Plant Disease • 2019 • 103:2494-2497 • https://doi.org/10.1094/PDIS-03-19-0454-A

RESOURCE ANNOUNCEMENT

The CpnClassiPhyR Is a Resource for *cpn*60 Universal Target-Based Classification of Phytoplasmas

Kevin Muirhead,[1] Edel Pérez-López,[2] Brian W. Bahder,[3] Janet E. Hill,[4] and Tim Dumonceaux[4,5,†]

[1] 1111 - 604 East Lake Blvd. NE, Airdrie, Alberta, Canada
[2] Department of Biology, University of Saskatchewan, Saskatoon, Saskatchewan, Canada
[3] University of Florida Fort Lauderdale Research and Education Center
[4] Department of Veterinary Microbiology, University of Saskatchewan, Saskatoon, Saskatchewan, Canada
[5] Agriculture and Agri-Food Canada, Saskatoon Research and Development Centre, Saskatoon, Saskatchewan, Canada

Abstract

Phytoplasmas are plant-pathogenic bacteria that are associated with yield losses in many crop plants worldwide. Phytoplasma strain differentiation is accomplished using *in silico* restriction fragment length polymorphism (RFLP) analysis of 16S ribosomal RNA-encoding gene sequences, which has resulted in the definition of ribosomal groups and subgroups of phytoplasmas. Due to limitations associated with this approach, a complementary classification scheme was recently developed based on RFLP analysis of the single-copy, protein-encoding gene chaperonin-60 (*cpn*60). We present the CpnClassiPhyR, software that facilitates phytoplasma strain classification using both RFLP and automated phylogenetic analysis of *cpn*60 sequences. This software is available through a web interface at http://cpnclassiphyr.ca.

Phytoplasmas ('*Candidatus* Phytoplasma' spp.) are plant- and insect-associated bacteria that have proven difficult to cultivate in axenic cultures (Hogenhout et al. 2008). Phytoplasmas are transmitted to plants via insect vectors and are associated with developmental abnormalities in infected plants that lead to altered inflorescence morphology, stunting, witches'-broom, yellowing, plant decline, and the formation of seeds and fruits with greatly reduced commercial value (Maejima et al. 2014). For this reason, phytoplasma detection, classification, and taxonomy are important to facilitate strain tracking and surveillance in cultivated and wild plants and within insect populations. Phytoplasma classification follows the criteria specified for uncultured microorganisms (IRPCM 2004), and more than 40 species have been identified within the genus '*Candidatus* Phytoplasma' (Miyazaki et al. 2018; Naderali et al. 2017). The detection and identification of phytoplasmas has been facilitated by the availability of PCR primers that amplify a portion of the 16S rRNA-encoding gene from any phytoplasma, using a nested PCR strategy that results in an amplicon of ~1.2 kb defined by the annealing sites of primers R16F2n/R16R2 (known as F2nR2) (Gundersen and Lee 1996).

In silico restriction fragment length polymorphism (RFLP) analysis of 16S-based F2nR2 sequences (Bertaccini et al. 2018) or protein-encoding genes (Martini et al. 2019) is a currently used, universally accepted, and widely applied method for typing phytoplasmas (Zhao et al. 2013). While 16S-based RFLP analysis is widely used for classifying phytoplasmas, concerns related to the biological meaning of mutations that result in changes to restriction patterns remain. Moreover, standards supporting the required quality of DNA sequences used for *in silico* RFLP analysis, along with requirements to verify such patterns

[†]Corresponding author: Tim J. Dumonceaux; tim.dumonceaux@canada.ca

The author(s) declare no conflict of interest.

Accepted for publication 17 April 2019.

Funding
This work was supported by Agriculture and Agri-Food Canada A-base Project 1668 - Sustainable Small Fruit Production. Establishment of this tool was enabled in part by support provided by the University of Saskatchewan (https://www.usask.ca/), Westgrid (https://www.westgrid.ca/), and Compute Canada (https://www.computecanada.ca).

Keywords
Prokaryotes, pathogen detection, pathogen diversity

with in vitro RFLP analysis, are suggested (https://plantpathology.ba.ars.usda.gov/cgi-bin/resource/iphyclassifier.cgi) but are rarely done, which may be a contributing factor in the noted proliferation of RFLP-based phytoplasma subgroups (Zhao and Davis 2016). Nevertheless, a key advantage of *in silico* RFLP analysis is that it facilitates the calculation of a similarity coefficient, or F-value, based on the number of bands generated by a specified set of restriction enzymes, as follows:

$$F = 2Nxy/(Nx + Ny)$$

where *Nx* and *Ny* correspond to the number of bands generated for a given restriction enzyme for strains x and y; and *Nxy* is the number of bands generated by that restriction enzyme that are common to strains x and y (Pérez-López et al. 2016a; Zhao et al. 2013). Pairwise determination of F-values based on RFLP analysis of 17 restriction sites within 16S rRNA-encoding gene sequences has resulted in the definition of >35 ribosomal groups, and many more subgroups, of phytoplasmas (Pérez-López et al. 2016b). The development and public availability of software for automatically calculating F2nR2 RFLP-based similarity coefficients, the *i*PhyClassifier (Zhao et al. 2009, 2013), has greatly advanced phytoplasma classification and taxonomy. However, the exclusive use of F2nR2 sequences for phytoplasma identification has limitations related to their ability to discern closely related strains, and complications arising from the fact that the two copies of the 16S rRNA-encoding gene found in phytoplasma genomes can provide conflicting typing results (Lee et al. 2010; Martini et al. 2007, 2019). Therefore, single-copy, protein-encoding genes have found increasing application in phytoplasma classification (Lee et al. 2010; Martini et al. 2007, 2019).

The gene encoding the 60 kDa chaperonin protein (chaperonin-60, *cpn*60) has been extensively used as a taxonomic marker for bacteria, and provides improved strain resolution compared with other markers, including 16S rRNA-encoding genes (Tian et al. 2016). Moreover, *cpn*60 sequences can be used to predict bacterial whole genome pairwise sequence similarities (Verbeke et al. 2011), and meet the criteria for a molecular barcode for bacteria (Links et al. 2012). PCR primers that amplify the *cpn*60 "universal target" (*cpn*60 UT) from phytoplasmas have been described, which provide access to *cpn*60 genes from a wide variety of phytoplasma species and ribosomal groups (Dumonceaux et al. 2014). Recently, a complementary phytoplasma classification system using *in silico* RFLP analysis and F-value calculation using *cpn*60 UT sequences was developed and validated (Pérez-López et al. 2016a). This scheme uses seven restriction enzymes, *Alu*I, *Bfa*I, *Hin*fI, *Hpa*I, *Mse*I, *Rsa*I, and *Taq*I, and requires manual strain-to-strain pairwise similarity coefficient determination, which can be tedious and error-prone. To facilitate the broader application of this classification scheme, we have developed CpnClassiPhyR, which accepts *cpn*60 sequences (Dumonceaux et al. 2014) in protein-coding or reverse complement orientations, with or without flanking sequences, and automatically trims the sequences to the *cpn*60 UT and calculates similarity coefficients compared with a reference database of *cpn*60 sequences. The software performs *in silico* RFLP analysis, determines the most closely related phytoplasma strain, and provides a downloadable virtual gel illustrating the results. In addition, in view of the known limitations of RFLP analysis, the site has been designed to analyze entire *cpn*60 UT sequences by performing automated phylogenetic analysis (using the neighbor-joining algorithm) in the context of the known diversity of phytoplasma *cpn*60 sequences. The site also provides all of the phytoplasma *cpn*60 reference sequences required for phylogenetic analysis to be performed offline by the user if preferred. At present, the reference dataset of validated, nonredundant *cpn*60 sequences consists of 30 sequences corresponding to 12 ribosomal groups, 25 ribosomal subgroups, and 15 '*Ca.* Phytoplasma' species. This database will expand to accommodate new *cpn*60 sequences as they are reported in the literature. New *cpn*60 sequences will only be added to the reference dataset when corresponding F2nR2 sequence information is available from the same strain, firmly rooting the *cpn*60 classification scheme within the existing 16S rRNA-based classification system.

To demonstrate the utility of the software for classifying phytoplasmas, we extracted the *cpn*60 sequence (full length) from the genome of '*Ca.* Phytoplasma' mali strain AT (GenBank accession NC_011047). With this sequence as input, the CpnClassiPhyR calculated the similarity coefficients of this sequence compared with all of the reference sequences. These similarity coefficients were compared with pairwise calculations determined manually,

demonstrating that the CpnClassiPhyR determined these values correctly (Table 1). Moreover, the CpnClassiPhyR performed an automated phylogenetic analysis of the input sequence compared with phytoplasma *cpn*60 reference sequences (not shown). These calculations identified strain AT as a group X phytoplasma, subgroup X-IA, consistent with the known classification of this strain as 16SrX-A using RFLP analysis of 16S rRNA-encoding gene sequences (Kube et al. 2008). F-values calculated by the CpnClassiPhyR for other groups and subgroups of phytoplasmas demonstrate the ability of the classification scheme to identify phytoplasmas in 16Sr groups I, X, and XIV (Table 1). The subgroups of '*Ca.* P. cynodontis' (16SrXIV) are not easily distinguished by their 16S rRNA RFLP profiles (Mitrović et al. 2015) yet possess much greater sequence variation using *cpn*60-based sequence analysis (Table 1).

A noted issue related to the use of 16S-encoding gene sequences for classifying phytoplasmas is the proliferation of novel subgroups and the need for a platform for registering new groups and subgroups (Zhao and Davis 2016). To address this need in the case of *cpn*60 UT sequences, we added a functionality for registering new groups and subgroups of phytoplasmas. To demonstrate this functionality, we examined the reported full-length *cpn*60 sequence from 'Cocos nucifera' phytoplasma (GenBank KY779619). This sequence is reportedly derived from a 16SrIV-A phytoplasma strain (Córdova et al. 2019). Analysis of this sequence using the CpnClassiPhyR identified this *cpn*60 sequence as a group IV phytoplasma derived from a potentially novel subgroup (Table 1). The CpnClassiPhyR also automatically trimmed the full-length sequence to the 552 bp UT and performed a phylogenetic analysis of the sequence, providing results that place the 'Cocos nucifera' phytoplasma in a

Table 1. F-values calculated manually or using the CpnClassiPhyR with different groups and subgroups of phytoplasmas. Detailed information on each strain, including accession numbers for the *cpn60* UT and 16S (F2nR2) sequences is available at cpnclassiphyr.ca. The closest matches to the RFLP pattern of a reference strain, and the F-values determined for the corresponding pairwise comparisons, are shown in bold. The complete matrix of F-values comparing all strains to all strains is also available at cpnclassiphyr.ca.

Phytoplasma query strain: phytoplasma reference strain	'*Ca.* P. mali' AT (16SrX-A)		'*Ca.* P. pyri' (16SrX-C)	'*Ca.* P. prunorum' (16SrX-F)	'*Ca.* P. asteris' O2L (16SrI-A)	'*Ca.* P. cynodontis' AL85/11 (16SrXIV-A)	'*Ca.* P. cynodontis' RS59/11 (16SrXIV-C)	'Cocos nucifera' phytoplasma (16SrIV-A)
	Classiphyr result	Manual calculation	Classiphyr result	Classiphyr result	Classiphyr result	Classiphyr result	Classiphyr result	Classiphyr result
cpn60 UT group-subgroup								
I-I(E/AI)AI	0.23	0.23	0.28	0.32	0.79	0.04	0.04	0.26
I-IA	0.14	0.14	0.20	0.24	**1.00**	0.00	0.00	0.18
I-IB	0.14	0.14	0.19	0.23	0.88	0.00	0.00	0.17
I-IC	0.24	0.24	0.34	0.33	0.74	0.04	0.04	0.27
I-IF	0.23	0.23	0.29	0.33	0.81	0.04	0.04	0.22
I-IIA	0.14	0.14	0.14	0.19	0.68	0.00	0.00	0.09
I-IIB	0.14	0.14	0.19	0.23	0.72	0.00	0.00	0.13
I-IIIB	0.14	0.14	0.19	0.23	0.75	0.00	0.00	0.13
I-IP	0.29	0.29	0.30	0.39	0.71	0.04	0.04	0.33
I-IVB	0.23	0.23	0.28	0.32	0.75	0.04	0.04	0.22
I-VB	0.14	0.14	0.19	0.23	0.85	0.00	0.00	0.18
II-IA	0.14	0.14	0.14	0.14	0.04	0.17	0.17	0.04
II-IC	0.12	0.12	0.13	0.17	0.04	0.19	0.19	0.04
IV-IC	0.37	0.37	0.43	0.42	0.21	0.13	0.09	0.4
IV-IE	0.30	0.30	0.31	0.45	0.14	0.18	0.18	**0.86**
IX-IA	0.26	0.26	0.26	0.31	0.28	0.05	0.05	0.29
IX-IB	0.10	0.10	0.20	0.15	0.18	0.04	0.00	0.14
IX-IJ	0.20	0.20	0.26	0.25	0.27	0.05	0.05	0.24
V-IC	0.25	0.25	0.21	0.29	0.15	0.19	0.15	0.24
VII-IA	0.23	0.23	0.24	0.33	0.21	0.17	0.17	0.27
X-IA	**1.00**	**1.00**	0.65	0.74	0.14	0.19	0.14	0.3
X-IC	0.65	0.65	**1.00**	0.65	0.20	0.15	0.15	0.31
X-IF	0.74	0.74	0.65	**1.00**	0.24	0.19	0.14	0.45
XI-ID	0.28	0.28	0.33	0.42	0.21	0.30	0.34	0.36
XII-I(B/C)B	0.29	0.29	0.20	0.43	0.30	0.09	0.04	0.32
XII-IA	0.15	0.15	0.15	0.20	0.13	0.09	0.00	0.19
XIII-(A/I)I	0.29	0.29	0.26	0.38	0.23	0.15	0.08	0.24
XIV-IA	0.19	0.19	0.15	0.19	0.00	**1.00**	0.87	0.18
XIV-IC	0.14	0.14	0.15	0.14	0.00	0.87	**1.00**	0.18
XV-IB	0.09	0.09	0.09	0.09	0.08	0.24	0.24	0.08

clade with other group IV *cpn*60 UT sequences (not shown). Since the 16Sr group of this strain is known, this *cpn*60 UT sequence can be used to register a novel subgroup of a group IV phytoplasma directly at the website (*cpn*60 UT group IV-IA). The registration service requires information on the 16Sr group and subgroup of the strain along with other metadata and also requires valid accession numbers for the 16S and *cpn*60 sequences. The service then automatically retrieves the sequences, performs a complete analysis of the *cpn*60 UT sequence, and adds the novel sequence to the reference database dynamically.

A full description and characterization of the CpnClassiPhyR, including the analysis of multiple groups and subgroups and a comparison with results obtained using 16S rRNA-encoding genes, has been reported (Muirhead et al. 2019). The routine determination and phylogenetic/RFLP analysis of 16S rRNA and *cpn*60 sequences from phytoplasmas will enhance strain tracking and phytoplasma surveillance efforts worldwide.

Data availability

The code that comprises the CpnClassiPhyR is available as a github link at https://github.com/kevmu/CpnClassiPhyR. The web interface is available at http://cpnclassiphyr.ca.

Literature Cited

Bertaccini, A., Paltrinieri, S., and Contaldo, N. 2018. Standard detection protocol: PCR and RFLP analyses based on 16S rRNA gene. Pages 83-95 in: Methods Mol. Biol.

Córdova, I., Oropeza, C., Harrison, N., Ku-Rodríguez, S., Puch-Hau, C., Narváez, M., and Sáenz, L. 2019. Simultaneous detection of coconut lethal yellowing phytoplasmas (group 16SrIV) by real-time PCR assays using 16Sr- and GroEL-based TaqMan probes. Pages 1-11 in: J. Plant Pathol.

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

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

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

IRPCM. 2004. '*Candidatus* Phytoplasma', a taxon for the wall-less, non-helical prokaryotes that colonize plant phloem and insects. Int. J. Syst. Evol. Microbiol. 54:1243-1255.

Kube, M., Schneider, B., Kuhl, H., Dandekar, T., Heitmann, K., Migdoll, A. M., Reinhardt, R., and Seemuller, E. 2008. The linear chromosome of the plant-pathogenic mycoplasma '*Candidatus* Phytoplasma mali'. BMC Genomics 9:306.

Lee, I. M., Bottner-Parker, K. D., Zhao, Y., Davis, R. E., and Harrison, N. A. 2010. Phylogenetic analysis and delineation of phytoplasmas based on *secY* gene sequences. Int. J. Syst. Evol. Microbiol. 60:2887-2897.

Links, M. G., Dumonceaux, T. J., Hemmingsen, S. M., and Hill, J. E. 2012. The chaperonin-60 universal target is a barcode for bacteria that enables *de novo* assembly of metagenomic sequence data. PLoS One 7:e49755.

Maejima, K., Oshima, K., and Namba, S. 2014. Exploring the phytoplasmas, plant pathogenic bacteria. J. Gen. Plant Pathol. 80:210-221.

Martini, M., Bottner-Parker, K. D., and Lee, I. M. 2019. PCR-based sequence analysis on multiple genes other than 16s rRNA gene for differentiation of phytoplasmas. Pages 97-115 in: Methods Mol. Biol.

Martini, M., Lee, I. M., Bottner, K. D., Zhao, Y., Botti, S., Bertaccini, A., Harrison, N. A., Carraro, L., Marcone, C., Khan, A. J., and Osler, R. 2007. Ribosomal protein gene-based phylogeny for finer differentiation and classification of phytoplasmas. Int. J. Syst. Evol. Microbiol. 57:2037-2051.

Mitrović, J., Smiljković, M., Seemüller, E., Reinhardt, R., Hüttel, B., Büttner, C., Bertaccini, A., Kube, M., and Duduk, B. 2015. Differentiation of '*Candidatus* Phytoplasma cynodontis' based on 16S rRNA and groEL genes and identification of a new subgroup, 16SrXIV-C. Plant Dis. 99:1578-1583.

Miyazaki, A., Shigaki, T., Koinuma, H., Iwabuchi, N., Rauka, G. B., Kembu, A., Saul, J., Watanabe, K., Nijo, T., Maejima, K., Yamaji, Y., and Namba, S. 2018. '*Candidatus* Phytoplasma noviguineense', a novel taxon associated with Bogia coconut syndrome and banana wilt disease on the island of New Guinea. Int. J. Syst. Evol. Microbiol. 68:170-175.

Muirhead, K., Pérez-López, E., Bahder, B. W., Hill, J. E., and Dumonceaux, T. 2019. The CpnClassiPhyR facilitates phytoplasma classification and taxonomy using cpn60 universal target sequences. In: Sustainable Management of Phytoplasma Diseases in Crops Grown in the Tropical Belt. C. Olivier, T. Dumonceaux, and E. Perez-Lopez, eds. In press.

Naderali, N., Nejat, N., Vadamalai, G., Davis, R. E., Wei, W., Harrison, N. A., Kong, L., Kadir, J., Tan, Y.-H., and Zhao, Y. 2017. '*Candidatus* Phytoplasma wodyetiae', a new taxon associated with yellow decline disease of foxtail palm (*Wodyetia bifurcata*) in Malaysia. Int. J. Syst. Evol. Microbiol. 67:3765-3772.

Pérez-López, E., Luna-Rodríguez, M., Olivier, C. Y., and Dumonceaux, T. J. 2016b. The underestimated diversity of phytoplasmas in Latin America. Int. J. Syst. Evol. Microbiol. 66:492-513.

Pérez-López, E., Olivier, C. Y., Luna-Rodríguez, M., and Dumonceaux, T. J. 2016a. Phytoplasma classification and phylogeny based on *in silico* and *in vitro* RFLP analysis of *cpn60* universal target sequences. Int. J. Syst. Evol. Microbiol. 66:5600-5613.

Tian, Q., Zhao, W., Lu, S., Zhu, S., and Li, S. 2016. DNA barcoding for efficient species- and pathovar-level identification of the quarantine plant pathogen *Xanthomonas*. PLoS One 11:e0165995.

Verbeke, T. J., Sparling, R., Hill, J. E., Links, M. G., Levin, D., and Dumonceaux, T. J. 2011. Predicting relatedness of bacterial genomes using the chaperonin-60 universal target (*cpn60* UT): Application to *Thermoanaerobacter* species. Syst. Appl. Microbiol. 34:171-179.

Zhao, Y., and Davis, R. E. 2016. Criteria for phytoplasma 16Sr group/subgroup delineation and the need of a platform for proper registration of new groups and subgroups. Int. J. Syst. Evol. Microbiol. 66:2121-2123.

Zhao, Y., Wei, W., Lee, I.-M., Shao, J., Suo, X., and Davis, R. E. 2009. Construction of an interactive online phytoplasma classification tool, iPhyClassifier, and its application in analysis of the peach X-disease phytoplasma group (16SrIII). Int. J. Syst. Evol. Microbiol. 59:2582-2593.

Zhao, Y., Wei, W., Lee, I. M., Shao, J., Suo, X., and Davis, R. E. 2013. The iphyclassifier, an interactive online tool for phytoplasma classification and taxonomic assignment. Pages 329-338 in: Methods Mol. Biol.