Arch Virol (2003) 148: 1523–1541
DOI 10.1007/s00705-003-0109-6

Archives of Virology
Printed in Austria

The complete nucleotide sequence of two distinct geminiviruses infecting cucurbits in Vietnam*

P. A. Revill[1], **C. V. Ha**[2], **S. C. Porchun**[1], **M. T. Vu**[2], and **J. L. Dale**[1]

[1] Plant Biotechnology, Science Research Centre, Queensland University of Technology, Brisbane, Australia
[2] Department of Plant Pathology, Hanoi Agriculture University, Gia Lam, Hanoi, Vietnam

Received September 4, 2002; accepted March 7, 2003
Published online June 5, 2003

Summary. We have characterised two distinct geminiviruses that infect cucurbit cultivars in Vietnam. The genomes of both viruses consisted of two circular ssDNA components (DNA-A and DNA-B), with a genome arrangement and coding sequence typical of viruses in the *Begomovirus* genus in the family *Geminiviridae*. The sequence of DNA-A of one of the viruses was approximately 97% similar to *Squash leaf curl virus-China* (SLCV-Ch), for which a DNA-B has yet to be identified. We have named this virus Squash leaf curl virus-Vietnam (SLCV-Vn). The intergenic region of the SLCV-Vn DNA-B contained a 40 nt deletion between the putative AC1 TATA box and the stem loop. A second virus isolated from loofa in southern Vietnam was only 80% similar to SLCV-Vn over the complete DNA-A sequence, however the nucleotide sequence in the coat protein coding regions was 95% similar. We have named this virus Loofa yellow mosaic virus-Vietnam (LYMV-Vn). Other regions of the SLCV-Vn and LYMV-Vn genomes differed markedly, suggesting the coat protein coding region was recombinant. The DNA-B of both viruses were only 60% similar over the complete nucleotide sequence, although the encoded amino acid sequence of the BC1 gene was 90% identical.

*The GenBank accession numbers for DNA-A and DNA-B of SLCV-Vn isolate B are AF509743 and AF509742 respectively. The DNA-A sequence of SLCV-Vn isolate K has been assigned the accession number AF509741. The DNA-A sequence of LYMV-Vn isolate M has been assigned AF509739, and the LYMV-Vn DNA-B sequence has been assigned AF509740.

Introduction

The *Geminiviridae* is a family of plant viruses with circular single-stranded DNA genomes encapsidated in twinned particles. Geminiviruses are currently classified into one of four genera, *Mastrevirus*, *Curtovirus*, *Topocuvirus*, and *Begomovirus* [6, 35], based on their genome arrangement and biological properties. Begomoviruses are transmitted by whiteflies to a wide range of dicotyledonous plants and most have bipartite genomes, the two components known as DNA-A and DNA-B. DNA-A has either one or two open reading frames (ORF) in the virion sense (AV1, AV2), and up to five ORFs in the complementary sense (AC1, AC2, AC3, AC4 and/or AC5). These ORFs are involved in replication (AC1, AC3, AV2), encapsidation (AV1) and transcriptional activation of virion sense genes (AC2). The role of the AC4 gene product is unknown, and the AC5 gene product is non-functional [15]. The DNA-B component has one major ORF in each of the virion (BV1) and complementary (BC1) orientations, involved in viral movement within the plant [reviewed by 10, 11]. *Agrobacterium*-mediated inoculation studies have shown that viral replication and movement can occur in the absence of DNA-B, although levels are higher when DNA-B is present [28]. The begomovirus DNA-A and DNA-B components share little sequence similarity, except for ∼170 nts of sequence in the intergenic region, termed the common region (CR). The CR contains the viral origin of replication, that interacts with the replication initiation (Rep) protein to initiate rolling circle replication of each component [8, 9, 24]. A DNA-B component has not been identified for some begomoviruses, including *Ageratum yellow vein virus* (AYVV) [31], and some isolates of *Tomato leaf curl virus* (TLCV) [5], and *Tomato yellow leaf curl virus* (TYLCV) [14, 22, 23]. In these monopartite begomoviruses, the movement and nuclear shuttle functions are encoded by DNA-A.

There is considerable evidence that many geminivirus species have arisen by recombination [25] and recombinant geminiviruses have now been identified in cotton [37], cassava [3, 7, 27, 38], and tomato [16, 21]. It is probable that recombination events have played a role in the emergence of new geminiviral diseases, such as cotton leaf curl in Pakistan, and it has been demonstrated that severe outbreaks of cotton leaf curl disease in Pakistan were the result of recombination between *Cotton leaf curl virus* (CLCuV) and *Okra yellow vein mosaic virus* (OYVMV) [37]. Evidence of recombination between old and new world geminiviruses suggests recombination events predate current geographical barriers [25]. The location and length of the recombinant sequence are variable and recombination occurs throughout the viral genome [25]. Sequences of the viral Rep [21], AC2/AC3 [7], AC4 [3], and BC1 [7] genes, as well as the origin of replication [37], have all been identified as recombinant regions in geminiviruses.

During surveys of cucurbit crops in Vietnam we have observed geminivirus-like symptoms on various cucurbit species. These included pumpkin (*Cucurbita maxima, Cucurbita* spp), zucchini (*Cucurbita pepo* var Pepo), green cucurbit/waxy gourd (*Benicasa hispida*), and in southern Vietnam, loofa (*Luffa acutangula*). Symptoms consisted of a bright yellow mosaic on the leaves, with stunting

evident only in zucchini. These symptoms are similar to those described for *Squash leaf curl virus-China* (SLCV-Ch), a whitefly-transmitted geminivirus that was first identified in Chinese *Cucurbita pepo* field plantings [13]. Analysis of the nucleotide sequences encoding the viral coat protein showed that SLCV-Ch was only 67% similar to another squash-infecting begomovirus, *Squash leaf curl virus* (SLCV) [17] from America. The Chinese sequence was more similar to geminiviruses from the old world (Europe, Asia and Africa) than those from the new world (The Americas) and although symptoms on squash were similar to SLCV, the low degree of coat protein sequence similarity indicated SLCV-Ch was a different virus species. The complete sequence of SLCV-Ch DNA-A is available (GenBank Acc No AB0274465), however there are no reports of a DNA-B. A DNA-B has, however, been identified for SLCV [17].

Here we report the identification of two distinct geminiviruses infecting cucurbits in Vietnam, and provide preliminary evidence of recombination between them. One virus that we have called Squash leaf curl virus-Vietnam (SLCV-Vn), infected a range of cucurbit species throughout Vietnam, and the other virus, which we have called Loofa yellow mosaic virus-Vietnam (LYMV-Vn) was only detected in loofa in the Ho Chi Minh City district in southern Vietnam. The distribution, sequence variability, and similarity of the viruses infecting cucurbits in Vietnam to other geminiviruses is discussed.

Methods

Source of material and extraction of DNA

Fifteen cucurbit samples exhibiting geminivirus symptoms were collected from throughout Vietnam (Table 1), and preserved as dried leaf samples on silica gel. Plant material was stored at room temperature, and DNA was extracted using the DNeasy (Qiagen) kit, according to the manufacturer's instructions.

Degenerate begomovirus PCR

Total DNA was used as template in a PCR with degenerate primers Krusty (5′CCNMRDGG HTGTGARGGNCC3′) and Homer (5′SVDGCRTGVGTRCANGCCAT3′) designed to amplify 580 nts of the begomovirus coat protein gene. PCRs contained 1.5 mM $MgCl_2$ buffer (Roche), 10 pmol dNTPs, 40 pmol of each primer, and 2.5 U Taq polymerase (Roche). The cycling parameters were 94 °C for 5 min, followed by 35 cycles of 94 °C for 30 sec, 55 °C for 30 sec, 72 °C for 30 sec, then 72 °C for 10 min.

PCR products were electrophoresed through 1% agarose gels, purified by the QIAX II (Qiagen) procedure and ligated into plasmid vector pGEM-T Easy (Promega), as recommended by the manufacturer. Plasmids were transformed into *E. coli* strain JM109 (Promega) using the heat shock method, according to the manufacturer's protocol. Plasmids were purified using the alkaline lysis protocol [30], and sequencing was performed using the ABI Prism® BigDye™ Terminator Cycle Sequencing Ready Reaction Kit (PE Applied Biosystems) according to the manufacturer's protocol. Cloned inserts were sequenced at the Australian Genomic Research Facility (University of Queensland). Initial sequences were determined with universal forward (M13-20), and reverse (M13 Reverse) primers.

Table 1. The location, and type, of cucurbit samples collected throughout Vietnam

Sample	Plant Type	Location	Region	Virus
A	pumpkin	Soc Son	north	SLCV-Vn
B	waxy gourd	Soc Son	north	SLCV-Vn
C	pumpkin	Soc Son	north	SLCV-Vn
D	zucchini	Me Linh	north	SLCV-Vn
E	pumpkin	Vin Phuc	north	SLCV-Vn
F	pumpkin	Hue	central	SLCV-Vn
G	pumpkin	Do Luong	central	SLCV-Vn
H	pumpkin	Dien Chau	central	SLCV-Vn
I	pumpkin	Than Hoa	central	SLCV-Vn
J	pumpkin	Than Hoa	central	SLCV-Vn
K	pumpkin	Buon Ma Thuot	central	SLCV-Vn
L	pumpkin	Hue	central	SLCV-Vn
M	loofa	Ho Chi Minh City	south	LYMV-Vn
N	loofa	Ho Chi Minh City	south	LYMV-Vn
O	waxy gourd	Ho Chi Minh City	south	SLCV-Vn

Amplification of genomic length component A

Outwardly extending adjacent primers Gem14FOR (5′ATCGTGATCGATATCAAGTGATG CGC3′) and Gem14Rev (5′GCATATTCTTCACCGTTGCAGTGCTCGG3′) were designed from the coat protein sequence to amplify genomic length DNA-A. An additional set of outwardly extended oligonucleotides (GCplus 5′ATACTTTGGACACCAAATGGC3′ and GCminus 5′ATATGAGACCCCATACCCC3′) were designed in the large intergenic region, using the sequence obtained from the Gem14FOR/Gem14Rev PCR product. Total DNA from four isolates (B, F, K, M) was used as template in the Long Template PCR system (Roche), as per the manufacturers instructions. PCRs contained 1 × buffer 1, 17.5 pmol dNTPs, 30 pmol of each primer, 1 ul template DNA, and 2.6 U Expand polymerase (Roche). The cycling parameters were 94 °C for 4 min, followed by 10 cycles of 94 °C for 10 sec, 60 °C for 10 sec, 68 °C for 2.5 min; 20 cycles of 94 °C for 10 sec, 60 °C for 10 sec, 68 °C for 2.5 min with 20 sec increments; then 68 °C for 10 min. Genomic length viral DNA was cloned into pGem-T Easy, transformed into *E. coli* and sequenced as described above. Consensus sequences were determined using the SeqMan program (DNASTAR), and nucleotide and deduced amino acid sequences from two clones for each isolate were analysed using EditSeq (DNASTAR), Vector NTI, and the programs and databases available through Australian National Genome Information Service (ANGIS) (Sydney University).

Sequence variability of DNA-A

To determine the level of SLCV-Vn DNA-A sequence variability throughout Vietnam, sequences encoding the coat protein and Replication initiation (Rep) genes, and the intergenic region, from an additional eleven cucurbit samples (samples A, C, D, E, G, H, I, J, L, N, O) (Table 1,) were amplified by PCR using the conditions described above, with oligonuleotides AFor6 (5′ACACGACATAATCAAGGTGCAGG3′) and ARev6 (5′ACATGTGGGATCCACT TATGCACG3′) (coat protein), AFor3 (5′CGTGAATTTCATTGGTCGAGGGCCC3′) and ARev2 (5′TCACTCCTAACCTATCCAGG3′) (Rep), or AFor3 and ARev5 (5′TGCATTGGAA

CTTTCCTTCGAACTGG3′) (intergenic region). Sequences from an additional LYMV-Vn isolate (sample N) from southern Vietnam were amplified with oligonucleotides AFor6 (5′ACACGACATAATCAAGGTGCAGG3′) and ARev6 (5′ACATGTGGGATCCACTTATG CACG3′) (coat protein), 609FOR2 (5′AGCACCCGATTCAATATGCCACG3′) and 609REV1 (5′AAGGCCTGGACAAACAGGCCG3′) (Rep), 609Rev3 (5′ATGTTTCGAAGGAGTTCG AGC3′) and AFor2 (5′ACGTCGGGACTTCTATACATCCTG3′) (intergenic region). The resultant PCR products were cloned and sequenced as described above, and sequences were aligned and bootstrapped using the ClustalX program [32]. Nucleotide and amino acid sequence similarity was determined using the MegAlign program (DNASTAR) and unrooted neighbour-joining trees were constructed using the program Treeview [26]. The accession numbers of the geminivirus DNA-A sequences used for comparison were as follows: AYVV (X74516), *Bean golden mosaic virus* (BGMV, M88686), *Cabbage leaf curl virus* (CabLCV, U65529), *Cowpea golden mosaic virus* (CGMV, AF029217), *Tomato yellow leaf curl Thailand virus* [1] (TYLCV-Th, X63016), *Tomato leaf curl Vietnam virus* (TLCV-Vn, AF264063), *African cassava mosaic virus* (ACMV, X17095), CLCuV (AJ002455), *Pepper huasteco yellow vein virus* (PHV, X70418), *Papaya leaf curl virus*, (PLCV, Y15934), OYVMV (AJ002451), *Potato yellow mosaic virus* (PYMV, D00940), *Sida golden mosaic virus* (SGMV, AF049336), and SLCV-Ch (AB0274465). The sequence of a geminivirus isolated from cucurbits in Thailand, (Thai, Dr. T. Burns, unpublished data) was also included in the analysis.

Recombination analysis

Phylogenetic analysis of the coat protein and Rep sequences from the cucurbit-infecting geminiviruses in Vietnam indicated that one or more of the sequences was a recombinant. To identify putative recombinant regions of the viral genomes, genomic length DNA-A sequences of SLCV-Vn isolates B, F, and K, LYMV-Vn isolates M and N, SLCV-Ch, and the Thai sequence were aligned using ClustalX, and analysed using two recombination programs, RDP [18] and PhylPro (version beta 0.8; [36]). Both programs use pairwise scanning of aligned sequences to search for recombinant sites and default parameters were used for all analyses.

Amplification of genomic length DNA-B

A 600 bp fragment of DNA-B was amplified from samples B (SLCV-Vn) and M (LYMV-Vn) using the method described by [29]. A degenerate sense oligonucleotide located in the BC1 gene (5′GCCTCTGCAGCARTGRT CKATCTTCATACA3′) was used with the specific GCminus antisense oligonucleotide described above (for northern and central isolates B, F, and K of SLCV-Vn), or oligonucleotide 609GCminus (5′TATAAATAGGCACCCAAAGCAC CCC3′) for LYMV-Vn sample M, in a PCR with 1 × PCR buffer (Roche), 20 pmol each oligonucleotide, 10 pmol dNTPs, and 2.5 U Taq polymerase (Roche). The cycle parameters were 94 ° for 5 min; 35 cycles of 94 ° for 1 min, 50 ° for 1 min, 72 ° for 1 min; and 72 ° for 10 min. The PCR products were cloned and sequenced as described above, and from the resultant sequence two, DNA-B outwardly extending adjacent oligonucleotides Btrueplus (5′GAGCGTACTCTACACGCTCCG3′) and Btrueminus (5′TCACAAACGACGCTGCGGA GG3′) (sample B), or B609BtoBplus (5′AAGCCCAATTTTTACAAATTCGGC3′) and B609BtoBminus (5′ACAACATATGGAGATCCAGTCCAC3′) (LYMV-Vn, sample M) were designed. Genomic length DNA-B was amplified and cloned and sequenced, using the same conditions and protocols as those described above for DNA-A.

Confirmation of nucleotide deletions in the intergenic region of SLCV-Vn DNA-B

The sequence of the DNA-B component from sample B had a 47 nt deletion in the intergenic region when compared with the DNA-A. To determine if this deletion was present in DNA-B of additional SLCV-Vn isolates, the intergenic region was amplified by PCR from samples F and K, using the DNA-B specific oligonucleotides, BFor8 (5′ATGTCGTGAATCTCAATTACC AC3′) and BRev6 (5′AGGTCAGTTGGGGTCCTGGATGCGG3′). The PCR contained 1 × PCR buffer (Roche), 20 pmol each oligonucleotide, 10 pmol dNTPs, and 2.5 U Taq polymerase (Roche), and the cycle parameters were 94 ° for 5 min; 35 cycles of 94 ° for 1 min, 50 ° for 1 min, 72 ° for 1 min; and 72 ° for 10 min. The PCR products were cloned and sequenced as described above.

To confirm that all DNA-B molecules in sample B had this deletion, PCRs were performed using the conditions described above, with either one of two oligonucleotides located in the deleted portion of the DNA-A intergenic region (IR1: 5′TACCATTTTGCCATTTGGTG3′ or IR2: 5′ATTTTGAATTAAAGTAATTATTTG3′), paired with a DNA-B specific oligonucleotide (BFOR8). Absence of a PCR product from the IR1/IR2 BFOR8 PCR would confirm the apparent deleted sequence was not present in other DNA-B molecules. PCR with IR1/IR2 and a specific DNA-A oligonucleotide (ARev4: 5′ACGAGTTCTTGCGGAACTCTGTCG3′) was performed as a control.

Results

Characterisation of DNA-A

Degenerate begomovirus oligonucleotides designed in the AV1 gene amplified a product of the predicted size (580 bp) from 15 cucurbit samples collected throughout Vietnam. Nucleotide sequencing and BLAST searches showed that these PCR products were most similar to the SLCV-Ch coat protein sequence.

Genomic length DNA-A was amplified from four of these cucurbit samples, and BLAST searches using the GenBank database confirmed that the coat protein gene (AV1) nucleotide sequences were most similar to that of SLCV-Ch (GenBank Acc No: AB0274465). The complete DNA-A nucleotide sequences of three of the viruses isolated from pumpkin in northern and central Vietnam were almost identical to SLCV-Ch, however the sequence of a virus isolated from loofa in southern Vietnam differed markedly from SLCV-Ch in the complementary sense ORFs, AC1 to AC4. However, the coat protein gene sequences of all cucurbit-infecting geminiviruses in Vietnam were almost identical (Fig. 1). The DNA-A sequences from pumpkin in northern and central Vietnam (B, F, K) were 2736 nts, whereas DNA-A isolated from loofa in southern Vietnam (M) was 2742 nucleotides.

Although nucleotide sequence comparisons showed that the viruses isolated from pumpkin in northern and central Vietnam (B, F, K) were 95% similar to the SLCV-Ch sequence, the DNA-A of the virus isolated from loofa in southern Vietnam (M) was only 80% similar to SLCV-Ch over the complete nucleotide sequence. This difference indicated that there were two different viruses infecting cucurbits in Vietnam, one almost identical to SLCV-Ch, the other with a similar coat protein, but distinct throughout the rest of its DNA-A sequence. The similarity

A

Isolates	DNA-A	AV1	AV2	AC1	AC2	AC3	AC4	AC5	IR
SLCV-Vn	4.7	4.2	4.4	7.7	5.2	4.2	3.5	3.5	8.9
SLCV-Vn c.f. LYMV-Vn	21.7	5.4	7.6	33.4	17.7	16.9	36.1	5.4	52.4

B

Isolates	DNA-B	BV1	BC1
SLCV-Vn c.f. LYMV-Vn	39.1	22.6	19.6

Fig. 1. Maximum percentage difference of nucleotide sequences for (**A**) DNA-A and (**B**) DNA-B amplified from geminiviruses infecting cucurbits in Vietnam

of one group of viruses to SLCV-Ch suggested they were strains of this virus, and we have named these viruses Squash leaf curl virus-Vietnam (SLCV-Vn). The virus isolated from loofa in southern Vietnam was sufficiently different to SLCV-Vn to be considered a separate species, and we have named this virus Loofa yellow mosaic virus-Vietnam (LYMV-Vn).

The genome arrangement of SLCV-Vn and LYMV-Vn were similar, each having six major ORFs that encoded proteins of greater than 100 amino acids (Fig. 2). Two of these ORFs were in the virion sense (AV1 and AV2), and four were in the complementary sense (AC1, AC2, AC3, AC5). All isolates also contained a small AC4 ORF in the complementary sense, although the size of the predicted gene product in one the SLCV-Vn isolates (B) differed due to a C to G nucleotide change at nt 2259. In almost all isolates sequenced, the putative AC4 gene product was 58 amino acids in length, however the AC4 gene product of SLCV-Vn isolate B and LYMV-Vn was 85 amino acids in length. Blast searches revealed that the AC4 gene product was most similar to the AC4 gene product encoded by some isolates of TYLCV (Acc Nos: P36283, P27272, P27271, and P38612) and *Indian cassava mosaic virus* (ICMV, Acc No: Q08588, equivalent to AL0 in ICMV). The AC5 gene product of SLCV-Vn and LYMV-Vn was most similar to that encoded by ICMV (equivalent to AL4 in ICMV) and PHV.

Additionally, in two of the SLCV-Vn isolates (samples B and F), the first AV2 ATG codon was located 5′ of the stem-loop sequence in the origin of replication, whereas the first in-frame ATG codon of AV2 for SLCV-Vn isolate K was located 3′ of the stem loop, in an identical position to that of SLCV-Ch. The first AV2 ATG

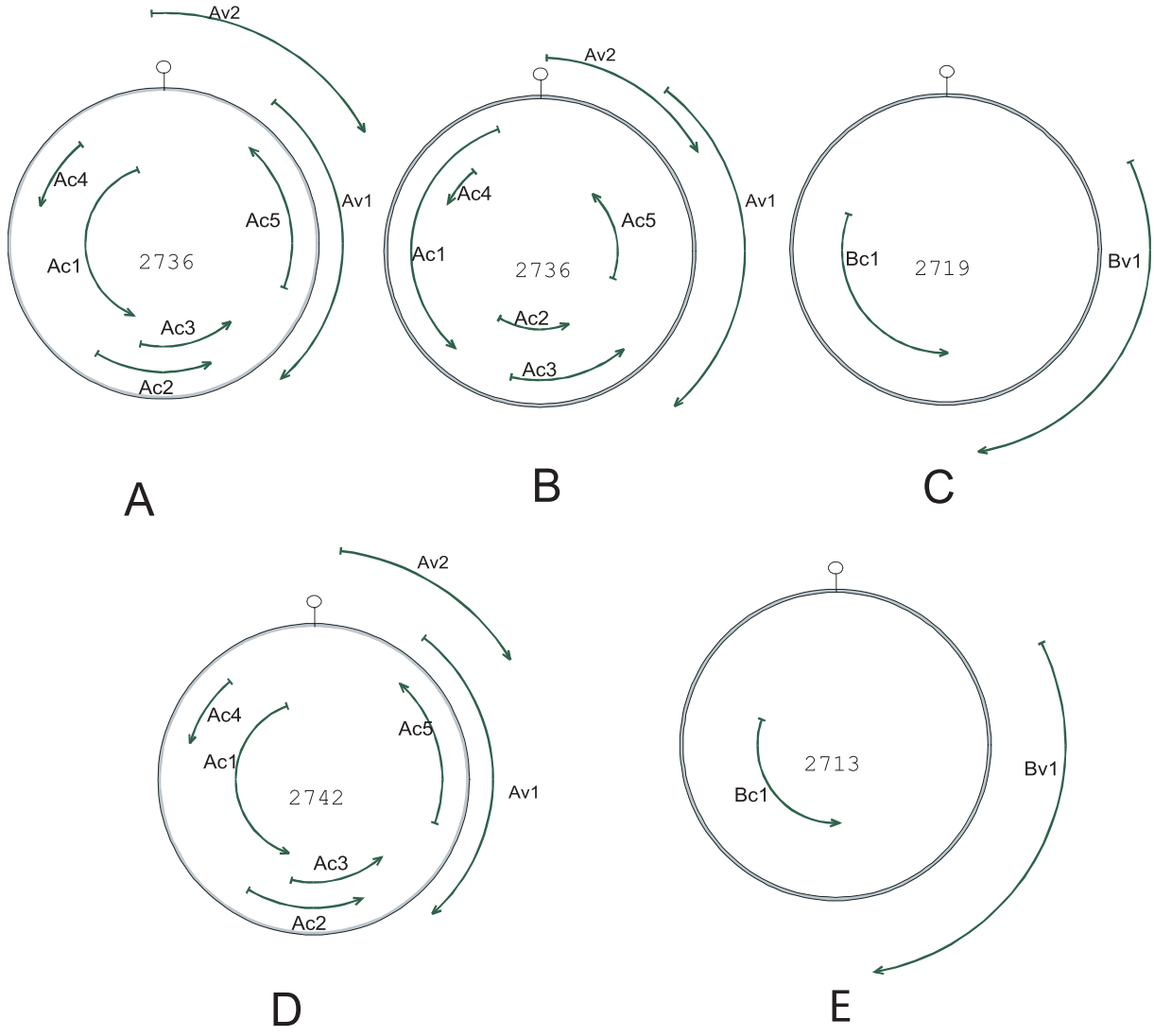

Fig. 2. The genome arrangement of SLCV-Vn and LYMV-Vn. **A** and **B**, DNA-A for two isolates of SLCV-Vn (samples B and K). In sample B the first ATG of ORF AV2 is 5′ of the stem loop in the origin of replication, whereas in sample K it is 3′ of the stem loop sequence. **C** SLCV-Vn DNA-B; **D** and **E** LYMV-Vn DNA-A and DNA-B (sample M). ORF designations are indicated, as is the number of nucleotides for each sequence and the stem loop sequence in the origin of replication (♁)

of LYMV-Vn was also located 3′ of the stem loop, but was an additional 27 nts downstream of the comparable SLCV-Ch sequence. The sequence differences in the SLCV-Vn isolates were confirmed by sequencing PCR products amplified using DNA-A specific primers (data not shown).

In addition to a similar genome arrangement, the nucleotide and deduced amino acid sequences of the SLCV-Vn and LYMV-Vn AV1 ORF (coat protein) was very similar, with a maximum nucleotide difference of 5.4% (Fig. 1). However

the nucleotide sequences of the complementary sense ORFs differed markedly, with AC1, AC2, AC3, and AC4 of SLCV-Vn and LYMV-Vn differing by 33.4%, 38.7%, 16.9%, and 37% respectively. The complementary sense AC5 ORF only differed by 5.4%, however this ORF is internal to AV1, in the reverse-complement orientation. Similar levels of variability were observed for the deduced amino acid sequences (data not shown). The intergenic regions of SLCV-Vn and LYMV-Vn differed by up to 52.4%.

Sequence variability of DNA-A

To investigate sequence variability among isolates of SLCV-Vn and LYMV-Vn in Vietnam, the AV1, AC1, and intergenic regions were sequenced from an additional 11 samples collected throughout Vietnam. These comprised ten SLCV-Vn isolates in a number of different cultivars, and one additional LYMV-Vn isolate collected from loofa in southern Vietnam (Table 1).

The nucleotide and deduced amino acid sequences of the additional isolates showed that sequences within isolates of SLCV-Vn and LYMV-Vn were conserved. The SLCV-Vn AV1 sequences differed by a maximum of 4.2% at the nucleotide level, and there was only 1.4% difference between the AV1 nucleotide sequences of LYMV-Vn isolates M and N (data not shown). Most variation in SLCV-Vn sequences was observed in the AC1 sequence, with a maximum difference of 7.7%, and the AC1 sequences of the two LYMV-Vn isolates differed by only 1.1% (data not shown). Similar levels of variability were observed for the amino acid sequences (data not shown). The intergenic regions of SLCV-Vn isolates differed by a maximum of 8.9%.

Analysis of the DNA-A intergenic region

Analysis of the SLCV-Vn intergenic region revealed the presence of three putative iterons upstream of a TATA sequence 5′ of the AC1 initiation codon (Figs. 3(A) and 3(B)). The consensus sequence for the putative core of these iterons was YGGGGT. The three putative iterons in the LYMV-Vn intergenic region were less conserved (YGGGKKC), although the core sequence of the second iteron was identical to that of the second and third iterons of SLCV-Vn (TGGGGT). Complementary iterons were not observed downstream of the TATA box for either virus. Interestingly, the SLCV-Vn and LYMV-Vn AC1 N-terminal amino acid residues, identified as a putative iteron-related domain [2] were different (Fig. 3(B)). The TAATATTAC sequence conserved in the stem-loop sequence of all geminiviruses was conserved in all Vietnamese sequences.

Phylogenetic analysis of DNA-A

Phylogenetic comparison of the complete SLCV-Vn and LYMV-Vn DNA-A sequences with DNA-A of other begomoviruses showed that SLCV-Vn closely aligned with SLCV-Ch, whereas LYMV-Vn was more similar to the unpublished

A

SLCV-Vn (A) .CAAAACGCG CGTTTTGA*AT CGGGGTCT*CT CCAAAACTAT GGTGT*ATTGG*
SLCV-Vn (B) TTCTCTCTCT AGACCCCA*AT TGGTGTCC*CT TCAAAACTAT GGTAT*ATTGG*

SLCV-Vn (A) *GGTATGGGGT CT*CATA**TATA** CTTTGGACAC CAAATGGCAA AATGGTAATT
SLCV-Vn (B) *GGTATGGGGT CT*TATA**TATA** CCTTAGA... GAAAT..... ....TT

SLCV-Vn (A) ATGCAAATAA TTACTTTAAT TCAAAATGAA TAAAGCGGCC ATTCGTATAA
SLCV-Vn (B) ATG..... .. ..GCGGCC TTTCGTTTAA

SLCV-Vn (A) TATTACCGAA TGGCCGCGCG ATTTTTT.
SLCV-Vn (B) TATTACCGAA AGGCCGCGCT TTTTTTTT

B

SLCV-Vn (A) CGGGGT TGGGGT TGGGGT...**TATA** (MAPPKHFKISAK)
SLCV-Vn (B) TGGTGT TGGGGT TGGGGT...**TATA**

LYVV-Vn (A) TCGGGT TGGGGT TGGGTG...**TATA** (MPRTNQFQVKAK)
LYVV-Vn (B) TCGGGT TGGGGT TGGGTG...**TATA**

Fig. 3. A ClustalX alignment of the DNA-A (A) and DNA-B (B) intergenic sequences of SLCV-Vn. Putative iterons are in italics. The TATA box is in bold type, and the TAATATTAC stem-loop sequence conserved amongst all geminiviruses is underlined. **B** Putative iteron core sequences from SLCV-Vn and LYMV-Vn in Vietnam, and the first 12 amino acid residues of the Rep gene from each virus (parentheses). (A) = DNA-A sequence, (B) = DNA-B sequence

Thai sequence (Fig. 4). Comparison of the SLCV-Vn and LYMV-Vn AC1 nucleotide sequences clearly showed they were distinct viruses, with the two loofa-infecting viruses (LYMV-Vn isolates M and N) grouped separately from SLCV-Vn isolates (A to L, O) (Fig. 5(A)). However, this distinction was less obvious for the AV1 nucleotide sequences, with clustering influenced by geography as well as isolate or host. That is, although AV1 sequences from LYMV-Vn isolates did

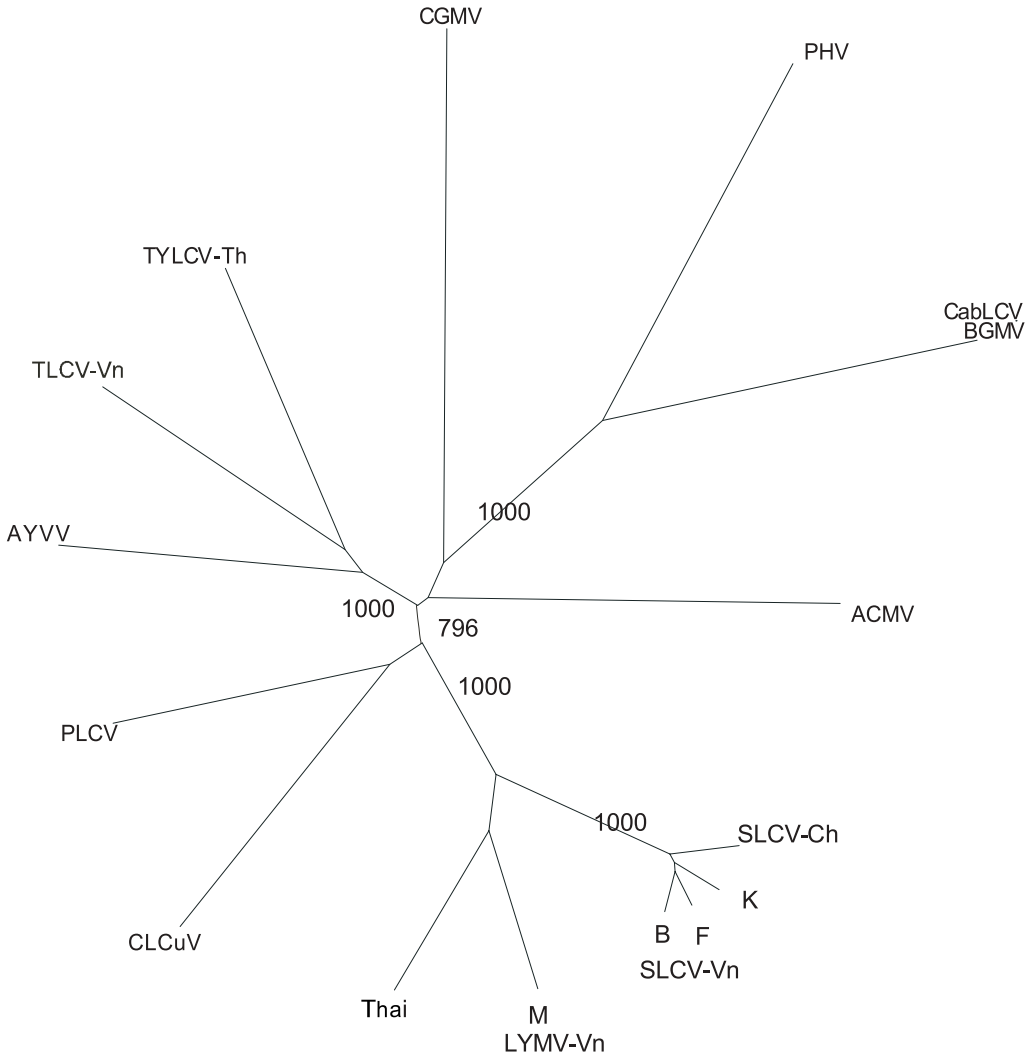

Fig. 4. An unrooted neighbour joining tree of the complete DNA-A nucleotide sequences from SLCV-Vn (samples B, F, and K) and LYMV-Vn (sample M) in Vietnam compared with other geminiviruses. The Thai sequence was obtained from Dr. T. Burns, Kasesaart University, Thailand. Bootstrap values (1000 replicates) are shown at the major clades

cluster together, SLCV-Vn sequences from central and southern Vietnam (F, K, L, O) clustered more closely to LYMV-Vn isolates (southern Vietnam, M, N), than they did to SLCV-Vn isolates from northern Vietnam (Fig. 5(B)). One SLCV-Vn sequence isolated from waxy gourd in southern Vietnam (sample O) was collected from the same location as the LYMV-Vn isolates (samples M and N), yet had an AC1 sequence almost identical with SLCV-Vn sequences amplified from waxy gourd and pumpkin collected elsewhere in Vietnam (Fig. 5(A)).

Comparison of the AV1 nucleotide sequences with AV1 sequences of other begomoviruses showed that the LYMV-Vn and Thai sequences clustered tightly with the Vietnamese SLCV-Vn isolates (Fig. 6A), however identical analysis

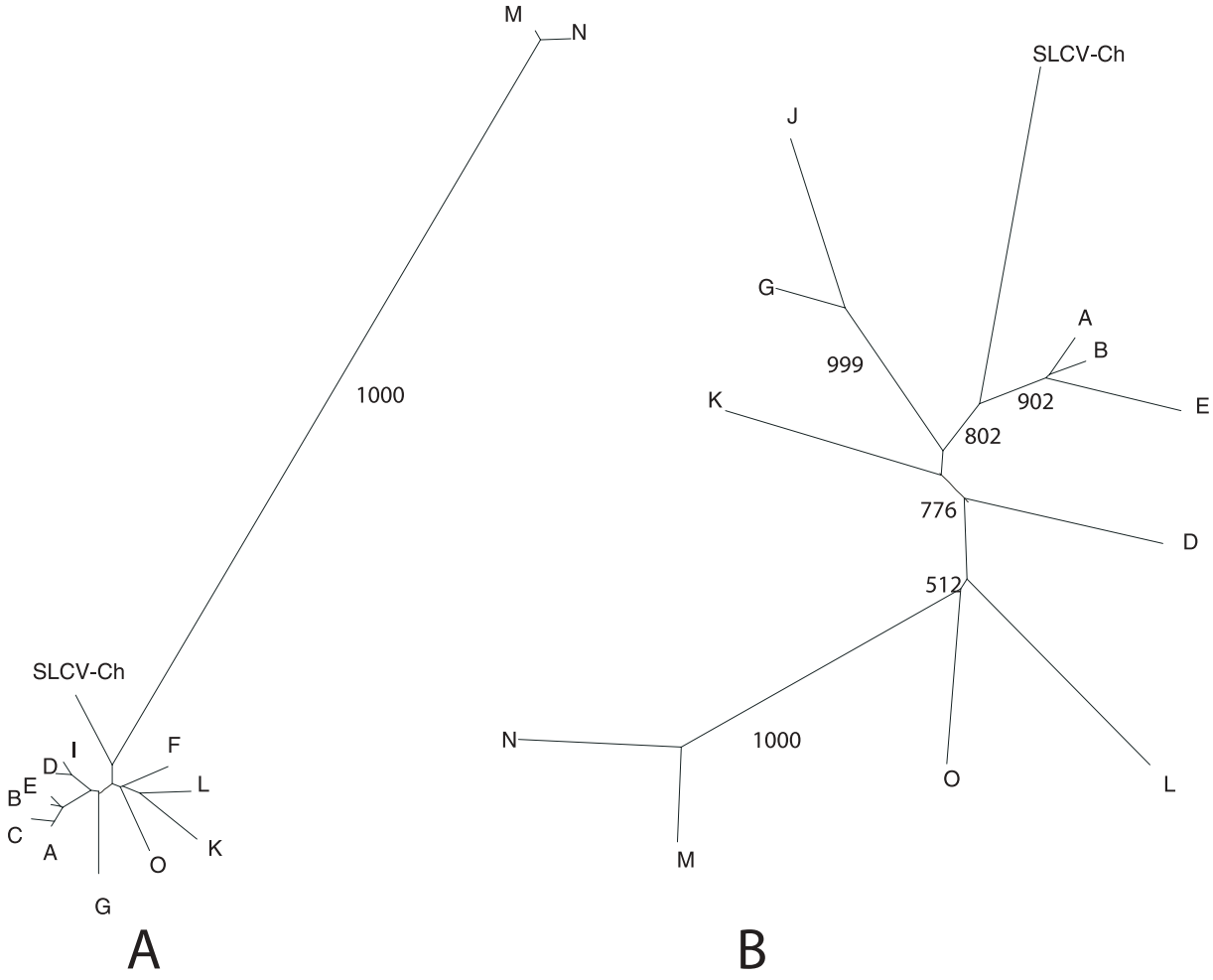

Fig. 5. Unrooted neighbour joining trees of the complete AC1 (**A**) and AV1 (**B**) nucleotide sequences from SLCV-Vn (samples A to L, O) and LYMV-Vn (samples M and N) compared to the SLCV-Ch sequence (Acc No: AB0274465). Samples were collected from northern (A, B, C, D, E), central (F, G, I, J, K, L), and southern (M, N) Vietnam. Bootstrap values (1000 replicates) are shown at the major clades

of the AC1 sequences showed that LYMV-Vn and the Thai sequence clustered separately (Fig. 6(B)). Similar topologies were obtained using the deduced amino acid sequences, and phylogenetic analysis of the intergenic region produced a similar topology to AC1 (data not shown).

Recombination analysis

The similarity of SLCV-Vn and LYMV-Vn AV1 sequences, and the marked differences observed in the complementary sense genes, indicated that the AV1 region in the viral genome may have had a recombinant origin. This was supported by recombination analysis using both PhylPro and RDP, which identified a putative

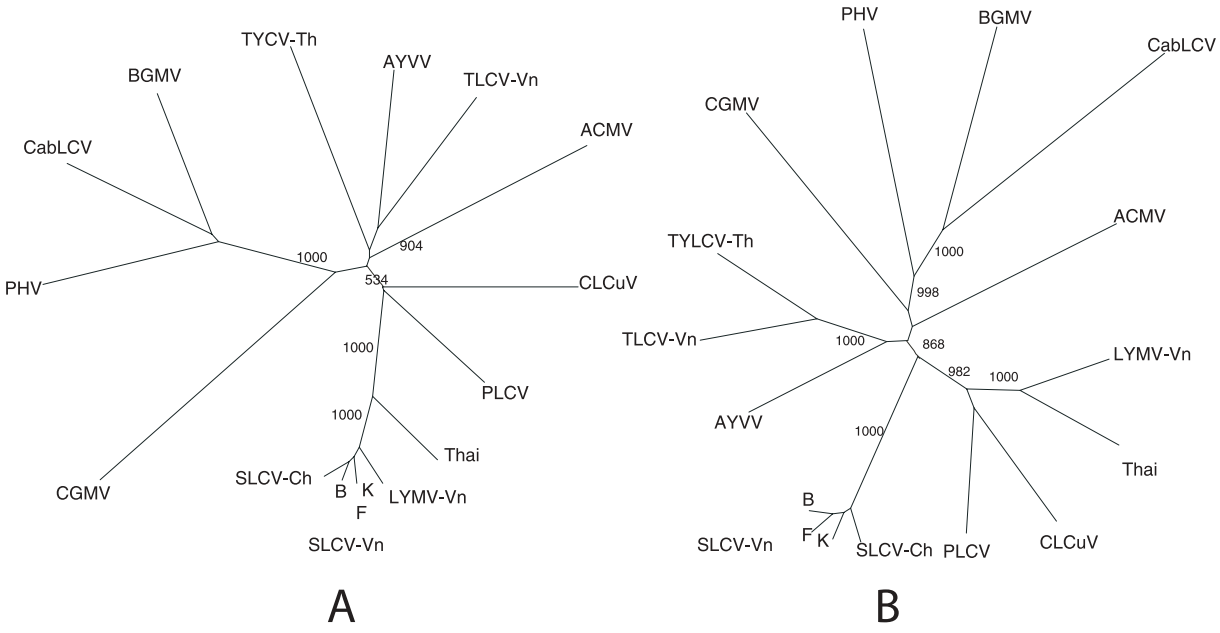

Fig. 6. Unrooted neighbour joining trees of the AV1 (**A**) and AC1 (**B**) nucleotide sequences from SLCV-Vn (samples B, F, K) and LYMV-Vn (sample M) in Vietnam compared with other geminiviruses. The Thai sequence was obtained from Dr. T. Burns, Kasesaart University, Thailand. Bootstrap values (1000 replicates) are shown at the major clades

550 bp recombinant region in the AV1 sequence between nucleotides 460 and 1010 (Fig. 7(A) and (B) respectively). However, the small dataset analysed prevented identification of "parent" and "daughter" sequences.

Sequence variability of DNA-B

DNA-B components were amplified from isolates of both SLCV-Vn and LYMV-Vn. The genomic length clones of the SLCV-Vn DNA-B and LYMV-Vn DNA-B were 2719 nts and 2713 nts respectively, and each encoded two major ORFs, one in the positive sense (BV1), and one in the complementary sense (BC1) (Fig. 1). This genome arrangement was typical of begomovirus DNA-B components. SLCV-Vn DNA-B and LYMV-Vn DNA-B were only 60% similar over the complete sequence, however nucleotide sequences of the ORFs were approximately 80% similar. The deduced amino acid sequences of the BC1 ORF were 90% similar.

BLAST searches of the deduced amino acid sequences showed that the SLCV-Vn BC1 gene product was most similar to the BC1 gene of ICMV (67% identity over 136 amino acids, Acc No: Q08594) and SLCV (54% identity over 173 amino acids, Acc No: P21936). The amino acid sequences encoded by BV1 were also most similar to those encoded by BV1 of ICMV and SLCV, although the percentage identity was lower than for the BC1 gene (39% over 173 amino acids, and 42% over 56 amino acids respectively). The LYMV-Vn BC1 and BV1 gene

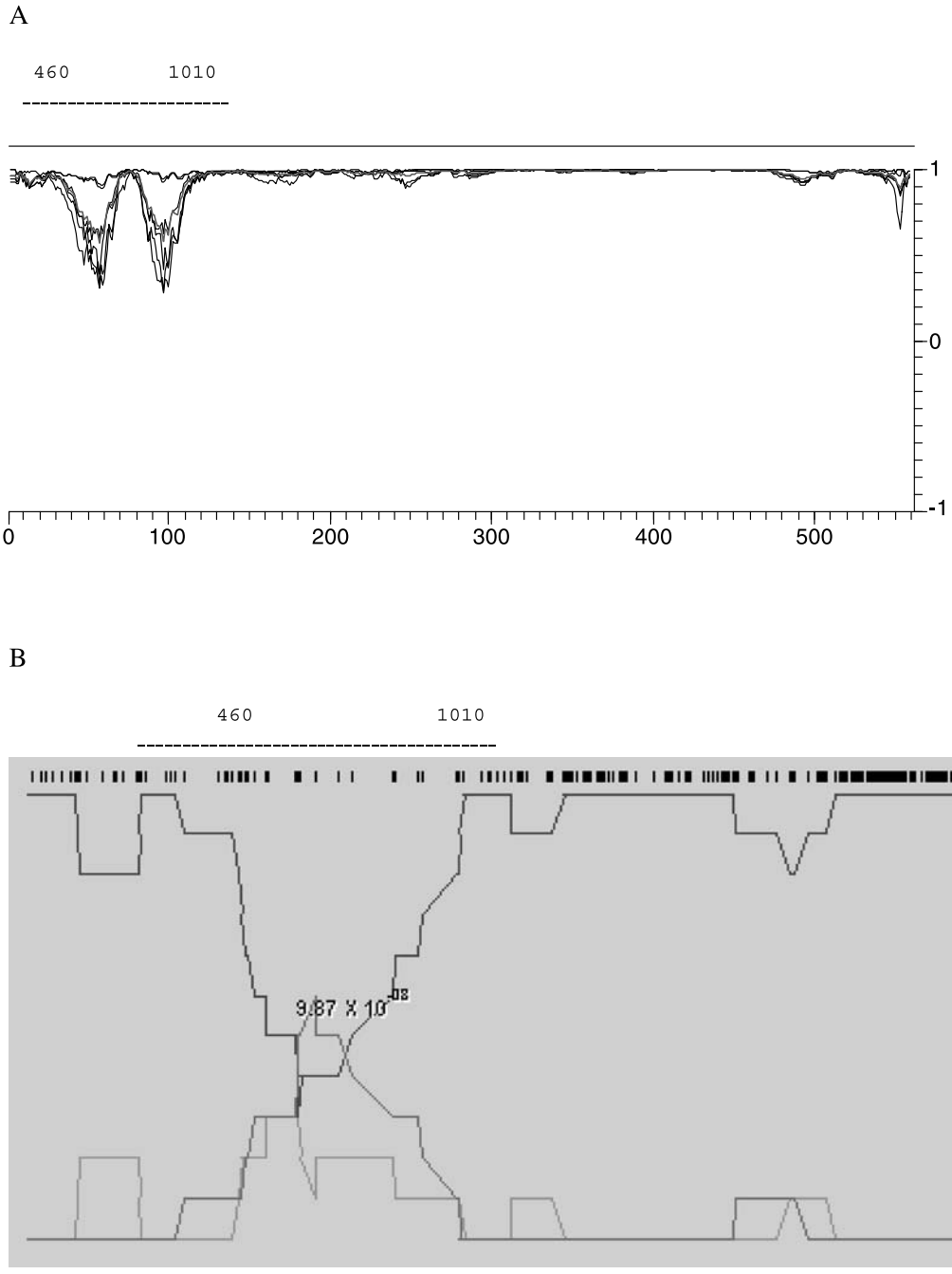

Fig. 7. Phylogenetic profiles generated by PhylPro (**A**) and RDP (**B**) of SLCV-Vn, LYMV-Vn, and Thai DNA-A sequences. In **A**, recombination signals appear as sharp downward peaks in areas of low phylogenetic correlation. The phylogenetic correlation (y-axis) was obtained from pairwise distance analyses of all aligned sequences and ranges from +1 (perfectly correlated) to 0 (unrelated). In **B**, the y-axis illustrates pairwise identity, and the putative recombinant region is shown by the crossover in the left hand portion of the panel. The probability that the sequences at this crossover could appear recombinant simply due to chance convergence is shown (9.87×10^{-08}). The viral coat protein regions (approx location 250 nt to 1050 nt) are indicated by dashes above both profiles

products were most similar to ICMV, with 65% identity over 136 amino acids, and 40% identity over 172 amino acids respectively.

Comparison of the SLCV-Vn and LYMV-Vn intergenic regions

Alignment of the DNA-A and DNA-B intergenic regions from SLCV-Vn revealed there was a 47 nucleotide deletion in the DNA-B intergenic region, upstream of the stem loop (Fig. 3(A)). This deletion resulted in only 66% identity over the length of the intergenic sequence. Putative iterons were identified upstream of the putative BC1 TATA sequence. The most 5′ iteron had two base changes compared to the comparative DNA-A iteron, however the two iterons immediately upstream of the TATA sequence were identical in DNA-A and DNA-B. Alignment of the matching portions of the DNA-A and DNA-B intergenic regions showed there was 88% identity over 113 nucleotides. The intergenic regions of LYMV-Vn DNA-A and DNA-B molecules contained no deletion, and were 98% identical to each other.

PCR using oligonucleotides in the deleted portion of the DNA-A intergenic region (IR1 or IR2) only amplified a PCR product when paired with oligonucleotides specific for DNA-A (data not shown). Pairing IR1 or IR2 with oligonucleotides specific to DNA-B failed to produce a PCR product, confirming that the SCLV-Vn DNA-B component contained a deletion in the intergenic region.

Discussion

This is the first report of the complete nucleotide sequence of SLCV-Vn, and shows that the virus genome is bipartite, consisting of both DNA-A and DNA-B components. The nucleotide sequences of DNA-A from viruses isolated throughout Vietnam were virtually identical to the SLCV-Ch DNA-A sequence on GenBank, suggesting that they are all strains of the same virus [20].

The DNA-A isolated from loofa in southern Vietnam (LYMV-Vn) was only 80% similar to the other sequences isolated from throughout Vietnam, over the complete nucleotide sequence. LYMV-Vn was more similar to a geminivirus isolated from cucurbits in Thailand than to SLCV-Vn. Geminiviruses are classified as different species if the complete nucleotide sequence of DNA-A differs by more than 90% [34]. Since the complete DNA-A nucleotide sequence of the virus isolated from loofa in southern Vietnam was only 80% similar to SLCV-Vn, it warrants classification as a distinct virus species. Consequently, we have named this virus Loofa yellow mosaic virus-Vietnam (LYMV-Vn). However the taxonomy of this virus is ambiguous, as [34] also state there should be less than 90% identity in the coat protein sequence to distinguish geminivirus species. Using these criteria, LYMV-Vn would be classified as an isolate of SLCV-Vn, as the deduced coat protein sequences were 95% identical. However, we believe it is preferable to distinguish LYMV-Vn and SLCV-Vn as separate species, due to the high degree of difference in other regions of the genome.

Recombination analysis showed that a 550 nt region of the SLCV-Vn and LYMV-Vn AV1 (coat protein) sequence was a putative recombination site. This

was supported by sequence alignments and phylogenetic analysis of the AV1 and AC1 sequences that showed the SLCV-Vn and LYMV-Vn AV1 sequences were nearly identical, whereas in the remainder of the genome their sequences were markedly different. Recombinant geminiviruses have now been identified infecting a range of crops, including cotton [37], cassava [3, 7, 27, 38], and tomato [16, 21]. Recombinant viruses have been identified within each geminivirus genus [25], with recombinant fragments distributed across the genome, including the coat protein. All of the cucurbit geminiviruses we have isolated from Vietnamese samples had nearly identical coat protein sequences. Yet, they infected different host species, as we only detected LYMV-Vn in loofa in southern Vietnam. The reasons for this are unclear. It has been proposed that viral coat protein sequences evolve in response to the vector, whereas replication proteins evolve in response to the host [19]. We have identified possible recombination within the coat protein gene in viruses that are transmitted by the same vector species, but infect different hosts. SLCV-Vn infected a range of cucurbit species, including pumpkin, bottle gourd, and zucchini, but was not found in loofa, whereas LYMV-Vn was only isolated from loofa. The importance of the coat protein in determining the specificity of vector transmission of geminiviruses has previously been demonstrated [4]. The coat protein sequences of SLCV-Vn and LYMV-Vn were nearly identical, however there was a high degree of sequence difference in the complementary-sense replication-associated genes. It is also possible that we did not detect LYMV-Vn in loofa in northern Vietnam due to geographical separation from southern isolates.

The genome arrangement for all the SLCV-Vn and LYMV-Vn DNA-A sequences from throughout Vietnam were virtually identical, with one exception. In two SLCV-Vn isolates (B and F), a deletion in the intergenic region shifted the first ATG codon in the AV2 ORF to 5′ of the stem-loop sequence in the common region. This is unusual, as in all geminiviruses sequenced to date, the first ATG codon of this ORF is located downstream of the stem loop sequence, as is the case for SLCV-Vn isolate K and LYMV-Vn. The location of the first methionine in the AV2 sequence of isolate K, however, was identical to SLCV-Ch. Interestingly, the first in-frame ATG encoded by AV2 in SLCV-Ch and SLCV-Vn isolate K was located in the stem of the stem loop sequence in the origin of replication. It is unknown if this site is transcriptionally active, as the first in-frame ATG of all geminivirus AV2 sequences characterised to date commences downstream of the stem loop sequence. It is possible that the second AV2 ATG is the AV2 initiation codon for SLCV-Vn and SLCV-Ch, and this corresponds to the first ATG in the AV2 of LYMV-Vn. However, 5′ RACE studies are required to determine the actual start codon of AV2 in these viruses.

SLCV-Vn and LYMV-Vn DNA-A sequences encoded two ORFs in the virion sense and five ORFs in the complementary sense. The presence of both AC4 and AC5 ORFs is unusual, but has been recorded for a number of begomoviruses including ICMV [12], AYVV [31], and *Watermelon chlorotic stunt virus* (WmCSV) [15]. The presence of an AC5 ORF internal to the AV1 ORF has also been reported for PHV [29]. The AC5 ORF was shown to be non-functional in WmCSV [15].

Interestingly, the amino acid sequences encoded by the SLCV-Vn DNA-B were also most similar to those encoded by DNA-B of ICMV and PHV, and those of LYMV-Vn were most similar to ICMV.

The nucleotide sequence of the SLCV-Vn DNA-B contained a 47 nt deletion in the intergenic region, upstream of the stem-loop, although in the remainder of the intergenic region the DNA-A and DNA-B sequences were almost identical. The three iterons 5′ of the TATA box were almost identical in both components. The presence of the deletion in the SLCV-Vn DNA-B intergenic region is unique, as in all other begomovirus DNA-B sequences reported to date, the intergenic regions of DNA-A and DNA-B are nearly identical. This deletion was not present in the intergenic region of LYMV-Vn, but was present in all isolates of SLCV-Vn tested. A complementary iteron was not identified downstream of the TATA box in any Vietnamese isolate, although a short sequence with some complementary to sequences surrounding the iterons was present.

The absence of a complementary iteron downstream of the TATA box is unusual, as this iteron is present in most eastern hemisphere geminiviruses. This arrangement is similar to that of a Thailand isolate of TYLCV (TYLCV-T, [1, 28], which, like the cucurbit-infecting geminiviruses in this study, lacks a complementary iteron 3′ of the TATA box. The putative SLCV-Vn and LYMV-Vn iterons share some similarities, and the core sequence of the second iteron is identical in both viruses. In many geminiviruses, there is a correlation between the iteron core sequence (GGNNN) and the N-Terminal Rep amino acid sequence, identified as a possible iteron-related domain (IRD) [2]. There was, however, no such correlation between the iteron and the N-Terminal Rep amino acid sequence of either SLCV-Vn or LYMV-Vn. This is similar to a number of other viruses that have atypical Rep-IRD sequences, including an isolate of TYLCV and *East African cassava mosaic virus* (EACMV, [2]). Experiments are underway to determine *in vitro* whether the Rep protein of SLCV-Vn can initiate replication of LYMV-Vn and vice-versa.

Acknowledgements

The authors thank D. Martin and G. Weiler for the use of the RdP and PhyloPro recombination programs respectively, and for assistance in the analysis of the recombination data. The authors also thank Dr. T. Burns for use of the Thai sequence prior to its publication. This research was funded by the Australian Centre for International Research (ACIAR).

References

1. Arguello-Astorga GR, Guevara-Gonzalez RG, Herrera-Estrella LR, Rivera-Bustamante, RF (1994) Geminivirus replication origins have a group-specific organization of iterative elements: A model for replication. Virology 203: 90–100
2. Arguello-Astorga GR, Ruiz-Medrano R (2001) An iteron-related domain is associated to Motif 1 in the replication proteins of geminiviruses: identification of potential interacting amino acid-base pairs by a comparative approach. Arch Virol 146: 1465–1485

3. Berrie LC, Rybicki EP, Rey MEC (2001) Complete nucleotide sequence and host range of South African cassava mosaic virus: further evidence for recombination amongst begomoviruses. J Gen Virol 82: 53–58
4. Briddon RW, Pinner MS, Stanley J, Markham PG (1990) Geminivirus coat protein replacement alters insect specificity. Virology 177: 85–94
5. Dry IB, Rigden JE, Krake LR, Mullineaux PM, Rezaian MA (1993) Nucleotide sequence and genome organization of Tomato leaf curl geminivirus. J Gen Virol 74: 147–151
6. Fauquet CM, Maxwell DP, Gronenborn B, Stanley J (2000) Revised proposal for naming geminiviruses. Arch Virol 145: 1733–1761
7. Fondong VN, Pita JS, Rey MEC, de Kochko A, Beachy RN, Fauquet CM (2000) Evidence of synergism between African cassava mosaic virus and a new double-recombinant geminivirus infecting cassava in Cameroon. J Gen Virol 81: 287–297
8. Fontes EPB, Eagle PA, Sipe PS, Luckow VA, Hanley-Bowdoin L (1994a) Interaction between a geminivirus replication protein and origin DNA is essential for viral replication. J Biol Chem 269: 8459–8465
9. Fontes EPB, Galdfelter HJ, Schaffer RL, Petty ITD, Hanley-Bowdoin L (1994b) Geminivirus replication origins have a modular organization. Plant Cell 6: 405–416
10. Guitierrez C (1999) Geminivirus DNA replication. Cell Mol Life Sci 56: 313–329
11. Hanley-Bowdoin L, Settlage SB, Orozco BM, Nagar S, Robertson D (1999) Geminiviruses: Models for plant DNA replication, transcription, and cell cycle regulation. Crit Rev Plant Sci 18: 71–106
12. Hong YG, Robinson DJ, Harrison BD (1993) Nucleotide sequence evidence for the occurrence of three distinct whitefly-transmitted geminiviruses in cassava. J Gen Virol 74: 2437–2443
13. Hong Y, Wang X, Tian B (1995) Chinese squash leaf curl virus: a new whitefly-transmitted geminivirus. Sci China (Series B) 38: 179–186
14. Kheyr-Pour A, Bendahmane M, Matzeit V, Accotto GP, Crespi S, and Gronenborn B (1991) Tomato yellow leaf curl virus from Sardinia is a whitefly-transmitted monopartite geminivirus. Nucleic Acids Res 19: 6763–6769
15. Kheyr-Pour A, Bananej K, Dafalla GA, Caciagli P, Noris E, Ahoonamesh A, Lecoq H, Gronenborn B (2000) Watermelon chlorotic stunt virus from the Sudan and Iran: Sequence comparisons and identification of a whitefly-transmission determinant. Phytopathology 90: 629–635
16. Kirthi N, Maiya SP, Murthy MRN, Savithri HS (2002) Evidence for recombination among the tomato leaf curl virus strains/species from Bangalore, India. Arch Virol 147: 255–272
17. Lazarowitz SG, Lazdins IB (1991) Infectivity and complete nucleotide sequence of the cloned genomic components of a bipartite Squash leaf curl geminivirus with a broad host range phenotype. Virology 180: 58–69
18. Martin M, Rybicki E (2000) RDP: detection of recombination amongst aligned sequences. Bioinformatics 16: 562–563
19. Matthews REF (1991) Plant Virology 3rd edn. Academic Press, California, USA
20. Mayo MA, Pringle CR (1998) Virus taxonomy 1997. J Gen Virol 79: 649–657
21. Navas-Castillo J, Sanchez-Campos SS, Noris E, Louro D, Accotto GP, Moriones E (2000) Natural recombination between tomato yellow leaf curl virus-Is and tomato leaf curl virus. J Gen Virol 81: 2797–2801
22. Navot N, Pichersky E, Zeidan M, Zamir D, Czosnek H (1991) Tomato yellow leaf curl virus: A whitefly-transmitted geminivirus with a single genomic component. Virology 185: 151–161

23. Norris E, Hidalgo E, Accotto GP, Moriones E (1994) High similarity among the tomato yellow leaf curl isolates from the West Mediterranean Basin: the nucleotide sequence of an infectious clone from Spain. Arch Virol 135: 165–170
24. Orozco B, Gladfelter HJ, Settlage SB, Eagle PA, Gentry RN, Hanley-Bowdoin L (1998) Multiple cis elements contribute to geminivirus origin function. Virology 242: 346–356
25. Padidam M, Sawyer S, Fauquet CM (1999) Possible emergence of new geminiviruses by frequent recombination. Virology 265: 218–225
26. Page RDM (1996) Treeview: An application to display phylogenetic trees on personal computers. Comput Appl BioSci 12: 357–358.
27. Pita JS, Fondong VN, Sangare A, Otim-Nape GW, Ogwal S, Fauquet CM (2001) Recombination, pseudorecombination and synergism of geminiviruses are determinant keys to the epidemic of severe cassava mosaic disease in Uganda. J Gen Virol 82: 655–665
28. Rochester DE, DePaulo JJ, Fauquet CM, Beachy RN (1994) Complete nucleotide sequence of the geminivirus tomato yellow leaf curl virus, Thailand isolate. J Gen Virol 75: 477–485
29. Rojas MR, Gilbertson RL, Russell DR, Maxwell DP (1993) Use of degenerate primers in the polymerase chain reaction to detect whitefly-transmitted geminiviruses. Plant Dis 77: 340–347
30. Sambrook J, Fritsch EF, Maniatis T (1989) Molecular Cloning: A laboratory manual. Cold Spring Harbour Laboratory Press, Cold Spring Harbour, New York
31. Tan PHN, Wong SM, Wu M, Bedford ID, Saunders K, Stanley J (1995) Genome organization of Ageratum yellow vein virus, a monopartite whitefly-transmitted geminivirus isolated from a common weed. J Gen Virol 76: 2915–2922
32. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res 24: 4876–4882.
33. Torres-Pacheco I, Garzon-Tiznado JA, Herrera-Estrella L, Rivera-Bustamante RF (1993) Complete nucleotide sequence of Pepper huasteco virus: analysis and comparison with bipartite geminiviruses. J Gen Virol 74: 2225–2231
34. Van Regenmortel MHV, Bishop DHL, Fauquet CM, Mayo MA, Maniloff J, Calisher CH (1997) Guidelines to the demarcation of virus species. Arch Virol 142: 1505–1518
35. Van Regenmortel MHV, Mayo MA, Fauquet CM, Maniloff J (2000) Virus nomenclature: consensus versus chaos. Arch Virol 145: 2227–2232
36. Weiler GF (1998) Phylogenetic profiles: A graphical method for detecting genetic recombinations in homologous sequences. Mol Biol Evol 15: 326–335
37. Zhou X, Liu Y, Robinson DJ, Harrison BD (1998) Four DNA-A variants among Pakistani isolates of Cotton leaf curl virus and their affinities to DNA-A of geminivirus isolates from Okra. J Gen Virol 79: 915–923
38. Zhou X, Robinson DJ, Harrison BD (1998) Types of variation in DNA-A among isolates of East African cassava mosaic virus from Kenya, Malawi, and Tanzania. J Gen Virol 79: 2835–2840

Author's address: Dr. Peter Revill, Plant Biotechnology, Science Research Centre, Queensland University of Technology, Gardens Point, Brisbane, Queensland, Australia; e-mail: p.revill@qut.edu.au