Identification of genes in the tomato big bud phytoplasma and comparison to those in sweet potato little leaf-V4 phytoplasma

Claire Streten and Karen S. Gibb

Faculty of Education, Health and Science, Northern Territory University, Darwin, 0909, NT, Australia

Genetic relatedness of phytoplasmas is commonly defined on the basis of differences in the highly conserved 16S rRNA gene, which may not resolve closely related phytoplasmas. An example of this is the closely related tomato big bud (TBB) and sweet potato little leaf strain V4 (SPLL-V4) phytoplasmas, which cannot easily be differentiated by their 16S rRNA gene sequences. This study aimed to identify genes on the TBB phytoplasma chromosome which could be used to examine genetic variation between these two closely related phytoplasmas. Random clones generated from TBB phytoplasma genomic DNA were sequenced and characterized by database analysis. Twenty-three genes were identified within 19 random clones, which contained approximately 18-0 kbp of TBB phytoplasma genomic DNA. Half of the TBB phytoplasma genes identified were involved in DNA replication, transcription and translation. The remaining TBB phytoplasma genes were involved in protein secretion, cellular processes and energy metabolism. Phylogenetic analysis of representative genes showed that the TBB phytoplasma grouped with the mycoplasmas with the exception of the TBB phytoplasma secA gene, which grouped with the onion yellows phytoplasma. PCR primers were designed based on the new genes and tested on isolates of the TBB and SPLL-V4 phytoplasmas. Most primers amplified a product from TBB and SPLL-V4 phytoplasma samples. When amplified products were subjected to RFLP analysis, the restriction patterns were the same as the respective original clones. This result confirmed that the same sequence had been amplified by PCR and showed that these isolates were indistinguishable using the new genes. This study showed that in fact the TBB and SPLL-V4 phytoplasmas are closely related even with the analysis of new genes. These new genes have, however, provided insight into the biology of the TBB and SPLL-V4 phytoplasmas.

INTRODUCTION

Phytoplasmas are plant pathogens that are associated with numerous diseases worldwide (McCoy et al., 1989). We have limited knowledge of the biochemical properties of phytoplasmas because they cannot be cultured in vitro. Phytoplasma genes identified to date include those for SecA (Kakizawa et al., 2001), ribosomal proteins (Lim & Sears, 1992; Gundersen et al., 1994), nitro-reductase, (Jarausch et al., 1994), elongation factor Tu (Schneider et al., 1997), elongation factor G (Berg & Seemu¨ller, 1999), immuno-dominant membrane proteins (Berg et al., 1999; Blomquist et al., 2001), RNase P RNA (Wagner et al., 2001), a major membrane protein (Barbara et al., 2002) and an antigenic protein (Yu et al., 1998). These genes provide some information on the possible biochemical pathways associated with phytoplasma metabolism, but further information is still needed.

Classification and differentiation of phytoplasmas is based on their 16S rRNA gene sequences (Seemu¨ller et al., 1998). Due to the highly conserved nature of this gene, closely related phytoplasmas may be perceived as identical (Fox et al., 1992). Based on previous studies, which have identified greater genetic diversity using non-ribosomal genes (Jarausch et al., 1994), we postulated that the identification of new, possibly less conserved, genes may reveal genetic variation between phytoplasmas that are indistinguishable on the basis of their 16S rRNA genes. The tomato big bud (TBB) phytoplasma is associated with a wide range of plant diseases throughout Australia (Davis et al., 1997). The sweet potato little leaf (SPLL) phytoplasma was originally detected in, and isolated from, Ipomoea batatas growing in...
Northern Australia (Gibb et al., 1995). RFLP analysis of the 16S rRNA gene of TBB and SPLL phytoplasmas, isolated from their natural host plants, showed no polymorphisms (Gibb et al., 1995). TBB phytoplasma causes phyllody and an absence of normal flower production in Catharanthus roseus (periwinkle), while SPLL phytoplasma when transmitted from sweet potato to periwinkle, causes virescence and smaller flowers (Padovan et al., 2000). RFLP analysis of the 16S rRNA gene of SPLL phytoplasma isolated from periwinkle showed slight variation in AluI and RsaI patterns compared to TBB phytoplasma (Schneider et al., 1999). The periwinkle isolate of SPLL phytoplasma was designated SPLL-V4 phytoplasma to distinguish it from field samples of SPLL phytoplasma (Padovan et al., 2000). RFLP analysis at the chromosome level revealed genetic diversity between TBB and SPLL-V4 phytoplasmas (Padovan et al., 2000), but these differences are not reflected in their 16S rRNA genes. We do not know which phytoplasma genes will be the best indicators of the genetic variation observed at the chromosome level. To identify phytoplasma genes that may reflect the heterogeneity between TBB and SPLL-V4 phytoplasmas at the chromosome level we characterized genomic sequences of the TBB phytoplasma. Oligonucleotide primers were designed based on the nucleotide sequences of the TBB phytoplasma genes identified and used in PCR to confirm that the characterized genes were not host plant genes. RFLP analysis of the genes amplified from TBB and SPLL-V4 phytoplasmas was then used to examine genetic diversity between these phytoplasmas.

METHODS

DNA sequencing and database analysis. Symptomatic plant samples that tested positive for the TBB or SPLL-V4 phytoplasmas were collected in Darwin (periwinkle), Adelaide (grape) and Katherine (papaya), Australia. The DNA was extracted from symptomatic tissue of diseased papaya and grapevine plants using a phytoplasma enrichment procedure (Ahrens & Seemüller, 1992). TBB and SPLL-V4 phytoplasma DNA was extracted from the mid-veins of symptomatic periwinkle leaves, maintained at the Northern Territory University, using a small-scale procedure (Doyle & Doyle, 1990).

TBB phytoplasma random clones were obtained as described previously (Schneider et al., 1999). Random clones were grown overnight in Luria–Bertani (Bacto-Tryptone, yeast extract and sodium chloride) broth. Plasmids were purified using the Mini-Prep Quick Spin Kit (Qiagen), according to the manufacturer’s protocol. The quantity of purified plasmid DNA was estimated by comparison to a DNA mass ladder on a 1% agarose gel stained with ethidium bromide and visualized on a UV transilluminator.

Plasmid DNA (300 ng per reaction tube) was sequenced using the Big Dye Terminator Reaction kit (Applied Biosystems). M13 primers (Promega) were used to sequence the start of clonal inserts. All steps were performed according to the manufacturer’s protocol. Tubes were cycled in a Corbett FTS-320 Thermocycler (Corbett Research). Sequencing reactions were analysed at the Australian Genomic Research Facility. The data were initially analysed using MACVECTOR (Eastman Kodak) and ASSEMBLYJNG (Eastman Kodak). Further sequencing primers were designed using the Primer3 program (http://www.genome.wi.mit.edu/genome_software/other/primer3.html) based on melting temperature and their location on the sequence. Sequencing primers were synthesized by Genset Pacific Oligos.

The complete nucleotide sequences of the clonal inserts were analysed using FASTA and FASTX (Pearson & Lipman, 1988) database search engines which were accessed through Entigen (http://www.entigen.com). The FLIPORF program (Bionet Software; Entigen) was used to identify the location of putative genes along the nucleotide sequence. The genes were translated using TRANSLATE [Genetics Computer Group (GGG) package, Oxford Molecular Group].

Database programs used to analyse the TBB phytoplasma amino acid sequences were accessed through the Entigen website (http://www.entigen.com), unless otherwise specified. BLOCK SEARCHER (Henikoff & Henikoff, 1994), MOTIFS (GCG) and HMM PPAM (http://hmmer.wustl.edu/) programs were used to identify conserved sequences, motifs and signature sequences on the putative genes. The GARNIER program (Pearson & Lipman, 1988) was used to predict the secondary structure of the proteins. Proteins with secondary structures similar to those predicted for the TBB phytoplasma gene products were identified using BLASTP (Altschul et al., 1997) searches of the PDB structural database.

A representative TBB phytoplasma gene was selected from each class of the protein functional groups (with the exception of the miscellaneous group) to determine whether the protein class influenced the phylogenetic position of the phytoplasmas. The TBB phytoplasma gene with the highest similarity to other amino acid sequences in the SWISS-PROT + TremBL database (within each protein function category) was used for phylogenetic analysis. The results from FASTA searches based on the amino acid sequences of the TBB phytoplasma genes were used to select 10 species of eubacteria for phylogenetic analysis. The highest match to the TBB phytoplasma genes based on FASTA searches was always included in phylogenetic analysis. The amino acid sequences were aligned using CLUSTAL W (accurate) (Thompson et al., 1994). A bootstrapped consensus tree was generated using the computer programs of Felsenstein (1989) (SEQBOOT, PROTPARS, NEIGHBOR and CONSENSE).

Signal peptide sequences and their cleavage sites were identified with PSORT version 6.4 (http://psort.nibb.ac.jp/). The database does not have entries for the class Mollicutes, but since members of this class are thought to have originated from Gram-positive bacteria, the TBB phytoplasma amino acid sequences were nominated as originating from genes of Gram-positive bacteria for PSORT analysis. The topology and orientation of the membrane protein was predicted using the TMHRED 2 program (http://psort.nibb.ac.jp/; http://bioweb.pasteur.fr/seqanal/interfaces/tmred.html). The amino acid sequences were further analysed for transmembrane regions using the TMHRED program (Hoffman & Stoffel, 1993; http://www.ch.embnet.org/software/TMPRED_form.html). TMHRED compares the query sequence to a database of transmembrane proteins and TOPRED 2 predicts the topology of membrane proteins based on hydrophobicity. Amino acid sequences of membrane proteins were aligned using PILEUP (GGG package 1998) and PRETTYBOX (GGG package 1998) programs.

PCR primer design and testing. PCR primers were designed based on representative genes: the peptide releasing factor 1, DNA polymerase beta III, FtsH and ATP-dependent RNA helicase. The Primer3 program was used to design PCR primers based on conserved regions of the phytoplasma, mycoplasma and other bacterial genes of the same function. The ability of each of the primer pairs to amplify the gene of interest from TBB phytoplasma isolates was determined by PCR. Initially, reactions were cycled through 95 °C for 1 min, 42 °C for 3 min, 72 °C for 3 min for 30 cycles followed by 72 °C for 10 min (Schneider et al., 1997). Cycling conditions were then optimized by increasing the annealing temperature by
Table 1. Functions assigned to ORFs identified within TBB phytoplasma random clones (similarity matches are based on FASTX search results)

<table>
<thead>
<tr>
<th>Clone</th>
<th>Size of insert (kb)</th>
<th>Protein encoded by gene</th>
<th>Similarity to highest match (%)</th>
<th>Source</th>
<th>Motifs</th>
<th>GenBank accession no.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>0.3</td>
<td>Preprotein translocase SecA subunit</td>
<td>62.0</td>
<td>Onion yellows phytoplasma</td>
<td>Thermophilic metalloprotease signature</td>
<td>AF494511</td>
</tr>
<tr>
<td>32</td>
<td>1.7</td>
<td>Cell division protein (FtsH)</td>
<td>37.8</td>
<td>Synechocystis sp.</td>
<td>AAA-protein family</td>
<td>AF494533</td>
</tr>
<tr>
<td>78</td>
<td>0.8</td>
<td>Cell division protein (FtsH)</td>
<td>32.3</td>
<td>Helicobacter pylori</td>
<td>AAA-protein family</td>
<td>AF494532</td>
</tr>
<tr>
<td>31</td>
<td>0.2</td>
<td>NADH oxidase</td>
<td>44.4</td>
<td>Mycoplasma capricolum</td>
<td>Lyso phospholipase catalytic domain</td>
<td>AF494522</td>
</tr>
<tr>
<td>51</td>
<td>0.2</td>
<td>2,3-Biphosphoglycerate-independent phosphoglycerate mutase</td>
<td>64.2</td>
<td>Bacillus halodurans</td>
<td>Metalloenzyme family</td>
<td>AF494526</td>
</tr>
<tr>
<td>14</td>
<td>0.7</td>
<td>DNA polymerase beta III</td>
<td>56.0</td>
<td>Acholeplasma laidlawii (Fasta)</td>
<td>DNA polymerase III beta subunit</td>
<td>AF494512</td>
</tr>
<tr>
<td>15</td>
<td>2.7</td>
<td>Uracil DNA glycosylase</td>
<td>51.8</td>
<td>Buchnera aphidicola</td>
<td>Uracil DNA glycosylase proteins</td>
<td>AF494513</td>
</tr>
<tr>
<td>22</td>
<td>1.7</td>
<td>Endonuclease IV</td>
<td>44.4</td>
<td>Bacillus subtilis</td>
<td>AP endonuclease family 2 protein</td>
<td>AF494517</td>
</tr>
<tr>
<td>24</td>
<td>1.8</td>
<td>DNA gyrase subunit b</td>
<td>52.0</td>
<td>Bacillus subtilis</td>
<td>DNA gyrase subunit B signature</td>
<td>AF494521</td>
</tr>
<tr>
<td>29</td>
<td>2.5</td>
<td>ATP-dependent RNA helicase</td>
<td>31.7</td>
<td>Bacillus subtilis</td>
<td>DEAD-box subfamily</td>
<td>AF494520</td>
</tr>
<tr>
<td>16</td>
<td>0.5</td>
<td>Glucose-inhibited division protein</td>
<td>96.0</td>
<td>Sweet potato witches’broom</td>
<td>Glucose-inhibited division protein A family</td>
<td>AF494515</td>
</tr>
<tr>
<td>25</td>
<td>2.7</td>
<td>Glucose-inhibited division protein</td>
<td>53.7</td>
<td>Bacillus subtilis</td>
<td>Glucose-inhibited division protein A family</td>
<td>AF494519</td>
</tr>
<tr>
<td>38</td>
<td>0.6</td>
<td>Glucose-inhibited division protein</td>
<td>44.4</td>
<td>Bacillus subtilis</td>
<td>Glucose-inhibited division protein</td>
<td>AF494528</td>
</tr>
<tr>
<td>87</td>
<td>1.3</td>
<td>Primosomal protein</td>
<td>29.0</td>
<td>Bacillus subtilis</td>
<td>Plant-CLC-chloride channel</td>
<td>AF494527</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Transcription</td>
<td>15</td>
<td>2.7</td>
<td>Ribonuclease HIII</td>
<td>41.8</td>
<td>Ureaplasma parvum</td>
<td>Ribonuclease HIII</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Translation</td>
<td>22</td>
<td>1.7</td>
<td>tRNA/rRNA methyl transferase</td>
<td>32.4</td>
<td>Escherichia coli</td>
<td>Ap_Nuclease_F2_1Hx[2]Y(I)(1)N(L)(A)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Miscellaneous</td>
<td>15</td>
<td>2.7</td>
<td>HIT protein</td>
<td>45.6</td>
<td>Borrelia burgdorferi</td>
<td>HIT family proteins</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Genetic relatedness of phytoplasmas
RESULTS

DNA sequencing and database analysis

Nineteen random TBB clones were sequenced and 23 genes were identified. Results for motif (Table 1), protein family (not shown), three-dimensional structural analysis (not shown) and FASTX searches (Table 1) all identified the same gene as the best match, indicating agreement across different analytical methods. All sequence analysis results (database searches, motif, protein family and structural) for each gene supported the assignment of specific protein function to the genes identified in the TBB phytoplasma clones (Table 1). Function could not be assigned to genes in clones 26 (AF494529), 49 (AF494530) and 28 (AF494525) because the sequence similarity was lower than the threshold value and no motifs were identified in the sequences. The random clone 42 gene sequence (AF494531) was similar to that for the gene for NADH dehydrogenase in the chloroplast DNA of *Reidia* sp.

Of the 23 genes identified, nine were involved in replication, including DNA polymerase beta III, uracil DNA glycosylase, endonuclease IV and gidA genes (Table 1), while others had roles in transcription, ribonuclease HIII and two helicase protein genes (Table 1). Three genes encoded proteins involved in translation, tRNA/rRNA methyl transferase, RF-1 and 50S ribosomal protein L17 (Table 1). NADH oxidase (clone 31) and 2,3-biphosphoglycerate-independent phosphoglycerate mutase (clone 51) are associated with energy metabolism (Table 1) while the secA gene product (clone 10) was the only protein identified as being involved in protein secretion (Table 1). The majority of the highest matches to the TBB phytoplasma genes were not genes of the A + T rich mollicutes, but rather genes in other bacteria. For the majority of searches, corresponding genes in members of the Mollicutes subclade were not within the top ten matches to the TBB phytoplasma genes (FASTA and FASTX results). *Bacillus subtilis* genes were identified as the best match for 7 of the 23 TBB genes, specifically the primosomal protein Nreplication factor, gidA2, gidA3, endonuclease IV, DNA gyrase subunit b, ATP-dependent RNA helicase and replicative DNA helicase genes.

Phylogenetic analysis showed that the TBB phytoplasma grouped with the mycoplasmas. These results are represented by the phylogenetic tree constructed based on the amino acid sequence of the TBB phytoplasma secA gene (Fig. 1). These results contrasted with the findings from FASTX and FASTA searches.

![Fig. 1. Phylogenetic grouping of the TBB phytoplasma secA gene (amino acid sequence). Bootstrapped values on the tree indicate the number of trees in a hundred for which the organism was located on that particular branch.](image)

The TMPRED and TOPPRED programs predicted that three of the 23 TBB phytoplasma proteins were membrane-bound (Table 2). These were the primosomal protein Nreplication factor Y, a tRNA/rRNA methyl transferase and an ATPase subunit 6. TMPRED and PSORT analysis of the histidine triad (HIT) and ribonuclease genes indicated that these proteins were also associated with the membrane, but TOPPRED 2 analysis found no transmembrane helices (Table 2). TMPRED and TOPPRED 2 analysis showed that the uracil DNA glycosylase and GidA2 proteins had membrane-spanning domains (Table 2), but the PSORT program was unable to predict the location of the uracil DNA glycosylase protein in the phytoplasma (Table 2). PSORT results for the gidA2 gene predicted that its protein product was located in the cytoplasm (Table 2). The remaining TBB phytoplasma genes encoded proteins that were predicted to be located in the cytoplasm.

PILEUP analysis showed that the TBB phytoplasma ftsH gene sequence was probably incomplete because it was shorter than corresponding genes of *Mycoplasma genitalium* and *Helicobacter pylori* 26695 (data not shown). The membrane-spanning regions of the FtsH proteins from *M. genitalium* and *H. pylori* were located in the missing region of the TBB phytoplasma gene.

Confirmation that the characterized genes were not from host

PCR primers were designed to amplify the DNA beta polymerase III (pTBB14), peptide chain release factor 1 (pTBB23), ATP-dependent RNA helicase (pTBB29) and cell division protein genes (pTBB32) (Table 3). The primers based on TBB phytoplasma clone 14 amplified a product from all TBB phytoplasma samples, but only one of the
two SPLL-V4 phytoplasma samples, while primers based on clone 23 amplified a product from both SPLL-V4 samples phytoplasma, but from only two of the TBB phytoplasma samples (Table 4). A PCR product was not amplified from TBB or SPLL-V4 phytoplasmas isolated from papaya using the pTBB29 forward and reverse primers, but a product was obtained for all the other TBB and SPLL-V4 phytoplasma samples (Table 4). Primers based on clone 32 amplified products from SPLL-V4 and TBB phytoplasma samples isolated from periwinkle and from TBB phytoplasma from grape (Table 4).

RFLP analysis of the products amplified from both TBB and SPLL-V4 phytoplasmas by primers designed from TBB phytoplasma gene sequences using digestion with AluI, RsaI or HpaI showed no variation in banding patterns between the different phytoplasmas isolated from different hosts (data not shown).

**DISCUSSION**

Eighteen kbp of the TBB phytoplasma genome was sequenced from 19 random clones. This represents about 3% of the TBB phytoplasma genome, based on an estimated chromosome size of 660 kb (Padovan et al., 2000). Half of the genes identified within the 18 kb region of the TBB phytoplasma chromosome are involved in DNA replication, transcription and translation, while only two genes have metabolic roles. Although only a fraction of the TBB phytoplasma genome has been sequenced, it is possible that relatively few metabolic genes will be discovered if the parasitic nature of phytoplasmas and their reliance on host metabolic products has led to a relatively low number of genes involved in metabolism, as has been reported for *M. genitalium* (Fraser et al., 1995).

To ascribe function to genes, amino acid similarity searches were conducted in conjunction with searches of motif databases, protein structural databases and protein family databases. These searches produced similar results, which supports the functions currently assigned to the TBB phytoplasma genomic clones. TBB genes that matched proteins isolated from plant tissue (pTBB34 and pTBB42) were not assigned a function in case the clone was of plant DNA. However, this is unlikely because the random clone library was screened with healthy plant DNA by Southern blot hybridization (Schneider et al., 1999). In the current study we screened diseased and healthy plants by PCR

---

**Table 2. Prediction of transmembrane helices and their orientation**

Transmembrane helices (TM)s and orientation predicted using TOPPRED2 and TMPRED analysis. The localization sites of the protein in the TBB phytoplasma structure are shown based on PSORT analysis. The amino acid sequences of the TBB phytoplasma genes were used for all secondary structure programs. Membrane proteins are highlighted in bold type.

<table>
<thead>
<tr>
<th>TBB phytoplasma gene</th>
<th>TMs (TOPPRED2)</th>
<th>TMs and orientation (TMPRED)</th>
<th>Location of protein</th>
</tr>
</thead>
<tbody>
<tr>
<td>Replication</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Uracil DNA glycosylase</td>
<td>1</td>
<td>1 N terminus outside</td>
<td>Uncertain</td>
</tr>
<tr>
<td>Glucose-inhibited division protein</td>
<td>1 or 2</td>
<td>2 N terminus outside</td>
<td>Cytoplasm</td>
</tr>
<tr>
<td>Primosomal protein N’ replication factor Y</td>
<td>1</td>
<td>1 N terminus outside</td>
<td>Membrane protein</td>
</tr>
<tr>
<td>Transcription</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ribonuclease HIII</td>
<td>0</td>
<td>1 N terminus outside</td>
<td>Membrane protein</td>
</tr>
<tr>
<td>Translation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>tRNA/rRNA methyl transferase</td>
<td>1</td>
<td>1 N terminus inside</td>
<td>Membrane protein</td>
</tr>
<tr>
<td>Miscellaneous</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ATPase subunit 6 protein</td>
<td>1</td>
<td>1 N terminus outside</td>
<td>Membrane protein</td>
</tr>
</tbody>
</table>

---

**Table 3. Nucleotide sequences of PCR primers**

Primers were designed based on PILEUP analysis of TBB phytoplasma random clones, mycoplasma and bacterial gene sequences.

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence (5’–3’)</th>
<th>Position</th>
<th>Gene product</th>
</tr>
</thead>
<tbody>
<tr>
<td>pTBB14F</td>
<td>TTCTGTATTTAGAAATGACTCTTG</td>
<td>143–169</td>
<td>DNA polymerase beta III chain gene</td>
</tr>
<tr>
<td>pTBB14R</td>
<td>AAGCTGACCAATTTCATCG</td>
<td>822–802</td>
<td>DNA polymerase beta III chain gene</td>
</tr>
<tr>
<td>pTBB23F</td>
<td>AACAATTCCTCGTGATTTAAA</td>
<td>1–24</td>
<td>Peptide chain release factor (RF-1)</td>
</tr>
<tr>
<td>pTBB23R</td>
<td>GGTAGCCGTTAGCGATGGT</td>
<td>913–893</td>
<td>Peptide chain release factor (RF-1)</td>
</tr>
<tr>
<td>pTBB29F</td>
<td>TCTACGTGGGTTCAACAGG</td>
<td>384–405</td>
<td>ATP-dependent RNA helicase</td>
</tr>
<tr>
<td>pTBB29R</td>
<td>AGGAATTCGGGAGATAGTG</td>
<td>1763–1745</td>
<td>ATP-dependent RNA helicase</td>
</tr>
<tr>
<td>pTBB32F</td>
<td>GGTTAACACCTCGATTCGGAC</td>
<td>182–205</td>
<td>FtsH</td>
</tr>
<tr>
<td>pTBB32R</td>
<td>GACGTCCGTTTATTCCTGG</td>
<td>840–819</td>
<td>FtsH</td>
</tr>
</tbody>
</table>
using primers designed from these newly discovered genes. Amplicons were obtained from diseased plants only and this was taken to indicate that these genes were of phytoplasma origin.

Although the majority of database searches found that B. subtilis genes were the closest matches, phylogenetic analysis of representative genes from different categories of protein function demonstrated that the TBB phytoplasma grouped with the mycoplasmas. Phylogeny based on the amino acid sequence of genes can sometimes cause species to be positioned incorrectly on a tree (Simmons, 2000). This type of error arises when single nucleotide substitutions are missed because the new nucleotide arrangement still encodes the same amino acid (Simmons, 2000). It is unlikely that this caused the TBB phytoplasma to group with the mycoplasmas in preference to B. subtilis because the TBB and mycoplasma genes were not closely related (similarity 40–50 %) (Simmons, 2000). When examining highly divergent sequences, the amino acid sequence should be the basis for phylogeny in preference to the nucleotide sequence (Simmons, 2000). The different results obtained for phylogenetic analysis and database searches may be due to the different parameters used by the programs to determine the closest match (Simmons, 2000). FASTX searches identify the closest match as the sequence that has the highest identity over the greatest number of amino acids of the query sequence (Pearson & Lipman, 1988). The PROPARS program determines the most closely related gene as the sequence with the highest amino acid identity. It also determines the amino acid substitutions that are more likely to occur in the query sequence (Felsenstein, 1989). Phylogenetic analysis is more likely to identify the genera closest to the phytoplasma because this analysis accounts for amino acid changes that possibly caused the divergent gene sequences. Therefore, when no corresponding phytoplasma gene has been characterized the genera closest to the phytoplasma because this analysis accounts for phytoplasma origin.

Table 4. PCR results for the amplification of selected TBB phytoplasma genes with primers based on random clones

Refer to Table 3 for primer sequences. + indicates that a PCR product was observed.

<table>
<thead>
<tr>
<th>Host</th>
<th>Symptoms</th>
<th>Phytoplasma</th>
<th>pTBB14F/R</th>
<th>pTBB23F/R</th>
<th>pTBB29F/R</th>
<th>pTBB32F/R</th>
</tr>
</thead>
<tbody>
<tr>
<td>Periwinkle</td>
<td>Virsence</td>
<td>SPLL-V4</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Periwinkle</td>
<td>Phyllody</td>
<td>TBB</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Papaya</td>
<td>Yellow crinkle</td>
<td>SPLL-V4</td>
<td>–</td>
<td>+</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Papaya</td>
<td>Yellow crinkle</td>
<td>TBB</td>
<td>+</td>
<td>+</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Grape</td>
<td>Late season leaf curl</td>
<td>TBB</td>
<td>+</td>
<td>–</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

**Cellular processes**

An ftsH gene was identified in TBB phytoplasma clones 32 and 78. The nucleotide sequences were identical, which indicates that the same gene was cloned twice. FtsH is a membrane-associated protein thought to be essential for cell growth (Kihara et al., 1998). PILEUP analysis of the ftsH genes from M. genitalium, H. pylori and TBB phytoplasma showed that the transmembrane domains of M. genitalium and H. pylori were located upstream from amino acids with similarity to the TBB phytoplasma sequence, indicating that the ftsH gene sequences from the TBB phytoplasma were partial sequences. This was further supported by the TMPRED and TOPPRED results which indicated an absence of transmembrane domains and the absence of zinc-binding sites at the C terminus of the TBB FtsH protein.

**Energy metabolism**

NADH oxidase mediates electron transfer from NADH to oxygen (Pollack et al., 1997). Oxidation of NADH produces NADH, which is a crucial intermediate in the Embden–Meyer–Parnas (EMP) pathway. The presence of NADH oxidase is not sufficient evidence to state that phytoplasmas have the EMP pathway, but all mollicutes have the EMP pathway to some degree (Pollack et al., 1997), so it may be reasonable to suggest that phytoplasmas do as well. The presence of the gene encoding biphosphoglycerate mutase (TBB phytoplasma clone 51), an enzyme associated with the EMP pathway, further supports the suggestion that phytoplasmas utilize the glycolytic pathway.

Partial sequences for several genes encoding the glucose-inhibited division A (GidA) protein were identified. Further analysis is required to determine whether the characterized TBB phytoplasma gidA genes align to form one complete gene sequence. The GidA protein stimulates cell division from Mycoplasma capricolum. The agreement between our results and those of Kakizawa et al. (2001) supports the annotation of the ORF in TBB phytoplasma clone 10 as secA. Members of the Sec protein group are involved in protein secretion and SecA is a membrane protein thought to form a channel to secrete pathogenicity factors and other proteins. The membrane association of the SecA protein and its biological role suggests that it is a potential drug target (Dandekar et al., 2000).

**Protein and peptide secretion**

The secA gene identified in random clone 10 was 62 % similar to a secA gene from the onion yellows phytoplasma (Kakizawa et al., 2001) and 52 % similar to a secA gene from Acholeplasma laidlawii. The agreement between our results and those of Kakizawa et al. (2001) supports the annotation of the ORF in TBB phytoplasma clone 10 as secA. Members of the Sec protein group are involved in protein secretion and SecA is a membrane protein thought to form a channel to secrete pathogenicity factors and other proteins. The membrane association of the SecA protein and its biological role suggests that it is a potential drug target (Dandekar et al., 2000).
in response to glucose (Ogawa & Okazaki, 1994). It is not known if the gidA gene utilizes the sugar or whether it is an environmental trigger (Ogawa & Okazaki, 1994).

Replication, transcription and translation

The TBB phytoplasma DNA polymerase beta III gene nucleic acid sequence was most similar to a gene in *Acholeplasma laidlawii*. This supports the current phylogenetic positioning of the phytoplasma subclade based on 16S rDNA sequences (Lim & Sears, 1992). The amino acid identity between the β subunit of DNA polymerase III (DnaN) of TBB phytoplasma and the corresponding gene in *A. laidlawii* was 56.0%. This lower than expected identity is consistent with other studies which have found that *dnaN* gene sequences are not highly conserved (Madiraju et al., 1999).

FASTX searches identified the closest match to the TBB phytoplasma *gyrB* gene as the corresponding gene from *B. subtilis*. However, phylogenetic analysis of *gyrB* genes showed that the TBB phytoplasma grouped with *Acholeplasma laidlawii* (bootstrap value of 100). The contrasting FASTX and phylogeny results were mostly likely due to the different program parameters.

The DNA polymerase subunit beta III is located near the origin of replication in both the mycoplasmas and eubacteria. Based on this it may be possible to locate the TBB phytoplasma origin of replication by hybridization of the gene encoding DNA polymerase subunit beta III to the TBB phytoplasma chromosome.

Peptide releasing factor 1 directs the termination of translation in response to the peptide chain termination codons UAG and UAA. It is located in the cytoplasm and belongs to the family of prokaryotic and mitochondrial release factors. The phytoplasma translation termination proteins and the genes associated with autonomous replication of DNA were 45-3% similar to those in bacteria, suggesting that phytoplasmas have DNA and protein synthesis pathways similar to other prokaryotic organisms. The gene for peptide releasing factor 2 (RF-2) was identified in clone 28. Function was not assigned to this clone because the similarity was below the threshold value. However, it should be noted that, unlike the mycoplasmas, the phytoplasmas have retained UGA as a stop codon (Razin et al., 1998) and therefore require the RF-2 protein for termination of translation. Only a partial gene was identified in this clone, so this may account for the low similarity.

**Confirmation that the characterized genes were not from host**

We designed PCR primers to amplify representative genes from the translation (peptide releasing factor-1), cellular processes (FtsH) and replication (DNA polymerase beta III and ATP-dependent RNA helicase) protein function categories. Two genes were selected from the replication group because the majority of the TBB genes identified were assigned to this group (Table 1). No PCR primers were designed for energy metabolism and protein secretion categories because the sequences were only 200 bp in length (Table 1). Furthermore, the grouping of the TBB phytoplasma *secA* gene with the onion yellows phytoplasma (Fig. 1) indicated that the clonal insert was not of plant origin. No genes were selected from the transcription and miscellaneous categories because the proteins encoded by these genes are involved in similar biological activities to the translation genes. A PCR product of correct size was amplified from the DNA of diseased plant samples and not from DNA of healthy plants, indicating that the genes identified in the random clones were from the TBB phytoplasma. The inability to consistently amplify a product from the TBB and SPLL-V4 phytoplasmas in papaya and grapevine may have been due to low titre, poor quality DNA or because the phytoplasma-based primers shared some similarity with papaya and grapevine plant genes. Products digested with *Alul* had identical RFLP patterns, indicating that the same gene was amplified from the TBB and SPLL-V4 phytoplasma isolates from different hosts. Enzymes *Hpal* and *Rsal* did not cut the PCR products amplified by the pTBB14, pTBB23, pTBB29 and pTBB32 primers. Further analysis of the PCR products by the inclusion of more restriction enzymes in RFLP analysis or sequencing of the PCR products may more accurately assess of the variability between the DNA polymerase beta III, RF-1, ATP-dependent helicase and *fisH* genes of the TBB and SPLL-V4 phytoplasma isolates.

**Conclusions**

The large number of genes identified in a small region of the TBB chromosome suggests that these organisms have an efficient genome. Assignment of function to the TBB phytoplasma ORFs was based on database analysis, so the results can only be as comprehensive as the sequence data available at the time. This has been demonstrated by work on the *M. genitalium* genome. Initial sequence analysis of the *M. pneumoniae* genome assigned two genes as glucose-inhibited division cell proteins (Gid) (Himmelreich et al., 1996). Reannotation of the genome showed these genes to be NADH oxidoreductase and methyltransferase (Dandekar et al., 2000). Three genes on the TBB phytoplasma chromosome were assigned the function of Gid proteins, but reannotation of these genes at a later date may change the function assigned to these genes. Functionality could not be assigned to any of the predicted ORFs and reannotation of some of these genes may provide insight into the proteins they encode. The amplification of genes from the TBB phytoplasma isolates and not DNA from healthy plants confirmed that the characterized genes were not from the host. The identical RFLP patterns of the genes of SPLL-V4 and TBB phytoplasmas indicated that these primers amplified the corresponding gene from the closely related phytoplasma SPLL-V4. The characterized TBB genes will form the basis for future research that will aim to isolate the corresponding genes from more distantly related...
phytoplasmas. Furthermore, the PCR primers designed to amplify these new TBB phytoplasma genes may be used to study the genetic diversity of phytoplasmas considered identical based on their 16S rRNA gene.

ACKNOWLEDGEMENTS

This research was funded by the Cooperative Research Center for Tropical Plant Protection, Brisbane, Australia. We thank Dr Bernd Schneider and Dr Anna Padovan for advice and mentoring.

REFERENCES


