Construction of an interactive online phytoplasma classification tool, iPhyClassifier, and its application in analysis of the peach X-disease phytoplasma group (16SrIII)

Yan Zhao, Wei Wei, Ing-Ming Lee, Jonathan Shao, Xiaobing Suo and Robert E. Davis

Molecular Plant Pathology Laboratory, USDA-Agricultural Research Service, Beltsville, MD 20705, USA

Phytoplasmas, the causal agents of numerous plant diseases, are insect-vector-transmitted, cell-wall-less bacteria descended from ancestral low-G+C-content Gram-positive bacteria in the Bacillus–Clostridium group. Despite their monophyletic origin, widely divergent phytoplasma lineages have evolved in adaptation to specific ecological niches. Classification and taxonomic assignment of phytoplasmas have been based primarily on molecular analysis of 16S rRNA gene sequences because of the inaccessibility of measurable phenotypic characters suitable for conventional microbial characterization. In the present study, an interactive online tool, iPhyClassifier, was developed to expand the efficacy and capacity of the current 16S rRNA gene sequence-based phytoplasma classification system. iPhyClassifier performs sequence similarity analysis, simulates laboratory restriction enzyme digestions and subsequent gel electrophoresis and generates virtual restriction fragment length polymorphism (RFLP) profiles. Based on calculated RFLP pattern similarity coefficients and overall sequence similarity scores, iPhyClassifier makes instant suggestions on tentative phytoplasma 16Sr group/subgroup classification status and ‘Candidatus Phytoplasma’ species assignment. Using iPhyClassifier, we revised and updated the classification of strains affiliated with the peach X-disease phytoplasma group. The online tool can be accessed at http://www.ba.ars.usda.gov/data/mppl/iPhyClassifier.html.

Phytoplasmas are small, insect-transmitted, cell-wall-less bacteria that cause numerous diseases in economically and environmentally important plant species worldwide (McCoy et al., 1989; Lee et al., 2000; Seemüller et al., 2002; Hogenhout et al., 2008). In infected plants, phytoplasmas colonize enucleate sieve cells of phloem tissue and induce various systemic symptoms including yellowing, shoot proliferation, witches'-broom growth, phyllody and virescence. Phylogenetic studies suggest that extant phytoplasmas share a common evolutionary root and are descended from low-G+C-content Gram-positive bacteria in the Bacillus–Clostridium group (Weisburg et al., 1989; Gundersen et al., 1994; Sears & Kirkpatrick, 1994; Zhao et al., 2005; Wei et al., 2007). After evolutionary divergence from an Acholeplasma-like last common ancestor, phytoplasmas emerged as a discrete clade and a large number of widely divergent phytoplasma lineages have evolved in adaptation to a broad range of bio- and geo-ecological niches (Lee et al., 1992a, b, 1993a, 2000; Marcone et al., 1999; Davis et al., 2005; Martini et al., 2007; Wei et al., 2007, 2008a, b; Cai et al., 2008). Since phytoplasmas cannot yet be cultured successfully in a cell-free medium and measurable phenotypic characters suitable for conventional microbial characterization remain inaccessible, current knowledge on biodiversity and genetic interrelationships among phytoplasma strains has been derived mainly from aetiological studies, applications of serological and nucleic acid-based assay techniques and molecular analysis of evolutionarily conserved gene sequences.

As in other prokaryotes (Rosselló-Mora & Amann, 2001), genes encoding 16S rRNAs are highly conserved across the
In the present study, we devised an interactive online tool, iPhyClassifier, for real-time identification and classification of phytoplasmas. Besides implementing the concepts and programs that we described previously (Wei et al., 2007, 2008b), iPhyClassifier integrates additional functions that we developed in the present study. Such new functions include overall sequence comparison and similarity score calculation, intelligent trimming of input sequences and publication-ready virtual gel plotting. The newly developed virtual gel-plotting function is able to generate not only virtual gel images resulting from multiple enzyme analysis of a single 16S F2nR2 DNA sequence from any given strain but also virtual gel images resulting from a single enzyme digestion of multiple DNA sequences. iPhyClassifier also incorporates carefully curated databases of phytoplasma 16S rRNA gene sequences for up-to-date classification and comparative studies. A simple operation of iPhyClassifier on a user input sequence can quickly lead to identification of the phytoplasma strain under study, providing suggestions on its tentative 16Sr group/subgroup classification status and ‘Ca. Phytoplasma’ species (or related strain) assignment. As a case study for iPhyClassifier application, we revised and updated the classification status of phytoplasma strains affiliated with the peach X-disease phytoplasma group.

**Program components**

The current version of iPhyClassifier contains the following three program modules: a sequence similarity search and pairwise sequence similarity score calculation module (PM1), an intelligent sequence trimming and virtual RFLP analysis module (PM2) and a virtual electrophoresis gel image plotting module (PM3).

PM1 carries out two functions. Firstly, it performs pairwise nucleotide sequence comparisons (query against database entries) using the basic local alignment search tool (BLAST; Altschul et al., 1990) to identify a query’s phylogenetically close neighbours quickly and to determine whether or not a query sequence is of phytoplasma origin. Secondly, PM1 creates a global sequence alignment between the query sequence and sequences from the reference strain of each known ‘Ca. Phytoplasma’ species using the CLUSTAL W algorithm (Thompson et al., 1994) and calculates percentage nucleotide sequence similarity scores using the Myers–Miller algorithm (Myers & Miller, 1988).

PM2 consists of two Perl scripts, TrimF2nR2 and RFLP_pattern_comparison. The TrimF2nR2 script, developed in the present study, prepares input nucleotide sequences for simulated enzymic digestions. The script parses through input sequences for generic annealing sites of phytoplasmal universal primers R16F2n and R16R2 (Gundersen & Lee, 1996) and trims each input sequence to the full-length F2nR2 region, which includes the primer annealing sites (Wei et al., 2007). On each trimmed F2nR2 sequence, the RFLP_pattern_comparison script conducts simulated enzymic digestions, records the length of each...
restriction fragment and performs pairwise comparisons of the recorded fragment lengths. Based on summarized numbers of similar and dissimilar fragments, the script calculates a similarity coefficient (F) for each pair of phytoplasma strains, as described previously (Wei et al., 2008b).

PM3 consists of two Perl scripts, VGelME and VGelMS; both were developed in the present study. While VGelME generates virtual electrophoresis gel images resulting from in silico digestions of a single input sequence (an F2nR2 fragment from a single phytoplasma strain) by 17 individual enzymes, VGelMS produces gel images resulting from in silico digestions of multiple input sequences (F2nR2 fragments from multiple phytoplasma strains) by a single restriction enzyme. The latter helps to identify key restriction enzymes that distinguish different group and subgroup patterns.

16S rRNA gene sequence databases

The current version of iPhyClassifier incorporates three 16S rRNA gene sequence databases: DB1, a set of full- or near-full-length 16S rRNA gene sequences from reference strains of all formally described ‘Ca. Phytoplasma’ species, reference strains of ‘Ca. Phytoplasma’ species that were proposed by the IRPCM Phytoplasma/Spiroplasma Working Team – Phytoplasma taxonomy group (2004) but have not yet been formally described, reference strains of potentially novel ‘Ca. Phytoplasma’ species identified in our previous study (Wei et al., 2007) and all type strains of named prokaryotic species; DB2, a set of F2nR2 sequences from representative strains of established phytoplasma 16Sr groups and subgroups; and DB3, a set of F2nR2 sequences compiled from all phytoplasma 16S rRNA gene sequences currently deposited in GenBank, EMBL and DDBJ. The names of the non-phytoplasmal prokaryotic species in DB1 are those validly published in the International Journal of Systematic and Evolutionary Microbiology (formerly International Journal of Systematic Bacteriology) and were obtained from the List of Prokaryote Names with Standing in Nomenclature (http://www.bacterio.cict.fr/validation-lists.html; Euzéby, 1997) (last update 9 October 2008). All 16S rRNA gene sequences were downloaded from the NCBI nucleotide sequence database at http://www.ncbi.nlm.nih.gov/gquery/gquery.fcgi using the Entrez search and retrieval tool (Wheeler et al., 2005). All databases are maintained using MySQL.

Operational process

The overall operational process of iPhyClassifier is outlined in Fig. 1. The aim of the entire operation is to provide meaningful suggestions on tentative 16Sr group/subgroup classification status and ‘Ca. Phytoplasma’ species (or related strain) assignment for any phytoplasma strain under study. The operation starts by receiving a query sequence(s) from the user. The queries, in FASTA format, can be either uploaded as a precompiled file from the user’s computer to the iPhyClassifier web server or directly typed or pasted into the query sequence input window in the
The first step of the iPhyClassifier operation is to perform pairwise comparisons between each query sequence and the sequences in database DB1 for quick identification of the query’s phylogenetically close neighbours. In this initial stage of sequence comparison, the BLAST algorithm is used. If none of the ‘Ca. Phytoplasma’ species in DB1 appears among the top 50 hits returned from the BLAST search, or one or more ‘Ca. Phytoplasma’ species appears among the top hits but shares ≤91 % sequence similarity with the query, the operation will abort, warning that the query sequence is unlikely to originate from a phytoplasma. If at least one ‘Ca. Phytoplasma’ species is among the top hits returned from the BLAST search and shares ≥92 % sequence similarity with the query, the operation will proceed and the query sequence will be fed into the CLUSTAL W program for global alignment with all phytoplasma sequences in the database DB1Phy (a subset of DB1) and for sequence similarity score calculation. Such a combined search strategy aids identification of the query’s phylogenetically closest neighbour with a significantly reduced computing time (Chun et al., 2007). In accordance with the convention on 16S rRNA gene sequence-based prokaryotic species delineation (Murray & Schleifer, 1994; Stackebrandt & Goebel, 1994), iPhyClassifier implements the recommendation of the IRPCM Phytoplasma/Spiroplasma Working Team – Phytoplasma taxonomy group (2004) and presets 97.5 % 16S rRNA gene sequence similarity as the cut-off value for novel ‘Ca. Phytoplasma’ species recognition. Since the generally conserved 16S rRNA gene sequence contains pockets of hypervariable regions, the sequence similarity score calculation should be based upon comparison of full- or near-full-length 16S rRNA gene sequences. It requires that each query sequence covers at least 1200 positions within a 16S rRNA gene. The output of this operational step consists of the assignment of the query strain tentatively to an existing ‘Ca. Phytoplasma’ species as a related strain or the suggestion that the query represents a potentially novel ‘Ca. Phytoplasma’ species, depending on the sequence similarity scores.

The second step of the iPhyClassifier operation is to trim each query sequence to the full-length F2nR2 region using regular expressions that match primer pair R16F2n/R16R2. This step is critical because, in the 16S rRNA gene-based phytoplasma classification scheme, strains are classified into groups and subgroups based strictly on RFLP patterns derived from 16S rRNA gene F2nR2 fragments (Lee et al., 1998, 2000; Wei et al., 2007, 2008b).

The third step of iPhyClassifier operation is to simulate restriction digestions on trimmed F2nR2 fragments, compare the RFLP pattern types derived from each query strain to those derived from database DB2 and calculate pairwise RFLP pattern similarity coefficients. In this step, iPhyClassifier implements the criterion proposed in our previous work (Wei et al., 2008b), presetting 0.97 as the threshold similarity coefficient for delineation of a new subgroup RFLP pattern type within a given group. Thus, if the virtual F2nR2 RFLP pattern derived from a 16S rRNA gene of a phytoplasma strain under study has a similarity coefficient of 0.97 or less with 16S rRNA genes of all existing representative or reference strains of the given group, a new subgroup pattern type is recognized. Adoption of 0.97 as the threshold similarity coefficient for new subgroup delineation is warranted because it reflects precisely the existing subgroup classification scheme, in which as few as one restriction site difference can distinguish a new subgroup. A similarity coefficient of 0.85 or less with all previously recognized subgroups signals that the strain under study may represent a new 16Sr group, in agreement with all previously designated groups. RFLP patterns that have a similarity coefficient of 0.99 or 0.98 with the standard pattern type of the designated representative or reference member in a given subgroup are considered as variants of the standard pattern type. These variants or minor pattern types are denoted...
with one or two asterisks (* or **) following the corresponding subgroup letter, for example 16SrI-A* (F=0.99) and 16SrI-A** (F=0.98), as suggested previously (Wei et al., 2008b). Since similarity coefficient values are influenced by both the number and the particular set of restriction enzymes selected for RFLP analysis, the threshold similarity coefficients for new subgroup and group pattern type delineations are based strictly on the use of a specific set of 17 restriction enzymes originally established for classification of phytoplasmas using actual gel electrophoresis-based RFLP analysis (Lee et al., 1998).

The output of this operational step is the assignment of the strain under study into an existing subgroup or erection of a new subgroup. Because the presence of two heterogeneous rrn operons in individual phytoplasma strains is widespread (see the section on interoperon sequence heterogeneity below), final subgroup designation of strains with heterogeneous rrn operons should be based on composite patterns derived from both rrn operons. At the end of this operational step, the query sequence is added to database DB3.

Concomitant with similarity coefficient calculation, which generates numerical output of the RFLP pattern analysis, the i PhyClassifier also provides visual output, i.e. virtual gel images resulting from the RFLP pattern analysis. The gel images reveal informative sites or ‘visible’ genetic markers along the 16S rRNA gene sequences, transforming sequence information into accessible ‘virtual phenotypic characters’ for phytoplasma strain differentiation and classification.

**Critical issues**

Since the operation of i PhyClassifier is solely dependent upon sequence information, any error in a query (input) sequence that misrepresents the phytoplasma strain under study could result in erroneous group/subgroup classification and ‘Ca. Phytoplasma’ species assignment. While sequence errors may arise at various stages during PCR amplification, plasmid multiplication and DNA sequencing, they usually occur randomly and can be rectified by sample replications. To ensure credible operations of i PhyClassifier, we highly recommend that consistent sequence data be obtained from at least two independent samples, i.e. from two or more infected plants or insect individuals. If only one infected plant or insect sample is available for study, consistent sequence data from at least two independently cloned DNA segments derived from two separate PCRs must be obtained. Each clone (plasmid) should be sequenced in both directions and a minimum of triplicate coverage per base position achieved.

The genomes of all four completely sequenced phytoplasma strains and numerous reference strains of ‘Ca. Phytoplasma’ species harbour two rrn operons, rrnA and rrnB (IRPCM Phytoplasma/Spiroplasma Working Team – Phytoplasma taxonomy group, 2004; Oshima et al., 2004; Bai et al., 2006; Kube et al., 2008; Tran-Nguyen et al., 2008). In many strains, the sequences of the two rrn operons differ (Lee et al., 1993b, 1998; Firrao et al., 1996; Liefing et al., 1996; Davis & Sinclair, 1998; Jomantiene et al., 2002; Davis et al., 2003). For those phytoplasma strains with two heterogeneous rrn operons, if the sequence variations between the two operons fall within restriction sites within the 16S rRNA gene F2nR2 region, two different virtual 16Sr RFLP pattern types will result from i PhyClassifier operation. It is therefore important to distinguish between subgroup pattern types and final subgroup designation and to avoid erroneous assignment of the same strain into two different 16Sr subgroups.

In this regard, i PhyClassifier adopts the recommendation of Wei et al. (2008b) and uses a three-letter subgroup designation, where the first and second letters (in parentheses) denote the RFLP pattern types of rrnA and rrnB, respectively, and the third letter designates the 16Sr subgroup. For example, paulownia witches’-broom (PaWB) phytoplasma, a member of the previously delineated subgroup 16SrI-D (Lee et al., 1998), possesses two sequence-heterogeneous rRNA operons that display different 16Sr RFLP patterns, characteristic of subgroups 16SrI-B and 16SrI-D, respectively (Wei et al., 2008b); therefore, the subgroup status of PaWB is redesignated 16SrI-(B/D)D.

**An update of peach X-disease phytoplasma group (16SrIII) classification**

X-disease of peach in North America was first reported in the early 1930s (Stoddard, 1934); however, it wasn’t until the 1970s that the causal agent of the disease was identified to be a phytoplasma (Granett & Gilmer, 1971; Macbeath et al., 1972). Since then, numerous phytoplasmas affecting stone-fruit trees, as well as many other plants, were found to be closely related to the phytoplasma(s) associated with X-disease of peach in the western and eastern USA. On the basis of RFLP analyses of 16S rRNA gene sequences, these strains have been classified into a single RFLP group, 16SrIII (Lee et al., 1993b, 1998; Jomantiene et al., 2002). This group includes strains that differ in their geographical origin, plant host/insect vector relations and symptoms induced in infected plants. It has been found that such varied strains may carry distinct molecular markers in their 16S rRNA gene sequences (Lee et al., 1993b, 1998). Based on actual enzymic RFLP analysis of 16S rRNA genes F2nR2 fragments, 14 subgroups were delineated (Lee et al., 1993b, 1998; Jomantiene et al., 2002), each subgroup consisting of strains that share identical or nearly identical RFLP patterns.

In the present study, using i PhyClassifier, we updated the classification status of the phytoplasma strains affiliated with the peach X-disease phytoplasma group. Standard virtual 16Sr rRNA gene RFLP patterns were generated for representative strains of the 14 previously delineated subgroups (Table 1 and Fig. 3). Composite RFLP patterns that reflect the presence of two heterogeneous rrn operons in the representative strains of three subgroups, 16SrIII-
Table 1. Classification of phytoplasmas in the peach X-disease phytoplasma group (16SrIII) based on RFLP patterns derived from 16S rRNA gene F2nR2 sequences

Novel subgroups delineated in the present study are highlighted in bold. Asterisks indicate variants of the designated subgroup or pattern type. In the Subgroup column, letters in parentheses denote the RFLP pattern types of \textit{rrnA/rrnB}. Representative (or reference) strains are the first strains for which a given subgroup pattern type was recognized.

| Subgroup | Virtual pattern | GenBank accession no. | Representative strain | Original source | Notes |
|----------|-----------------|-----------------------|-----------------------|-----------------|-------|
| 16SrIII-A | 16SrIII-A | L33733 | Peach X-disease phytoplasma CX | Peach (Canada) | No heterogeneous \textit{rrn} reported so far |
| 16SrIII-B | 16SrIII-B | AF175304 | Clover yellow edge phytoplasma CYE-C | Clover (Canada) | No heterogeneous \textit{rrn} reported so far |
| 16SrIII-C | 16SrIII-C | FJ376626 | Pecan bunch phytoplasma PB1 | Pecan (Georgia, USA) | No heterogeneous \textit{rrn} reported so far |
| 16SrIII-D | 16SrIII-D | FJ376627 | Goldenrod yellows phytoplasma GR1 | Goldenrod (New York, USA) | No heterogeneous \textit{rrn} reported so far |
| 16SrIII-(A*/?)E | 16SrIII-A*, \textit{rrnA} | AF190228 | Spirea stunt phytoplasma SP1 | Spirea (New York, USA) | Incomplete; \textit{rrnB} sequence unavailable |
| 16SrIII-F | 16SrIII-F | AF510724 | Milkweed yellows phytoplasma MW1 | Milkweed (New York, USA) | No heterogeneous \textit{rrn} reported so far |
| 16SrIII-(G/A*)G | 16SrIII-G, \textit{rrnA} | AF190226 | Walnut witches' broom phytoplasma WWB | Walnut (Georgia, USA) | Composite pattern is available (Fig. 4) |
| 16SrIII-(A*/?)I | 16SrIII-A*, \textit{rrnA} | AF060875 | Virginia grapevine yellows phytoplasma VGYIII | Grapevine (Virginia, USA) | Incomplete; \textit{rrnB} sequence unavailable |
| 16SrIII-J | 16SrIII-J | AF147006 | Chayote witches' broom phytoplasma ChWBIII (Ch10) | Chayote (Brazil) | No heterogeneous \textit{rrn} reported so far |
| 16SrIII-K | 16SrIII-K | AF274876 | Strawberry leafy fruit phytoplasma SLF | Strawberry (Maryland, USA) | No heterogeneous \textit{rrn} reported so far |
| 16SrIII-L | 16SrIII-L | EU169138 | Poinsettia exuberant flower-inducing phytoplasma EF-MM | Poinsettia (Mexico) | No heterogeneous \textit{rrn} reported so far |
| 16SrIII-M | 16SrIII-M | FJ226074 | Montana potato purple top phytoplasma PPT-MT17-1 | Potato (Montana, USA) | No heterogeneous \textit{rrn} reported so far |
| 16SrIII-N | 16SrIII-N | FJ376629 | Alaska potato purple top phytoplasma PPT-AK6 | Potato (Alaska, USA) | No heterogeneous \textit{rrn} reported so far |
| 16SrIII-(P/O)P | 16SrIII-O, \textit{rrnB} | AF370120 | Dandelion virescence phytoplasma DanVir | Dandelion (Lithuania) | Composite pattern is available (Fig. 4) |
| 16SrIII-Q | 16SrIII-Q | AF302841 | Black raspberry witches’ broom phytoplasma BRWB | Black raspberry (Oregon, USA) | No heterogeneous \textit{rrn} reported so far |
| 16SrIII-(R/B)R | 16SrIII-R, \textit{rrnA} | AF370119 | Dandelion virescence phytoplasma DanVir | Dandelion (Lithuania) | Composite pattern is available (Fig. 4) |
| 16SrIII-S | 16SrIII-S | L04682 | Western peach X-disease phytoplasma WX | Peach (California, USA) | No heterogeneous \textit{rrn} reported so far |
(A*/G)G, 16SrIII-(O/P)P and 16SrIII-(B/R)R, were also generated (Fig. 3).

Four new 16SrIII subgroup lineages were delineated in the present study. These new subgroups include subgroup 16SrIII-L, represented by poinsettia exuberant flower-inducing phytoplasma strain EF-MM (GenBank accession no. EU169138), subgroup 16SrIII-M, represented by Montana potato purple top phytoplasma strain PPT-MT117-1 (FJ226074), and subgroup 16SrIII-N, represented by Alaska potato purple top phytoplasma strain PPT-AK6 (FJ376629) (Table 1, Fig. 3 and Supplementary Table S1, available in IJSEM Online). Just as various strains classified in a given subgroup would be expected to share common

Fig. 3. Virtual RFLP patterns derived from in silico digestions of 16S rRNA gene F2nR2 fragments from representative strains of novel and previously delineated 16SrIII subgroups. Recognition sites for the following 17 restriction enzymes were used in the simulated digestions: AluI, BamHI, BglII, BstUI (ThaI), DraI, EcoRI, HaeIII, HhaI, Hinfl, HpaI, HpaII, KpnI, Sau3AI (MboI), MseI, RsaI, SspI and TaqI. Lanes MW, φX174 DNA digested with HaeIII.
biological characteristics, strains classified in different subgroups may have distinguishing biological properties, as apparently illustrated by the newly delineated subgroups III-L, III-M and III-N. We recognize that, in many cases, distinguishing biological characteristics may currently be unknown.

Results from the current study revealed that the RFLP patterns derived from three 16S rRNA gene sequences of peach X-disease phytoplasma strains that originated from the western USA (GenBank accession nos L04682, AF533231 and EU168790) are identical and novel (Fig. 3). This novel pattern has a similarity coefficient of 0.97 with "PhyClassifier, an online phytoplasma classification tool."
the pattern from the 16SrIII-A representative strain Canada peach X-disease phytoplasma (CX; GenBank accession no. L33733), and has similarity coefficients of less than 0.97 with the patterns from the representative strains of all other 16SrIII subgroups (Supplementary Table S1). Therefore, these western X-disease phytoplasma strains may constitute a distinct lineage within the X-disease phytoplasma group. We recommend that a new subgroup, 16SrIII-S, be erected to accommodate these strains and to designate strain WX (GenBank accession no. L04682) as the representative member of the subgroup. Previous aetiological and epidemiological studies suggested that WX and CX possessed distinct biological properties. Strain WX was prevalent in the western USA and was found to be transmitted predominantly by Colladosus montanus, while CX was found to be transmitted mainly by Paraphlepsius irratus and to occur in eastern USA and eastern Canada (Sinha & Chiykowski, 1980; Chiykowski & Sinha, 1990; Kirkpatrick et al., 1990; Liefting & Kirkpatrick, 2003). Interestingly, these distinguishing characteristics correlate with the classification of these strains into two separate subgroups in this study. Thus, the significance of subgroup delineations lies in the potential of the subgroup-level molecular markers to distinguish closely related strains that may differ in subtle but biologically significant properties.

In addition, the current study also identified a new pattern type that is a variant of standard 16SrIII-A pattern (with a similarity coefficient of 0.99). 16S rRNA gene sequences that exhibited this variant pattern (16SrIII-A*; Fig. 3) include those from phytoplasma strains that were classified in subgroups 16SrIII-E, 16SrIII-H and 16SrIII-I (Table 1 and Supplementary Table S1), indicating that these phytoplasma strains may possess two sequence-heterogeneous rRNA operons, as does the walnut witches’ broom phytoplasma [16SrIII- (A*/G)G] (Supplementary Table S1). The representative strains of the novel 16SrIII subgroups delineated using iPhyClassifier are clustered phylogenetically with representative strains of previously recognized 16SrIII subgroups, forming a subclade (Fig. 4). The tree topology indicated that clustering of 16Sr RFLP groups is consistent with 16S rRNA gene sequence-based phylogeny. However, we note that, not surprisingly, multiple strains belonging to a single 16Sr subgroup may not necessarily cluster together in a phylogenetic subtree (not shown), since subgroup classification is based upon a subset (i.e. recognition sites of 17 restriction enzymes) of the characters that are used in phylogenetic analysis. According to iPhyClassifier analysis, novel subgroup patterns delineated in the present study can be distinguished from previously recognized patterns by in silico digestions with a few key enzymes. For example, subgroup pattern 16SrIII-L can be distinguished from others by TaqI digestion, separation of subgroup pattern 16SrIII-M from all other subgroup pattern types can be achieved by MseI digestion, distinction between subgroup pattern 16SrIII-N and other subgroup pattern types can be made by either Alul or TaqI digestion and subgroup pattern 16SrIII-S can be differentiated from other pattern types by MseI and BstUI digestions (Fig. 5).

In conclusion, RFLP profiling of PCR-amplified 16S rRNA gene fragments has served as a primary means for differentiation and classification of phytoplasmas over the past 15 years. We anticipate that 16S rRNA gene RFLP patterns will continue to be exploited as molecular genetic markers for identification of known phytoplasmas and discovery of novel phytoplasmas in the foreseeable future. iPhyClassifier provides a user-friendly platform to identify such molecular genetic markers quickly in diverse phytoplasma strains and to transform them into virtual phenotypic characters in the form of sequence similarity scores. RFLP pattern similarity coefficients and virtual gel images followed by instant suggestions on tentative 16Sr group/subgroup classification status and ‘Ca. Phytoplasma’ species assignment for the phytoplasma strains under study. Rapid delineation of novel phytoplasma lineages affiliated with the peach X-disease phytoplasma group demonstrated the feasibility and effectiveness of the iPhyClassifier operation. In addition, since computer-generated RFLP patterns (Wei et al., 2007, 2008b; Cai
et al., 2008; this paper) faithfully replicate the classical, authoritative patterns that had been established by conventional RFLP analysis (Lee et al., 1993b, 1998). iPhyClassifier can serve as an assistant, by providing reference strain RFLP patterns, to researchers who prefer to perform conventional RFLP analysis for identification and classification of unknown or novel phytoplasmas. The framework of iPhyClassifier can easily be expanded to accommodate the full-length 16S rRNA gene, other phytoplasma genes and multilocus virtual RFLP analyses when more phytoplasma genomic DNA sequences become available.

References

Al-Saady, N. A., Khan, A. J., Calari, A., Al-Subhi, A. M. & Bertaccini, A. (2008). ‘Candidatus Phytoplasma omanense’, associated with witches’-broom of Cassia italica (Mill.) Spreng. in Oman. Int J Syst Evol Microbiol 58, 461–466.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990). Basic local alignment search tool. J Mol Biol 215, 403–410.

Arocha, Y., Antesana, O., Montellano, E., Franco, P., Plata, G. & Jones, P. (2007). ‘Candidatus Phytoplasma lycopersici’, a phytoplasma associated with ‘hoja de perejil’ disease in Bolivia. Int J Syst Evol Microbiol 57, 1704–1710.

Bai, X. D., Zhang, J. H., Ewing, A., Miller, S. A., Radek, A. J., Schevchenko, D. V., Tsukerman, K., Walunas, T., Lapidus, A. & other authors (2006). Living with genome instability: the adaptation of phytoplasmas to diverse environments of their insect and plant hosts. J Bacteriol 188, 3682–3696.

Cai, H., Wei, W., Davis, R. E., Chen, H. & Zhao, Y. (2008). Genetic diversity among phytoplasmas infecting Opuntia species: virtual RFLP analysis identifies new subgroups in the peanut witches’-broom phytoplasma group. Int J Syst Evol Microbiol 58, 1448–1457.

Chiykowski, L. N. & Sinha, R. C. (1990). Differentiation of MLO diseases by means of symptomatology and vector transmission. Zentralbl Bakteriol Hyg Suppl 20, 280–287.

Chun, J., Lee, J. H., Jung, Y., Kim, M., Kim, S., Kim, B. K. & Lim, Y. W. (2007). EzTaxon: a web-based tool for the identification of prokaryotes based on 16S ribosomal RNA gene sequences. Int J Syst Evol Microbiol 57, 2259–2261.

Davis, R. E. & Sinclair, W. A. (1998). Phytoplasma identity and disease etiology. Phytopathology 88, 1372–1376.

Davis, R. E., Jomantiene, R., Kalvelyte, A. & Dally, E. L. (2003). Differential amplification of sequence heterogeneous ribosomal RNA genes and classification of the ‘Fragaria multicipita’ phytoplasma. Microbiol Res 158, 229–236.

Davis, R. E., Jomantiene, R. & Zhao, Y. (2005). Lineage-specific decay of folate biosynthesis genes suggests ongoing host adaptation in phytoplasmas. DNA Cell Biol 24, 832–840.

Euzéby, J. P. (1997). List of Bacterial Names with Standing in Nomenclature: a folder available on the Internet. Int J Syst Bacteriol 47, 590–592.

Firrao, G., Carraro, L., Gobbi, E. & Locci, R. (1996). Molecular characterization of a phytoplasma causing phyllody in clover and other herbaceous hosts in northern Italy. Eur J Plant Pathol 102, 817–822.

Firrao, G., Cibb, K. & Streten, C. (2005). Short taxonomic guide to the genus ‘Candidatus Phytoplasma’. J Plant Pathol 87, 249–263.

Granett, A. L. & Gilmer, R. M. (1971). Mycoplasma associated with X-disease in various Prunus species. Phytopathology 61, 1036–1037.

Gundersen, D. E. & Lee, I.-M. (1996). Ultrasensitive detection of phytoplasmas by nested-PCR assays using two universal primer pairs. Phytopathol Mediterr 35, 144–151.

Fig. 5. Key restriction enzymes that distinguish novel 16SrIII subgroup pattern types. Lanes MW, φX174 DNA digested with HaeIII.
Y. Zhao and others

Gundersen, D. E., Lee, I.-M., Rehner, S. A., Davis, R. E. & Kingsbury, D. T. (1994). Phylogeny of mycoplasmalike organisms (phytoplasmas): a basis for their classification. J Bacteriol 176, 5244–5254.

Hogenhout, S. A., Oshima, K., Ammar, el-D., Kakizawa, S., Kingdom, H. N. & Namba, S. (2008). Phytoplasmas: bacteria that manipulate plants and insects. Mol Plant Pathol 9, 403–423.

IRPCM Phytoplasma/Spiroplasma Working Team – Phytoplasma taxonomy group (2004). ‘Candidateis Phytoplasma’, a taxon for the wall-less, non-helical prokaryotes that colonize plant phloem and insects. Int J Syst Evol Microbiol 54, 1243–1255.

Jomantiene, R., Davis, R. E., Valiunas, D. & Alminaita, A. (2002). New group 16SrIII phytoplasma lineages in Lithuania exhibit interoperon sequence heterogeneity. Eur J Plant Pathol 108, 507–517.

Jung, H. Y., Sawayanagi, T., Kakizawa, S., Nishigawa, H., Miyata, S., Oshima, K., Ugaki, M., Lee, J. T., Hibí, T. & Namba, S. (2002). ‘Candidateis Phytoplasma castaneae’, a novel phytoplasma taxon associated with chestnut witches’-broom disease. Int J Syst Evol Microbiol 52, 1543–1549.

Kirkpatrick, B. C., Fisher, G. A., Fraser, J. D. & Purcell, A. H. (1990). Epidemiological and phylogenetic studies on western X-disease mycoplasmalike organisms. Zentralbl Bakteriol Hyg Suppl 20, 287–297.

Kube, M., Schneider, B., Kuhl, H., Dandekar, T., Heitmann, K., Migdoll, A. M., Reinhardt, R. & Seemu¨ ller, E. (2008). The linear chromosome of the plant-pathogenic mycoplasm ‘Candidateis Phytoplasma mali’. BMC Genomics 9, 306.

Lee, I.-M., Davis, R. E., Chen, T.-A., Chiykowski, L. N., Fletcher, J., Hiruki, C. & Schaff, D. A. (1992a). A genotype-based system for identification and classification of mycoplasmalike organisms (MLOs) in the aster yellows MLO strain cluster. Phytopathology 82, 977–986.

Lee, I.-M., Gundersen, D. E., Davis, R. E. & Chiykowski, L. N. (1992b). Identification and analysis of a genomic strain cluster of mycoplasmalike organisms associated with Canadian peach (eastern) X-disease, Western X-disease, and clover yellow edge. J Bacteriol 174, 6694–6698.

Lee, I.-M., Davis, R. E. & Hsu, H.-T. (1993a). Differentiation of strains in the aster yellows mycoplasmalike organism strain cluster by serological assays with monoclonal antibodies. Plant Dis 77, 815–817.

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

Lee, I.-M., Gundersen-Rindal, D. E., Davis, R. E. & Bartoszyk, I.-M. (1998). Revised classification scheme of phytoplasmas based on RFLP analysis of 16S rRNA and ribosomal protein gene sequences. Int J Syst Bacteriol 48, 1153–1169.

Lee, I.-M., Davis, R. E. & Gundersen-Rindal, D. E. (2000). Phytoplasma: phytopathogenic mollicutes. Annu Rev Microbiol 54, 221–255.

Lee, I.-M., Martini, M., Macone, C. & Zhu, S. F. (2004a). Classification of phytoplasma strains in the elm yellows group (16SIV) and proposal of ‘Candidateis Phytoplasma ulmi’ for the phytoplasma associated with elm yellows. Int J Syst Evol Microbiol 54, 337–347.

Lee, I.-M., Gundersen-Rindal, D. E., Davis, R. E., Bottner, K. D., Marcone, C. & Seemu¨ ller, E. (2004b). ‘Candidateis Phytoplasma asteris’, a novel phytoplasma taxon associated with aster yellows and related diseases. Int J Syst Evol Microbiol 54, 1037–1048.

Lee, I.-M., Bottner, K. D., Secor, G. & Rivera-Varas, V. (2006). ‘Candidateis Phytoplasma americanum’, a phytoplasma associated with a potato purple top wilt disease complex. Int J Syst Evol Microbiol 56, 1593–1597.

Liefting, L. W. & Kirkpatrick, B. C. (2003). Cosmid cloning and sample sequencing of the genome of the uncultivable mollicute, Western X-disease phytoplasma, using DNA purified by pulsed-field gel electrophoresis. FEMS Microbiol Lett 221, 203–211.

Liefting, L. W., Andersen, M. T., Beever, R. E., Gardner, R. C. & Foster, L. S. (1996). Sequence heterogeneity in the two 16S rRNA genes of Phormium yellow leaf phytoplasma. Appl Environ Microbiol 62, 3133–3139.

Macbath, J. H., Nyland, G. & Spurr, A. R. (1972). Morphology of mycoplasmalike bodies associated with peach X-disease in Prunus persica. Phytopathology 62, 935–937.

Marcone, C., Neimark, H., Ragozzino, A., Lauer, U. & Seemu¨ ller, E. (1999). Chromosome sizes of phytoplasmas composing major phylogenetic groups and subgroups. Phytopathology 89, 805–810.

Martini, M., Lee, I.-M., Bottner, K. D., Zhao, Y., Botti, S., Bertaccini, A., Harrison, N. A., Carraro, L., Marcone, C. & Osler, R. (2007). Ribosomal protein gene-based phylogeny for finer differentiation and classification of phytoplasmas. Int J Syst Evol Microbiol 57, 2057–2051.

McCoy, R. E., Caudwell, A., Chang, C. J., Chen, T. A., Chiykowski, L. N., Cousin, M. T., Dale, J. L., de Leeuw, G. T. N., Golino, D. A. & other authors (1989). Plant diseases associated with mycoplasmalike organisms. In The Mycoplasmas, vol. 5, pp. 545–560. Edited by R. F. Whitcomb & J. G. Tully. New York: Academic Press.

Murray, R. G. E. & Schleifer, K. H. (1994). Taxonomic notes: a proposal for recording the properties of putative taxa of prokaryotes. Int J Syst Bacteriol 44, 174–176.

Myers, E. W. & Miller, W. (1988). Optimal alignments in linear space. Comput Appl Biosci 4, 11–17.

Oshima, K., Kakizawa, S., Nishigawa, H., Jung, H.-Y., Wei, W., Suzuki, S., Arashida, R., Nakata, D., Miyata, S. & other authors (2004). Reductive evolution suggested from the complete genome sequence of a plant-pathogenic phytoplasma. Nat Genet 36, 27–29.

Quaglino, F., Zhao, Y., Bianco, P. A., Wei, W., Casati, P., Durante, G. & Davis, R. E. (2009). New 16Sr subgroups and distinct SNP lineages among grapevine Bois noir phytoplasma populations. Ann Appl Biol 154, 279–289.

Roselló-Mora, R. & Amann, R. (2001). The species concept for prokaryotes. FEMS Microbiol Rev 25, 39–67.

Sears, B. B. & Kirkpatrick, B. C. (1994). Unveiling the evolutionary relationships of plant pathogenic mycoplasmalike organisms. ASM News 60, 307–312.

Seemu¨ ller, E., Garnier, M. & Schneider, B. (2002). Mycoplasmas of plants and insects. In Molecular Biology and Pathogenicity of Mycoplasmas, pp. 91–115. Edited by S. Razin & R. Herrmann. New York: Kluwer Academic/Plenum.

Sinha, R. C. & Chiykowski, L. N. (1980). Transmission and morphological features of mycoplasmalike bodies associated with peach X-disease. Can J Plant Pathol 2, 119–124.

Stackebrandt, E. & Goebel, B. M. (1994). Taxonomic note: a place for DNA-DNA reassociation and 16S rRNA sequence analysis in the present species definition in bacteriology. Int J Syst Bacteriol 44, 846–849.

Stoddard, E. M. (1934). Progress report of investigations on a new peach trouble. Conn Pomol Soc Proc 43, 115–117.

Tamura, K., Dudley, J., Nei, M. & Kumar, S. (2007). MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. Mol Biol Evol 24, 1596–1599.

Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22, 4673–4680.
Tran-Nguyen, L. T., Kube, M., Schneider, B., Reinhardt, R. & Gibb, K. S. (2008). Comparative genome analysis of ‘*Candidatus Phytoplasma australiense*’ (subgroup tuf-Australia I; rp-A) and ‘*Ca. Phytoplasma asteris*’ strains OY-M and AY-WB. *J Bacteriol* **190**, 3979–3991.

Valiunas, D., Staniulis, J. & Davis, R. E. (2006). ‘*Candidatus Phytoplasma fragariae*’, a novel phytoplasma taxon discovered in yellows diseased strawberry, *Fragaria × ananassa*. *Int J Syst Evol Microbiol* **56**, 277–281.

Wei, W., Davis, R. E., Lee, I.-M. & Zhao, Y. (2007). Computer-simulated RFLP analysis of 16S rRNA genes: identification of ten new phytoplasma groups. *Int J Syst Evol Microbiol* **57**, 1855–1867.

Wei, W., Davis, R. E., Jomantiene, R. & Zhao, Y. (2008a). Ancient, recurrent phage attacks and recombination events shaped dynamic sequence-variable mosaic structures at the root of phytoplasma genome evolution. *Proc Natl Acad Sci U S A* **105**, 11827–11832.

Wei, W., Lee, I.-M., Davis, R. E., Suo, X. & Zhao, Y. (2008b). Automated RFLP pattern comparison and similarity coefficient calculation for rapid delineation of new and distinct phytoplasma 16Sr subgroup lineages. *Int J Syst Evol Microbiol* **58**, 2368–2377.

Weisburg, W. G., Tully, J. G., Rose, D. L., Petzel, J. P., Oyaizu, H., Yang, D., Mandelco, L., Sechrest, J., Lawrence, T. G. & other authors (1989). A phylogenetic analysis of the mycoplasmas: basis for their classification. *J Bacteriol* **171**, 6455–6467.

Wheeler, D. L., Barrett, T., Benson, D. A., Bryant, S. H., Canese, K., Church, D. M., DiCuccio, M., Edgar, R., Federhen, S. & other authors (2005). Database resources of the National Center for Biotechnology Information: update. *Nucleic Acids Res* **33**, D39–D45.

Zhao, Y., Davis, R. E. & Lee, I. M. (2005). Phylogenetic positions of ‘*Candidatus Phytoplasma asteris*’ and *Spiroplasma kunkelii* as inferred from multiple sets of concatenated core housekeeping proteins. *Int J Syst Evol Microbiol* **55**, 2131–2141.