Taro vein chlorosis virus: characterization and variability of a new nucleorhabdovirus

Peter Revill,† Xuan Trinh, James Dale and Rob Harding

Science Research Centre, Queensland University of Technology, GPO Box 2434, Brisbane, Queensland 4001, Australia

Sequencing of the monopartite RNA genome of a Fijian isolate of Taro vein chlorosis virus (TaVCV) confirmed that it is a definitive rhabdovirus with most similarity to members of the genus Nucleorhabdovirus. The TaVCV 12 020 nt negative-sense RNA genome contained six ORFs in the antigenomic sequence, equivalent to the N, P, 3, M, G and L genes that have been identified in other rhabdoviruses. The putative gene products had highest similarity to those of the nucleorhabdovirus Maize mosaic virus. A characteristic 3’-AAUUCUUUUGGGUUGU/A-5’ sequence was identified in each of the intergenic regions and the TaVCV leader and trailer sequences comprised 140 and 61 nt, respectively. Assignment of TaVCV to the genus Nucleorhabdovirus was supported by thin-section electron microscopy of TaVCV-infected taro leaves, which identified virions budding from nuclear membranes into the perinuclear space. Variability studies identified high levels of TaVCV sequence diversity. Within the L gene of 20 TaVCV isolates from Fiji, the Federated States of Micronesia, New Caledonia, Papua New Guinea, Solomon Islands and Vanuatu, maximum variability at the nucleotide level was 27-4 %. Within the N gene, maximum variability among 15 isolates at the nucleotide level was 19-3 %. The high level of TaVCV variability observed suggested that the introduction of TaVCV to the Pacific Islands was not a recent occurrence.

INTRODUCTION

Taro vein chlorosis virus (TaVCV) is a putative rhabdovirus that infects taro [Colocasia esculenta (L.) Schott], a staple food crop in many Pacific Island countries. TaVCV has not been characterized, but electron microscopy of sap dips has shown that virions have a bullet-shaped morphology (~ 210 x 65 nm) that is typical of rhabdoviruses (Pearson et al., 1999). TaVCV-infected plants exhibit a striking leaf-vein chlorosis symptom, particularly at the leaf margin, which often leads to necrosis of the affected tissue (Pearson et al., 1999). TaVCV is serologically distinct from another putative rhabdovirus that infects taro, Colocasia bobone disease virus (CBDV), initially named Taro large bacilliform virus (James et al., 1973; Pearson et al., 1999; Shaw et al., 1979). Based on electron microscopy and symptoms, TaVCV is distributed widely in the Pacific Islands, having been recorded in Fiji, Vanuatu and Solomon Islands (Pearson et al., 1999). TaVCV is not mechanically transmissible and the vector is unknown.

Plant-infecting rhabdoviruses are classified into two genera, Nucleorhabdovirus and Cytorhabdovirus, in the family Rhabdoviridae, order Mononegavirales (Walker et al., 2000). Members of the genus Nucleorhabdovirus replicate in the cell nucleus and bud from the nuclear membrane to accumulate in the perinuclear space, whereas members of the genus Cytorhabdovirus are more similar to animal-infecting rhabdoviruses in that they replicate and accumulate in the cytoplasm (Jackson et al., 1987). The complete nucleotide sequences of only five plant-infecting rhabdoviruses are available, namely the nucleorhabdoviruses Rice yellow stunt virus (RYSV) (Chen et al., 1998; Fang et al., 1994; Huang et al., 2003; Luo & Fang, 1998; Luo et al., 1998; Wang et al., 1999; Zhu et al., 1997), Sonchus yellow net virus (SYNV) (Choi et al., 1992; Goldberg et al., 1991; Heaton et al., 1987, 1989; Hillman et al., 1990; Scholthof et al., 1994), Maize fine streak virus (MFSV; GenBank accession no. AY618417) and Maize mosaic virus (MMV; AY618418) and the cytorhabdovirus Northern cereal mosaic virus (NCMV) (Tanno et al., 2000). Partial nucleotide sequences are also available for Lettuce necrotic yellows virus (LNYV) (Wetzel et al., 1994a, b) and Strawberry crinkle virus (SCV) (Posthuma et al., 2002). At least 50 putative plant rhabdoviruses, including TaVCV and CBDV, remain unassigned, as there are no sequence or cytological data to enable further classification.

Rhabdovirus genomes encode at least five major proteins:
the nucleocapsid (N), phosphoprotein (P), matrix protein (M), glycoprotein (G) and the RNA-dependent RNA polymerase (L). The N, P and L proteins interact with the genomic RNA to form a ribonucleoprotein core that is essential for virus replication, whereas the G and associated M protein constitute the major structural component of the virion shell. Rhabdovirus gene junctions contain near-identical short sequences of approximately 20 nt, which contain signals for transcription initiation, termination and polyadenylation of viral mRNA (reviewed by Neumann et al., 2002).

The genomes of all plant-infecting rhabdoviruses sequenced to date contain more than five ORFs. LNYV, MMV, RYSV and SYNV contain one additional ORF between the P and M genes, known as 4b (LNYV) (Wetzel and SYNV contain one additional ORF between the P and M genes, whereas NCMV contains four additional ORFs (genes 3–6) (Tanno et al., 2000). MPSV contains an additional ORF between gene 3 and the M gene (gene 4), and RYSV contains an additional ORF between the G and L genes (gene 6), the product of which has been detected in purified virions (Huang et al., 2003). In the nucleorhabdovirus SYNV, the sc4 protein (Scholthof et al., 1994) and G protein (Goldberg et al., 1991) have been identified in the phospholipid membrane surrounding the ribonucleoprotein core.

TaVCV is one of a number of viruses that infect taro and restrict the international movement of germplasm and for which a sensitive diagnostic test is a necessity. In addition to the putative rhabdoviruses TaVCV and CBDV, taro may also be infected with the potyvirus *Dasheen mosaic virus* (DsMV) (Maino, 2003) and/or the badnavirus *Taro bacilliform virus* (TaBV) (Yang et al., 2003a). A putative reovirus has also recently been identified in taro from Papua New Guinea (Devitt et al., 2001). The development of a reliable PCR-based diagnostic test requires knowledge of sequence variability to enable the design of primers that will detect variant virus isolates. Variability studies on Pacific Island isolates of TaBV identified up to 23% nucleotide (14% amino acid) variability in the reverse transcriptase/ribonuclease H-coding region and 31% nucleotide (20% amino acid) variability in the coat protein (Yang et al., 2003b). Maino (2003) observed up to 21% nucleotide and deduced amino acid variability in the coat protein sequences of Pacific Island isolates of DsMV. Despite the high level of variability, Yang et al. (2003b) developed degenerate TaBV PCR primers that enabled detection of TaBV isolates from all Pacific Island countries. The only variability study of plant rhabdoviruses has been on SCV, in which Klersk et al. (2004) identified up to 11% nucleotide and deduced amino acid variability in a 1.6 kbp fragment of the L gene sequence. PCR protocols were subsequently developed that detected all SCV isolates. There have been no other variability studies of plant rhabdoviruses and the sequence diversity of TaVCV in the Pacific Islands is unknown.

In this paper, we have presented the complete nucleotide sequence of a Fijian isolate of TaVCV and identified the site of TaVCV maturation in infected cells. We have also reported sequence variability in the L and N genes among Pacific Island isolates and compared the TaVCV sequence with that of other rhabdoviruses.

**METHODS**

**Virus purification, RNA isolation and cDNA synthesis.** Leaves were collected from taro plants exhibiting characteristic vein chlorosis symptoms from near Suva, Fiji. To eliminate the possibility that the plants were co-infected with another rhabdovirus such as CBDV, leaves were tested by RT-PCR using in-house PCR protocols (M. L. Dowling, P. Revill, J. Dale and R. Harding, unpublished data). Electron microscopy of sap dips from infected leaves also confirmed the presence of a homogeneous virus population. Virions were purified at 4°C by using the method of Hsu & Black (1968) and RNA was isolated by adding SDS (1% final concentration) and incubating at 37°C for 1 h. Following extraction in acid phenol, phenol/chloroform and chloroform, the purified RNA was precipitated and resuspended in water. cDNA copies of TaVCV RNA were generated with random primers by using an Invitrogen Superscript II cDNA synthesis kit, following the manufacturer’s instructions. The resultant dsDNA was ligated into Smal-digested pUC18, and *Escherichia coli* JM109 cells were transformed by using standard molecular techniques. Cloned inserts were sequenced by an ABI Prism Big Dye Terminator Cycle Sequencing Ready Reaction kit (PE Applied Biosystems) at the Australian Genomic Research Facility, University of Queensland, Australia. Initial sequences were determined with universal forward (M13-20) and reverse (M13 reverse) primers. Sequences were analysed by using the BLAST X program available on the National Center for Biotechnology Information website (http://www.ncbi.nlm.nih.gov/).

**RT-PCR.** Two clones from the cDNA library showed similarity to sequences encoded by nucleorhabdoviruses, namely the G and L genes of RYSV and SYNV, respectively. Based on the RYSV and SYNV genomes, these clones were assumed to be approximately 4 kb apart on the TaVCV genome and were designated G1 and L1. To confirm that the L1 clone was TaVCV-specific, two primers, TaVCV1 and TaVCV2 (Table 1), were designed 220 nt apart in the L1 sequence and used in Titan (Roche) one-step RT-PCRs, using total RNA extracted from TaVCV-symptomatic and asymptomatic taro (RNeasy kit; Qiagen) as template. Reactions contained 1· Titan

<table>
<thead>
<tr>
<th>Primer name</th>
<th>Sequence (5′→3′)</th>
</tr>
</thead>
<tbody>
<tr>
<td>TaVCV1</td>
<td>ATATCCAGCTTATTCATTCACTGAC</td>
</tr>
<tr>
<td>TaVCV2</td>
<td>TG CCTG GGGCTTCTGGAGTATCTG</td>
</tr>
<tr>
<td>Glycoplus2</td>
<td>TAAC AGA ATC GAG CT A C T G T CC</td>
</tr>
<tr>
<td>Glycominus2</td>
<td>ACAC ATG A C T C T CG T T G G CAT CC</td>
</tr>
<tr>
<td>190Grminus</td>
<td>GCA C A C T C C C G C C G AT C G G G CTAT</td>
</tr>
<tr>
<td>160Grminusnest</td>
<td>GTC G C T T G G G G C T CAT C C C A A T</td>
</tr>
<tr>
<td>1572GrPlus</td>
<td>GGC A A AT G T G G G G C G CT G G A GA G A</td>
</tr>
<tr>
<td>1572GRNested</td>
<td>TGG AA G A C T G G T G G G A T C T G T G CA</td>
</tr>
<tr>
<td>Pol2A1</td>
<td>AAT ATG C T C C C A G T T C A C C C</td>
</tr>
<tr>
<td>Pol2A2</td>
<td>AGGT G CT C A T A T AG C T C A G T T C C</td>
</tr>
<tr>
<td>Cap2A</td>
<td>AC G A G C C A C T AT A G C T C C G A T C G</td>
</tr>
<tr>
<td>Cap2B</td>
<td>T T G G G G A T G T C G C C T C C C G G</td>
</tr>
</tbody>
</table>

Table 1. TaVCV-specific primers used in this study
reaction buffer, 1 mM dNTPs, 20 pmol each primer and 0.5 μl Titan polymerase mix. The reactions were incubated at 42 °C for 30 min, denatured at 94 °C for 2 min and subjected to 35 cycles of 94 °C for 30 s, 55 °C for 30 s and 68 °C for 2 min, with a final extension of 68 °C for 10 min. Amplicons were separated by electrophoresis through a 2% TAE/agarose gel and purified by using a QIAx II (Qiagen) gel extraction kit. Amplicons were ligated into pGEM-TE vectors and plasmids were cloned into E. coli JM109 cells, purified and sequenced.

To amplify intervening sequences between G1 and L1, a sense PCR primer located at the 3' end of the G1 sequence (Glycominus2; Table 1) was used in a Titan RT-PCR with primer TaVCV2, located in the L1 sequence. TaVCV virion RNA was used as template and reaction mixes were set up as described previously. The reaction mix was incubated at 42 °C for 30 min, denatured at 94 °C for 2 min and then subjected to 10 cycles of 94 °C for 30 s, 45 °C for 30 s and 68 °C for 2 min, followed by 35 cycles of 94 °C for 30 s, 50 °C for 30 s and 68 °C for 2 min, with a final extension of 68 °C for 10 min. Amplicons were purified, cloned and sequenced as described previously.

Amplification of additional TaVCV sequences. Although the predicted ~4 kbp amplicon spanning the G1 and L1 sequences was not obtained, smaller TaVCV-specific amplicons of up to 2.5 kbp were amplified and cloning and sequencing showed that some of these products shared sequence similarity with RYSV and SYNV. One clone was also obtained that had similarity to the RYSV and SYNV L gene, downstream of the region in RYSV and SYNV that was analogous to the L1 sequence. This downstream sequence was assumed to have been amplified by mispriming. To determine whether additional TaVCV sequences could be obtained by mispriming, an antisense primer (Glycominus2; Table 1), designed by using the G1 sequence, was used alone in a Titan RT-PCR as described above. Additional contiguous TaVCV sequences were obtained upstream and downstream of the G1 sequence and mispriming with sense or antisense primers designed at the termini of each newly obtained sequence was subsequently used to obtain additional TaVCV sequences. Intervening sequences up to 3 kbp in length were also amplified by using paired TaVCV-specific primers.

Leader and trailer sequences. The 5′-trailer sequence was obtained by using 5′ RNA ligase-mediated rapid amplification of cDNA ends (RACE) from 1 μg TaVCV virion RNA, which had been dephosphorylated and decapped by using a GeneRacer kit (Invitrogen) according to the manufacturer’s instructions. The RNA was ligated to a GeneRacer RNA oligonucleotide, precipitated with ethanol and 5 μl was used as template in a cDNA synthesis reaction with 20 pmol primer 1572GRPlus (Table 1). The RNA was hydrolysed with NaOH and the cDNA was precipitated with ethanol. The cDNA was resuspended in 10 μl water and 1 μl was used as template in a PCR with 10 pmol 1572GRPlus primer and 30 pmol GeneRacer 5′ RACE primer. Nested RACE was subsequently performed by using 5 μl of the first-round PCR product as template, with 30 pmol GeneRacer oligo 3′-nested primer and 10 pmol primer 160GRminusnest (Table 1). The 5′-trailer and 3′-leader amplicons were cloned and sequenced as described previously.

Sequence analysis. Each sequence was confirmed in at least three clones in both orientations and a contiguous sequence was generated by using the Seqman program (DNASTAR). Nucleotide and deduced amino acid sequences were analysed by using Editseq (DNASTAR) and Vector NTI software and programs available at the EXPASY website (http://au.expasy.org).

The TaVCV L gene sequence was compared with L and polymerase gene sequences from a number of viruses with negative-sense RNA genomes by using the CLUSTAL_X alignment program (Thompson et al., 1997). An unrooted neighbour-joining tree was constructed by using the TREEVIEW program (Page, 1996). In addition, the complete TaVCV nucleotide sequence was compared with the nucleotide sequences of the five characterized plant-infecting rhabdoviruses, MFNV, MMV, RYSV, SYNV and NCMV.

Amino acid sequencing of virion-associated proteins. Virion polypeptides were separated by SDS-PAGE and transferred to ProBlott membrane (Applied Biosystems). The membrane was stained with Coomassie blue and a major band with an approximate molecular mass of 70 kDa, typical of that expected for a rhabdoviral glycoprotein, was excised and sequenced at the Department of Biochemistry and Molecular Biology, University of Queensland, Australia.

Electron microscopy. Frozen taro leaf material stored at −80 °C was thawed in 3% glutaraldehyde (in cacodylate buffer, pH 7.2) and processed according to standard methods (Hall & Hawes, 1991). Ultrathin sections were examined and photographed with a JEOL 1200EX transmission electron microscope.

TaVCV variability studies. Taro leaves showing TaVCV symptoms, as well as asymptomatic controls, were collected from the Federated States of Micronesia, Fiji, New Caledonia, Papua New Guinea, Solomon Islands and Vanuatu.

Total RNA was extracted by using an RNeasy (Qiagen) RNA extraction kit, according to the manufacturer’s protocol. TaVCV-specific primers, designed to amplify a 1 kbp fragment of the TaVCV L gene (Pol2A1/Pol2A2; Table 1) or 1·1 kbp fragment of the N gene (Cap2A/Cap2B; Table 1) were used in Titan (Roche) one-step RT-PCRs as described previously. The cycling parameters were 42 °C for 35 min and 94 °C for 2·5 min, followed by 35 cycles of 94 °C for 30 s, 55 °C for 30 s, 68 °C for 30 s and then 68 °C for 10 min.

PCR products were gel-purified, cloned and sequenced as described previously. Consensus sequences from two clones for each isolate were obtained by using the SeqMan program and pairwise sequence variability was determined by using the MEGALIGN program (DNASTAR).

RESULTS

Nucleotide sequence and coding regions

Initial sequencing of a cDNA library prepared from purified TaVCV RNA identified two clones (G1 and L1) with deduced amino acid sequences similar to plant rhabdovirus sequences, namely the G and L genes of RYSV and SYNV. RT-PCR using primers designed to amplify 220 bp of the L1 sequence only amplified the predicted amplicon from taro plants exhibiting veinal chlorosis symptoms (data not shown), indicating that the L1 cDNA clone was viral in
origin. Additional TaVCV sequences were amplified by RT-PCR using TaVCV-specific primers, either individually in mispriming reactions or in pairs that flanked intervening sequences. In total, 272 overlapping sequences were generated and at least three clones for each new sequence were sequenced in both orientations. When the consensus sequences of the clones were aligned, the TaVCV genome comprised 12,020 nt and contained six major ORFs in negative polarity (Fig. 1). By analogy to other plant-infecting rhabdoviruses, the TaVCV genome arrangement (in coding polarity) was 3'-leader-N-P-3-M-G-L-trailer-5'.

The six ORFs comprised 1506 (N), 813 (P), 861 (3), 705 (M), 1764 (G) and 5784 (L) nt.

The deduced amino acid sequences of the TaVCV ORFs had 63-1 (N), 46-1 (P), 43-4 (gene 3), 46-4 (M), 49-9 (G) and 67-9 (L) % identity with the deduced amino acid sequences of the respective MMV ORFs. Identities with the deduced amino acid sequences of other rhabdoviruses were no higher than 35-1 % (RYSV, L gene product) and ranged from 10-5 to 26-4 % for the N, G and L gene products. Although the deduced amino acid sequences of the N, G and L ORFs did have similarity to those of other rhabdoviruses, the P, gene 3 and M gene products were only similar to those of MMV.

In addition to sharing sequence similarity with MMV, the deduced amino acid sequences of all TaVCV ORFs, except ORF3, shared some features with the gene products of other rhabdoviruses. For example, the putative 55-4 kDa ORF1 gene product had three groups of basic amino acids at its carboxyl terminus (402RGTKR406, 421HPTKKTWRK429 and 462RGKHWR466) that were similar to the nuclear localization signals encoded by the N genes of other nucleorhabdoviruses. ORF2 encoded a putative 30-4 kDa polypeptide and comparison of the deduced amino acid sequence with the analogous P gene sequence from SYNV showed that both proteins had a conserved hydrophilic core that, in TaVCV, lay between residues 108 and 200 and contained eight serine and 11 threonine residues. Although the putative 31-9 kDa ORF3 polypeptide was 43-4 % identical to gene 3 of MMV, it had no similarity to proteins encoded by other viruses. The putative 26 kDa TaVCV ORF4 gene product encoded a cluster of basic residues (69HIIIRNK75) in the amino-terminal region of the gene upstream of a YXG motif and a basic region at the carboxyl terminus of the gene, similar to the deduced M gene products of both animal- and plant-infecting rhabdoviruses. ORF5 encoded a putative 65-6 kDa polypeptide and the deduced amino acid sequence contained six glycosylation signals (N-X-S/T). Direct N-terminal protein sequencing of a 70 kDa virion-associated polypeptide revealed that the N-terminal amino acid sequence of TaVCV ORF5 at a position commencing 24 residues downstream of the first methionine (MSILAVILPILSVEPVNAG-LVVDLRRN), indicating that the 70 kDa protein was

![Fig. 1. Genome organization of TaVCV, in coding orientation.](Image)

![Fig. 2. Neighbour-joining phylogenetic tree of the entire TaVCV L gene deduced amino acid sequence compared with the complete or partial (SCV) polymerase sequences of selected viruses with negative-sense RNA genomes. Bootstrap values (from 1000 replicates) are shown at select nodes. Plant-infecting rhabdoviruses are indicated by the vertical line and nucleorhabdoviruses are circled. The virus names and GenBank accession numbers of sequences used in the analysis were: ABLV (Australian bat lysavirus), AAN05310; APMV-6 (Avian paramyxovirus 6), NP_150063; BDV (Borna disease virus), CAC70659; IHNV (Infectious hematopoietic necrosis virus), CAA52076; LBVV, AB075039; MFSV, AY618417; MMV, AY618418; NCMV, NP_597914; NDV (Newcastle disease virus), AAR99852; RABV, BAC53869; RYSV, NC_003746; SCV, AAQ97583; SHRV (Snakehead rhabdovirus), NP_050585; SYN1, NP_042286; TaVCV, AY674964; VSIV (Vesicular stomatitis Indiana virus), AAN16984.)

[Diagram of Plant rhabdoviruses]
encoded by ORF5. The putative 217-3 kDa ORF6 gene product contained the conserved $^{710}$GDN$^{712}$ motif present in the L gene of all viruses with negative-sense RNA genomes. It had closest similarity to rhabdovirus L gene products and also shared 24-6 % similarity with the RNA-dependent RNA polymerase gene product of Lettuce big-vein virus (LBVV), a virus with a negative-sense bipartite genome in the genus Varicosavirus (Sasaya et al., 2004).

**Phylogenetic analysis**

Phylogenetic analysis of the complete TaVCV L gene deduced amino acid sequence showed that it was related most closely to the L gene of plant nucleorhabdoviruses, particularly MMV (Fig. 2). All nucleorhabdoviruses grouped together as one clade, although the inclusion of MFSV in this group had low bootstrap support (32-5 %). The two cytorhabdoviruses, SCV and NCMV, formed a separate clade, which had high bootstrap support (99-7 %). Phylogenetic analysis of the complete TaVCV nucleotide sequence compared with the five sequenced plant rhabdoviruses (MMV, MFSV, RYSV, SYNv, NCMV) also showed that TaVCV was most similar to MMV (data not shown).

**Leader and trailer sequences**

The 3'-leader and 5'-trailer sequences comprised 140 and 61 nt, respectively, and 9 of the 13 nt at each terminus of the genome were complementary, with the terminal 3 nt (UCU/AGA) being exact complements. The leader and trailer sequences had low overall sequence identity to those of other rhabdoviruses, with closest identity to SYNv (34-4 %, leader) and RYSV (30-5 %, trailer), respectively. The TaVCV 5'-trailer sequence is the smallest identified to date, being closest in size to that of MMV (93 nt). As the 3'-leader sequence was obtained by using artificially polyadenylated virion RNA as template, it was not possible to determine the terminal nucleotide as it was assumed that all As had been added during the RACE procedure. The tetranucleotide sequence UGUU found in the 3'-leader sequences of all rhabdoviruses was located at nt 9–12.

**Gene junctions**

A repeat sequence of AAUUCUUUUGGGUUGU/A was identified between each of the TaVCV ORFs except at the junction of the L gene and the 5' trailer, where the sequence was AAUUCUUUUGGG. This was almost identical to the MMV intergenic sequence and was also similar to the intergenic sequences of the other nucleorhabdoviruses, SYNv and RYSV (Table 2).

**Electron microscopy**

Thin-section analysis of TaVCV-infected leaf tissue showed that large numbers of virions with sizes of approximately $200 \times 70$ nm accumulated in the nucleus, budding into the perinuclear space between the inner and outer membranes of the nuclear envelope (Fig. 3a and b).

<table>
<thead>
<tr>
<th>Virus</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>TaVCV</td>
<td>AAUUCUUUUGGGUUGU/A</td>
</tr>
<tr>
<td>MMV</td>
<td>AAUUCUUUUGGGUUGW</td>
</tr>
<tr>
<td>SYNv</td>
<td>AUUCUUUUGG-UUGW</td>
</tr>
<tr>
<td>RYSV</td>
<td>AUUCUUUUGGGUUGUG</td>
</tr>
<tr>
<td>NCMV</td>
<td>AUUCUUUUGACUCUAGU</td>
</tr>
</tbody>
</table>

**Table 2. Consensus intergenic sequences of characterized plant rhabdoviruses in virion-sense orientation**

W represents the nucleotide A or U.

Fig. 3. Electron micrographs of cross-sections through a TaVCV-infected cell (a) showing the nucleus (N) containing TaVCV virions (V). Bar, 500 nm. The boxed area is enlarged in (b) and shows TaVCV virions (V) in the perinuclear space between the inner membrane (M) and outer membrane (E) of the nuclear envelope. Bar, 200 nm.
Variability studies

A 1 kbp fragment of the TaVCV L gene, encompassing a region containing the conserved GDN motif, was amplified from 19 isolates collected from the Federated States of Micronesia (n = 2), Fiji (n = 4), New Caledonia (n = 1), Papua New Guinea (n = 4), Solomon Islands (n = 3) and Vanuatu (n = 5), in addition to the Fijian genomic-length sequence (GenBank accession no. AY674964). A 1-1 kbp fragment of the TaVCV N gene, representing approximately 75 % of the gene and including conserved sequences that shared similarity with other plant-infecting rhabdoviruses, was amplified from 14 isolates derived from Fiji (n = 3), New Caledonia (n = 3), Papua New Guinea (n = 2), Solomon Islands (n = 2) and Vanuatu (n = 4).

Comparison of the L gene sequences of all Pacific Island isolates identified a maximum variability of 27.4 % at the nucleotide level, between isolates FM47 from the Federated States of Micronesia and V4 from Vanuatu (Table 3). The maximum variability of the deduced amino acid sequences was 11.3 % (data not shown). The N gene was less variable, with maximum nucleotide variability of 19.3 % between isolates V40 and GenBank accession no. AY674964 (6.3 % variability at the amino acid level; data not shown). When the L gene sequence variability of TaVCV isolates within each country was examined, most variability was observed in Papua New Guinea and Vanuatu, with up to 24.2 % and 23.9 % nucleotide difference, respectively. Within Solomon Islands, samples varied by up to 14.3 %, whilst the two isolates from the Federated States of Micronesia varied by 16.1 %. The least variability was observed in Fijian isolates, where most isolates varied by no more than 3 %. The exception was isolate F20 from Viti Levu, which was up to 13.9 % different from the other Fiji isolates and was related more closely to isolates S5 and S10 from Solomon Islands. Phylogenetic analysis of the entire TaVCV L gene nucleotide sequences showed that, although isolates generally grouped according to geographical location, there were numerous exceptions, particularly with isolates from Papua New Guinea, Vanuatu and Solomon Islands (Fig. 4).

DISCUSSION

This is the first report of the TaVCV genome sequence and confirms that it is a rhabdovirus in the genus Nucleorhabdovirus. The classification of TaVCV in this genus was supported by thin-section electron microscopy, which identified virions accumulating in the nuclei of infected cells and budding into the perinuclear space between the nuclear membrane and the nuclear envelope.

The similarities of the TaVCV deduced amino acid and intergenic sequences with those of MMV indicated that TaVCV and MMV appeared to be the two most closely related plant rhabdoviruses sequenced to date. Most ORFs shared at least twice the amino acid sequence identity compared with the next most similar rhabdovirus, with a minimum identity of 43.1 %. In addition, gene 3 and the

<table>
<thead>
<tr>
<th>Table 3. Percentage nucleotide sequence identity of TaVCV L gene sequences from Pacific Island isolates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Origin of isolates: GenBank accession no. AY674964, F2, F16 and F20 (Viti Levu, Fiji); F2918 (Vanua Levu, Fiji); FM47 and FM92 (Federated States of Micronesia); NC33 (New Caledonia); P30, P44, P52 and P71 (Papua New Guinea); S5, S10 and S12 (Guadalcanal, Solomon Islands); V4 and V9 (Efate, Vanuatu); V16 (Tanna, Vanuatu); V36 and V40 (Santo, Vanuatu).</td>
</tr>
</tbody>
</table>

| F16 | F2 | F20 | F2918 | FM47 | FM92 | NC33 | P30 | P44 | P52 | P71 | S10 | S12 | S5 | V16 | V36 | V4 | V40 | V9 |
| 97-6 | 97-8 | 86-1 | 97-9 | 75-4 | 75-5 | 78-2 | 75-5 | 87-5 | 88-3 | 85-6 | 97-9 | 85-6 | 77-0 | 75-2 | 76-9 | 87-3 | 75-5 | AY674964 |
| 99-3 | 86-9 | 99-4 | 76-5 | 76-6 | 78-7 | 75-7 | 88-3 | 89-2 | 89-9 | 86-4 | 94-9 | 86-4 | 97-3 | 75-6 | 76-9 | 87-0 | 75-9 |
| 87-1 | 99-7 | 76-4 | 76-7 | 79-0 | 76-0 | 88-5 | 89-4 | 89-4 | 86-6 | 99-7 | 86-6 | 77-7 | 75-8 | 77-6 | 86-3 | 76-2 | 76-4 |
| 87-2 | 78-0 | 77-8 | 77-3 | 76-1 | 86-2 | 83-9 | 85-3 | 98-7 | 87-2 | 98-7 | 76-9 | 75-8 | 76-8 | 85-5 | 75-4 | F20 |
| 87-4 | 76-7 | 78-9 | 75-9 | 88-6 | 89-5 | 86-7 | 99-8 | 86-7 | 77-6 | 75-7 | 77-6 | 88-3 | 76-1 | F2918 |
| 83-9 | 76-4 | 74-5 | 76-7 | 77-1 | 76-2 | 78-6 | 75-6 | 76-5 | 78-6 | 72-8 | 74-3 | 72-6 | 76-0 | 73-8 | 73-9 | 73-8 | 73-9 |
| 78-5 | 76-9 | 77-5 | 77-9 | 78-1 | 78-6 | 76-8 | 76-8 | 76-9 | 77-5 | 76-8 | 77-3 | 76-6 | 73-8 | FM47 |
| 86-4 | 76-7 | 76-5 | 77-6 | 77-5 | 77-8 | 79-0 | 77-8 | 88-8 | 86-4 | 88-6 | 76-1 | 86-2 | 73-4 | 73-3 | NC33 |
| 76-2 | 75-8 | 76-8 | 76-8 | 76-0 | 76-8 | 72-7 | 87-4 | 99-4 | 87-1 | 76-1 | 95-1 | P30 |
| 90-6 | 94-0 | 87-3 | 88-6 | 87-3 | 77-6 | 75-9 | 77-5 | 93-0 | 75-6 | 75-6 | 75-6 | P44 |
| 90-9 | 84-4 | 89-5 | 84-4 | 76-5 | 75-7 | 76-4 | 89-8 | 76-8 | 73-8 | 75-8 | 75-8 | P52 |
| 86-2 | 89-5 | 86-2 | 81-7 | 78-1 | 78-0 | 96-7 | 76-2 | P71 |
| 86-7 | 100 | 77-7 | 76-5 | 75-6 | 86-2 | 75-7 | 87-0 | S10 |
| 86-7 | 77-7 | 75-8 | 77-8 | 88-3 | 76-2 | 76-2 | S12 |
| 77-7 | 76-5 | 77-6 | 86-2 | 75-7 | S5 |
| 87-2 | 99-9 | 76-9 | 86-7 | 76-1 | 76-1 | 85-6 | V16 |
| 87-1 | 76-1 | 95-2 | V36 |
| 76-8 | 86-6 | V4 |
| 76-2 | V40 |
M gene of TaVCV and MMV were over 40% identical, whereas those of other plant rhabdoviruses showed no relationship to any sequences in GenBank. Although both taro and maize are monocots, taro is non-graminaceous and is vegetatively propagated. The MMV vector is the leafhopper, *Peregrinus maidis*, although it is unknown whether taro is a host for *P. maidis*. The vector of TaVCV is also unknown, although another taro-infecting putative rhabdovirus, CBDV, is spread by the taro planthopper, *Tarophagus proserpina* (Shaw et al., 1979). The close phylogenetic relationship of TaVCV and MMV, as well as the high level of deduced amino acid sequence identity, suggests that these viruses may have shared a common ancestor relatively recently. Both viruses are present in the Pacific Islands, with MMV first recorded in Hawaii in 1927 (Brunt et al., 1990). We have also observed MMV-like symptoms on maize growing in Vanuatu in areas near TaVCV-infected taro. The phylogeny of rhabdovirus L proteins suggests a monophyletic origin and Hogenhout et al. (2003) speculated that an insect was the primary host of the rhabdovirus ancestor. The similarity of the deduced amino acid sequences of TaVCV and MMV suggests that an ancestral leafhopper/planthopper-transmitted virus may have been their progenitor.

The deduced amino acid sequences of the N, P, M, G and L genes shared a number of other features with analogous rhabdovirus genes. These included a series of basic residues in the carboxyl-terminal portion of the N gene, similar to the nuclear localization signals identified in the N gene of SYNV and RYSV (Goodin et al., 2001), and a central core of hydrophilic residues in the P gene, similar to SYNV (Heaton et al., 1987).

The position of the TaVCV gene 3 was analogous to that of the sc4 gene of SYNV and gene 3 of MMV and RYSV. Melcher (2000) proposed that the analogous sc4 gene in SYNV was a putative movement protein, based on its secondary structure and similarity to proteins in the 30K superfamily of viral movement proteins. The possible role of the sc4 gene product as a movement protein was supported by Goodin et al. (2002), who showed that sc4 fusion proteins were targeted primarily to the periphery of epithelial cells. It is unknown whether the TaVCV gene 3 encodes a movement protein, as it had no obvious similarity to proteins in the 30K superfamily. The presence of a gene in this position in the genome of the *Drosophila*-infecting rhabdovirus Sigma virus (Teninges et al., 1993), which does not require a movement protein, shows that it is not possible to attribute a function to TaVCV gene 3 based on its genome position alone.

The TaVCV M gene shared some features with the M gene products of both animal- and plant-infecting rhabdoviruses. This included a cluster of basic residues in the amino-terminal region of the gene upstream of a YXG motif, which was present in the animal-infecting rhabdoviruses *Rabies virus* (RABV) (Rayssiguier et al., 1986) and *Vesicular stomatitis virus* (new Jersey serotype; VSNJV) (Gill & Banerjee, 1986), but was not present in the RYSV or SYNV M gene (Hillman et al., 1990; Luo et al., 1998). However, like SYNV and RYSV, the TaVCV M gene product had a basic region at the carboxyl terminus, which was absent from VSNJV and RABV.

Based on the presence of six N-X/S/T glycosylation signals in the deduced amino acid sequence, the TaVCV ORF5 is presumed to encode the G protein. An identical number of glycosylation signals is present in the deduced glycoproteins of SYNV (Goldberg et al., 1991), MMV and NCMV, whereas that of RYSV contains 10 signals (Luo & Fang, 1998). The TaVCV N-terminal G sequence was identified 24 residues downstream of a signal peptide sequence in the deduced ORF5 sequence, similar to the nucleorhabdovirus RYSV, where the amino terminus of the mature glycoprotein was identified immediately downstream of a 32 aa signal peptide sequence (Luo & Fang, 1998).

TaVCV ORF6 encoded an L protein with a molecular mass almost identical to that of MMV (217 kDa). The deduced amino acid sequence of the TaVCV L protein shared large regions of conserved amino acids with the polymerase genes of rhabdoviruses and other viruses in the order Mononegavirales, including the conserved GDN motif, which is equivalent to the GDD motif present in
the polymerases of viruses with positive-sense genomes (Koonin, 1991; Koonin & Dolja, 1993). Interestingly, despite the high level of conservation, degenerate PCR primers designed by Posthuma et al. (2002) to enable amplification of a ~700 bp region of the L gene from SCV failed to amplify the predicted product from all TaVCV isolates (data not shown), although the primer sequences were present in the isolate with GenBank accession no. AY674964. This suggested the presence of considerable sequence diversity in the TaVCV L gene, which was supported by our variability studies. However, it was not possible to determine the extent of TaVCV variability at the primer-binding sites of Posthuma et al. (2002), as they were located downstream of the region analysed in our variability studies. It should also be noted that Posthuma et al. (2002) failed to amplify SCV from all infected samples tested and Klers et al. (2004) identified up to 11% sequence variability in isolates from Germany and the Netherlands.

Although the TaVCV 3′-leader sequence showed low sequence identity to sequences of other rhabdoviruses, it contained a conserved rhabdovirus UUGU tetranucleotide sequence, although its position (nt 9) was closer to the 3′ terminus than the tetranucleotides of other nucleorhabdoviruses (SYNV, nt 22; RYSV, nt 62; Wang et al., 1999). The TaVCV 61 nt 5′-leader sequence was the shortest of all plant-infecting rhabdoviruses sequenced to date, which range from 93 (MMV) to 273 (NCMV) nt. It was most similar in length to the 5′-trailer of VSNJV (59 nt), although the 3′-leader sequences of TaVCV and VSNJV were markedly different in length (47 nt for VSNJV).

TaVCV variability studies showed that, within the partial (1·1 kbp) N gene sequence, maximum variability of 19·3 and 6·3% was observed at the nucleotide and amino acid levels, respectively, whilst within the 1 kbp fragment of the L gene, the variability was even greater, at 27·4 and 11·3%, respectively. This amino acid variability was almost identical to that observed for SCV (Klers et al., 2004), although the TaVCV L gene nucleotide variability was much higher. It remains to be determined whether the level of variability observed in the putative TaVCV polymerase and capsid genes is reflected throughout the genome. Taro is an ancient vegetatively propagated crop, having been cultivated in Papua New Guinea since the early Holocene period (Denham et al., 2003). The high level of TaVCV sequence variability observed within each country suggests that TaVCV has also been in these countries for a considerable period. Phylogenetic analysis showed that TaVCV isolates generally grouped according to geographical location, although there were exceptions. This probably reflects the movement of virus-infected taro-propagating material throughout the Pacific Islands, a common practice for hundreds of years. The variability studies undertaken herein will enable the development of degenerate primers for use in a PCR-based diagnostic test to detect variant isolates of TaVCV throughout the Pacific Islands. This will allow the international distribution of virus-indexed germplasm and improve both the quality and range of taro available to all Pacific Island countries.

ACKNOWLEDGEMENTS

This work was funded by the Australian Centre for International Agricultural Research (ACIAR). We wish to thank Grahame Jackson (Australia), Geoff Wiles (Papua New Guinea), Jimi Sealea (Solomon Islands), Benuel Tarilongi (Vanuatu), Kay Kumar (Fiji), Didier Varin (New Caledonia), Albert Peters (Samoa), Apefa’I Taifane (American Samoa) and their staff for assistance with collecting plant samples. We also thank the many farmers who allowed us access to their gardens and Jacqui Wright (Secretariat of the Pacific Community) for providing samples from the Federated States of Micronesia. We also thank Dr Deb Stenzel [Analytical Electron Microscopy Unit, Queensland University of Technology (QUT)] for the electron microscope work, Dr Jonathon Harris (QUT) for assistance with the protein analysis and Mr Luke Devitt (QUT) for technical assistance.

REFERENCES


