Detection and identification of the heterogeneous novel subgroup 16SrXIII-(A/I)I phytoplasma associated with strawberry green petal disease and Mexican periwinkle virescence

Edel Pérez-López1 and Tim J. Dumonceaux2,3

1Instituto de Biotecnología y Ecología Aplicada (INBIOTECA), Universidad Veracruzana, Avenida de Las Culturas Veracruzanas, Xalapa, Veracruz, Mexico
2Agriculture and Agri-Food Canada, Saskatoon Research Centre, Saskatoon, Saskatchewan, Canada
3Department of Veterinary Microbiology, University of Saskatchewan, Saskatoon, Saskatchewan, Canada

Phytoplasmas (species of the genus ‘Candidatus Phytoplasma’) are insect-vectored phytopathogenic bacteria associated with economically and ecologically important crop diseases. Strawberry production represents an important part of agricultural activity in Mexico and elsewhere, and infection of plants with phytoplasma renders the fruit inedible by altering plant development, resulting in virescence and phyllody. In this study we examined samples taken from four strawberry plants showing symptoms associated with strawberry green petal disease and from two periwinkle plants showing virescence, sampled in different areas of Mexico. Analysis of the 16S rRNA-encoding sequences showed that the plants were infected with a phytoplasma previously identified as Mexican periwinkle virescence (MPV; 16SrXIII). Examination of bacterial sequences from these samples revealed that two distinct 16S rRNA gene sequences were present in each sample along with a single chaperonin-60 (cpn60) sequence and a single rpoB sequence, suggesting that this strain displays 16S RNA gene sequence heterogeneity. Two distinct rrn operons, identified with subgroup 16SrXIII-A and the newly described subgroup 16SrXIII-I, were identified from the six samples analyzed, delineating the novel subgroup 16SrXIII-(A/I)I, following the nomenclature proposed for heterogeneous subgroups.

Phytoplasmas are members of the Class Mollicutes, wall-less bacteria transmitted by phytophagous insects, mainly leafhoppers, and are associated with a diverse array of economically important crops around the world (Lee et al., 2000; Pérez-López et al., 2016a). Phytoplasmas are host-dependent microorganisms and are classified in the Genus ‘Candidatus Phytoplasma’, according to established taxonomic conventions for unculturable microorganisms (Murray & Stackebrandt, 1995; IRPCM, 2004). Phytoplasmas are further classified into groups and subgroups based on restriction

Abbreviations: MGP, Mexican green petal phytoplasma; MPV, Mexican periwinkle virescence; SbGP, strawberry green petal phytoplasma.

The GenBank/EMBL/DBJ accession numbers for 16S sequences of samples SbGP–S26, SbGP S27, SbGP–S31, SbGP–S289, MPV–S83, and MPV–S86 are KU896184 (16SrXIII-A) and KU896183 (16SrXIII-I), KU896185 (16SrXIII-A) and KU896186 (16SrXIII-I), KT444665 (16SrXIII-A) and KT444664 (16SrXIII-I), KU896182 (16SrXIII-A) and KU896181 (16SrXIII-I), KT444662 (16SrXIII-A) and KT444663 (16SrXIII-I) and KU896188 (16SrXIII-A) and KU896187 (16SrXIII-I), respectively. Those for the cpn60 UT sequences of SbGP–S26 are KU896201–KU896202, those for SbGP–S27 are KU896203–KU896205, that for SbGP–S31 is KT444666, those for SbGP–S289 are KT444667 and KU896209, those for MPV–S83 are KT444668 and KU896206, those for MPV–S86 are KT444669 and KU896207–KU896208. Those for the rpoB partial gene sequence of SbGP–S26 are KU896192–KU896195, that for SbGP–S31 is KU896196, that for SbGP–S289 are KU896197, those for MPV–S83 are KU896198–KU896200, those for MPV–S86 is KU896191.

The novel subgroup 16SrXIII-(A/I)I is represented by the strain Mexican green petal phytoplasma (MGP), KU896182 (16SrXIII-A) and KU896181 (16SrXIII-I).

One supplementary figure is available with the online Supplementary Material.
fragment length polymorphism (RFLP) analysis of their 16S rRNA-encoding loci. A phytoplasma classification scheme has been established using RFLP analysis of the PCR-amplified F2nR2 fragment of 16S rRNA-encoding genes with a set of 17 restriction endonucleases (Lee et al., 1993, 1998, 2000, 2004). Recently, the number of groups identified has been expanded up to 16SrXXXIV (Pérez-López et al., 2016a), indicative of the vast diversity among phytoplasma strains worldwide.

However, targets other than the F2nR2 fragment of the 16S rRNA-encoding gene have been used as additional markers. Some of the additional markers used to differentiate phytoplasmas are the 23S rRNA region, rp (ribosomal protein) operon, tuf, secY, secA and rpOB genes (Marcone et al., 2000; Botti & Bertacini, 2003; Lee et al., 2006; Hodggetts et al., 2008; Valiunas et al., 2013). Another gene used as an additional marker to improve phytoplasma classification is groEL or cpn60 (Mitrovic et al., 2011, 2015). The cpn60 universal target (cpn60 UT), a sequence of approximately 550 bp, is located within the Cpn60-encoding gene (Goh et al., 1996). This sequence has been identified as a molecular barcode for the domain Bacteria and has been used as a taxonomic marker to characterize microbial communities (Dumonceaux et al., 2006; Links et al., 2013, 2014; McKenny et al., 2014). Recently, tools to access the cpn60 UT have been described for phytoplasmas, and the cpn60 UT has been shown to be a marker capable of differentiating a broad range of phytoplasmas while functioning as a suitable diagnostic target (Dumonceaux et al., 2014; Sugawara et al., 2012).

One of the economically important crops that is seriously affected by phytoplasma is the strawberry (Fragaria ananassa), in which strawberry lethal yellows disease (SYL), strawberry red leaf disease (StrawRL), and strawberry green petal disease (SbGP) (Gundersen et al., 1996; Jomantiene et al., 1998; Padovan et al., 2000; Fernández et al., 2013) have been described. Strawberry green petal disease has been reported in Canada, the Czech Republic, Australia and Italy, and is typically associated with phytoplasmas of the Aster Yellows (‘Ca. Phytoplasma’ asteris, 16SrI) group (Gundersen et al., 1996; Honetšlegrová et al., 1996; Contaldo et al., 2012). However, strawberry plants are also known to be a natural host of phytoplasmas belonging to groups 16SrI, 16SrIII, 16SrVI, 16SrVII, 16SrXII and 16SrXIII (Harrison et al., 1997; Jomantiene et al., 1998, 1999, 2001; Padovan et al., 2000; Fernández et al., 2013; Valiunas et al., 2006). In Latin America the unspeciated group 16SrXIII is commonly represented by subgroups 16SrXIII-A, 16SrXIII-B, and 16SrXIII-F found to be associated with strawberry plants (Jomantiene et al., 1998; Fernández et al., 2015).

Mexico is one of the largest strawberry producers in the world (Bouffard, 2012), and it is therefore extremely important for Mexican agricultural producers to undertake disease surveillance and identify the causal agent of any disease detected in strawberry plants in commercial orchards. In Mexico, Michoacan state, with 121 942 tons of strawberries produced for export and additional 153 398 tons for its national market, is without doubt the nation’s biggest strawberry producer. During the 2014–2015 fruit production season (November–February), four strawberry plants showing the symptoms described above for SbGP disease were detected in Michoacan state. As part of ongoing surveillance efforts to support effective disease management, we undertook the identification of the causative pathogen. We also analyzed periwinkle plants showing virescence symptoms collected in San Luis Potosi, over 400 km away from Michoacan, as a possible alternative host for 16SrXIII phytoplasma. In pursuit of this objective, we obtained the 16S rRNA-encoding gene F2nR2 sequences, along with the cpn60 UT and rpOB sequences, facilitating better characterization.

### Symptomology of affected plants

All six strawberry plants showed green flowers and fruit with green structures, but only two of the five plants showed leaves with red margins (Fig. S1a–d, available in the online Supplementary Material). The samples were identified as SbGP-S31, SbGP-S26, SbGP-S27 and SbGP-S289. Both of the periwinkle plant samples (MPV-S83 and MPV-S86) showed symptoms of virescence and phyllody, consistent with infection by phytoplasma (Fig. S1e and f).

### Determination of the 16S rRNA-encoding gene sequences

DNA was extracted from 0.3 g of plant tissue using a cetyltrimethylammonium bromide (CTAB)-based DNA extraction method as previously described (Pérez-López et al., 2016b). All DNA extracts were diluted 1:10 with 10 mM Tris-Cl, pH 8.5, and used as templates in PCR with primers R16F2n/R16R2 (Gundersen & Lee, 1996) to amplify the 16S rRNA gene. In all PCR assays we included as negative controls one reaction without DNA, and as positive controls DNA extracted from periwinkle plants affected by an aster yellows phytoplasma strain previously described (Olivier et al., 2006). Positive reactions were confirmed through the visualization in a 1% agarose gel of a 1.25 kb fragment corresponding to the 16S rRNA-based F2nR2 sequence (which is defined as the sequence between the annealing sites of primers R16F2n and R16R2). PCR products generated from all samples were cloned into the vector pGEM-T Easy (Promega) according to the manufacturer’s recommendations, then plasmids were transformed into chemically competent Escherichia coli TOP10 (Life Technologies), and clones were sequenced using plasmid-targeted primers T7/SP6.

The F2nR2 sequences were analyzed using the iPPhylclassifier tool (Zhao et al., 2009) to determine the group and subgroup of the strains detected. The in silico RFLP pattern obtained was thereby automatically compared with the RFLP pattern from previously identified strains.
The F2nR2 sequences obtained from strawberry and periwinkle plants showed 99% nucleotide identity with the reference strain for the 16SrXIII group, the Mexican periwinkle virescence (MPV, GenBank accession no. AF248960), through BLAST. This result was supported by in silico RLFP analysis using iPhyClassifier, which identified the strains as members of group 16SrXIII. Two distinct F2nR2 sequences (1244 and 1245 bp) were retrieved from each of the six samples. One of the sequences from each sample was identified as 16SrXIII-A by RFLP digestion pattern (Fig. 1, Table 1), and the coefficients of similarity for these sequences were 0.98 to 1.00 with the reference strain MPV (16SrXIII-A, GenBank accession number AF248960). The other F2nR2 sequences from each sample were most closely related to 16SrXIII-D showing a coefficient of similarity of 0.95 with the reference strain SINPV (16SrXIII-D, GenBank accession number FJ914647). With a unique RFLP digestion pattern (Fig. 1), and a coefficient of similarity value less than 0.97 but higher than 0.85 (not shown), this sequence belonged to a novel 16Sr subgroup, which is identified here as 16SrXIII-I and is represented by the MGP phytoplasma strain.

Table 1. F2nR2 sequences obtained for each sample in this study

<table>
<thead>
<tr>
<th>Samples</th>
<th>F2nR2 sequence GenBank accession number</th>
<th>Nucleotide differences</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>rrrA (16SrXIII-A) (1244 bp)</td>
<td>rrrB (16SrXIII-I) (1245 bp)</td>
</tr>
<tr>
<td>SbGP–S26</td>
<td>KU896184</td>
<td>KU896183</td>
</tr>
<tr>
<td>SbGP–S27</td>
<td>KU896185</td>
<td>KU896186</td>
</tr>
<tr>
<td>SbGP–S31</td>
<td>KT444665</td>
<td>KT444664</td>
</tr>
<tr>
<td>SbGP–S289</td>
<td>KU896182</td>
<td>KU896181</td>
</tr>
<tr>
<td>MPV–S83</td>
<td>KT444662</td>
<td>KT444663</td>
</tr>
<tr>
<td>MPV–S86</td>
<td>KU896188</td>
<td>KU896187</td>
</tr>
</tbody>
</table>
Phylogenetic analysis was conducted for the F2nR2 sequences using the maximum-parsimony method with MEGA v6.0 (Tamura et al., 2013), and bootstrapping 1000 times to estimate stability. The results obtained using iPhyClassifier were confirmed by phylogenetic analysis. The sequences classified as 16SrXIII-A grouped with the reference strain MPV, while the sequences classified as closely related to the 16SrXIII-D subgroup grouped with the reference strain SINPV (Fig. 2). This sequence was classified as the novel subgroup 16SrXIII-I.

The detection of two distinct 16SrXIII subgroups within a single sample taken from both strawberry and periwinkle plants was indicative of the presence of a heterogeneous 16S rRNA operon, which is typically present in two identical copies in the phytoplasma genome (Zhao et al., 2014). Alternatively, this result could be interpreted as evidence of a mixed infection in all six geographically distinct plants (Wei et al., 2016) with two strains possessing related but distinct F2nR2 sequences. Evidence for 16S rRNA-encoding gene heterogeneity has been obtained from other phytoplasma-infected samples, based on the detection of different RFLP patterns obtained from F2nR2 sequences amplified from the same sample (Liefing et al., 1996; Jomantiene et al., 2002; Davis et al., 2003; Mello et al., 2006). In other cases, 16S operon heterogeneity has been shown using complementary analysis of single-copy protein-encoding genes from the same samples (Marcone et al., 2000; Davis et al., 2013). To clarify whether the two F2nR2 sequences from each sample were the result of heterogeneity or a mixed infection, we amplified and characterized the cpn60 UT and partial rpoB sequences of the phytoplasma detected in strawberry and periwinkle plants, as both of these genes are present in a single copy in the phytoplasma genome.

**Fig. 2.** Phylogenetic tree (maximum parsimony) of the 16S rRNA-gene-based F2nR2 gene sequences from the six samples analyzed in this study (including both sequences detected in each plant) and reference sequences retrieved from Genbank. Accession numbers are indicated in parentheses, and the strain acronym is given if applicable. R, reference species; RS, reference strain. Acholeplasma laidlawii PG8 was used as the outgroup. The phylogenetic tree was bootstrapped 1000 times to achieve reliability. Bar, two substitutions per 100 positions.
Characterization of the cpn60 UT and rpoB sequences

To obtain the cpn60 UT and partial rpoB sequences, we used the DNA extracts from infected plants as template in a PCR with primers H279p/H280p : D0317/D0318 (1 : 7 ratio) and the conditions described by Dumonceaux et al. (2014), while the rpoB gene was amplified using primers rpoBF1/rpoBR3 following the methods described by Valiunas et al. (2013).

We amplified the 1373 bp rpoB sequences and the 605 bp cpn60 UT sequences from all of the samples. Amplicons were cloned as described above and clones were sequenced with primers T7/SP6. Sequences were assembled using the Staden package (Bonfield & Whitwham, 2010), and compared with reference sequences from GenBank using BLAST.

A total of 56 cpn60 UT clones were sequenced from all plants, and sequences from 21 clones were determined for rpoB (Table 2). A global sequence alignment (CLUSTALW) was used for pairwise sequence similarity calculation. The alignment included the cpn60 UT sequences from all samples, and in a second alignment we included all the rpoB sequences obtained from the samples. Pairwise sequence similarity scores showed that cpn60 UT sequences and rpoB gene shared no less than 99.0% sequence identity, ranging from 99 to 100%.

BLAST analysis showed that cpn60 UT sequences had the highest nucleotide identity (84%) with members of the 16SrXII group such as ‘Ca. Phytoplasma australiense’ (16SrXII-B, GenBank accession number NC010544). Similarly, rpoB sequences shared 82% sequence identity with ‘Ca. Phytoplasma fragariae’ (16SrXII-E, GenBank accession number KC663420).

Phylogenetic trees were reconstructed with the cpn60 UT and rpoB sequences from each plant, using the procedure described above. The results obtained by BLAST were confirmed through the phylogenetic analysis. In both trees, the sequences generated in this study grouped closely with those of 16SrXIII-related strains (Fig. 3). This is the first time, to our knowledge, that protein-encoding genes have been amplified from a phytoplasma belonging to the 16SrXIII group.

Evidence for heterogeneity rather than mixed infection

Sequence analysis of individual clones corresponding to the 16S-based F2nR2 fragment clearly indicated the presence of two distinct sequence types that differed at 8–12 nucleotide positions in each infected plant (Table 1). This was confirmed both by in silico RFLP analysis (Fig. 1) and by phylogenetic analysis (Fig. 2). In contrast, individual clone sequences generated from each plant through the amplification of the single-copy protein-encoding genes cpn60 and rpoB were much more similar to one another, with nucleotide differences observed in 0–3 locations in these amplified sequences.

### Table 2. cpn60 UT and rpoB clone sequences obtained for each sample in this study

<table>
<thead>
<tr>
<th>Sample</th>
<th>Number of clones examined/number of distinct sequences observed per plant</th>
<th>Maximum nucleotide differences within a single plant</th>
<th>GenBank accession no. (observed frequency)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>cpn60 UT</td>
<td>rpoB</td>
<td>cpn60 UT (552 bp)</td>
</tr>
<tr>
<td>SbGP–S26</td>
<td>8/2</td>
<td>4/4</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SbGP–S27</td>
<td>11/3</td>
<td>2/1</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SbGP–S31</td>
<td>10/1</td>
<td>3/1</td>
<td>0</td>
</tr>
<tr>
<td>SbGP–S289</td>
<td>7/2</td>
<td>4/3</td>
<td>1</td>
</tr>
<tr>
<td>MPV–S83</td>
<td>7/2</td>
<td>4/3*</td>
<td>1</td>
</tr>
<tr>
<td>MPV–S86</td>
<td>13/3</td>
<td>4/1</td>
<td>2</td>
</tr>
<tr>
<td>Totals</td>
<td>56/8</td>
<td>21/8</td>
<td>–</td>
</tr>
</tbody>
</table>

*One of the rpoB clones from MPV–S83 contained an in-frame stop codon and was excluded from further analysis.
New heterogeneous 16SrXIII subgroup in Mexico

**Fig. 3.** Phylogenetic tree (maximum parsimony) of (a) *cpn60* UT sequences and (b) partial *rpoB* sequences from the six samples analyzed in this study and reference sequences obtained from Genbank. Accession numbers are indicated in parentheses, and the strain acronym is given if applicable. Clone frequencies are indicated along with the total number of clones analyzed in each host plant. *Acholeplasma laidlawii* PG8 was used as an outgroup. The phylogenetic tree was bootstrapped 1000 times to achieve reliability. Bars: (a) two substitutions per 100 positions; (b) five substitutions per 100 positions.
Fig. 4. Distinctive RFLP patterns obtained with pDRAW32 from in silico digestion of (a) cpn60 UT sequences and (b) partial rpoB sequences from the strains used in the phylogenetic analysis and the novel 16SRIII subgroup delineated in this study (indicated by §). Group and subgroups of the strains used are indicated in parentheses. In the computer-simulated digestions, we used (a) TaqI, HhaI and TruI (MseI), and (b) AluI, BfaI, HhaI, HinII, RsaI and TaqI. Lanes labelled MW represent HaeIII-digested phage φX174 DNA.
New heterogeneous 16SrXIII subgroup in Mexico

fragments (Table 2). At least part of this apparent microheterogeneity in the protein-encoding gene sequences may have been due to Taq error, as one of the rpoB clones had an in-frame stop codon but was otherwise identical at the amino acid sequence level to the other clones (data not shown). Of the sequences analyzed, only rpoB could reliably be used to discern the sequences amplified from strawberry and periwinkle (Fig. 3b), on the basis of two nucleotide differences: position 575 (T in periwinkle; C in strawberry), and position 749 (G in periwinkle; A in strawberry). This observation indicates that the two host plants, which were geographically separated by over 400 km, were not infected with the same strain(s) of MPV phytoplasma and a double infection with two closely related phytoplasma strains in both host plants is unlikely.

RFLP analysis has been used to analyze cpn60 sequences from phytoplasma strains in the 16SrI group using restriction endonucleases TaqI, HhaI and Trul (Msal) (Mitrović et al., 2011); similarly, RFLP analysis of rpoB sequences with Alul, BflI, HhaI, HindIII, Rsal and TaqI has been used (Valiunas et al., 2013). Using pDRAW32 (AcaClone Software, http://www.acaclone.com), the 56 cpn60 UT sequences generated in this study were subjected to in silico analysis with the endonucleases suggested by Mitrović et al. (2011), and the 21 rpoB sequences were subjected to the same procedure with the endonucleases suggested by Valiunas et al. (2013). A single RFLP pattern was determined from all of the cpn60 UT sequences, as well as with the rpoB sequences (Fig. 4). These results indicate that the different nucleotides detected among the sequences were not related with more than one cpn60 UT or rpoB subgroup. Moreover, protein-encoding genes such as cpn60 and rpoB tend have higher sequence diversity than 16S rRNA-encoding genes (Mitrović et al., 2011; Verbeke et al., 2011; Dumonceaux et al., 2014; Links et al., 2014; Valiunas et al., 2013), so the presence of lower sequence diversity in these two genes compared with F2nR2 indicates that a single strain of phytoplasma is present in all of the plants analyzed. Taken together, these observations indicate that the two F2nR2 sequences observed in each plant reflect 16S rRNA-encoding gene heterogeneity rather than co-infection with two closely related strains of MGP phytoplasmas.

Novel phytoplasma subgroup 16SrXIII-(A/I)

Based on the presence of two F2nR2 sequences from each infected plant, which corresponded to distinct 16SrXIII subgroups, it was difficult to classify the MPV and SbGP phytoplasmas using the typical 16Sr classification scheme. Therefore, following the three-letter nomenclature recommended by Wei et al. (2008), we suggest that the MGP phytoplasmas may be classified as members of a novel subgroup, 16SrXIII-(A/I) I. Heterogeneous rrr operons in a single strain (referred to as rrrA and rrrB) have been previously observed in other phytoplasmas (Lee et al., 1993, 1998; Liefting et al., 1996; Marcone et al., 2000; Jomantiene et al., 2002; Davis et al., 1998, 2003; Torres et al., 2004), and this fact is an acknowledged weakness of the use of the 16S rRNA-encoding gene to identify and differentiate phytoplasmas (Wei et al., 2008; Zhao et al., 2009; Valiunas et al., 2013). Similarly, plants of the genera Catharanthus and Fragaria have been previously observed to be infected by phytoplasmas possessing heterogeneous rrr operons (Torres et al., 2004; Davis et al., 2003), but this study is the first report, to our knowledge, of 16SrXIII phytoplasmas showing rrr operon heterogeneity.

The Mexican periwinkle phytoplasma group has been found mainly in countries within Latin America, from Mexico to Argentina (Lee et al., 1998; Fernández et al., 2015). Subgroups 16SrXIII-B and 16SrXIII-F have been detected affecting strawberry plants in Florida and Argentina, respectively (Jomantiene et al., 1998; Fernández et al., 2015); subgroup 16SrXIII-A was detected in periwinkle plants in Mexico (Lee et al., 1998); subgroup 16SrXIII-C was found in chinarberry plants in Bolivia (Harrison et al., 2003); and subgroup 16SrXIII-D was detected in samples collected from potato plants in Mexico (Santos-Cervantes et al., 2010). Recently the diversity within this group has been expanded with the identification of novel subgroups 16SrXIII-G and 16SrXIII-H among strains reported in Latin America (Pérez-López et al., 2016a). This study provides additional evidence of the high phytoplasma diversity found in the 16SrXIII group in Latin America.

Acknowledgements

This work was supported by the Genomic Research and Development Initiative for the shared priority project on quarantine and invasive species. We thank Douglas Rodríguez for supplying the samples.

References


