Phylogenetic analysis and delineation of phytoplasmas based on secY gene sequences

I.-M. Lee,1 K. D. Bottner-Parker,1 Y. Zhao,1 R. E. Davis1 and N. A. Harrison2

1Molecular Plant Pathology Laboratory, USDA, ARS, Beltsville, MD 20705, USA
2FLREC, University of Florida, Fort Lauderdale, FL 33314, USA

The secY gene sequence is more variable than that of the 16S rRNA gene. Comparative phylogenetic analyses with 16S rRNA and secY gene sequences from 80 and 83 phytoplasma strains, respectively, were performed to assess the efficacy of these sequences for delineating phytoplasma strains within each 16Sr group. The phylogenetic interrelatedness among phytoplasma taxa inferred by secY gene-based phylogeny was nearly congruent with that inferred by 16S rRNA gene-based phylogeny. Phylogenetic analysis based on the secY gene permitted finer differentiation of phytoplasma strains, however. The secY gene-based phylogeny not only readily resolved 16Sr subgroups within a given 16Sr group, but also delineated distinct lineages irresolvable by 16S rRNA gene-based phylogeny. Such high resolving power makes the secY gene a more useful genetic marker than the 16S rRNA gene for finer differentiation of closely related phytoplasma strains based on RFLP analysis with selected restriction enzymes. Such strains were readily identified by collective secY RFLP patterns. The genetic interrelationships among these strains were determined by pattern similarity coefficients, which coincided with delineations by phylogenetic analysis. This study also revealed two heterogeneous spc operons present in the phytoplasma clade. This latter finding may have significant implications for phytoplasma evolution.

INTRODUCTION

Phytoplasmas, formerly termed mycoplasma-like organisms (MLOs), are minute, cell-wall-less prokaryotes that are associated with diseases in several hundred plant species (Lee et al., 2000; Bertaccini, 2007; Hogenhout et al., 2008). Extensive phylogenetic studies have placed phytoplasmas in the class Mollicutes. Because it has proved impossible to obtain pure cultures of any phytoplasma and the paucity of accessible phenotypic criteria, it is difficult to classify phytoplasmas according to the criteria applied to taxonomy of the Mollicutes (Weisburg et al., 1989; Murray & Schleifer, 1994; Vandamme et al., 1996; Razin et al., 1998), and phytoplasma taxonomy is inevitably based heavily on molecular characteristics and phylogeny. The 16Sr RNA gene sequence has been used as the primary parameter for differentiation and classification of phytoplasmas (Lee et al., 1993b, 1998a; Schneider et al., 1995; Seemüller et al., 1994, 1998). So far, a total of 19 distinct groups based on actual RFLP analysis of PCR-amplified 16Sr RNA gene sequences, termed 16Sr RNA groups (16Sr groups), or 30 groups based on in silico RFLP analysis have been identified (Lee et al., 1998a, 2000; Wei et al., 2007). Eighty-nine 16Sr subgroups were further classified by in silico analysis. The latter RFLP-based grouping is highly consistent with phylogenetic relationships based on analysis of 16Sr rRNA gene sequences. It was proposed that each 16Sr rRNA gene RFLP group represents at least one phytoplasma species (Gundersen et al., 1994). At present, species designation is based primarily on dissimilarity of 16Sr rRNA gene sequences among phytoplasmas. An arbitrary threshold of 2.5% dissimilarity was applied as a guideline for erecting a novel species (IRPCM Phytoplasma/ Spiroplasma Working Team – Phytoplasma Taxonomy Group, 2004). Because of the highly conserved nature of the 16Sr rRNA gene, this guideline may exclude many ecologically or biologically distinct phytoplasma strains, some of which may warrant designation as novel taxa.

Classification of distinct strains below the species level has been based primarily on RFLP analysis of 16Sr rRNA gene sequences. Over the last decade, epidemiological studies have revealed that diverse phytoplasma strains, which are very closely related based on analysis of 16Sr rRNA gene sequences, are involved in similar diseases associated with different cultivars of a given plant species grown in the same or different geographical regions (Martini et al., 2002; Lee et al., 2003, 2004c, 2006b). It is not uncommon that closely related phytoplasma strains have unique ecological niches encompassing both plant host range and insect vectors (Lee et al., 1998b, 2000).
Table 1. Phytoplasma strains used in this study

'Ca. P.', 'Candidatus Phytoplasma'.

<table>
<thead>
<tr>
<th>Strain</th>
<th>Associated disease(s)</th>
<th>Geographical origin</th>
<th>16S rRNA gene RFLP classification</th>
<th>GenBank accession number(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BB</td>
<td>Tomato big bud</td>
<td>Arkansas, USA</td>
<td>16SrI-A</td>
<td>AY180955, AY803178</td>
</tr>
<tr>
<td>NJAY</td>
<td>New Jersey aster yellows</td>
<td>New Jersey, USA</td>
<td>16SrI-A</td>
<td>–</td>
</tr>
<tr>
<td>GD1</td>
<td>Grey dogwood stunt</td>
<td>New York, USA</td>
<td>16SrI-A</td>
<td>DQ112021, AY803171</td>
</tr>
<tr>
<td>CHRYM</td>
<td>Chrysanthemum yellows</td>
<td>Germany</td>
<td>16SrI-A</td>
<td>AY265214, AY803170</td>
</tr>
<tr>
<td>HYDP</td>
<td>Hydrangea phyllody</td>
<td>Belgium</td>
<td>16SrI-A</td>
<td>AY265215, AY803181</td>
</tr>
<tr>
<td>PVM</td>
<td>Plantago virescence</td>
<td>Germany</td>
<td>16SrI-A</td>
<td>AY265216, AY803185</td>
</tr>
<tr>
<td>'Ca. P. asteris' AYWB</td>
<td>Aster yellow witches'-broom</td>
<td>Ohio, USA</td>
<td>16SrI-A</td>
<td>NC_007716, NC_007716</td>
</tr>
<tr>
<td>HY</td>
<td>Maryland aster yellows</td>
<td>Maryland, USA</td>
<td>16SrI-B</td>
<td>AF322644, AY803177</td>
</tr>
<tr>
<td>HyPH</td>
<td>Chrysanthemum yellow</td>
<td>Italy</td>
<td>16SrI-B</td>
<td>EF634457, DQ787851</td>
</tr>
<tr>
<td>PRVC</td>
<td>Primrose virescence</td>
<td>Germany</td>
<td>16SrI-B</td>
<td>AY265210, AY803173</td>
</tr>
<tr>
<td>OAY</td>
<td>Oenothera virescence</td>
<td>Michigan, USA</td>
<td>16SrI-B</td>
<td>M30790, –</td>
</tr>
<tr>
<td>AV2192</td>
<td>Aster yellows</td>
<td>Germany</td>
<td>16SrI-B</td>
<td>AY180957, AY803167</td>
</tr>
<tr>
<td>MBS</td>
<td>Maize bushy stunt</td>
<td>Mexico</td>
<td>16SrI-B</td>
<td>AY265208, AY803175; GU004341*</td>
</tr>
<tr>
<td>CPh-a</td>
<td>Clover phyllody</td>
<td>Canada</td>
<td>16SrI-C</td>
<td>AF222065, AY803179; GU004339*</td>
</tr>
<tr>
<td>KVG</td>
<td>Clover phyllody</td>
<td>Germany</td>
<td>16SrI-C</td>
<td>AY265218, AY803183</td>
</tr>
<tr>
<td>KVE</td>
<td>Clover phyllody</td>
<td>France</td>
<td>16SrI-C</td>
<td>AY265217, AY803174</td>
</tr>
<tr>
<td>PaWB</td>
<td>Paulownia witches'-broom</td>
<td>Taiwan</td>
<td>16SrI-D</td>
<td>AY265206, AY803184; GU004337*</td>
</tr>
<tr>
<td>BBS3</td>
<td>Blueberry stunt</td>
<td>Michigan, USA</td>
<td>16SrI-E</td>
<td>AY265213, AY803169; GU004357*</td>
</tr>
<tr>
<td>ACLR-AY</td>
<td>Apricot chlorotic leaf roll</td>
<td>Spain</td>
<td>16SrI-F</td>
<td>AY265221, AY803166</td>
</tr>
<tr>
<td>CVB</td>
<td>Leafhopper borne</td>
<td>Germany</td>
<td>16SrI-F</td>
<td>AY265212, AY803172</td>
</tr>
<tr>
<td>STRAWB2</td>
<td>Strawberry multiplier</td>
<td>Florida, USA</td>
<td>16SrI-K</td>
<td>U96616, AY803180</td>
</tr>
<tr>
<td>AVUT</td>
<td>Aster yellows</td>
<td>Germany</td>
<td>16SrI-M</td>
<td>AY265209, AY803168</td>
</tr>
<tr>
<td>IOWB</td>
<td>Ipomoea obscura witches'-broom</td>
<td>Taiwan</td>
<td>16SrI-N</td>
<td>AY265205, AY803182</td>
</tr>
<tr>
<td>PnWB</td>
<td>Peanut witches'-broom</td>
<td>Taiwan</td>
<td>16SrI-A</td>
<td>L33765, GU004331*</td>
</tr>
<tr>
<td>SPWB</td>
<td>Sweet potato witches'-broom</td>
<td>Taiwan</td>
<td>16SrI-A</td>
<td>L33770, GU004321*</td>
</tr>
<tr>
<td>SEPN</td>
<td>Sesame phyllody</td>
<td>Thailand</td>
<td>16SrI-A</td>
<td>EF193357, GU004362*</td>
</tr>
<tr>
<td>SEPT</td>
<td>Sesame phyllody</td>
<td>Thailand</td>
<td>16SrI-A</td>
<td>GU004373, GU004322*</td>
</tr>
<tr>
<td>CaWBYN16</td>
<td>Cactus witches'-broom YN16</td>
<td>China</td>
<td>16SrI-C</td>
<td>EU099561, GU004323*</td>
</tr>
<tr>
<td>SOYP</td>
<td>Soybean phyllody</td>
<td>Thailand</td>
<td>16SrI-C</td>
<td>EF193353, GU004324*</td>
</tr>
<tr>
<td>CrP</td>
<td>Crotalaria phyllody</td>
<td>Unknown</td>
<td>16SrI-C</td>
<td>EF193355, GU004349*</td>
</tr>
<tr>
<td>CoP</td>
<td>Cotton phyllody</td>
<td>Unknown</td>
<td>16SrI-C</td>
<td>EF186827, GU004350*</td>
</tr>
<tr>
<td>PEP</td>
<td>Pectis echioides phyllody</td>
<td>Italy</td>
<td>16SrI-E</td>
<td>Y16393, GU004348*</td>
</tr>
<tr>
<td>TBB</td>
<td>Australian tomato big bud</td>
<td>Australia</td>
<td>16SrI-D</td>
<td>Y08173, GU004347*</td>
</tr>
<tr>
<td>CX</td>
<td>Peach X-disease</td>
<td>Canada</td>
<td>16SrI-A</td>
<td>L33733, GU004327*</td>
</tr>
<tr>
<td>WX</td>
<td>X-disease</td>
<td>Canada</td>
<td>16SrI-A</td>
<td>L04682, GU004354*</td>
</tr>
<tr>
<td>CYE</td>
<td>Clover yellow edge</td>
<td>Lithuania</td>
<td>16SrI-B</td>
<td>AF173558, GU004332*</td>
</tr>
<tr>
<td>Vac</td>
<td>Vaccinium witches'-broom</td>
<td>Germany</td>
<td>16SrI-F</td>
<td>X76430, GU004360*</td>
</tr>
<tr>
<td>PBT</td>
<td>Pecan bunchy top</td>
<td>Georgia</td>
<td>16SrI-C</td>
<td>GU004371, GU004361*</td>
</tr>
<tr>
<td>GR1</td>
<td>Goldenrod yellows</td>
<td>New York, USA</td>
<td>16SrI-D</td>
<td>GU004372, GU004364*</td>
</tr>
<tr>
<td>SP1</td>
<td>Spirea stunt</td>
<td>New York, USA</td>
<td>16SrI-E</td>
<td>AF190228, GU004326*</td>
</tr>
<tr>
<td>MW1</td>
<td>Milkweed yellows</td>
<td>New York, USA</td>
<td>16SrI-F</td>
<td>AF510724, GU004340*</td>
</tr>
<tr>
<td>AKpot7</td>
<td>Potato purple top-AK</td>
<td>Alaska, USA</td>
<td>16SrI-F</td>
<td>GU004370, GU004358*</td>
</tr>
<tr>
<td>WWB</td>
<td>Walnut witches'-broom</td>
<td>Georgia, USA</td>
<td>16SrI-G</td>
<td>AF190227, GU004325*</td>
</tr>
<tr>
<td>JRI1Ph</td>
<td>Poinsettia branch inducing</td>
<td>USA</td>
<td>16SrI-H</td>
<td>AF190223, GU004328*</td>
</tr>
<tr>
<td>MT117</td>
<td>Potato purple top-MT</td>
<td>Montana, USA</td>
<td>16SrI-M</td>
<td>FJ226074, GU004333*</td>
</tr>
<tr>
<td>AKpot6</td>
<td>Potato purple top-AK</td>
<td>Alaska, USA</td>
<td>16SrI-N</td>
<td>GU004365, GU004359*</td>
</tr>
<tr>
<td>LYF (LYF-C5)</td>
<td>Coconut lethal yello</td>
<td>Florida, USA</td>
<td>16SrI-A</td>
<td>AF498308, GU004320*</td>
</tr>
</tbody>
</table>
Often, such strains cannot be readily differentiated by analysis of 16S rRNA gene sequences, hindering the development of targeted and efficient disease control measures and underscoring the need to seek additional markers that permit finer differentiation of closely related strains. In the last decade, several other conserved genes or specific genomic DNA fragments have been studied in the search for molecular markers for finer strain differentiation (Lee et al., 2006a; Marcone et al., 2000; Martini et al., 2007).

Earlier studies on differentiation of phytoplasma strains within the aster yellows group (16SrI) (Lee et al., 2006a) and the elm yellows group (16SrV) (Lee et al., 2004a)
indicated that, because of its high resolving power, the secY gene, encoding a protein translocase subunit, represents one of the most promising markers for finer differentiation of phytoplasma strains and for delineating biologically and/or ecologically distinct strains that often cannot be readily resolved by analysis of the 16S rRNA gene alone. Phylogenetic interrelationships based on secY gene analysis have not been explored for strains constituting the majority of phytoplasma groups. In the present study, comparative phylogenetic analyses were performed and phylogenetic trees constructed based on sequence analyses of 16S rRNA genes from representative phytoplasma strains in order to evaluate the efficacy of secY gene sequences for differentiation of closely related phytoplasma strains.

**METHODS**

**Phytoplasma strains and nucleic acid preparation.** Phytoplasma strains used in this study are listed in Table 1; they are representative of 12 16Sr groups. Many of the phytoplasma strains were available in our inventory (as extracted total nucleic acid preparations). For those that were unavailable, total nucleic acid was extracted using leaf midribs or other tissues from periwinkle or original hosts that were infected by these phytoplasma strains according to the method described by Lee et al. (1993a) or Green et al. (1999). The strains were previously characterized and identified by RFLP or sequence analysis of the 16S rRNA gene (Bertaccini, 2007; Jacobs et al., 2003; Lee et al., 1993b, 1998a, 2000; Marcone et al., 1997a, b, c, 2000; Martini et al., 2007; Schneider et al., 1997; Seemüller et al., 1998; Cai et al., 2008).

**Primer design.** Degenerate primer pair L15F1/MapR1 (Table 2) was initially designed based on conserved regions of the ribosomal protein gene rpl15 (or rplO) and methionine aminopeptidase gene map in the spc ribosomal protein operon, which contains the rpl15 gene, the adenylate kinase gene (adk), a protein translocase gene (secY) and the map gene (Suh et al., 1996). Conserved regions were identified by alignment of sequences of the rpl15 and map genes from four phytoplasma and Acholeplasma strain genome sequences available in GenBank and the LY phytoplasma sequence (N. A. Harrison, unpublished) using CLUSTAL V from the LaserGene software MEGALIGN program (DNASTAR). Additional primers were designed within the L15F1/MapR1 ampiclon to facilitate sequencing of the spc operon (Table 2). Subsequently, more specific primers were designed based on alignment of sequences from partial spc operons amplified from representative 16Sr group phytoplasmas using the primer pair L15F1/MapR1. Specific primer pairs were designed for amplification of DNA fragments (1700–1850 bp) from particular 16Sr phytoplasma groups (Table 2) for use in computer-simulated RFLP analysis (Wei et al., 2008). Primer positions are indicated in Fig. 1.

**PCR amplification, cloning and sequencing of the partial spc operon.** The secY gene was amplified from some of the phytoplasma strains listed in Table 1 belonging to different phytoplasma groups and sequenced to determine the nucleotide sequence of the previously characterized and identified phytoplasma strains.

---

**Table 2.** Primers designed in this study for amplification of the partial spc operon and DNA fragments containing the complete secY gene and partial sequence of adjacent genes for virtual RFLP analysis.

<table>
<thead>
<tr>
<th>Primer(s)</th>
<th>Sequence(s) (3’–3’)</th>
<th>16Sr group(s) amplified</th>
<th>Annealing temperature (°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>L15F1/MapR1</td>
<td>CCTGGTGAGTGGYAMTGGWAAAAC/ATTARRAATATARGGTYTCTCTCRTG</td>
<td>I–VIII, X, XII, XIII, XVIII</td>
<td>50</td>
</tr>
<tr>
<td>L15F1a/MapR1a-a</td>
<td>TGGWAAAACCTKCBGWAAARGG/AAGMTKYACCRATDCCATG</td>
<td>I, XII, XIII</td>
<td>55</td>
</tr>
<tr>
<td>L15F1b/MapR1a-b</td>
<td>GGWAAAACCTSHGGYMRVGGHCATAAAGG/CCWATMCCRTGWCCDGWAAAA</td>
<td>II, III, IV</td>
<td>64</td>
</tr>
<tr>
<td>L15F1a(I)/MapR1a(I)</td>
<td>CTTCTGTTAAAGGACATAAAAGG/GTTTCTCTGCAAAAGATGTCAC</td>
<td>I</td>
<td>52</td>
</tr>
<tr>
<td>L15F1a(II)/MapR1a(II)</td>
<td>CGCTGGTGCGGGCCATAAAAGG/GTTTCTCTGTAAGATTTCAC</td>
<td>II</td>
<td>60</td>
</tr>
<tr>
<td>L15F1a(III)/MapR1a(III)</td>
<td>CTTCTGTTAAAGGACATAAAAGG/GTTTCTCTGGAATGGCAACC</td>
<td>III</td>
<td>55</td>
</tr>
<tr>
<td>L15F1a(VI)/MapR1a(VI)</td>
<td>CCTGGTAGTGGYAMTGGWAAAAC/ATTARRAATATARGGTYTCTCTCRTG</td>
<td>VI</td>
<td>50</td>
</tr>
<tr>
<td>L15F64-a</td>
<td>TGGAARGAMCAAGGACACAT</td>
<td>I, XII, XIII</td>
<td>53</td>
</tr>
<tr>
<td>L15F-696-b</td>
<td>DMAAGAATGGMRWRAAACAAG</td>
<td>II, III, IV</td>
<td>53</td>
</tr>
<tr>
<td>MapR-703-a</td>
<td>TYYWGGTTTTCATCTYKTGTTTG</td>
<td>I, XII, XIII</td>
<td>53</td>
</tr>
<tr>
<td>MapR-696-b</td>
<td>AGMHCRRCCDACMAYAAATTA</td>
<td>II, III, IV</td>
<td>48</td>
</tr>
<tr>
<td>L15F-806-a</td>
<td>SCTTTAATBGCWRCTGCWGG</td>
<td>I, XII, XIII</td>
<td>57</td>
</tr>
<tr>
<td>secY gene</td>
<td>GAGCAATGACWGCGTACGG/CAGGACTAACCTCCCTTGA</td>
<td>XII</td>
<td>55</td>
</tr>
<tr>
<td>SecYF1a/SecYR1(XII)</td>
<td>CCAGGTATAATTGTTTAGGTCG/ATTAGGCTTATTTCCACACCAAG</td>
<td>II</td>
<td>55</td>
</tr>
<tr>
<td>SecYF1b/SecYR1(XII)</td>
<td>TGGATGGTGICAACTCTCT/CTGGCCATTCTTCTATTAGCG</td>
<td>II</td>
<td>55</td>
</tr>
<tr>
<td>SecYF1c/SecYR1(XII)</td>
<td>CTGGAGCAGCTTTTGAAGG/GACCGTCTTTCTATTAGCG</td>
<td>III</td>
<td>50</td>
</tr>
<tr>
<td>SecYF2/SecYR1(XII)</td>
<td>GAGGGGAGCTCAAGG/CAGCCCGAACAACATTGATAAC</td>
<td>VI</td>
<td>50</td>
</tr>
<tr>
<td>SecYF2/SecYR1(XII)</td>
<td>ATYYGAGGGGYYCAACAC/GAACRCACAAAACCTGTAATT</td>
<td>VI</td>
<td>55</td>
</tr>
<tr>
<td>SecYF3/SecYR1(XII)</td>
<td>GGTGGTGTTGACGCTATTGG/AGGATAACCTAGAACAACCT</td>
<td>X</td>
<td>50</td>
</tr>
<tr>
<td>SecYF2a/SecYR1(XII)</td>
<td>CTCTCCGMCCYGGGTGGTGAAGG/CAGGAACTAACCTCCCTTGG</td>
<td>XII</td>
<td>55</td>
</tr>
</tbody>
</table>
using the primer pair L15F1/MapR1. The amplicon size is 2.8 kb for groups 16SrI, 16SrXII, 16SrXIII and 16SrXVIII and 2.2 kb for groups 16SrII, 16SrIII, 16SrIV, 16SrV, 16SrVI, 16SrVII and 16SrX. For PCR amplification, 35 cycles were conducted in an automated thermal cycler (MJ Research DNA thermal cycler PTC-200) with TaKaRa LA Taq polymerase (Takara Mirus Bio). PCR was performed in mixtures containing 1 μl DNA extract (approx. 10–30 ng), 400 μM each dNTP, 0.8 μM each primer and 2.5 U Taq polymerase. The following conditions were used: denaturation at 94 °C for 30 s (1 min for the first cycle), annealing for 1 min at 50 °C and primer extension for 5 min at 68 °C (10 min in the final cycle, at 72 °C). A negative control devoid of DNA template in the reaction mixture was included in all PCRs. Aliquots of the PCR products (3 μl) were electrophoresed through a 1% agarose gel, stained with ethidium bromide and visualized with a UV transilluminator. The amplicons were purified using PCR Kleen spin columns (Bio-Rad) or, for amplicons with multiple bands, the QIAquick gel extraction kit (Qiagen) according to the manufacturers' instructions. The majority of purified products were cloned into *Escherichia coli* TOP10 by using the TOPO TA cloning kit (Invitrogen) according to the manufacturer's instructions and sequenced with an automated DNA sequencer (ABI Prism model 3730) at the Center for Biosystems Research (University of Maryland, College Park, MD, USA) using SP6 and T7 promoter primers. For purified products with a low concentration, direct sequencing with L15F1/MapR1 primers was performed. Additional walking primers were needed to determine the complete nucleotide sequence. For 16Sr groups I, XII, XIII and XVIII, the walking primers used were L15F-646-a, L15F-806-a and MapR-703-a. For all other 16Sr groups, the walking primers used were L15F-696-b and MapR-696-b (Table 2). Other primers listed in Table 2 were used to amplify additional 16Sr phytoplasma groups for sequencing according to the procedure described above. All sequences were submitted to GenBank; accession numbers are listed in Table 1.

**Phylogenetic analysis.** *secY* nucleotide and deduced amino acid sequences and 16S rRNA gene sequences from 83 and 80 phytoplasma strains, respectively, representative of 12 distinct phytoplasma 16Sr groups (consisting of more than 40 16Sr subgroups) and from two
Acholeplasma strains and Bacillus subtilis 168 were aligned using CLUSTAL V from the LaserGene software MEALIGN program (DNASTAR). Cladistic analyses were performed with PAUP version 4.0 (Swofford, 1998) on a Power Mac G4. Uninformative characters were excluded from analyses. Phylogenetic trees were constructed by a heuristic search (or neighbour-joining algorithm) by random stepwise addition, implementing the tree bisection and reconnection branch-swapping algorithm to find the optimal tree(s) (Gunderson et al., 1994). B. subtilis 168 was used as the outgroup to root the trees. The analysis was replicated 1000 times. Bootstrapping was performed to estimate the stability and support for the inferred clades. Full-length secY and near full-length (1.43 kb) 16S rRNA gene sequences of phytoplasma strains used for phylogenetic analysis were either sequenced in this work or obtained from the GenBank database.

Strain differentiation by computer-simulated RFLP analysis of secY sequences. To demonstrate the efficacy of using the secY gene sequence for finer differentiation of phytoplasma strains, DNA fragments ranging from 1700 to 1850 bp containing the entire secY gene and partial sequences of adjacent genes (rpl15 and map) were amplified using nested PCR and sequenced from phytoplasma strains in groups 16SrI, 16SrIII and 16SrV using specific primers listed in Table 2. These sequences were trimmed to about 1.4, 1.34 and 1.37 kb for groups 16SrII, 16SrIII and 16SrVI, respectively, each of which contains the complete secY gene and partial rpl15 and map genes or intergenic spacer regions, for in silico RFLP analysis (Fig. 1b). Collective RFLP pattern types of full-length phytoplasma secY gene sequences were based on analysis by 18 restriction enzymes. In silico restriction digestion and pairwise virtual RFLP pattern comparison were performed using a modified Perl program developed previously (Wei et al., 2008), with the inclusion of an additional restriction enzyme, Tsp509I. Virtual gel images were generated using the program VGelME. Key enzymes that distinguish different (group/subgroup) pattern types were identified by using the program VGelMS. The programs VGelME and VGelMS were described previously (Zhao et al., 2009).

RESULTS AND DISCUSSION

Amplification of the partial phytoplasma spc operon

A partial spc ribosomal protein operon from each of 50 phytoplasma strains representing 12 16Sr groups was amplified by PCR and sequenced. Annotated sequences revealed the presence of two distinct spc operons among the phytoplasma strains analysed. Strains belonging to groups 16SrI, 16SrII, 16SrIII and 16SrXVIII had an spc operon with the gene order rpl15–secY–adk–map, while strains belonging to groups 16SrII, 16SrIII, 16SrIV, 16SrV, 16SrVI, 16SrVII, 16SrVIII and 16SrX had an operon with the gene order rpl15–secY–map, lacking the gene adk. The distinctness of the two spc operon types separated phytoplasmas into two major subclades, which seem to coincide with the divergence of the two major phylogenetic groups from an Acholeplasma-like ancestor early in the course of evolution, as shown on phylogenetic trees constructed previously and in this study based on 16S rRNA, ribosomal protein and secY gene sequences. This finding of two heterogeneous spc operons in ‘Candidatus Phytoplasma’ may have implications for phytoplasma evolution. Members of the genus Acholeplasma, the closest known relatives of phytoplasmas, and strains in the Clostridium/Lactobacillus group, believed to contain the common ancestor of the class Mollicutes, have an spc operon with a gene order identical to that found in phytoplasma strains belonging to groups 16SrI, 16SrXII, 16SrXIII and 16SrXVIII. This implies that phytoplasma strains in groups lacking the adk gene may have diverged from the phytoplasma common ancestor, which is most closely related to Acholeplasma species, during the course of phytoplasma evolution. Phylogenies based on secY, 16S rRNA and ribosomal protein gene sequences also support this suggestion.

Phylogenetic interrelationships and genetic variations of phytoplasmas based on nucleotide sequences of the secY gene

The phylogenetic analyses resulted in 77 equally parsimonious trees based on secY nucleotide sequences and 82 trees based on deduced SecY amino acid sequences. The trees inferred from the secY gene had a topology similar to that based on deduced SecY amino acid sequences (not shown). One representative parsimonious tree inferred from the secY gene was selected (Fig. 2). There were two subclades within the phytoplasma clade: one comprised strains in groups 16SrX and the other contained strains in groups 16SrII, 16SrIII, 16SrIV, 16SrV, 16SrVI, 16SrVII, 16SrVIII and 16SrIX. The interrelatedness among phytoplasma taxa inferred from secY gene-based phylogeny was nearly congruent with that inferred from 16S rRNA gene-based phylogenies shown in this study and reported previously (Lee et al., 1998a, 2000; Seemüller et al., 1998; Jung et al., 2002; Martini et al., 2007). Clearly, phylogenetic analysis based on secY gene sequences permitted finer differentiation of phytoplasma strains. The secY gene-based phylogeny readily resolved subgroups within a given 16Sr group, and also delineated distinct lineages irresolvable by 16S rRNA gene-based phylogeny. For example, members in each of groups 16SrV, 16SrII and 16SrIII clearly represented two distinct genetically divergent lineages, as supported by high bootstrap

Fig. 2. Phylogenetic trees constructed by parsimony analyses of partial 16S rRNA (a) and full secY (b) gene sequences from phytoplasma strains representative of 12 distinct phytoplasma 16Sr groups (consisting of more than 40 16Sr subgroups) and from two Acholeplasma strains and Bacillus subtilis. B. subtilis 168 was used as the outgroup to root the trees. Branch lengths are proportional to the number of inferred character state transformations. Bootstrap values are shown on the main branches. Bars, 5 (a) and 50 (b) inferred character state changes. 16Sr groups are indicated; subgroup affiliations are indicated in parentheses in (a). Phytoplasma strain abbreviations are defined in Table 1.
values, which were not well resolved by 16S rRNA gene-based phylogenetic analysis.

**Genetic variations of phytoplasmas assessed by comparative sequence analyses of secY and 16S rRNA genes**

The secY gene nucleic acid sequence similarity between members of different phytoplasma groups ranged from 53.5 to 77.9%, compared with 85.1–96.9% for 16S rRNA gene sequences. Examples are members of the PnWB group (16SrII) analysed in this study, where the sequence similarity ranged from 87.7 to 100% for the secY gene and 97.4 to 99.6% for the 16S rRNA gene. Among members of the X-disease (16SrIII) group, sequence similarities ranged from 93.6 to 99.8% for the secY gene and 99.0 to 99.8% for the 16S rRNA gene, whereas, among members of the clover proliferation (16SrVI) group, the sequence similarities ranged from 94.4 to 99.8% for the secY gene and 98.4 to 99.2% for the 16S rRNA gene. The greater sequence variability indicates that the secY gene is a more informative molecular tool for classification of closely related phytoplasma strains.

**Strain differentiation by virtual RFLP analysis of secY sequences**

The relatively variable secY gene sequences should be highly useful in RFLP analysis for finer differentiation of ecologically or biologically distinct strains within a given group. Table 3 summarizes the profiles produced by virtual RFLP analyses of the secY gene from phytoplasma 16Sr groups II, III and VI.

<table>
<thead>
<tr>
<th>Strain Subgroup</th>
<th>Group 16SrII</th>
<th>Group 16SrIII</th>
<th>Group 16SrVI</th>
</tr>
</thead>
<tbody>
<tr>
<td>PnWB A</td>
<td>1 1 1 1 1 1 1</td>
<td>1 1 1 1 1</td>
<td>1 1 1 1 1 1 1</td>
</tr>
<tr>
<td>SPWB A</td>
<td>1 2 2 1 1 1 1 1</td>
<td>1 1 1 1 1 1</td>
<td>1 1 1 1 1 1 1 1</td>
</tr>
<tr>
<td>SEPT A</td>
<td>1 1 2 1 1 1 1 1 1</td>
<td>1 1 1 1 1</td>
<td>1 1 1 1 1 1 1</td>
</tr>
<tr>
<td>SEPN A</td>
<td>1 1 2 1 1 1 1 1 1 1</td>
<td>1 1 1 1 1</td>
<td>1 1 1 1 1 1</td>
</tr>
<tr>
<td>CacWBRYN16 C</td>
<td>2 3 3 2 2 2 2 2 2 2 1 2</td>
<td>1 1 1 1</td>
<td>1 1 1 1 1</td>
</tr>
<tr>
<td>SOYP C</td>
<td>2 3 3 2 2 2 2 2 2 2 3 2 1 2</td>
<td>1 1 1 1</td>
<td>1 1 1 1 1</td>
</tr>
<tr>
<td>CrP C</td>
<td>2 3 3 2 2 3 2 2 2 2 2 1 2</td>
<td>1 1 1 1</td>
<td>1 1 1 1 1</td>
</tr>
<tr>
<td>CoP C</td>
<td>3 4 4 1 2 2 2 2 2 2 2 1 2</td>
<td>1 1 1 1</td>
<td>1 1 1 1 1</td>
</tr>
<tr>
<td>TBB D</td>
<td>1 1 1 1 1 1 1 1 1 1 1 1</td>
<td>1 1 1 1</td>
<td>1 1 1 1 1</td>
</tr>
<tr>
<td>PEP E</td>
<td>4 4 5 3 3 2 3 3 4 3 2 3</td>
<td>1 1 1 1</td>
<td>1 1 1 1 1</td>
</tr>
<tr>
<td>Group 16SrIII</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CX A</td>
<td>1 1 1 1 1 1 1 1 1 1 1</td>
<td>1 1 1 1 1</td>
<td>1 1 1 1 1</td>
</tr>
<tr>
<td>WX A</td>
<td>1 1 1 1 1 1 1 2 1 1 1 1</td>
<td>1 1 1 1</td>
<td>1 1 1 1 1</td>
</tr>
<tr>
<td>CYE B</td>
<td>2 2 2 1 1 2 3 2 1 2 1 2</td>
<td>1 1 1 1</td>
<td>1 1 1 1 1</td>
</tr>
<tr>
<td>Vac F</td>
<td>3 2 1 2 3 2 4 1 1 2 1 3</td>
<td>1 1 1 1</td>
<td>1 1 1 1 1</td>
</tr>
<tr>
<td>PBT C</td>
<td>4 2 1 1 4 3 5 1 1 2 1 4</td>
<td>1 1 1 1</td>
<td>1 1 1 1 1</td>
</tr>
<tr>
<td>GR1 D</td>
<td>3 2 1 2 3 2 4 1 1 2 1 3</td>
<td>1 1 1 1</td>
<td>1 1 1 1 1</td>
</tr>
<tr>
<td>SP1 E</td>
<td>5 3 1 1 1 1 6 3 1 1 1 5</td>
<td>1 1 1 1</td>
<td>1 1 1 1 1</td>
</tr>
<tr>
<td>MW1 F</td>
<td>6 2 1 1 3 2 4 1 1 2 1 3</td>
<td>1 1 1 1</td>
<td>1 1 1 1 1</td>
</tr>
<tr>
<td>AKpot7 F</td>
<td>3 2 1 2 3 2 4 1 1 2 1 6</td>
<td>1 1 1 1</td>
<td>1 1 1 1 1</td>
</tr>
<tr>
<td>WWB G</td>
<td>1 1 1 1 1 1 7 1 1 1 1 7</td>
<td>1 1 1 1</td>
<td>1 1 1 1 1</td>
</tr>
<tr>
<td>JR1Ph H</td>
<td>2 1 1 1 1 1 1 1 1 1 1 1</td>
<td>1 1 1 1</td>
<td>1 1 1 1 1</td>
</tr>
<tr>
<td>MT117 M</td>
<td>3 2 1 2 3 2 4 1 1 2 1 3</td>
<td>1 1 1 1</td>
<td>1 1 1 1 1</td>
</tr>
<tr>
<td>AKpot6 N</td>
<td>3 2 1 2 3 2 4 1 1 2 1 7</td>
<td>1 1 1 1</td>
<td>1 1 1 1 1</td>
</tr>
<tr>
<td>Group 16SrVI</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CP A</td>
<td>1 1 1 1 1 1 1 1 1 1 1 1 1 1</td>
<td>1 1 1 1</td>
<td>1 1 1 1 1</td>
</tr>
<tr>
<td>PWB A</td>
<td>1 1 1 1 1 1 1 1 1 1 1 1 2 1</td>
<td>1 1 1 1</td>
<td>1 1 1 1 1</td>
</tr>
<tr>
<td>AKpot1 A</td>
<td>1 1 1 1 1 1 1 1 1 1 1 1 1 1</td>
<td>1 1 1 1</td>
<td>1 1 1 1 1</td>
</tr>
<tr>
<td>AKpot4 A</td>
<td>1 1 1 1 1 1 1 2 2 1 2 1 1 1</td>
<td>1 1 1 1</td>
<td>1 1 1 1 1</td>
</tr>
<tr>
<td>AKpot5 A</td>
<td>1 1 1 1 1 1 1 2 2 1 2 1 1 1</td>
<td>1 1 1 1</td>
<td>1 1 1 1 1</td>
</tr>
<tr>
<td>AKpot2 A</td>
<td>1 1 1 1 1 1 1 2 2 1 2 1 1 1</td>
<td>1 1 1 1</td>
<td>1 1 1 1 1</td>
</tr>
<tr>
<td>LUM A</td>
<td>2 1 1 1 1 1 1 3 1 1 1 3 2</td>
<td>1 1 1 1</td>
<td>1 1 1 1 1</td>
</tr>
<tr>
<td>DBPh2 A</td>
<td>3 1 1 1 1 2 1 4 1 1 3 1 3</td>
<td>1 1 1 1</td>
<td>1 1 1 1 1</td>
</tr>
<tr>
<td>DBPh3 A</td>
<td>3 1 1 1 1 2 1 4 1 1 3 1 3</td>
<td>1 1 1 1</td>
<td>1 1 1 1 1</td>
</tr>
<tr>
<td>VR A</td>
<td>3 1 1 1 1 2 1 4 1 1 3 4 3</td>
<td>1 1 1 1</td>
<td>1 1 1 1 1</td>
</tr>
<tr>
<td>BLL D</td>
<td>2 2 1 1 1 1 1 2 1 1 1 1 4</td>
<td>1 1 1 1</td>
<td>1 1 1 1 1</td>
</tr>
</tbody>
</table>
16Sr phytoplasma group. To evaluate the efficacy of the secY gene for finer differentiation of phytoplasma strains, computer-simulated virtual RFLP patterns were generated using iPhyClassifier (Zhao et al., 2009). Table 3 summarizes the collective pattern types of strains analysed in groups 16SrII, 16SrIII and 16SrVI using 12 key restriction enzymes among the 18 used for analysis. Biologically distinct strains within a given 16Sr group and, in most cases, 16Sr subgroup were readily differentiated based on collective patterns. For example, ten members of subgroup 16SrVI-A were differentiated into seven secY(VI) genotypes represented by the following individual strains or strain clusters: CP and AKpot1; AKpot2, AKpot4 and AKpot5; DBPh2 and DBPh3; PWB; LUM; BLL; and VR. Four members of subgroup 16SrII-A (SEPT, SEPN, PnWB and SPWB) were differentiated into three secY(II) genotypes, with two strains, SEPT and SEPN, having the same genotype. Four members of subgroup 16SrII-C (CaWBYN16, SOYP, CrP and CoP) were differentiated into four secY(II) genotypes. However, in one case, two different 16Sr subgroups were found to have similar secY genotypes; AKpot6 and AKpot7, which are classified as

Fig. 3. Computer-simulated virtual RFLP patterns derived from in silico digestions of phytoplasma secY gene fragments (about 1.4, 1.34 and 1.37 kb, respectively, for groups 16SrII, 16SrIII and 16SrVI) from representative strains of groups 16SrII, 16SrIII and 16SrVI with six key enzymes: Alul, BfaI, Msel, Sau3AI, TaqI and Tsp509I. Lanes MW, HaeIII digest of φX174 RFI DNA; fragment sizes (bp) from top to bottom: 1353, 1078, 872, 603, 310, 281, 271, 234, 194, 118, 72. Phytoplasma strain abbreviations are defined in Table 1. Subgroup affiliations: PnWB, SPWB, SEPT and SEPN, 16SrII-A; CaWBYN16, SOYP, CrP and CoP, 16SrII-C; TBB, 16SrII-D; PEP, 16SrII-E; CX and WX, 16SrIII-A; CYE, 16SrIII-B; PBT, 16SrIII-C; GR1, 16SrIII-D; SP1, 16SrIII-E; Vac, MW1 and AKpot7, 16SrIII-F; WWB, 16SrIII-G; JRP1, 16SrIII-H; MT117, 16SrIII-M; AKpot6, 16SrIII-N; CP, PWB, AKpot1, AKpot4, AKpot5, AKpot2, LUN, DBPh2, DBPh3 and VR, 16SrVI-A; BLL, 16SrVI-D.
gene in combination with an additional marker that displays moderate genetic variability in phylogenetic analysis overcomes the limitations of the 16S rRNA gene for classification of phytoplasma strains below the species level. The present study and our previous studies (Lee et al., 2006a; Martini et al., 2002, 2007) indicate that phylogenetic analyses based on moderately variable genes, such as secY or ribosomal protein genes, increased the resolving power substantially for delineation of 16S rRNA groups and subgroups and for closely related but biologically distinct strains. For example, strain MBS, strains FD-C, ALY and SpaWB and strains DBPh2, DBPh3 and VR can be differentiated readily from other members of groups 16SrIB, 16SrV and 16SrVI, respectively. These strains inhabit mutually distinct ecological niches (Table 1). Emerging multiple-gene-based classification systems should provide molecular criteria for improved delineation of species and strains.

**ACKNOWLEDGEMENTS**

We thank all the individuals who provided phytoplasma strains used in this study. We thank Wei Wei for her assistance in preparing in silico RFLP images. We also thank Prachi Bagadia for her technical assistance in sequence assembly and phylogenetic analysis.

**REFERENCES**


