species and four strains detected in sesame and lettuce with phyllody symptom was determined (Table 1). Three16S rDNA sequence variants, sharing >99.8% identity, were identified among phytoplasma strains infecting *Prunus* species. They also showed >99.8% identity with the 16S rRNA gene sequence of the '*Ca*. Phytoplasma Phoenicium' strains A4 and 21, indicating their relatedness to this '*Ca*. Phytoplasma' species. Virtual RFLP analysis using *i*PhyClasssifier revealed that all of the AlmWB phytoplasma strains, detected in diverse *Prunus* species, belonged to subgroup 16SrIX-B or its variants (similarity coefficient 1 versus reference strains '*Ca*. Phytoplasma phoenicium' A4 or '*Ca*. Phytoplasma phoenicium' 21) (Tables 1 and S1). The phytoplasma strains associated with sesame phyllody (SEJ2 and WIY2) and lettuce phyllody (LET1 and LET2) belonged, respectively, to the subgroups 16SrIX-C and 16SrIX-J with similarity coefficient of 1. The virtual RFLP patterns obtained using analysis by *i*Phyclassifier were also confirmed by actual RFLP analysis using key restriction enzymes (data not shown). As also

evidenced by the phylogenetic analysis of the 16S rRNA gene, AlmWB phytoplasma strains couldn't be differentiated based on their host plant or country of origin. However, all five AlmWB strains belonging to the variants of subgroup 16SrIX-B (AlGN1, AlGN2, A21, Urmia and Bardseer) were from Iran and clustered together (Fig.2).

On the other hand, the SNP analysis of 16S rDNA fragments showed high sensitivity to resolve the affiliation of AlmWB phytoplasma strains, examined in this study, to three SNP lineages a, a2 and f5 (Table 2). ALKI1/ALKI2 strains were identified as the genetic lineage a. This genetic lineage appeared to be the original SNP genetic lineage present both in Lebanon and Iran. Other lineages represented 16S rDNA SNPs that were found only among Lebanese or Iranian strains; it means that some mutations in Iranian strains were never found in Lebanese strains, and vice versa.

Phytoplasma characterization by sequence analysis of variable genes

The *rplV-rpsC* (rp) gene amplicons (1000 bp) were obtained by nested PCR analysis from all '*Ca*. Phytoplasma phoenicium'-infected plants examined in this study (including 40 symptomatic *Prunus* plants, and four periwinkle plants infected by WIY2, SEJ2 LET1 and LET2 phytoplasma strains). Of these, the *rplV-rpsC* amplicons from 18 phytoplasma strains including 14 strains of AlmWB phytoplasma, and SEJ2, WIY2, LET1 and LET2 strains were sequenced (Table 1). The *rplV-rpsC* nucleotide sequence identity among Iranian AlmWB phytoplasma strains was > 99.2% that represents six variants (Table 1). The lowest rp sequence identity (~90%) was between strains of 16SrIX-B (AlmWB phytoplasma strains) and those of 16SrIX-C/-J. Phylogenetic tree generated using *rplV-rpsC* gene sequences had enough resolution power to classify AlmWB phytoplasma strains (from Iran and Lebanon; accession numbers listed in Table 1 and S1) in two subpopulations defined as Iranian (Ir) and Lebanese (Lb) subpopulations (Fig.3A). Members of these two

subpopulations could be distinguished by RFLP analysis of the *rplV-rpsC* amplicon using *Alu*I restriction enzyme (data not shown). Apparently, the geographical origin had a great impact on differentiation of AlmWB phytoplasmas into these subpopulations. The subpopulation Ir was composed only of Iranian AlmWB phytoplasma strains and the subpopulation Lb was composed of Lebanese strains together with Iranian strains ALGN1, ALGN2, AlKI1, and ALKI2 (Fig.3) Sequence identity analysis also revealed that ALGN1/ALGN2 and ALKI1/ALKI2 strains had *rplV-rpsC* sequence identity of, respectively, 99.9% (one SNP) and 100 % to the AlmWB phytoplasma strain SA213 from Lebanon. The other Iranian AlmWB phytoplasma strains examined had lower *rplV-rpsC* identities (99.2- 99.6%) to the same strain.

Phytoplasma strains from sesame phyllody (SEJ2 and WIY2; 16SrIX-C) and lettuce phyllody (LET1 and LET2, 16SrIX-J) were well resolved in two clusters closely related to strains KAP and PEY, respectively (Fig. 2). The remarkable *rplV-rpsC* sequence variability between AlmWB phytoplasma strains (16SrIX-B and its variant) and strains of other 16SrIX subgroups was also evident in the phylogenetic tree (Fig. 3A).

The *imp* gene sequence from 20 phytoplasma strains (listed in Table 1) examined in this study was sequenced and characterized. As shown in Table 1, nine imp sequence variants, of which six variants belonged to the AlmWB phytoplasma strains (16SrIX-B), were identified. The size of the *imp* gene was between 456 and 463 bp. NCBI BLASTP analysis identified the putative Imps (annotated as hypothetical proteins) of phytoplasma strains SA213 (16SrIX-B) and ChiP (16SrIX-C) as the closest relatives of the Imps from phytoplasmas identified in the present study. Sequence analysis using TOPCONS confirmed that all *imp* genes identified in this study code for a membrane anchored protein. The Imps of the phytoplasmas in 16SrIX group were predicted to contain a short cytoplasmic anchor of about 20 amino acids (aa) at N-terminus, connected to a

transmembrane domain of about 20 aa, while leaving the main part of the protein at C-terminus on the outside of the phytoplasma cell surface. As shown by the phylogenetic tree (Fig. 3B), the *imp* gene from all AlmWB phytoplasma strains were resolved in a cluster comprising two well-distinguished subpopulations, Ir and Lb. This clustering was similar, with better resolution, to that proposed by *rplV-rpsC* phylogenetic analysis. In this regard, subpopulation Ir comprised only of Iranian AlmWB phytoplasma strains and the subpopulation Lb was composed of ALKI1/AlKI2 and ALGN1/ALGN2 strains from Iran and the strain SA213 from Lebanon. The *imp* genes from the phytoplasma strains SEJ2, WIY2 and LET1/LET2 were resolved on three highly divergent lineages (Fig. 3B).

The lowest *imp* identity score among phytoplasmas of the 16SrIX (subgroups 16SrIX-B, IX-C and IX-J) was 62.3 % that was lower than that calculated among members of some other 16Sr groups (Table 3). Also the *imp* sequence identity between AlmWB phytoplasma strains (16SrIX-B) and strains of other 16SrIX subgroups (16SrIX-C or 16SrIX-J) was very low ranged from 62.3 to 69.1. This accords with the results obtained by *rplV-rpsC* sequence analysis, suggesting that AlmWB phytoplasma strains (16SrIX-B subgroup) were only distantly related to other phytoplasmas within the group 16SrIX. Moreover, the *imp* identity score among AlmWB phytoplasmas was as low as 94.1%; showing the high variability of this gene to study the genetic diversity of AlmWB phytoplasmas. As with rp sequence analysis, the Iranian AlmWB phytoplasma strains ALKI1/ALKI2 and ALGN1/ALGN2 had a higher *imp* sequence identity (>98.7%) to the Lebanese strain SA213 than to other Iranian AlmWB phytoplasma strains. This is in accordance with *imp* phylogenetic analysis, corroborating the divergence of AlmWB phytoplasma strains in two subpopulations. A large degree of *imp* sequence variability was also observed among phytoplasma strains of 16SrIX-C (e.g. between SEJ2 and WIY2).

Selection pressure on the *imp* gene

Imp gene sequences from phytoplasmas of the group IX with variable nucleotide sequences (10 *imp* variants; listed in Table 4) were examined for the presence of positive selection pressure. A significant positive selection (*dN>dS*) was found in eight *imp* gene sequence pairs with values ranging from 2.576 to 7.8. As shown in Table 4, the positive selection was only found between members of the two AlmWB phytoplasma subpopulations, established by *imp*- or rp-based phylogenetic analysis. This may also support the segregation of AlmWB phytoplasma strains in two subpopulations. Also, the excess number of non-synonymous (*N*) sites over synonymous sites (*S*) was found in nearly all *imp* gene sequence pairs (Table 4). This shows that 16SrIX phytoplasmas had some constraints to accumulate the nonsynonymous substitutions with higher rate than synonymous substitutions in their *imp* gene sequence.

A coiled-coil structure consisting of nine heptad repeats, positioned between the amino acids 60 and 123, was predicted in the Imp sequence of AlmWB phytoplasma strains. BLASTP database searches also identified homology between the Imp from AlmWB phytoplasmas of subpopulation Ir and numerous coiled-coil domain containing proteins. Interestingly the sequence homology was only found between the coiled-coil regions in the proteins. No coiled-coil domain was predicted in the Imp of phytoplasmas in the other 16SrIX subgroups examined in this study.

Discussion

In this study, genetic and biological characteristics, and phylogenetic relationships of 16SrIX Iranian phytoplasma strains were studied and compared with phytoplasmas from other countries (strains and accession numbers listed in Tables 2 and S1). Presence of the AlmWB phytoplasma

(16SrIX-B and its variant) was demonstrated in plants showing previously unreported symptoms including leaf rolling and yellowing, observed on peach and GF-677 rootstocks in central regions of Iran. These symptoms did not include witches'-broom or rosetting, typically observed on AlmWB-affected peach trees (Molino Lova *et al*., 2011). Similar symptoms were also observed on apricot trees affected by AlmWB phytoplasma in Iran (Salehi *et al*., 2018).

Even if the results of this and other studies identified mutually exclusive single nucleotide polymorphisms within 16S rDNA of Lebanese and Iranian AlmWB phytoplasma strain populations (Salehi *et al*., 2018, 2020), phylogenetic analyses did not clearly separate AlmWB phytoplasma strains with respect to distinct biological or ecological properties (i.e. host range, symptoms, geographic origin). Analysis of *rplV-rplC* and *imp* gene sequences revealed high heterogeneity among 16SrIX phytoplasmas, notably between AlmWB phytoplasmas (16SrIX-B) and those of other 16SrIX subgroups. This hypothesizes a long-time independent evolutionary history between AlmWB phytoplasma strains and phytoplasmas in other 16SrIX subgroups. Similarly, Martini *et al*. (2007) delineated AlmWB phytoplasmas in an rp-based phylogenetic subclade distinct from other members of the 16SrIX group. However, further investigation is required to determine if AlmWB phytoplasma strains could be delineated as a monophyletic lineage.

The majority of examined Iranian AlmWB phytoplasma strains (Iranian subpopulation Ir) could be distinguished from Lebanese strains. As also reported elsewhere (Kumar *et al*., 2019), this proposes that country of origin (geographical isolation) played a major role in the genetic diversity of AlmWB phytoplasma population. The exception, that is reported for the first time in this study, was phytoplasma strains AlGN1/ALGN2 and particularly ALKI1/ALKI2 (identified in Kermanshah province) that were genetically and phylogenetically closer to Lebanese strains.

Although this finding reveals a high sequence diversity within Iranian AlmWB phytoplasma population, it hypothesizes for a new introduction of AlmWB phytoplasma strains from Lebanon to Iran.

Based on the rp gene phylogenetic analysis, sesame phyllody (SEJ2 and WIY2; 16SrIX-C) and lettuce phyllody (LET1 and LET2; 16SrIX-J) phytoplasmas were separately clustered with reference strains of 16SrIX-C (KAP and PEY). This evidently shows the high heterogeneity of phytoplasmas classified within the ribosomal subgroup 16SrIX-C. In other words it reveals the high resolution power of rp gene to discriminate among phytoplasmas classified within the subgroup 16SrIX-C.

Imp is one of the three non-homologous antigenic membrane protein genes proposed to play determinant roles in biology of phytoplasmas. This gene has been identified in diverse phytoplasmas, hence, it was supposed to be existed in the ancestor of the phytoplasmas (Kakizawa *et al*., 2006). The Imp has been considered as a biologically important protein that engaged in mutualistic phytoplasma-host interactions (Siampour *et al*., 2011; Boonord *et al*., 2012). Findings of this study confirmed the suitability of *imp* gene in fine differentiation of closely related phytoplasmas within the group 16SrIX. Similarly, *imp* gene has been used in several studies as a genetic marker for differentiation of closely related phytoplasmas (Bohunická *et al*., 2018; Siampour *et al*., 2019b). The topology of the *imp* based phylogenetic tree was comparable to that depicted based on conserved rp gene. Due to the higher variability, however, the resolution power of the *imp* phylogenetic tree was higher than the rp-based phylogenetic tree.

Pairwise comparisons using Nei-Gojoboori method provided strong evidence that *imp* gene is under strong positive selection (high *dN/dS* values) in eight out of the 15 sequence pairs; only between members of the two AlmWB strain subpopulations Ir and Lb. Such high values indicates

that the comparisons were made between closely related but biologically distinct phytoplasma strains (Kakizawa et al., 2006). Thereby, it could be hypothesized that these two designated AlmWB subpopulations have developed biologically distinct characteristics. The coiled-coil domain present in Imp of AlmWB phytoplasmas further reveals the potential of the Imp to bind other proteins (e.g., host proteins). Altogether, it could be reasoned that the high variability and positive selection events in the *imp* gene were promoted by adaptation of AlmWB phytoplasmas to different environment of insect or plant hosts. In this regard, it has been shown that Imp and Amp (another phytoplasma antigenic membrane protein) play important role in phytoplasma transmissibility by insects (Galetto *et al*., 2011. Siampour *et al*., 2011). Conceivably, the significant *imp* sequence variability may explain why Iranian AlmWB strains (subpopulation Ir) were not transmitted by the leafhopper *Asymmetrasca decedens* known to transmit Lebanese strains (subpopulation Lb) (Abou-Jawdah *et al*., 2014; Taghizadeh and Salehi, 2002).

A very high *imp* sequence variation was also observed between sesame phyllody phytoplasma strains SEJ2 and WIY2 (16SrIX-C). This finding suggests that these strains, infecting the same plant host, could have distinct biological features leading to differential host adaptation. This may explain why 16SrIX-C phytoplasma strains were found in association with an AlmWB-like disease in Iran but not in other countries (Salehi *et al*., 2006; Casati *et al*., 2016). Thereby, results of this study suggest the use of biologically important genes such as *imp* for identification of biologically or ecologically distinct phytoplasma strains that could not be differentiated by conserved genes alone. In such fashion, Quaglino *et al*. (2015) used an integral membrane protein (*inmp*) gene as a marker to distinguish AlmWB phytoplasma strains isolated from different host plants in Lebanon.

The rp and *imp* gene sequences used in this study corroborated the differentiation of AlmWB phytoplasma strains in two subpopulations that may have distinct insect vector or plant host specificity. Based on the current sequence data it seemed that the predominant AlmWB phytoplasma strains in Iran and Lebanon had some unique genetic or biological features.

Acknowledgments

This work was supported by a grant from the Shahrekord University, Shahrekord, Iran

References

Abou-Jawdah Y., Abdel Sater A., Jawhari M., Sobh H., Abdul-Nour H., Bianco P.A., Molino Lova M., Alma A. (2014) Asymmetrasca decedens (cicadellidae, Typhlocybinae), a natural vector of '*Candidatus* phytoplasma phoenicium'. *Annals of Applied Biology*, **165,** 395–403.

Abou-Jawdah Y., Karakashian A., Sobh H., Martini M., Lee I.-M. (2002). An epidemic of almond witches'-broom in Lebanon: classification and phylogenetic relationships of the associated phytoplasma. *Plant Disease*, **86,** 477–484

Bernsel A., Viklund H., Hennerdal A., Elofsson A. (2009). TOPCONS: consensus prediction of membrane protein topology. *Nucleic Acids Research,* **37,** W465–W468

Bertaccini A., Lee I.M. (2018). Phytoplasmas: An Update. In: *Phytoplasmas: Plant Pathogenic Bacteria – I,* pp.1-29. Eds G. Rao, A. Bertaccini and L. Liefting. Springer, Singapore,

Bohunická M., Valentová L., Suchá J., Nečas T., Eichmeier A., Kiss T., Cmejla R. (2018). Identification of 17 '*Candidatus* Phytoplasma pyri'genotypes based on the diversity of the imp gene sequence. *Plant Pathology,* **67,** 971-977.

Boonrod K., Munteanu B., Jarausch B., Jarausch W., Krczal G. (2012). An immunodominant membrane protein (Imp) of '*Candidatus* Phytoplasma mali' binds to plant actin. *Molecular Plant Microbe Interaction,* **25,** 889-895.

Casati P., Quaglino F., Abou-Jawdah Y., Picciau L., Cominetti A., Tedeschi R., Jawhari M., Choueiri E., Sobh H., Lova M.M., Beyrouthy M. (2016). Wild plants could play a role in the spread of diseases associated with phytoplasmas of pigeon pea witches'-broom group (16SrIX). *Journal of Plant Pathology,* **98,** 71-81.

Deng S., Hiruki C. (1991). Genetic relatedness between two nonculturable mycoplasmalike organisms revealed by nucleic acid hybridization and polymerase chain reaction. *Phytopathology*, **81,** 1475– 1479.

Esmaeilzadeh-Hosseini S.A., Babaie G., Salehi M., Bertaccini A. (2018). Molecular differentiation of 16SrIX-I phytoplasmas detected inOnobrychis viciifolia leaf yellowing in Iran from phytoplasmas in 16SrIX-J subgroup. *Phytopathogenic Mollicutes*, **8**, 24-31.

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

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

Hogenhout S.A., Oshima K. Ammar E. Kakizawa S., Kingdom H.N., Namba S. (2008). Phytoplasmas: bacteria that manipulate plants and insects. *Molecular Plant Pathology,* **9,** 403– 423.

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

Kakizawa S., Oshima K., Namba S. (2006). Diversity and functional importance of phytoplasma membrane proteins. *Trends in Microbialogy.* **14,** 254-256.

Kumar S., Abou-Jawdah Y., Siampour M., Sobh H., Tedeschi R., Alma A., Bianco P.A., Quaglino F. (2019). Genetic diversity of 'Candidatus Phytoplasma phoenicium'strain populations associated with almond witches' broom in Lebanon and Iran. *Phytopathogenic Mollicutes*, **9**, 217-218.

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., Bottner-Parker K.D., Zhao Y., Bertaccini A., Davis R.E. (2012). Differentiation and classification of phytoplasmas in the pigeon pea witches'-broom group (16SrIX): an update based on multiple gene sequence analysis. *International Journal of Systematic and Evolutionary Microbiology,* **62,** 2279-2285.

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.

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

Martini M., Lee I.-M., Bottner K.D., Zhao Y., Botti S., Bertaccini A., Harrison N.A., Carraro L., Marcone C., Khan A.J., 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.

Messier W., Stewart C. B. (1997). Episodic adaptive evolution of primate lysozymes. *Nature,* **385,** 151–154.

Molino Lova M., Quaglino F., Abou-Jawdah Y., Choueiri E., Sobh H., Casati P., Tedeschi R., Alma A., Bianco P. (2011). Identification of new 16SrIX subgroups,-F and-G, among '*Candidatus* Phytoplasma phoenicium' strains infecting almond, peach and nectarine in Lebanon. *Phytopathologia Mediterranea,,* **50,** 273–282.

Muhire B.M., Varsani A., Martin D.P. (2014). SDT: a virus classification tool based on pairwise sequence alignment and identity calculation. *PLoS One*; **9**:e108277

Pérez-López E., Omar A.F., Al-Jamhan K.M., Dumonceaux T.J. (2018). Molecular identification and characterization of the new 16SrIX-J and cpn60 UT IX-J phytoplasma subgroup associated with chicory bushy stunt disease in Saudi Arabia. *International Journal of Systematic and Evolutionary Microbiology,* **68,** 518-522.

Quaglino F., Kube M., Jawhari M., Abou-Jawdah Y., Siewert C., Choueiri E., Sobh H., Casati P., Tedeschi R., Lova M.M., Alma A. (2015). '*Candidatus* Phytoplasma phoenicium'associated with almond witches'-broom disease: from draft genome to genetic diversity among strain populations. *BMC microbiology,* **15,** 148.

Salehi M., Haghshenas F., Khanchezar A., Esmailzadeh-Hosseini S.A. (2011). Association of '*Candidatus* Phytoplasma phoenicium' with GF-677 witches'-broom in Iran. *Bulletin of Insectology,* **64,** S113–S114.

Salehi M., Hosseini S.A.E., Salehi E., Quaglino F. and Bianco P.A. (2020). Peach witches'-broom, an emerging disease associated with '*Candidatus* Phytoplasma phoenicium'and '*Candidatus* Phytoplasma aurantifolia'in Iran. *Crop Protection*, **127,** 104946.

Salehi M., Hosseini S.E., Salehi E., Bertaccini A. (2017). Genetic diversity and vector transmission of phytoplasmas associated with sesame phyllody in Iran. *Folia microbiologica*, ***62*,** 99-109.

Salehi M., Izadpanah K., Heydarnejad J. (2006). Characterization of a new almond witches' broom phytoplasma in Iran. *Journal of Phytopathology,* **154,** 386–391.

Salehi M., Izadpanah K., Nejat N., Siampour M. (2007). Partial characterization of phytoplasmas associated with lettuce and wild lettuce phyllodies in Iran. *Plant Pathology,* **56,** 669-676.

Salehi M., Salehi E., Abbasian M., Izadpanah K. (2015). Wild almond (*Prunus scoparia*), a potential source of almond witches' broom phytoplasma in Iran. *Journal of Plant Pathology,* **97,** 377–381.

Salehi M., Salehi E., Siampour M., Quaglino F., Bianco P.A. (2018). Apricot yellows associated with '*Candidatus* Phytoplasma phoenicium' in Iran. *Phytopathologia Mediterranea,* **57,** 269–283.

Schneider B., Seemüller E., Smart C.D., Kirkpatrick B.C. (1995). Phylogenetic classification of plant pathogenic mycoplasma-like organisms orphytoplasmas. In *Molecular and diagnostic procedures in mycoplasmology.* pp. 369–80. Eds S. Razin and J.G. Tully. San Diego, CA, USA: Academic,.

Siampour M., Izadpanah K., Salehi M., Afsharifar A. (2019a). Occurrence and Distribution of Phytoplasma Diseases in Iran. In *Sustainable Management of Phytoplasma Diseases in Crops Grown in the Tropical Belt. Sustainability in Plant and Crop Protection,* vol. 12. pp. 47-86. Eds C. Olivier, T. Dumonceaux and T. Pérez-López. Springer.

Siampour M., Galetto L., Bosco D., Izadpanah K., Marzachì C. (2011). *In vitro* interactions between immunodominant membrane protein of lime witches' broom phytoplasma and leafhopper vector proteins. *Bulletin of Insectology,* **64,** S149-S150.

Siampour M., Izadpanah K., Martini M., Salehi M. (2019b). Multilocus sequence analysis of phytoplasma strains of 16SrII group in Iran and their comparison with related strains. *Annals of Applied Biology,* **175,** 83-97.

Taghizadeh M., Salehi M. (2002). Leafhoppers of subfamily Typhlocybinae found on almond (*Prunus amygdalus*) in Fars province. In: Proceedings, 15th Iranian Plant Protection Congress, pp. 92–93, 2002, Kermanshah, Iran.

Verdin E., Salar P., Danet J.L., Choueiri E., Jreijiri F.E., Zammar S., Gelie B., Bové J., Garnier M. (2003). '*Candidatus* Phytoplasma phoenicium' sp. nov, a novel phytoplasma associated with an emerging lethal disease of almond trees in Lebanon and Iran. *International Journal of Systematic and Evolutionary Microbiology,* **53,** 833–838.

Zhao Y., Wei W., Lee M., Shao J., Suo X., Davis R.E. (2009). Construction of an interactive online phytoplasma classification tool, iPhyClassifier, and its application in analysis of the peach X-disease phytoplasma group (16SrIII). *International Journal of Systematic and Evolutionary Microbiology,* **59,** 2582-2593

Supporting information

Table S1. Strains or species, subgroup attribution, geographical origin and 16S rRNA , *rplV-rpsC*, and *imp* accession numbers of phytoplasmas used in this study

Table S2. Symptomatic and asymptomatic *Prunus* trees sampled from three geographical location of Iran and tested for phytoplasma presence

Table S1. Strains, subgroup attribution, geographical origin and 16S rRNA , *rplV-rpsC*, and *imp* accession numbers of phytoplasmas used in this study

Phytoplasma strains	Strain or abbreviation	Subgroup	Country of origin	16S rRNA	*rplV-rpsC*	*imp*
Onion yellows	OY-M	16SrI-B	Japan	NC_005303.2	NC_005303.2	NC_005303.2
Aster yellows witches'-broom	AYWB	16SrI-A	USA	NC_007716.1	NC_007716.1	NC_007716.1
Gomphocarpus physocarpus witches'-broom '*Ca*. P. australiense'	CBWB	16SrXII-B	Australia	NC_010544.1	NC_010544.1	NC_010544.1
'Ca. P. solani' (strain 231/09)	STOL-231/09	16SrXII-A	Serbia	NC_022588	NC_022588	NC_022588
Apple proliferation (*'Ca*. P. mali')	AT	16SrX-A	Germany	NC_011047.1	NC_011047	NC_011047
'*Ca*. P. pyri'	PD1	16SrX-C	Germany	AJ542543	EF193370	
'*Ca*. P. pyri'	AA973		Czech Republic			MF374927
'*Ca*. P. prunorum'	LNS2		Italy	-	EF193369	
'*Ca*. P. prunorum'	ESFY G2		Germany	AJ542545	-	
'*Ca*. P. prunorum'	Azer10		Azerbaijan	-	-	FN600711
'Ca. P. pruni'	WX-95	16SrIII-A	USA	JQ044396		
'Ca. P. pruni'	CX	16SrIII-A	Canada		NZ_LHCF01000006	NZ_LHCF01000020
Poinsettia Branch-Inducing	JR1	16SrIII-H	USA	AKIK00000000	AKIK00000000	AKIK00000000
Italian clover phyllody	MA1	16SrIII-B	Italy	AKIM00000000	AKIM00000000	AKIM00000000
Vaccinium witches'-broom	VAC	16SrIII-F	Germany	AKIN00000000	AKIN00000000	AKIN00000000
Milkweed yellows	MW1	16SrIII-F	USA	AKIL00000000	AKIL00000000	AKIL00000000
Peanut witches'-broom	PnWB	16SrII-A	Taiwan	AMWZ00000000	AMWZ00000000	AMWZ00000000
'*Ca*. P. aurantifolia'	WBDL	16SrII-B	Oman	EF186828	EF186815	NZ MWKN01000023.1
Crotalaria phyllody	CrP	16SrII-C	Thailand	EF193355	EF186818	JQ745277
Cucumber phyllody	CuPh	16SrII-D	Iran	KY412986	KY365523	KY365515
Gliricidia little leaf	GLLhon	16SrIX-F	Honduras	AF361017	EF186800	
Knautia arvensis phyllody	KAP	16SrIX-C	Italy	EF186823	EF186801	
Picris echioides yellows	PEY	16SrIX-C	Italy	Y16389	EF186802	
Pigeon pea witches'-broom	PPWB	16SrIX-A	USA	AF248957	EF193383	
Chicory phyllody	ChiP	16SrIX-C	Italy	--	PUUG00000000	PUUG00000000.1
'*Ca*. P. phoenicium'	SA213	16SrIX-B	Lebanon	KM275491	NZ_JPSQ00000000	NZ_JPSQ00000000
'*Ca*. P. phoenicium'	AlmWB-A112	16SrIX-B	Lebanon		EF186803	
Almond witches'-broom	LBB001	16SrIX-B	Lebanon		HM745927	
'*Ca*. P. phoenicium'	A4	16SrIX-B	Lebanon	AF515636		
Almond witches'-broom	N28-1	16SrIX-B	Lebanon	HQ407526		
Almond witches'-broom	Al4	16SrIX-B	Lebanon	HQ407521		
Almond witches'-broom	Smilax13	16SrIX-B	Lebanon	KF583757		
Almond witches'-broom	Urmia	16SrIX-B*	Iran	MH363614		

Almond witches'-broom	KermanII	16SrIX-B	Iran	JN565013	
Almond witches'-broom	Borujerd	16SrIX-B	Iran	MH363615	
Almond witches'-broom	21	16SrIX-B*	Iran	AF515637	
Almond witches'-broom	Bardseer	16SrIX-B*	Iran	MH363613	
Almond witches'-broom	Naeen	16SrIX-B	Iran	MH363617	
Almond witches'-broom	G317	16SrIX-B	Italy	MK377252	
Brinjal little leaf	BLL	16SrVI	India		EF183489

* Variant of the subgroup 16SrIX-B

Table S2: Symptomatic and asymptomatic *Prunus* trees sampled from three geographical location of Iran tested for presence of phytoplasma

Plant host	Location (Province)	Main symptoms	Positive/No. tested
Peach	Chaharmahal-Bakhtiari	Leaf roll, mild yellowing	20/20
GF-677	Chaharmahal-Bakhtiari	Small leaves	4/4
Almond	Fars	Witches' -broom	5/5
Almond	Chaharmahal-Bakhtiari	Witches' -broom	1/1
Wild Almond	Fars	Witches' -broom	6/6
Almond	Kermanshah	Severe witches' -broom	4/4
Peach	Chaharmahal-Bakhtiari	Symptomless	1/7

Table 1. Strain, host, geographical location and accession numbers of gene sequences determined in this study

Strain (subgroup)	Natural host	Location	Sequence identity and Acc. No.[a]			Type[b]
			16S rRNA	*rplv-rpsC*	*imp*	
AlKI1 (IX-B)	almond	Kermanshah	a (MT845871)	a (MT845603)	a (MT845596)	B1
AlKI2 (IX-B)	almond	Kermanshah	a	a	a	B1
AlGN1 (IX-B*)	GF-677	Chaharmahal-Bakhtiari	b (MT845872)	b (MT845604)	b (MT845595)	B2
AlGN2 (IX-B*)	GF-677	Chaharmahal-Bakhtiari	b	b (MT845604)	b	B2
WAlmWB1 (IX-B)	wild almond	Fars	c	c (MT845601)	c	B3
WAlmWB2 (IX-B)	wild almond	Fars	c	not sequenced	c	
PEAB (IX-B)	peach	Chaharmahal-Bakhtiari	c	d	c	B4
PE5 (IX-B)	peach	Chaharmahal-Bakhtiari	c	d	d (MT845591)	B5
PC (IX-B)	peach	Chaharmahal-Bakhtiari	c	d	c	B4
PB (IX-B)	peach	Chaharmahal-Bakhtiari	c	d	c	B4
PA (IX-B)	peach	Chaharmahal-Bakhtiari	c (MT845873)	d (MT845602)	c (MT845594)	B4
PE2 (IX-B)	peach	Chaharmahal-Bakhtiari	c	d	c	B4
ALNS (IX-B)	almond	Fars	c	e (MT845600)	e (MT845592)	B6
ALCM (IX-B)	almond	Fars	c	d	f (MT845593)	B7
ALC (IX-B)	almond	Chaharmahal-Bakhtiari	a	d	c	B8
ALR (IX-B)	almond	Fars	c	not sequenced	c	
SEJ2 (IX-C)	sesame	Fars	d (MT845874)	f (MT845605)	g (MT845598)	C1
WIY2(IX-C)	sesame	Fars	d	g (MT845606)	h (MT845597)	C2
LET1(IX-J)	lettuce	Fars	e (MT845875)	h (MT845607)	i (MT845599)	J1
LET2(IX-J)	lettuce	Fars	e	h	i	J1

[a] One nucleotide sequence, representative of identical sequences of 16S rRNA , *rplV-rpsC* and *imp* genes, was deposited in GenBank. Identical 16S rRNA , *rplV-rpsC* and *imp* gene sequences are shown by the same character in each column.

[b] Type is determined by the combination of 16S rRNA , *rplV-rpsC* and *imp* gene sequence variants

Table 2. SNP lineages and positions in the R16F2n/R16R2 16S rDNA sequence of AlmWB phytoplasma strains

Strain	Origin	Lineage	SNPs in 16S rDNA [a] (position from the annealing site of the primer R16F2n)																				
			159	**287**	**460**	**473**	572	639	**646**	**690**	**691**	713	**759**	781	806	**863**	**901**	**940**	983	**1089**	**1107**	1152	1171
ALKI1 (MT845871)	Iran	a	T	A	T	G	G	G	-	T	G	A	A	A	C	-	-	C	G	T	A	T	T
ALGN1 (MT845872)	Iran	f5	C	A	T	G	G	G	-	T	G	A	A	A	C	-	-	C	G	G	A	T	T
PA (MT845873)	Iran	a2	T	A	T	G	G	G	-	T	G	A	A	A	C	-	-	C	G	G	A	T	T
Bardseer (MH363613)	Iran	f4	C	A	T	G	G	G	-	T	G	A	A	A	C	-	-	C	G	G	G	T	T
Urmia (MH363614)	Iran	f1	C	A	T	G	G	G	C	T	G	A	A	A	C	A	-	C	G	G	A	T	T
Borujerd (MH363615)	Iran	g	T	A	C	G	G	G	-	T	G	A	G	A	C	-	-	C	G	G	A	T	T
Chenaran (MH363616)	Iran	a1	T	G	T	G	G	G	-	T	G	A	A	A	C	-	-	C	G	T	A	T	T
Naeen (MH363617)	Iran	a1	T	G	T	G	G	G	-	T	G	A	A	A	C	-	-	C	G	T	A	T	T
Sanandaj (MH363618)	Iran	a1	T	G	T	G	G	G	-	T	G	A	A	A	C	-	-	C	G	T	A	T	T
Zarghan (MH363619)	Iran	g	T	A	C	G	G	G	-	T	G	A	G	A	C	-	-	C	G	G	A	T	T
A4 (AF515636)	Lebanon, Iran	a	T	A	T	G	G	G	-	T	G	A	A	A	C	-	-	C	G	T	A	T	T
N27-2 (HQ407532)	Lebanon	b1	T	A	T	G	C	G	-	T	G	A	A	A	C	-	-	C	G	T	A	T	T
N5 (HQ407522)	Lebanon	b2	T	A	T	G	C	G	-	T	G	A	A	A	C	-	-	C	G	T	A	G	T
N28-1 (HQ407526)	Lebanon	b3	T	A	T	G	G	G	-	T	G	A	A	A	C	-	-	C	G	T	A	G	T
N29-1 (HQ407529)	Lebanon	c1	T	A	T	G	G	T	-	T	G	A	A	A	C	-	-	C	G	T	A	T	T
Al4 (HQ407521)	Lebanon	c2	T	A	T	G	G	G	-	T	G	A	A	A	T	-	-	C	G	T	A	T	T
Smasp (KP851773)	Lebanon	c3	T	A	T	G	G	G	-	T	G	A	A	C	T	-	-	C	G	T	A	T	T
P3-1 (HQ407518)	Lebanon	c4	T	A	T	G	G	T	-	T	G	A	A	C	T	-	-	C	G	T	A	T	T
Smilax12 (KF583756)	Lebanon	d	T	A	T	G	G	G	-	T	G	A	A	A	C	-	-	C	A	T	A	T	T
Smilax13 (KF583757)	Lebanon	e	T	A	T	G	G	G	-	T	G	G	A	A	C	-	-	C	G	T	A	T	C
A21 (AF515637)	Iran	f1	C	A	T	G	G	G	C	T	G	A	A	A	C	A	-	C	G	G	A	T	T
Breijan (KY014991)	Iran	f2	C	A	T	A	G	G	-	C	T	A	A	A	C	-	-	C	G	G	A	T	T
Kavar (KM235725)	Iran	f3	C	A	T	G	G	G	-	T	G	A	A	A	C	-	G	C	G	G	G	T	T
Meymand (KM235727)	Iran	g	T	A	C	G	G	G	-	T	G	A	G	A	C	-	-	C	G	G	A	T	T
Moshkan (JN565017)	Iran	h1	T	A	T	G	G	G	-	T	G	A	A	A	C	-	-	A	G	G	A	T	T
Kerman II (JN565013)	Iran	h2	T	A	T	G	G	G	-	T	G	A	A	A	C	-	-	A	G	T	A	T	T

[a] Distinctive SNP site positions detected in Lebanese and Iranian AlmWB strains are shown in normal and bold characters

Table 3. *Imp* gene nucleotide sequence scores (%) in pairwise comparisons between selected phytoplasma strains identified in this study and from GenBank

#	Strain	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	AlKI1 (IX-B)	ID																	
2	AlGN1 (IX-B)*	98.9	ID																
3	PE5 (IX-B)	94.4	94.5	ID															
4	PA (IX-B)	94.6	94.8	99.8	ID														
5	ALNS (IX-B)	94.4	94.5	99.6	99.8	ID													
6	ALCM (IX-B)	94.4	94.5	99.6	99.8	99.6	ID												
7	SEJ2 (IX-C)	65.4	65.3	62.3	62.8	62.8	63	ID											
8	WIY2(IX-C)	68.4	69.1	62.9	62.9	62.9	63	77	ID										
9	LET1(IX-J)	65.3	65.5	64	64	63.8	64	77	74	ID									
10	SA213 (IX-B)	99.3	98.7	94.1	94.3	94.1	94.1	65.5	68.4	65.3	ID								
11	ChiP (IX-C)	65.1	65.3	63.3	63.3	63.1	63.5	77.1	74	99	65.3	ID							
12	CX (III-A)	62.3	62.3	64.7	64.9	64.4	65	62.2	67.3	63.4	62.5	63.4	ID						
13	MW1 (III-F)	60.4	58.9	63.2	63.2	63.2	64	62.1	63	62.6	59.9	62.5	90.2	ID					
14	AT (X-A)	63.5	63.5	62.8	63	62.8	63.2	65.2	68	66	63	65.9	61	64.9	ID				
15	Azer10 (X-F)	65.1	63.6	61.5	61.8	62	62	64	64	62.4	64.3	65	59.4	64.9	70.2	ID			
16	AA973 (X-C)	68.5	68.2	65.6	67	67.2	67	63.6	67.1	63.6	68	63.7	63	62.5	72.9	71.2	ID		
17	WBDL (II-B)	63.7	64.9	62.8	63	63.5	62.8	62.1	61	65	66	65.3	63	64.7	61.8	63	65.8	ID	
18	PnWB (II-A)	67.8	66	61.1	62.7	62.7	63	68.3	62.1	62.8	68	63.3	61.9	60.1	62.2	60.9	63.3	82	ID

*Variant of the subgroup 16SrIX-B

Table 4. Number of synonymous sites (*S*), number of non-synonymous sites (*N*) and the criterion *dN*/*dS* of the *imp* pairwise sequence comparisons calculated by Nei-Gojobori method.

Parameter	WAlmWB1	ALNS-F	PE5	ALGN1	ALKI1	SA213	WIY2	SEJ2	LET1
ALNS-F									
dN/dS	-								
No. of *N*	1								
No. of *S*	0								
PE5									
dN/dS	0	0.272							
No. of *N*	0	1							
No. of *S*	1	1							
ALGN1									
dN/dS	4.187**	4.375**	2.576*						
No. of *N*	22.5	23.5	22.5						
No. of *S*	1.5	1.5	2.5						
ALKI1									
dN/dS	3.285*	3.6*	2.156	1.2					
No. of *N*	23	24	23	4					
No. of *S*	2	2	3	1					
SA213									
dN/dS	7.5***	7.8***	3.5710**	---	0.6				
No. of *N*	25	26	25	6	2				
No. of *S*	1	1	2	0	1				
WIY2									
dN/dS	1.001	1.006	0.976	0.976	0.996	1.039			
No. of *N*	150.33	150.58	149.83	147	149	149			
No. of *S*	40.66	40.42	41.17	41	41	40			
SEJ2									
dN/dS	0.995	0.989	1.017	0.974	0.981	0.939	1.253		
No. of *N*	151.83	151.83	151.33	151.75	148.75	147.75	101.75		
No. of *S*	42.16	42.16	42.66	43.25	42.25	43.25	23.25		
LET1									
dN/dS	1.257	1.283	1.224	1.181	1.194	1.218	1.282	0.791	
No. of *N*	141.92	143.25	141.42	138	139	138.5	108	84.08	
No. of *S*	33.08	32.75	33.58	34	34	33.5	24	27.92	
ChiP									
dN/dS	1.360	1.391	1.326	1.278	1.294	1.319	1.357	0.869	1.4
No. of *N*	144.92	146.25	144.42	141	142	141.5	108	86.25	5
No. of *S*	32.08	31.75	32.58	33	33	32.5	23	26.75	1

The significant difference between *dN* and *dS* at 5% (*), 0.01 (**) and 0.001 (***) levels are shown

Figure legends

Fig. 1. Symptoms of AlmWb phytoplasma on different host in Iran. Typical severe witches'-broom symptoms on a branch of Almond tree in Kermanshah (A); shoot proliferation with appearance of witches'-broom on the main trunk on an almond tree in Chaharmahal-Bakhtiarri (B); reddening, leaf rolling and dieback of a peach tree in Chaharmahal-Bakhtiarri (C); Witches'-broom symptoms onold almond trees in Fars province (D); witches'-broom and perpendicular growth of the buds on the branches of wild almond in Fars province.

Fig. 2. Maximum likelihood phylogenetic tree of phytoplasmas reconstructed by 16S rRNA gene sequence analysis. Numbers on the branches show the bootstrap values >60 obtained from 1000 replicates. The 16SrIX phytoplasma strains whose 16Sr RNA gene sequences obtained in this study are shown in bold character. See table 2 and S1 for details of phytoplasma strains and accession numbers.

Fig. 3. Maximum likelihood phylogenetic tree reconstructed using *rplV-rpsC* (A) and *imp* gene sequences (B) of diverse phytoplasma strains. Numbers on the branches show the bootstrap values of >50 obtained from 1000 replicates. The 16SrIX phytoplasma strains whose the *rplv-rpsC* or *imp* gene sequences obtained in this study are shown in bold character See tables 2 and S1 for details of phytoplasma strains and accession numbers.

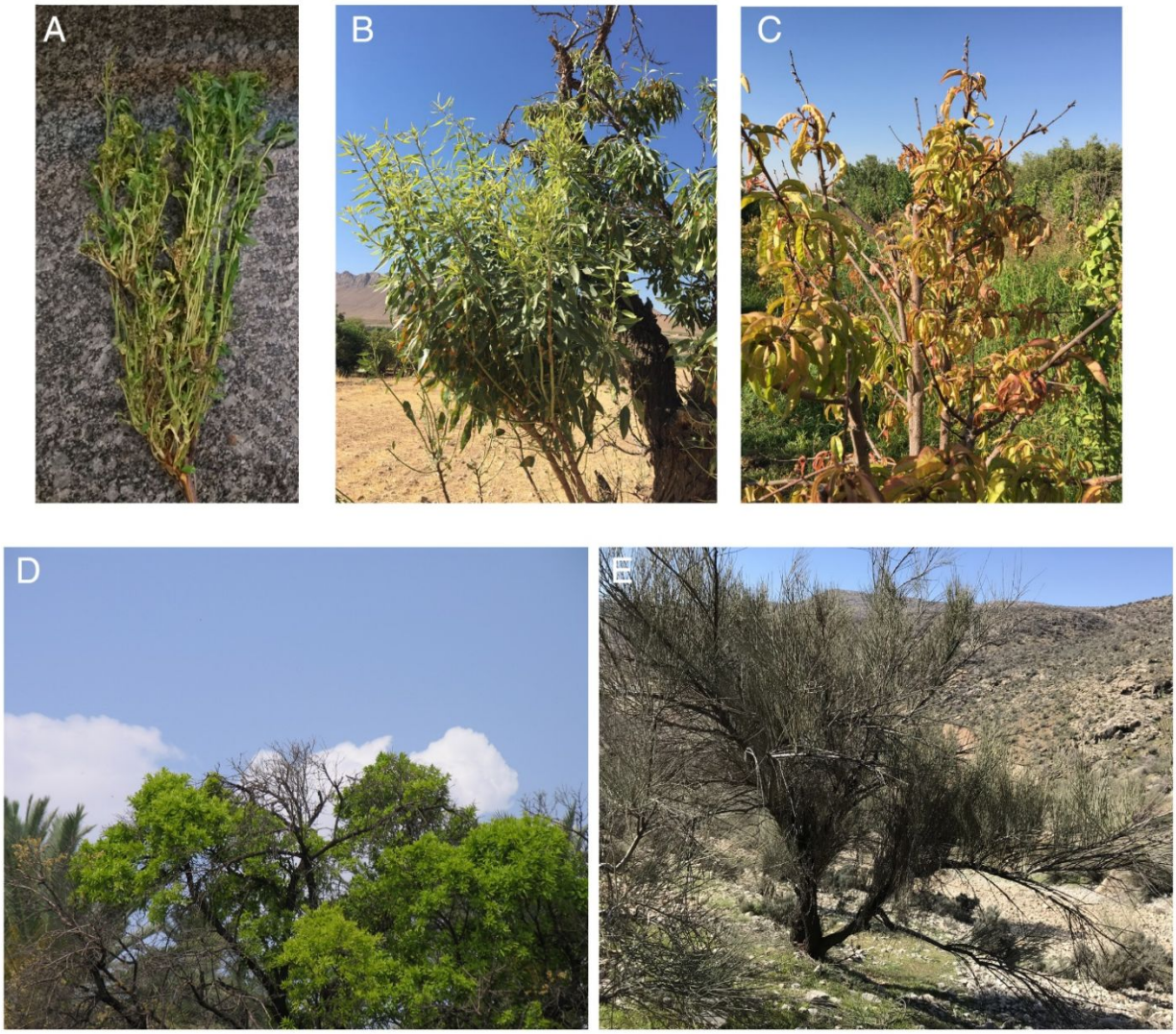

Fig. 1.

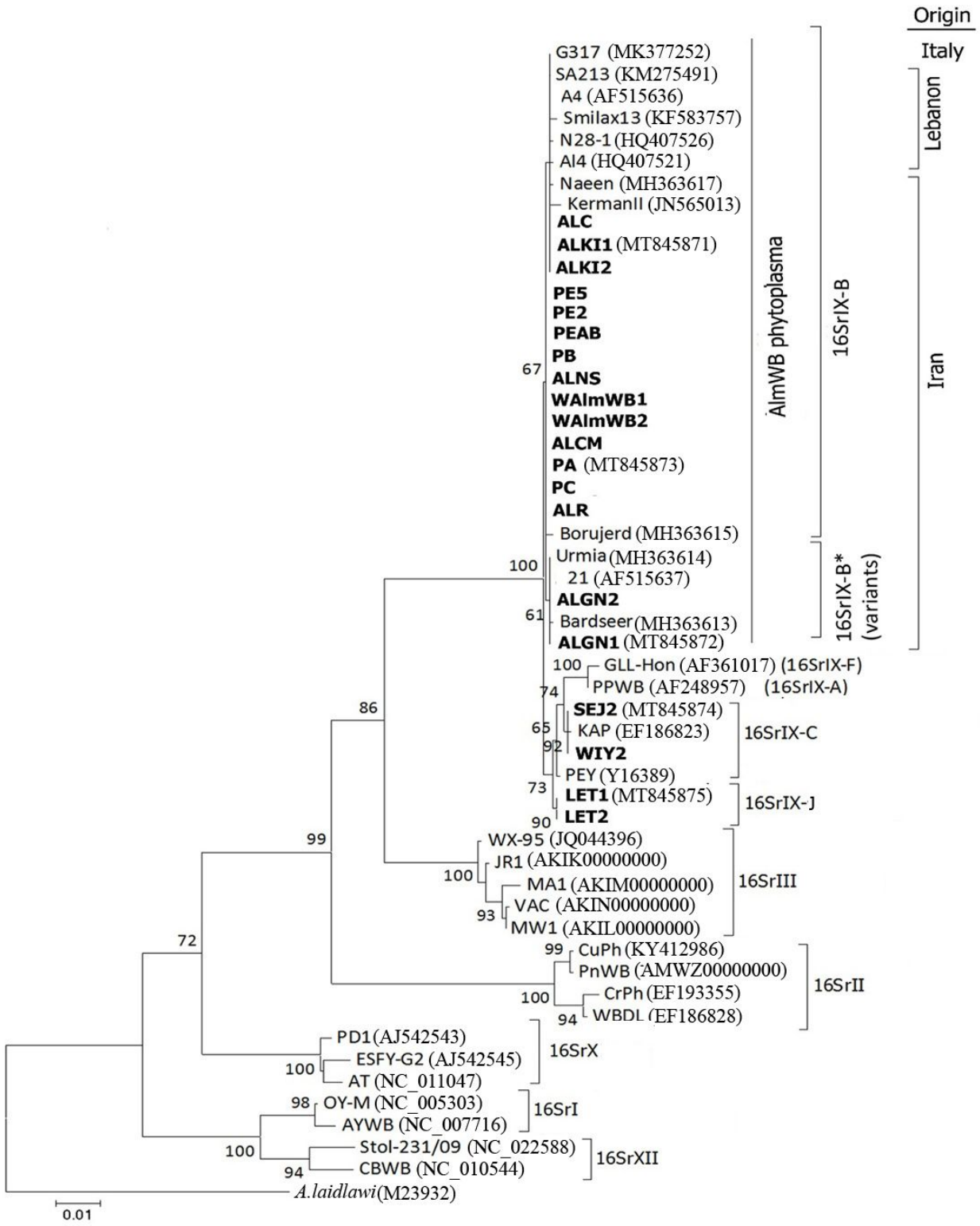
Origin
Italy
Lebanon
Iran
G317 (MK377252)
SA213 (KM275491)
A4 (AF515636)
Smilax13 (KF583757)
N28-1 (HQ407526)
Al4 (HQ407521)
Naeen (MH363617)
KermanII (JN565013)
ALC
ALKI1 (MT845871)
ALKI2
PE5
PE2
PEAB
PB
ALNS
WAlmWB1
WAlmWB2
ALCM
PA (MT845873)
PC
ALR
Borujerd (MH363615)
Urmia (MH363614)
21 (AF515637)
ALGN2
Bardseer (MH363613)
ALGN1 (MT845872)
AlmWB phytoplasma
16SrIX-B
16SrIX-B* (variants)
GLL-Hon (AF361017) (16SrIX-F)
PPWB (AF248957) (16SrIX-A)
SEJ2 (MT845874)
KAP (EF186823)
WIY2
PEY (Y16389)
16SrIX-C
LET1 (MT845875)
LET2
16SrIX-J
WX-95 (JQ044396)
JR1 (AKIK00000000)
MA1 (AKIM00000000)
VAC (AKIN00000000)
MW1 (AKIL00000000)
16SrIII
CuPh (KY412986)
PnWB (AMWZ00000000)
CrPh (EF193355)
WBDL (EF186828)
16SrII
PD1 (AJ542543)
ESFY-G2 (AJ542545)
AT (NC_011047)
16SrX
OY-M (NC_005303)
AYWB (NC_007716)
16SrI
Stol-231/09 (NC_022588)
CBWB (NC_010544)
16SrXII
A.laidlawi (M23932)
0.01

Fig.2

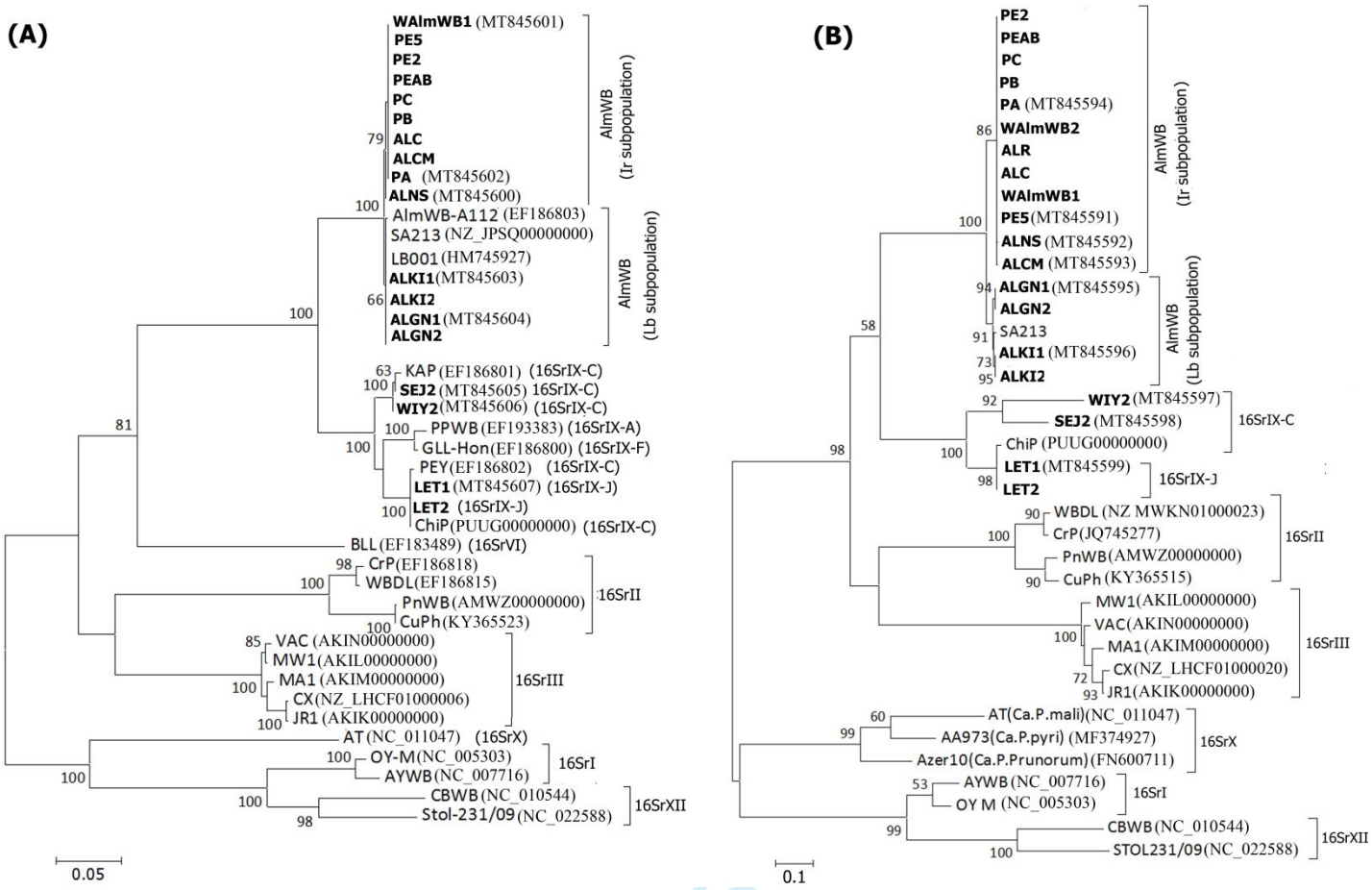
(A)
WAlmWB1 (MT845601)
PE5
PE2
PEAB
PC
PB
ALC
ALCM
PA (MT845602)
ALNS (MT845600)
AlmWB (Ir subpopulation)
AlmWB-A112 (EF186803)
SA213 (NZ_JPSQ00000000)
LB001 (HM745927)
ALKI1 (MT845603)
ALKI2
ALGN1 (MT845604)
ALGN2
AlmWB (Lb subpopulation)
KAP (EF186801) (16SrIX-C)
SEJ2 (MT845605) 16SrIX-C)
WIY2 (MT845606) (16SrIX-C)
PPWB (EF193383) (16SrIX-A)
GLL-Hon (EF186800) (16SrIX-F)
PEY (EF186802) (16SrIX-C)
LET1 (MT845607) (16SrIX-J)
LET2 (16SrIX-J)
ChiP (PUUG00000000) (16SrIX-C)
BLL (EF183489) (16SrVI)
CrP (EF186818)
WBDL (EF186815)
PnWB (AMWZ00000000)
CuPh (KY365523)
16SrII
VAC (AKIN00000000)
MW1 (AKIL00000000)
MA1 (AKIM00000000)
CX (NZ_LHCF01000006)
JR1 (AKIK00000000)
16SrIII
AT (NC_011047) (16SrX)
OY-M (NC_005303)
AYWB (NC_007716)
16SrI
CBWB (NC_010544)
Stol-231/09 (NC_022588)
16SrXII
0.05
(B)
PE2
PEAB
PC
PB
PA (MT845594)
WAlmWB2
ALR
ALC
WAlmWB1
PE5 (MT845591)
ALNS (MT845592)
ALCM (MT845593)
AlmWB (Ir subpopulation)
ALGN1 (MT845595)
ALGN2
SA213
ALKI1 (MT845596)
ALKI2
AlmWB (Lb subpopulation)
WIY2 (MT845597)
SEJ2 (MT845598)
16SrIX-C
ChiP (PUUG00000000)
LET1 (MT845599)
LET2
16SrIX-J
WBDL (NZ MWKN01000023)
CrP (JQ745277)
PnWB (AMWZ00000000)
CuPh (KY365515)
16SrII
MW1 (AKIL00000000)
VAC (AKIN00000000)
MA1 (AKIM00000000)
CX (NZ_LHCF01000020)
JR1 (AKIK00000000)
16SrIII
AT (Ca.P.mali) (NC_011047)
AA973 (Ca.P.pyri) (MF374927)
Azer10 (Ca.P.Prunorum) (FN600711)
16SrX
AYWB (NC_007716)
OY M (NC_005303)
16SrI
CBWB (NC_010544)
STOL231/09 (NC_022588)
16SrXII
0.1

Fig. 3