PHYLOGENY AND CHARACTERIZATION OF PHYTOPLASMAL NusA AND USE OF THE *nusA* GENE IN DETECTION OF GROUP 16SrI STRAINS

J. Shao[1], R. Jomantiene[1,2], E.L. Dally[1], Y. Zhao[1], I.-M. Lee[1], D.L. Nuss[3], and R.E. Davis[1]

[1] *Molecular Plant Pathology Laboratory, USDA-Agricultural Research Service, Beltsville, MD, USA 20705*
[2] *Fitovirus Laboratory, Institute of Botany, Zaliuju ezeru 49, Vilnius LT-08406, Lithuania*
[3] *Center for Biosystems Research, University of Maryland Biotechnology Institute, College Park, MD USA 20742*

SUMMARY

Phytoplasmas are cell wall-less prokaryotes that exist as obligate parasites and pathogens of insect vectors and plants. In their descent from walled ancestors in the *Bacillus/Clostridium* group, phytoplasmas evolved some of the smallest known bacterial genomes. In this study, we cloned and sequenced *nusA* transcription factor gene sequences from clover phyllody (CPh) and other phytoplasmas and from *Acholeplasma palmae*, a non-phytopathogenic wall-less bacterium.

The CPh *nusA* gene was flanked at its 5'-end by a hypothetical protein gene and *smpB* (small protein B), and at its 3'-end by a hypothetical protein gene that may be coordinately regulated with *nusA* and *infB* (translation initiation factor). The predicted 357-aa NusA protein of CPh phytoplasma was significantly smaller than those of *Mycoplasma* spp. and similar in size to NusA of *Clostridium* spp., *Bacillus* spp., and *A. palmae.* A phylogenetic tree based on NusA proteins indicated that phytoplasmal and acholeplasmal *nusA* genes diverged from a common ancestor. Amplification and RFLP analysis of *nusA* gene sequences, and phylogenetic analysis of NusA proteins indicated that use of *nusA* may assist in detection and differentiation of distinct lineages within group 16SrI, '*Candidatus* Phytoplasma asteris'-related strains.

Key words: Mollicutes, reductive evolution, yellows disease, phytoplasma, *nusA.*

INTRODUCTION

Cell wall-less prokaryotes that are known as phytoplasmas are obligate parasites and pathogens of insects and plants. In their descent from walled, low G+C Gram positive (*Firmicutes*) bacterial ancestors in the *Bacillus/Clostridium* group, phytoplasmas underwent massive reduction in genome size. Since they possess some of the smallest cellular genomes known in bacteria, ranging from ca. 0.5 to 1.1 Mb (Lim and Sears, 1991; Neimark and Kirkpatrick, 1993; Marcone *et al*., 1999), phytoplasmas appear to represent an evolutionary trend toward minimal sets of genes required for insect transmission and plant pathogenicity (Razin *et al*., 1998; Lee *et al*., 2000; Oshima *et al*., 2004; Davis *et al*., 2003, 2005). Disease symptoms induced in infected plants include the development of leafy structures in place of floral parts (phyllody), small flowers or loss of normal flower color and development of green petals (virescence), reduced size of leaves, sterility of flowers, yellowing and general stunting of plants, decline and death of branches in woody hosts, and abnormal proliferation of axillary buds to produce witches' broom-like growths (Lee *et al*., 2000). The molecular bases for such striking alterations of normal plant morphogenesis are not understood, and the reasons why phytoplasmas cannot be isolated and cultured *in vitro* have not been explained (Lee and Davis, 1986; Oshima *et al*., 2004).

Because phytoplasmas cannot be isolated and characterized in pure culture, molecular criteria have been adopted for classification, identification, and taxonomy of phytoplasmas.

Phylogenetic studies based on 16S rRNA gene sequences have indicated that phytoplasmas form a monophyletic clade that descended from *Acholeplasma*-like ancestors within class *Mollicutes* (Gundersen *et al*., 1994). The increasing availability of gene sequences from diverse phytoplasmas makes it possible to trace the evolutionary histories of individual genes and affords new opportunities for gaining insights into mutations, possible horizontal transfer, and recombinational events that resulted in the structure of extant phytoplasmal genomes.

In this study, we analyzed a cloned 4.6 kbp genomic DNA fragment from clover phyllody (CPh) phytoplasma containing five potential open reading frames (ORFs), including one encoding a potential 357-amino acid protein, NusA. The NusA protein is a transcription factor involved in elongation, transcription termination and anti-termination in prokaryotes (Greenblatt and Li, 1981a, 1981b; Greenblatt *et al*., 1981). To study the evolution of the NusA protein, we determined partial *nusA* gene se-

Corresponding author: R. Jomantiene
Fax: +301.504.5449
E-mail: jomantir@ba.ars.usda.gov

quences from four additional phytoplasma strains, sequenced the *nusA* gene from *Acholeplasma palmae*, and carried out comparative and phylogenetic analyses of NusA from the phytoplasmas, *A. palmae*, and other members of phylum *Firmicutes*. To assess *nusA* as a tool for detection and differentiation of phytoplasma lineages within group 16SrI, '*Candidatus* Phytoplasma asteris'-related strains, we carried out amplification of *nusA* gene sequences in the polymerase chain reaction (PCR), restriction fragment length polymorphism (RFLP) analysis of PCR products, and phylogenetic analysis of phytoplasmal NusA amino acid sequences.

MATERIALS AND METHODS

Phytoplasmas. '*Candidatus* Phytoplasma asteris'-related strains were as follows: CPh (clover phyllody phytoplasma), a member of phytoplasma 16S rDNA RFLP group 16SrI, subgroup 16SrI-C; BB (tomato big bud) and AY-WB (aster yellows witches' broom), members of subgroup 16SrI-A; AY (AY1, Maryland aster yellows), CY (chrysanthemum yellows), OY (OY-M onion yellows), and MBS (maize bushy stunt), members of subgroup 16SrI-B; PaWB (paulownia witches' broom), a member of subgroup 16SrI-D; and BBS3 (blueberry stunt), a member of subgroup 16SrI-E. Strains CPh, BB AY, CY, and PaWB were propagated in host plants of *Catharanthus roseus* L. (G. Don) (Madagascar periwinkle) that were maintained in an insect proof greenhouse. Strain MBS in corn (*Zea mays* L.) was kindly supplied by L.R. Nault. Strain BBS3 was from blueberry (Lee *et al*., 2004). Sources of sequence data for strains AY-WB and OY are given in Table 1. Group and subgroup classification are *sensu* Lee *et al*. (2004).

CPh phytoplasma DNA cloning and sequencing. CPh phytoplasma genomic DNA was extracted (Lee and Davis, 1988) from phytoplasma-enriched preparations of host plant sieve cells (Lee and Davis, 1983) and cloned as previously described (Davis *et al*., 2003). Automated DNA sequencing and primer walking were used to achieve at least 3-fold coverage per base position in sequencing both strands of a 4.6 kbp cloned DNA insert (CPh163) containing the *nusA* gene. Nucleotide sequence data were assembled using the SeqMan Option of the sequence analysis software suite Lasergene (DNASTAR, Madison, WI, USA).

Acholeplasma palmae. *A. palmae* strain J233 (Gundersen *et al*., 1994) was kindly provided by K.J. Hackett, USDA-Agricultural Research Service, Beltsville, MD, and was grown to late log phase in liquid medium LD59M incubated at 30ºC (Lee and Davis, 1989). As part of a project to survey the genome sequence of *A. palmae*, a genomic library was constructed of *A. palmae* DNA extracted from pure cultures. The extracted DNA was digested with *Tsp*509I and, using pUC19 vector, cloned by standard methods in *E. coli* TOP10 cells (INVITROGEN, Carlsbad, CA, USA). Automated sequencing of cloned DNA and sequence assembly followed by BLAST searches (Altschul *et al*., 1990) identified a contig containing the entire *nusA* gene of *A. palmae*.

Polymerase chain reactions (PCRs). For the detection of *nusA* gene sequences in diverse phytoplasmas, total DNA was extracted from tissues of phytoplasma-infected plants by the method of Lee and Davis (1988) and used as template in PCRs primed by primer pair PhNusF1/PhNusR5 to yield a 0.8 kbp *nusA* gene amplicon, or by primer pair PhNusF2/PhNusR5, designed to yield a 1.0 bp *nusA* gene amplicon. DNA extracted from healthy plants was used as control. Nucleotide sequences of the primers, designed on the basis of the *nusA* sequences from strains CPh and/or OY-M and AY-WB (Table 1), were: PhNusF1, 5'-aga gtc aac tcc cat tac tc-3'; PhNusR5, 5'-tct att tga gct tga tct aag g-3'; PhNusF2, 5'-aag acc aag tta ttg atg c-3'. The PCR was carried out under previously described conditions (Davis *et al*., 2003). DNA fragment size standard was 1 Kb DNA Ladder (Invitrogen, Carlsbad, CA, USA). PCR products were cloned in *E. coli* using the TOPO-TA Cloning Kit (Invitrogen, Carlsbad, CA, USA), and both strands of the cloned inserts were sequenced.

RFLP analysis of amplified phytoplasma DNA. Products from PCR primed by PhNusF2/PhNusR5 were analyzed by single enzyme digestion with *Mse*I and *Tsp*509I according to manufacturer's instructions (New England Biolabs, Beverly, MA, USA). The RFLP profiles of digested DNA were analyzed by electrophoresis through 12% polyacrylamide gel, staining with ethidium bromide, and visualization using a UV transilluminator. DNA fragment size standard was ϕX174 RFI DNA *Hae*III digest (Life Technologies, Gaithersburg, MD, USA).

Southern DNA hybridizations. For Southern blot analysis, fragments resulting from digestion of CPh phytoplasma genomic DNA with *Hind*III were separated by electrophoresis through a 0.8% agarose gel and transferred to a nylon membrane. For production of a hybridization probe, recombinant plasmid DNA was extracted from clone CPh163 using the BioRad Miniprep Kit (Biorad Laboratories, Hercules, CA, USA) and used as template in PCR primed by primer pair PhNusF1/PhNusR5. The membrane was probed with the digoxigenin (DIG)-labeled 0.8 kbp DNA PCR product under conditions previously described (Davis *et al*., 2003). Molecular weight marker was DNA Molecular Weight Marker III, DIG-labeled (Roche Applied Science, Indianapolis, IN, USA).

Nucleotide and amino acid sequence analyses. Sequence analysis was done using the sequence analysis software suite Lasergene (DNAstar). Nucleotide sequence assembly was done using the SeqMan program of Lasergene. *In silico* restriction mapping, sequence alignments, and analysis of DNA statistics were done using the Lasergene MapDraw, MegAlign, and EditSeq programs, respectively. BLAST searches (Altschul *et al*., 1990) were carried out at the National Center for Biotechnology Information (NCBI) web site at http://www.ncbi.nlm.nih.

Table 1. Microorganisms and the GenBank accession numbers of their NusA proteins or *nusA* genes, in parentheses ( ), and/or 16S rRNA genes, in brackets [ ], used in this work.

Acholeplasma palmae (DQ164210) [L33734]
Agrobacterium tumefaciens (NP_530798) [NC_003305]
Bacillus anthracis (AAP27682) [NC_003997]
B. cereus (NP_833536) [NC_003909]
B. halodurans (NP_243282) [AB043971]
B. subtilis (NP_389542) [Z99104]
B. thuringiensis (YP_037877) [AE017355]
Bordetella bronchiseptica (NP_889783) [NC_002927]
Bradyrhizobium japonicum (NP_767425) [NC_004463]
Buchnera aphidicola (AAM67919) [NC_004061]
Clostridium acetobutylicum (NP_348424) [NC_003030]
C. perfringens (NP_562605) [BA000016]
C. tetani (NP_781903) [AE015927]
Enterococcus faecalis (NP_814994) [NC_004668]
Erwinia carotovora (YP_048825) [NC_004547]
Escherichia coli (NP_312077) [NC_000913]
Haemophilus influenzae (NP_439435) [NC_000907]
Lactobacillus plantarum (NP_785568) [NC_004567]
Lactococcus lactis (NP_266920) [NC_002662]
Listeria innocua (NP_470695) [AL596164]
Mycobacterium tuberculosis (P0A5M2)
Mycoplasma genitalium (NP_072803) [L43967]
M. mycoides (NP_975347) [NC_005364]
M. penetrans (NP_757497) [BA000026]
M. pneumoniae (NP_109842) [AF132741]
M. pulmonis (NP_326034) [NC_002771]
Phytoplasma strain AY-WB, aster yellows witches' broom; Ohio, USA ()[a] [AY389828]
Phytoplasma strain BB, tomato big bud; Arkansas, USA (DQ370431) [AF222064]
Phytoplasma strain BBS3, blueberry stunt; Michigan, USA (DQ370429) [AY265213]
Phytoplasma strain CPh, clover phyllody; Canada (AAS20956) [AF222065]
Phytoplasma strain MBS, maize bushy stunt; Mexico (DQ370430) [AY265208]
Phytoplasma strain OY-M, onion yellows; Japan (NP_950400) [D12569]
Phytoplasma strain PaWB, paulownia witches' broom; Taiwan (DQ370432) [AY265206]
Pseudomonas aeruginosa (NP_253433) [NC_002516]
P. syringae (NP_794244) [AE016853]
Ralstonia solanacearum (NP_519409) [NC_003295]
Rickettsia typhi (YP_067491) [AE017197]
Spiroplasma kunkelii (DQ164211) []b
Staphylococcus aureus (NP_371790) [NC_002952]
Streptococcus mutans (NP_720866) [NC_004350]
Thermoanaerobacter tengcongensis (NP_623014) [NC_003869]
Ureaplasma parvum (NP_078152) [NC_002162]
Xanthomonas campestris (NP_637862) [NC_003902]
Xylella fastidiosa (NP_297527) [NC_004556]

[a]Phytoplasma strain AY-WB NusA amino acid and *nusA* nucleotide sequences were obtained at internet site http://www.oardc.ohio-state.edu/phytoplasma/.
[b]The *Spiroplasma kunkelii* 16S rRNA gene sequence was obtained at the Spiroplasma Genome Sequencing Project internet site at http://www.genome.ou.edu/spiro.html.

gov/. ORF analyses were carried out using ORF finder at the NCBI web site at http://www.ncbi.nlm.nih.gov/gorf/gorf.html. COGnitor searches (Tatusov *et al*., 1997) were carried out at the NCBI web site at http://www.ncbi.nlm.nih.gov/COG/. Delineation of genes was done using the Gene Mark/GeneMark.hmm program at http://opal.biology.gatech.edu/ GeneMark/ (Borodovsky and McIninch, 1993; Lukashin and Borodovsky, 1998), the FGENESB program at http://www.softberry.com/berry.phtml, and the Orpheus program at http://pedant.gsf.de/orpheus/ (Frishman *et al*., 1998). Ribosome binding sites (RBS) were identified by inspection of a purine-rich region upstream of the transcription start codon and confirmed using the RBSfinder program at http://www.tigr.org/software/. BLAST searches against the Protein Databank were carried out at the NCBI web site at http://www.ncbi.nlm.nih.gov/ BLAST/. The SMART (Simple Modular Architecture Research Tool) database (http://smart.embl-heidelberg. de/) (Schultz *et al*., 1998) was used to identify protein domains.

Phylogenetic analyses. Phylogenetic trees were constructed using 16S rDNA sequences and amino acid sequences of bacterial NusA proteins. Sequence alignments were performed using ClustalX1.81 (Thompson *et al*., 1997). Phylogenetic and molecular evolutionary analyses were conducted using *MEGA* version 2.1 (Kumar *et al*., 2001) Neighbor-Joining method and all default values, gaps excluded, with 5000 replications for bootstrap analysis. GenBank accession numbers and other sources of NusA and 16S rRNA gene sequences used in phylogenetic analyses are given in Table 1.

RESULTS AND DISCUSSION

The nucleotide sequence determined for cloned DNA fragment CPh163 was deposited in the GenBank database under Accession no. AY463357. Those determined for the *nusA* gene sequences from *A. palmae* and phytoplasma strains BB, BBS3, MBS, and PaWB were deposited in the GenBank database under Accession nos. listed in Table 1. The clone CPh163 4564 bp fragment contained five open reading frames (ORFs) (Table 2), and had an overall composition of 27.4 mol% G + C.

CPh phytoplasma *nusA* gene. A BLASTX search of NCBI's nonredundant protein database using the CPh163 sequence as a query revealed the presence of a region (ORF3; bases 2521..3594, 30.2 mol% G+C) encoding a putative 357-amino acid protein, NusA (Table 2). A 0.8 kbp DNA sequence was amplified in PCR primed by oligonucleotide pair PhNusF1 and PhNusR5 when DNA from CPh infected plants, but not healthy plants, was used as the template (Fig. 1), confirming that the cloned CPh163 DNA containing the *nusA* gene was derived from the phytoplasma.

A genomic DNA gel blot analysis was carried out, us-

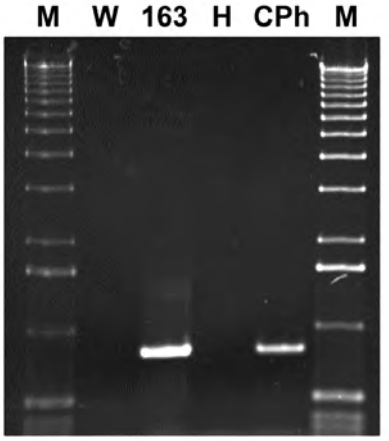

Fig. 1. Amplification of *nusA* gene sequences, 0.8 kbp in size, from clover phyllody (CPh) phytoplasma DNA. Polymerase chain reactions were primed by oligonucleotide pair PhNusF1/PhNusR5; annealing temperature was 55ºC. W, reaction mixture devoid of template DNA. 163, template was derived from DNA clone CPh163 containing an insert that included the *nusA* gene. H, healthy plant (*C. roseus*) DNA template. CPh, CPh phytoplasma infected plant DNA template. M, 1 Kb DNA Ladder.

Table 2. Properties of ORFs in the clone CPh163 DNA and their deduced products.

ORF	Endpoints (nt)	Ribosome-binding site/start codon[1]	Product size (aa/kDa)	Best BLAST hit[2] % identity (over aa)	Predicted function/ similar protein
ORF1	< 1..219	N/A	partial	NP_950398 47 (70)	small protein B
ORF2	419..2275	AGACAATAATTA**AGGACA**CATAA/ATG	618/73	NP_950399 86 (618)	hypothetical protein
ORF3	2521..3594	ATGCATTTACT**AGGAGG**TATCGT/TTG	357/40.3	NP_950400 98 (357)	NusA
ORF4	3600..3857	N/A	85/9.7	AAS22255 82 (85)	hypothetical protein
ORF5	4012.> 4564	TTAAAATTAA**AGGAGG**GATGTAA/ATG	partial	NP_950402 68 (140)	translation initiation factor, IF2

[1] Bolded, underlined nucleotides denote the purine-rich region within the putative ribosomal binding site.

[2] GenBank accession number or protein ID of the best BLAST hit, followed by the percent identity between the query and the best hit.

ing the 0.8 kbp CPh *nusA* gene fragment, amplified in PCR primed by PhNusF1 and PhNusR5, as a probe. A single-band hybridization pattern indicated that the phytoplasma's genome contains a single copy of the *nusA* gene (Fig. 2); no hybridization signal was observed with DNA from healthy plants, confirming that cloned DNA fragment CPh163 was derived from the phytoplasma.

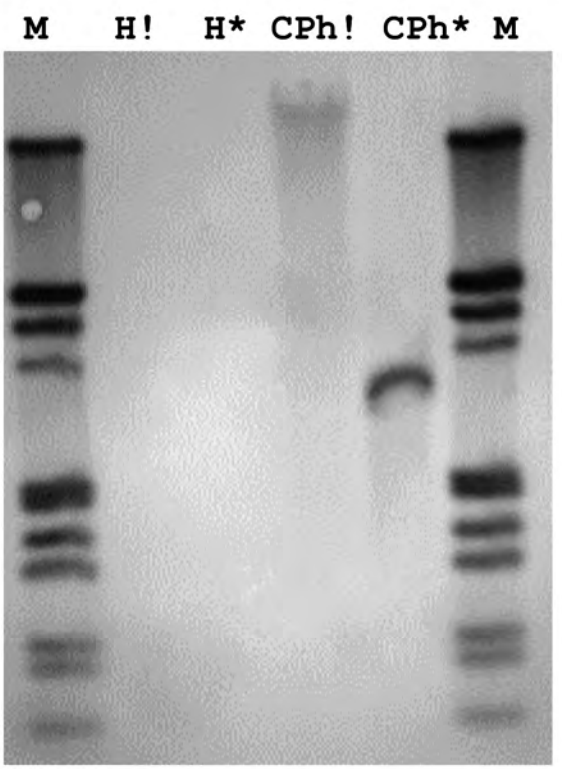

Fig. 2. Southern hybridization of the *nusA* gene in clover phyllody phytoplasma genome. Genomic DNA was probed with a digoxigenin (DIG)-labeled 0.8 kbp DNA fragment amplified, in the polymerase chain reaction primed by primer pair PhNusF1/PhNusR5, from CPh phytoplasma DNA. M, DIG-labeled molecular weight marker. H!, undigested healthy plant DNA control. H*, digested healthy plant DNA control. CPh!, undigested CPh phytoplasma genomic DNA. CPh*, digested CPh DNA. *, DNA digested with *Hind*III.

Genes flanking the *nusA* locus in clover phyllody phytoplasma. The *nusA* gene (ORF3) in clone CPh163 was flanked at the 5'-end by two protein coding regions, ORF1 and ORF2. ORF1 (bases < 1..219, gene *smpB*) was a partial ORF encoding the C-terminal portion of a putative small protein B (Table 2). Small protein B is a part of the SsrA-SmpB system for adding a tag-peptide to proteins translated from defective or shortened mRNA's; a tagged protein is recognized and degraded by proteases (Karzai *et al.*, 1999, 2000). ORF2 (bases 419..2275, 27.4 mol% G+C) encodes a putative 618-amino acid hypothetical protein (Table 2).

The *nusA* gene was flanked at the 3'-end by two protein coding regions, ORF4 and ORF5. ORF4 (bases 3600..3857, 27.9 mol% G+C) is located six bases downstream from the stop codon of ORF3 (*nusA*) and codes for an 85-amino acid hypothetical protein. This putative protein had a most significant BLASTP hit with YlxR of aster yellows witches'-broom phytoplasma (GenBank no. AAS22255). The ORF4 protein also exhibited similarity with GenBank no. ZP_00037542, a possible nucleic acid-binding protein involved in transcription termination in *Enterococcus faecium*. Using the SMART program, the ORF4-encoded protein matched PFAM entry DUF448 (protein of unknown function) with an E-value of 1.6 e-14. A similar putative protein was found by Bai *et al.* (2004) in aster yellows phytoplasma strain AY-WB. Given that NusA contains three RNA-binding domains, S1, KH1, and KH2 (Gopal *et al.*, 2001), it would not be surprising if the role of the ORF4-encoded hypothetical protein were to assist NusA or RNA polymerase in transcription termination. The putative ribosome binding site upstream of ORF3, the close proximity (six bases) of ORF3 and ORF4 to each other, and the possibility of ORF4 coding for a nucleic acid-binding protein suggest that these two ORF's may be coordinately regulated and may perform related functions.

ORF5 (bases 4012..4564, gene *infB*) codes for a putative translation initiation factor, IF2. IF2 is a GTP-binding protein that binds fMet-tRNA to the 30S ribosomal subunit (Kozak, 1999).

Phytoplasmal NusA protein. Alignment of the deduced amino acid sequences of phytoplasmal NusAs with other bacterial NusA proteins revealed distinct blocks of conserved regions (data not shown). The boundaries for three conserved RNA-binding domains (S1, KH1 and KH2) in the CPh, AY-WB, and OY-M phytoplasma NusAs were determined based on alignment with NusA from *Mycobacterium tuberculosis*, whose NusA crystal structure was previously solved by Gopal *et al.* (2001). Based on this analysis, the phytoplasmal NusA proteins were shown to contain the three conserved domains. However, the phytoplasmal NusA proteins are significantly smaller than NusA proteins from known *Mycoplasma* spp. For example, the phytoplasma strain CPh NusA is 357 amino acids in length, while the NusA proteins of *Mycoplasma genitalium* and *M. penetrans* are 531 and 597 amino acids, respectively (Fig. 3). Amino acid sequence alignment revealed that the CPh phytoplasma NusA is shorter than the *M. genitalium* NusA by about 30 amino acids at the N-terminal region and 150 amino acids at the C-terminal region (data not shown). Although the NusA of CPh phytoplasma is considerably smaller than the NusA proteins of *M. genitalium* and *M. penetrans*, it is likely functional, since all three essential domains were found in the CPh NusA in our study, and its size is similar to that of NusA

in *Clostridium* spp., *Bacillus* spp., and *A. palmae* (Fig. 3).

To gain further insight into phytoplasma evolution, we examined the frequency of lysine and arginine residues in NusA proteins of CPh phytoplasma and other members of class *Mollicutes*. These values differed from those in other bacterial groups (Table 3). For example, lysine (K) is preferred over arginine (R) (48K/8R) in the CPh and OY-M phytoplasma NusA proteins, as well as in NusA proteins of other members of class *Mollicutes*, whereas, arginine is preferred over lysine (9K/36R) in *M. tuberculosis* (This paper and Gopal *et al*., 2001). Neither lysine nor arginine appeared to be favored in NusA of *Bacillus subtilis* and *Lactococcus lactis.*

Like that of the *Mollicutes*, NusA of the walled bacterium *Clostridium perfringens* favored lysine over arginine (37K/11R). These amino acid preferences support the concept that phytoplasmas descended from a *Clostridium*-like ancestor.

Phylogenetic relationships of phytoplasmal NusA. Phylogenetic analysis of NusA amino acid sequences yielded a tree whose branching order was similar to that of a tree based on 16S rDNA in this study (Fig. 3) and elsewhere (Gundersen *et al*., 1994), and to trees based on ribosomal protein genes (Lee *et al*., 2004), *tuf* gene sequences (Schneider *et al*., 1997; Marcone *et al*., 2000), and glycoprotease amino acid sequences (Davis *et al*.,

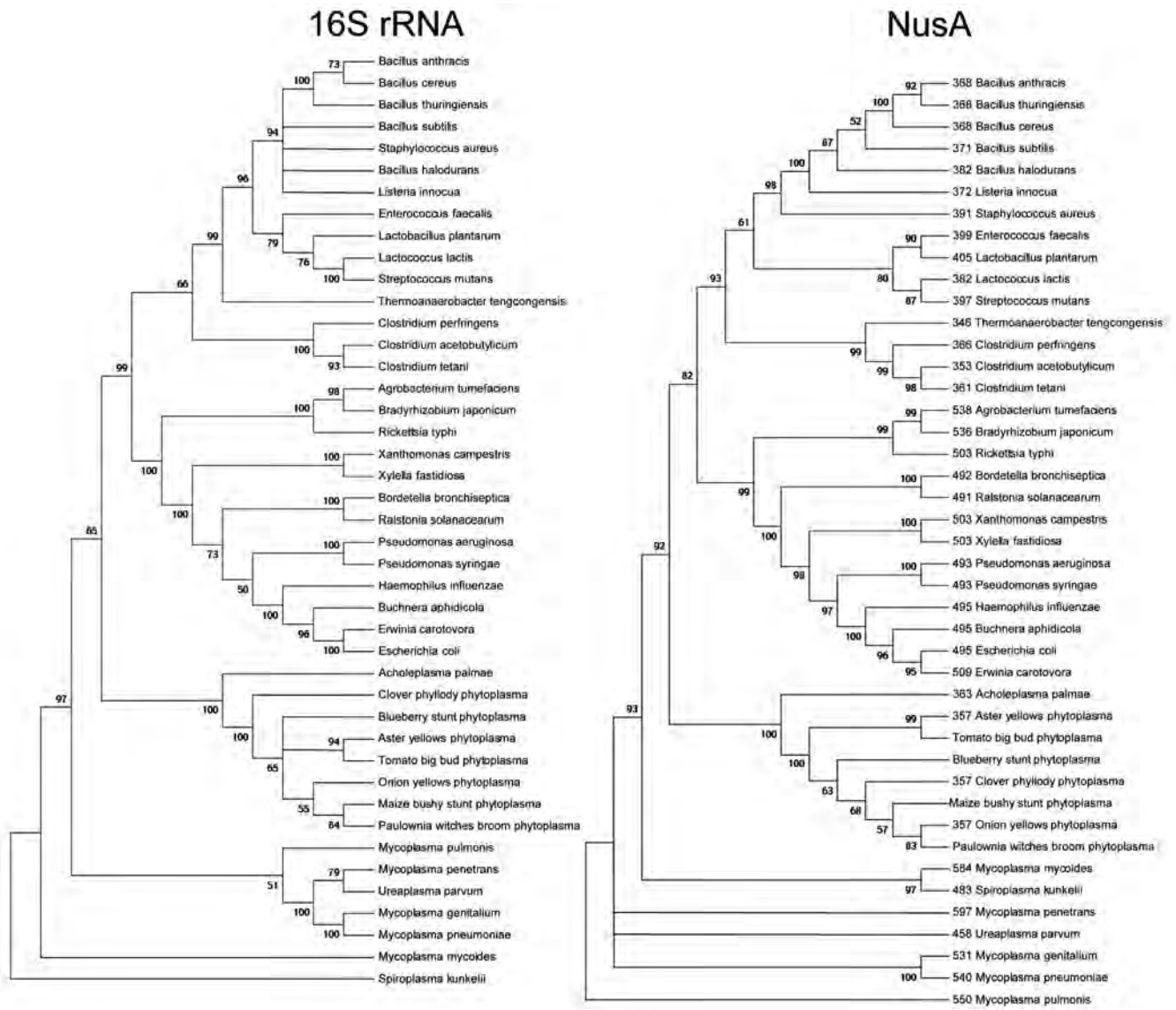

Fig. 3. Phylogenetic trees constructed by the Neighbor-Joining method of 16S rRNA gene sequences (left) and NusA amino acid sequences (right) from clover phyllody (CPh), onion yellows (OY-M), aster yellows (AY-WB), blueberry stunt (BBS3), tomato big bud (BB), paulownia witches' broom (PaWB), and maize bushy stunt (MBS) phytoplasmas; *Acholeplasma palmae*; *Spiroplasma kunkelii*, five *Mycoplasma* and one *Ureaplasma* species, and 28 walled bacteria. Numbers indicate numbers of amino acids in full length NusA proteins where known.

Table 3. Numbers of arginine (R) and lysine (K) residues in NusA proteins of selected mollicutes and walled bacteria.

Taxonomic group	Species	Number of residues	
		R	K
Mollicutes	Clover phyllody phytoplasma	8	48
	Onion yellows phytoplasma	8	48
	Acholeplasma palmae	15	38
	Spiroplasma kunkelii	12	44
	Mycoplasma genitalium	13	59
Clostridia	*Clostridium perfringens*	11	37
Bacilli	*Lactococcus lactis*	21	29
	Bacillus subtilis	23	26
Actinobacteria	*Mycobacterium tuberculosis*	36	9

2003). The NusA protein therefore appears to be useful for distinguishing diverse species of bacteria (Fig. 3, right panel).

By including *A. palmae* NusA, the present study is, to our knowledge, the first to explore *Mollicutes* phylogeny employing a non-viral and non-ribosomal acholeplasma protein in a phylogenetic reconstruction.

In the NusA tree, strongly supported (bootstrap value, 100) branching indicates that phytoplasmas form a sister clade with *A. palmae*, consistent with relationships seen in the 16S rDNA tree (Fig. 3). Comparison of the 16S rDNA and NusA trees indicated that use of NusA may assist in definition of some subgroups within group 16SrI (Fig. 3). For example in the NusA tree, grouping of strains within the phytoplasma clade was consistent with relationships based on the analysis of 16S rDNA. Thus, clustering of the subgroup 16SrI-A strains (AY-WB and BB) and the divergence of subgroup 16SrI-A from other subgroups were strongly supported in the NusA tree. While distinctions among other 16SrI subgroups were less strongly supported by bootstrap values, branching of the NusA tree was generally consistent with subgroup lineages recognized previously by RFLP analyses of other gene sequences (Gundersen *et al*., 1994; Schneider *et al*., 1997; Marcone *et al*., 2000; Lee *et al*., 2004).

Use of *nusA* for detection and differentiation of diverse phytoplasmas in group 16SrI. The CPh phytoplasma full length *nusA* gene shares 96.3% and 96.6% nucleotide sequence identity with *nusA* of AY-WB and OY-M phytoplasmas, respectively, clearly distinguishing CPh phytoplasma (member of subgroup 16SrI-C) from AY-WB and OY-M phytoplasmas (members of subgroups 16SrI-A and 16SrI-B). These values contrast with sequence similarities among 16S rDNAs of these phytoplasmas, which range from 99.3% to 99.4%, suggesting that study of *nusA* could contribute to multi-locus differentiation of distinct phytoplasma lineages.

The results from PCR and RFLP analyses also indicated that *nusA* may be useful in detection and differentiation of group 16SrI phytoplasmas. Sequences (~1 kbp) comprising nearly all of the *nusA* gene were amplified, in PCRs primed by oligonucleotide pair PhNusF2 and PhNusR5, from genomic DNAs of BB, AY, CY, MBS, and CPh phytoplasmas, members of three distinct subgroups in group 16SrI (Fig. 4). RFLP analysis, using *Mse*I and *Tsp*509I, of the PCR products differentiated subgroups 16SrI-A, 16SrI-B, and 16SrI-C (Fig. 5), further pointing

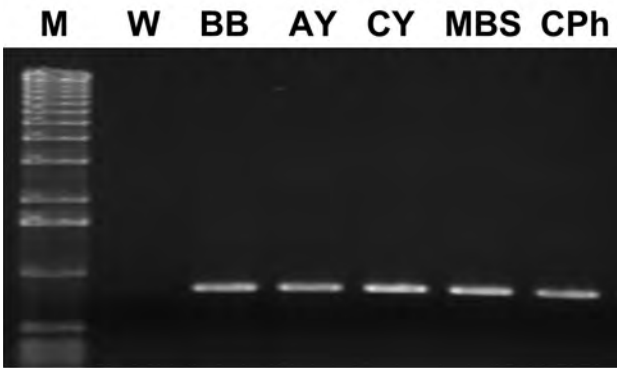

Fig. 4. Amplification of 1 Kbp *nusA* gene sequences from phytoplasmas belonging to group 16SrI, subgroups 16SrI-A (tomato big bud phytoplasma, BB), 16SrI-B (aster yellows, AY [AY1]; chrysanthemum yellows, CY; maize bushy stunt, MBS), and 16SrI-C (CPh, clover phyllody phytoplasma). M, 1 Kb DNA Ladder. W, polymerase chain reaction (PCR) mixture devoid of template DNA. Polymerase chain reactions were primed by oligonucleotide pair PhNusF2/PhNusR5.

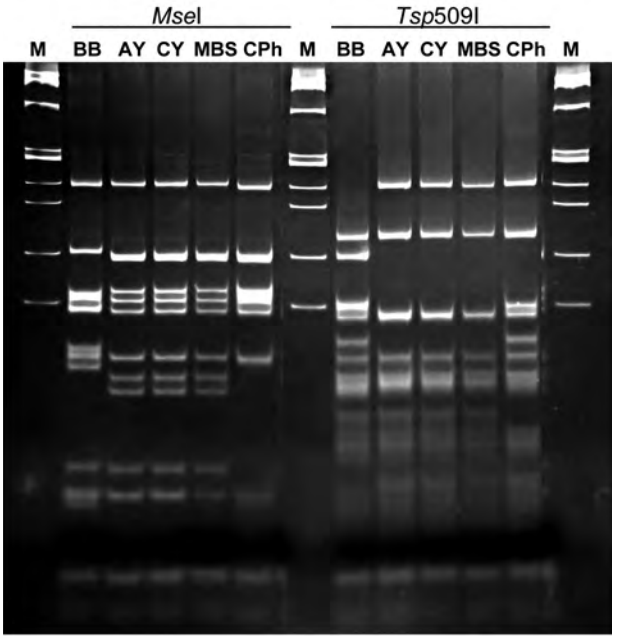

Fig. 5. RFLP analysis of *nusA* gene sequences amplified in polymerase chain reactions (PCRs) primed by oligonucleotide pair PhNusF2/PhNusR5 from phytoplasma strains, BB, AY (AY1), CY, MBS and CPh. Products from PCRs were digested with restriction endonucleases *Mse*I and *Tsp*509I. M, ϕX174 RFI DNA *Hae*III digest size standard, fragment sizes (bp) from top to bottom: 1353, 1078, 872, 603, 310, 281, 271, 234, 194, 118, 72.

to *nusA* as another molecular tool that can be used for detection and differentiation of closely related phytoplasma strains. Since NusA is a ubiquitous and universally conserved protein among bacteria (Borukhov *et al*., 2005), and a search of completed genomes stored at the NCBI website (http://www.ncbi.nlm.nih.gov/) indicated that *nusA* occurs as single copy in bacterial genomes, the *nusA* gene has potential to be useful in aiding phytoplasma identification.

ACKNOWLEDGEMENTS

The authors are grateful to Kristi Bottner, Kelly Davis, and Erika Pinto for excellent technical assistance, L.R. Nault for phytoplasma strain MBS, and Erika Pinto for careful proof reading of the manuscript.

REFERENCES

Altschul S.F., Gish W., Miller W., Myers W., Lipman D.J., 1990. Basic local alignment search tool. *Journal of Molecular Biology* **215**: 403-410.

Bai X., Zhang J., Holford I.R., Hogenhout S.A., 2004. Comparative genomics identifies genes shared by distantly related insect-transmitted plant pathogenic mollicutes. *FEMS Microbiology Letters* **235**: 249-258.

Borodovsky M., McIninch J., 1993. GENEMARK: parallel gene recognition for both DNA strands. *Computers & Chemistry* **17**: 123-134.

Borukhov S., Lee J., Laptenko O., 2005. Bacterial transcription elongation factors: new insights into molecular mechanism of action. *Molecular Microbiology* **55**: 1315-1324.

Davis R.E., Jomantiene R., Zhao Y., 2005. Lineage-specific decay of folate biosynthesis genes suggests ongoing host adaptation in phytoplasmas. *DNA and Cell Biology* **24**: 832-840.

Davis R.E., Jomantiene R., Zhao Y., Dally E.L., 2003. Folate biosynthesis pseudogenes, *ψfolP* and *ψfolK*, and an O-sialoglycoprotein endopeptidase gene homolog in the phytoplasma genome. *DNA and Cell Biology* **22**: 697-706.

Frishman D., Mironov A., Mewes H.-W., Gelfand M., 1998. Combining diverse evidence for gene recognition in completely sequenced bacterial genomes. *Nucleic Acids Research* **26**: 2941-2947.

Gopal B., Haire L.F., Gamblin S.J., Dodson E.J., Lane A.N., Papavinasasundaram K.G., Colston M.J., Dodson G., 2001. Crystal structure of the transcription elongation/anti-termination factor NusA from *Mycobacterium tuberculosis* at 1.7 Å resolution. *Journal of Molecular Biology* **314**: 1087-1095.

Greenblatt J., Li J., 1981a. Interaction of the sigma factor and the *nusA* gene protein of *E. coli* with RNA polymerase in the initiation-termination cycle of transcription. *Cell* **24**: 421-428.

Greenblatt J., Li J., 1981b. The *nusA* gene protein of *Escherichia coli*. Its identification and a demonstration that it interacts with the gene N transcription anti-termination protein of bacteriophage lambda. *Journal of Molecular Biology* **147**: 11-23.

Greenblatt J., McLimont M., Hanly S., 1981. Termination of transcription by *nusA* gene protein of *Escherichia coli*. *Nature* **292**: 215-220.

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. *Journal of Bacteriology* **176**: 5244-5254.

Karzai A.W., Roche E.D., Sauer R.T., 2000. The SsrA-SmpB system for protein tagging, directed degradation and ribosome rescue. *Nature Structural Biology* **7**: 449-455.

Karzai A.W., Susskind M.M., Sauer R.T., 1999. SmpB, a unique RNA-binding protein essential for the peptide-tagging activity of SsrA (tmRNA). *The EMBO Journal* **18**: 3793-3799.

Kozak M., 1999. Initiation of translation in prokaryotes and eukaryotes. *Gene* **234**: 187-208.

Kumar S., Tamura K., Jakobsen I.B., Nei M., 2001. MEGA2: Molecular evolutionary genetics analysis software. *Bioinformatics* **17**: 1244-1245.

Lee I.-M., Davis R.E., 1989. Serum-free media for cultivation of spiroplasmas. *Canadian Journal of Microbiology* **35**: 1092-1099.

Lee I.-M., Davis R.E., 1988. Detection and investigation of genetic relatedness among aster yellows and other mycoplasmalike organisms by using cloned DNA and RNA probes. *Molecular Plant-Microbe Interactions* **1**: 303-310.

Lee I.-M., Davis R.E., 1986. Prospects for in vitro culture of plant pathogenic mycoplasmalike organisms. *Annual Review of Phytopathology* **24**: 339-354.

Lee I.-M., Davis R.E., 1983. Phloem-limited prokaryotes in sieve elements isolated by enzyme treatment of diseased plant tissues. *Phytopathology* **73**: 1540-1543.

Lee I.-M., Davis R.E., Gundersen-Rindal D.E., 2000. Phytoplasma: phytopathogenic mollicutes. *Annual Review of Microbiology* **54**: 221-255.

Lee I.-M., Gundersen-Rindal D.E., Davis R.E., Bottner K.D., Marcone C., Seemüller E., 2004. '*Candidatus* Phytoplasma asteris', a novel phytoplasma taxon associated with aster yellows and related diseases. *International Journal of Systematic and Evolutionary Microbiology* **54**: 1037-1048.

Lim P.O., Sears B.B., 1991. The genome size of a plant-pathogenic mycoplasmalike organism resembles those of animal mycoplasmas. *Journal of Bacteriology* **173**: 2128-2130.

Lukashin A., Borodovsky M., 1998. GeneMark.hmm: new solutions for gene finding. *Nucleic Acids Research* **26**: 1107-1115.

Marcone C., Lee I.-M., Davis R.E., Ragozzino A., Seemüller E., 2000. Classification of aster yellows-group phytoplasmas based on combined analyses of rRNA and *tuf* gene sequences. *International Journal of Systematic and Evolutionary Microbiology* **50**: 1703-1713.

Marcone C., Neimark H., Ragozzino A., Lauer U., Seemüller E., 1999. Chromosome sizes of phytoplasmas composing

major phylogenetic groups and subgroups. *Phytopathology* **24**: 805-810.

Neimark H., Kirkpatrick B.C., 1993. Isolation and characterization of full-length chromosomes from non-culturable plant-pathogenic *Mycoplasma*-like organisms. *Molecular Microbiology* **7**: 21-28.

Oshima K., Kakizawa S., Nishigawa H., Jung H.Y., Wei W., Suzuki S., Arashida R., Nakata D., Miyata S., Ugaki M., Namba S., 2004. Reductive evolution suggested from the complete genome sequence of a plant-pathogenic phytoplasma. *Nature Genetics* **36**: 27-29.

Razin S., Yogev D., Naot Y., 1998. Molecular biology and pathogenicity of mycoplasmas. *Microbiology and Molecular Biology Reviews* **62**: 1094-1156.

Received 2 December 2005
Accepted 10 March 2006

Schneider B., Gibb K.S., Seemüller E., 1997. Sequence and RFLP analysis of the elongation factor Tu gene used in differentiation and classification of phytoplasmas. *Microbiology* **143**: 3381-3389.

Schultz J., Milpetz F., Bork P., Ponting C.P., 1998. SMART, a simple modular architecture research tool: Identification of signaling domains. *Proceedings of the National Academy of Sciences* **95**: 5857-5864.

Tatusov R.L., Koonin E.V., Lipman D.J., 1997. A genomic perspective on protein families. *Science* **278**: 631-637.

Thompson J.D., Gibson T.J., Plewniak F., Jeanmougin F., Higgins D.G., 1997. The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research* **24**: 4876-4882.