Kasetsart J. (Nat. Sci.) 28: 632-639 (2537)

Complete Nucleotide Sequence and Genome Analysis of Bipartite Tomato Yellow Leaf Curl Virus in Thailand

Supat Attathom[1], Pissawan Chiemsombat[1], Wichai Kositratana[1] and Nuchnard Sae-Ung[2]

ABSTRACT

The genome of Thai-isolate of tomato yellow leaf curl virus(TYLCV-THL), a bipartite geminivirus, was cloned and the nucleotide sequence was analysed. The genome designated component A and B are 2751 and 2737 nucleotides long, respectively. The component A contains seven open reading frames(ORFs) with coding potential for proteins of molecular weight more than 10,000 Daltons. The component B has totally three ORFs on both strands. Computer-based analysis to compare nucleotide sequence of TYLCV-THL DNAs and the predicted amino acids with other selected geminiviruses indicated high degree of similarity with geminivirus which infected tomato in Australia and Israel for the component A. The component B showed greater homology of amino acid sequence with those of cassava latent and squash leaf curl geminiviruses.

Key words : tomato yellow leaf curl virus, geminiviruses, nucleic acid sequence

INTRODUCTION

Tomato yellow leaf curl virus (TYLCV) in Thailand is classified into a member of geminivirus group, possesing two components of about 2.7 kb genomes (Attathom *et al.*, 1990, Rochester *et al.*, 1990). The virus is transmitted by whitefly and caused yellow leaf curl disease of tomato, *Lycopersicon esculentum* Mill. (Thongrit *et al.*, 1986). The disease was also recognized in the Mediterranean basin (Navot *et al.*, 1991), Sardinia and Italy (Kheyr-Pour *et al.*, 1991) but the causal virus had been characterized as monopartite geminivirus. Agroinoculation with the component A of TYLCV-Thailand isolate can cause systemic infection in plants, in the absence of the component B (Rochester *et al.*, 1990).

Here we report the DNA sequence and genome organization of TYLCV-Thailand isolate(TYLCV-THL). For sequencing the dideoxy chain termination method and an automated DNA sequencer were employed. The nucleotide sequence and amino acid sequence were compared with other published geminivirus genomes.

MATERIALS AND METHODS

Virus and DNA clones

TYLCV was purified from infected tomato tissues and viral single stranded DNA was extracted from virus particles. Circular double stranded DNA was synthesized and cloned into Bluescript plasmid at EcoRI site. Full length cloned DNAs of TYLCV components A and B were generated and designed as pTYLCA and pTYLCB (Attathom *et al.*, 1990; Chiemsombat *et al.*, 1990).

Nucleotide Sequencing

Double stranded DNA templates for nucleotide sequencing were prepared by CsCl purification. Fragments of DNAs for sequencing were obtained by using exonuclease deletion method and nucleotide sequences were determined by the dideoxy chain termination method (Sanger *et al.*, 1977) by Taq dye-primer sequencing kits (Perkin Elmer Cetus). The fluorescent-dye labeled DNA fragments were analysed by using an automated DNA sequencer, ABI model 373A. Sequences were further analysed and connected

1 Dept. of Plant Pathology, Faculty of Agriculture, Kasetsart University, Kamphaengsaen, Nakorn Pathom 73140, Thailand.
2 Plant Genetic Engineering Unit, Kasetsart University, Kamphaengsaen, Nakorn Pathom 73140, Thailand.

TYLCV-A SEQ.

```
          10         20         30         40         50         60
5' GGTCAATCGG GTCCTCTCAA ACTTGGCTAT GCAATCGGGG AATGGGTCCT TATTATAGTT
          70         80         90        100        110        120
   GTGGACCTAA ATGGCATAAT TGTAAATAAG TGTATGAAAT TCAAAATTTA AAATTCAATC
         130        140        150        160        170        180
   GTGGCCATCC GTATAATATT ACCGGATGGC CGCGATTTTT TTAAAGTGGT CCCCTTGATG
         190        200        210        220        230        240
   TGATATGTCA TCCAATCAGA ACACTCTTTG AAAGCCTAAT TATTTATGGT CCCCTATTTA
         250        260        270        280        290        300
   AGACTTAGTC CCCAAGTTTC GGCGAAATTC AAAATGTGGA TCCACTACTA AACGAATTTC
         310        320        330        340        350        360
   CAGAAAACGT CCACGGTTTC CGTTGTATGT TAGCGGTGAA GTATCTGCAA GCGGTCGAGA
         370        380        V2 start 400        410        420
   AGACTTATTC CCCTGATACT CTAGGGTTTG ATCTCATCCG TGATCTCATC GGTGTAATTC
         430        440        450        460        C5 stop 480
   GTGCGAAGAA CTATGTCGAA GCGTCCAGCA GATATTCTCA TTTCCACTCC CGTCTCGAAA
         490        V1 start   510        520        530        540
   GTACGTCGCC GTCTGAACTT CGACAGCCCA TACAACAGCC GTGCTGCTGT CCCCACTGTC
         550        560        570        580        590        600
   CGCGTCACAA AAGGGCAGGT ATGGAAGAAC CGACCTGCAT ACAGAAAGCC CAGGATCTAC
         610        620        630        640        650        660
   AGAATGTATA GAAGCCCTGA TGTCCCTAAG GGATGTGAGG GCCCATGTAA GGTCCAATCT
         V2 stop 680        690        700        710        720
   TTCGATGCGA AGAACGACAT TGGACATATG GGCAAGGTAA TATGTCTGTA TGACGTTACC
         730        740        750        760        770        780
   CGTGGTATTG GGCTTACTCA TCGAGTTGGC AAGCGTTTCT GTGTGAAGTC ACTTTATTTT
         790        C5 start 810        820        830        840
   GTCGGGAAGA TCTGGATGGA TGAAAATATT AAGGTTAAGA ATCATACTAA CACCGTTTTA
         850        860        870        880        890        900
   TTCTGGATAG TTAGGGATCG CAGTCCTACT GGAACGCCTT ATGATTTTCA GCAGGTCTTT
         910        920        930        940        950        960
   AATGTATATG ATAATGAACC CAGCACTGCT ACTGTGAAGA ACGACCAGCG TGATCGTTTC
         970        980        990        1000       1010       1020
   CAGGTTATAA GGAGGTTCCA GGCAACAGTT ACTGGTGGAC AATATGCAGC TAAGGAGCAG
        1030       1040       1050       1060       1070       1080
   GCGATTATTA GAAAGTTTTA TCGTGTTAAT AATTATGTAG TTTATAATCA CCAGGAAGCT
        1090       1100       1110       1120       1130       1140
   GGGAAGTATG AGAACCATAC TGAAAATGCT TTGTTGTTAT ATATGGCATG TACTCATGCC
        1150       1160       1170       1180       1190       1200
   TCTAACCCTG TGTATGCTAC TTTGAAAGTC AGGAGTTATT TCTATGACTC AGTGACGAAT
V1 stop 1210       1220       1230       1240       1250       1260
   TAATAAATAT TAAATTTTAT ATCATGTTCT TCAATTACAT CAATTGTTCC TTCTAATACA
C3 stop 1270       1280       1290       1300       1310       1320
   TTGTACAATA CATGAGATAT TGCCCTAATT ACATTGTTTA TACTAATCAC GCCTAATCTA
        1330       1340       1350       1360       1370       1380
   TTTAAATATT TATTACATTG ATATTTAAAT ACTCTTAAGA AACGCGAGGT CTGAGGATGT
        1390       1400       C2 stop 1420       1430       1440
   AAATGAGTCC AGATTTTGCA GGCTAGAAAA CATTTGTGTA TCCCCAACGC TTTCCTCAGG
        1450       1460       1470       1480       1490       1500
   TTGTAGTTGA ACTGGATTTG TAACGTGATT ATGTCGTGGT TCCTCTGGAA TGGTCTCTCT
        1510       1520       1530       1540       1550       1560
   AGGTGCCTGG TTATCTTGAA ATATAGGGGA TTTTTGACCG TCCAGATATA TACGCCACTC
        1570       1580       1590       1600       1610       1620
   TCTGATTGAG TTGCAGTGAG TAGTTCCCCG GTGCGTGAAT CCATTATTGT GACAATATGT
        1630       1640       1650       1660       CJ start   1680
   GACAGCTATG CACAAGTACG AACATCCACA AGGTTGATCA ACTCTCCGTC GTCTGGTTGT
        1690       1700       1710       C1 stop 1730       1740
   CCTCTTGGCT ATTCGGTGTT GCACCTTGAT AGGTACCTGA GTACAGTGGG CCTTCGAGGG
        1750       1760       1770       1780       1790       1800
   TGACGAAGAT CGCATTTTTT ATAGCCCAGT TTCTAAGTGC GGAGTTCTTT TCCTCTTCCA
        1810       C2 start   1830       1840       1850       1860
   AGTACTCTTT ATAGCTGGAG TTTGGTCCAG GATTGCAGAG GAAGATAGTG GGAATTCCGC
        1870       1880       1890       1900       1910       1920
   CTTTAATTTG AACTGGTTTG CCGTATTTGG TGTTGCTTTG CCAGTCCCTT TGGGCCCCCA
        1930       1940       1950       1960       1970       1980
   TGAACTCTTT AAAGTGTTTT AGATAATGCG GATCGACGTC ATCGATGACG TTGAACCACG
        1990       2000       2010       2020       2030       2040
   CATCATTATT ATACACTTTT GGACTTAAAT CTAAATGGCC ACACAGATAA TTATGTGGTC
        2050       2060       2070       2080       2090       2100
   CCAATGACCT AGCCCACATC GTCTTCCCTG TTCTGCTATC ACCCTCAATG ACTATAGTTA
        2110       2120       2130       2140       2150       2160
   TTGGTCTCAA TGGCCGCGCA CGACTGAGAA CATTATTACA AGCCCACTCT TCAAGTTCTT
        2170       2180       2190       2200       2210       2220
   CGGGAACTTG ATTGAAAGAA GAAGAAGAAA AGGGAGAAAT ATATTCCTCT ATTGGAGGAG
        2230       2240       2250       2260       2270       2280
   TAAAAATCCT ATCTAAATTA GAATTTAAAT TGTGAAATTG TAAAACAAAA TCTTTAGGGG
        2290       2300       2310       2320       2330       2340
   CCTTTTCCCT CAGTATATTG AGGGCCGAAG CCTTGGACCC TGAATTGATT GCCTCGGCAT
        2350       C4 stop    2370       2380       2390       2400
   ATGCGTCGTT GGCAGATTGG CAACCTCCTC TAGCCGATCT TCCATCGATT TGGAAAATTC
        2410       2420       2430       2440       2450       2460
   CATGATCAAG CACGTCTCCG TCTTTTTCCA TGTATGCTTT AACATCTGTT GAGCTTTTAG
        2470       2480       2490       2500       2510       2520
   CTCCCTGAAT GTTCGGATGG AAATGTGCTG ACCTGGTTGG GGATGTGAGA TCGAAGAATC
        2530       2540       2550       2560       2570       2580
   GGTTATTTTG GCATTTGAAT TTTCCTTCGA ATTGGATGAG GAGATGCAGG TGAGGAGTCC
        2590       2600       2610       2620       2630       2640
   CATCTTCATG GAGTTCCCTG CAGATTCTGA TGATTAATTT ATTAGTTGGA GTTGATAGGT
        C4 start   2660       2670       2680       2690       2700
   TTAATATTTG GGAGAGTGCT TCTTCTTCTT TGGTGAGTGA GCAGTGTGGG TATGTGAGGA
        2710       2720       2730       2740       2750
   AATAATTCTT GGCATTTATT AGAAATTTCT TTGAAGGAGG CATATTGACT T 3'
                                              C1 start
```

TYLCV-B SEQ.

```
          10          20          30          40          50          60
5' GGTCAATCGG TGTCTCTCAA ACTTGGCTAT GCAATCGGGT GCTGGTGTCT TATTTATACC
          70          80          90         100         110         120
   TGGACACCAA ATGGCATAAT TGTAATTTAG TAAATGTAAT TCAAAATTCA AAATCCAATC
         130         140         150         160         170         180
   GTGGCCATCC GTATAATATT ACCGGATGGC CGCGATTTTT TTCAAGTGGT CCCACCACTG
         190         200         210         220         230         240
   ACATCTCAGC CCCACTGGGA AAATCATAGC CGTTGAATCG TAACACGCGT TACATATCCA
         250         260         270         280         290         300
   ATTGATTTGC AGTGTATTTA TTCTGACTAT AAATGTGATG TTATGGCGTC ATTTGTTGTT
         310         320         330         340         350         360
   GTGAAATAGC ATGAGAGTTC CTATTAGACG ACCTTTTGCT TAAACATGAA CCCCTCCTCC
         370         380         390         400         V1 start    420
   GGCGGTTCCT ATGTCCGTTA TAATTATCCC TACGCCAATT ACTTTGGCAG GAGGGTTGGC
         430 V2 start 440         450         460         470         480
   AACCGTGTAT ATGGAATGCC CTTTGGTAGT ACCACCTCTG TCCGGCGACC AATTAGAAGT
         490         500         510         520         530         540
   ACAGTTCGGA GGAATTTATT TTCCGACCAG TCTTCTTCAG GTAACAAGAG CCGTAAAACT
         550         560         570         580         590         600
   ATTGAGGAAG TGCATGATGG TTCTGACTAT CTTCTTGGTA ATAACACTTC GAAGGTGTCG
         610         620         630         640         650         660
   TATATTAGTT ATCCTCCTCT TAGTCGGTCG GAATTTGGTA ACCGTCTTGA CGCATTTGTC
         670         680         690         700         710         720
   AAGATTTTGG GATTTAATGT TTCTGGTTCA GTTGCTGTGA GACATCTGGA ACAGCGTGCC
         730         740         750         V2 stop     770         780
   ACTGGAGCAA GTCAAGGTAT CCATGGCATA TATTGCACAG CTGTTGTACG TGACAAGCGA
         790         800         810         820         830         840
   CCATGTCAGT TCTCCGCTGT GGAGCCTATT GTGCCATTTC CGGAGTTATT TGGTCTGGAG
         850         860         870         880         890         900
   AAGATGGCAT GCTCCTCATT ACGTGTTAGG GATATTCATA GGAGTAGGTT TAGTCCAGTT
         910         920         930         940         950         960
   TACCAGAAGA AAGTTGTTGT TAACAGCTCC CTTCCGACCC ATGTTTTTAA GTTTAATATC
         970         980         990        1000        1010        1020
   CTGTTAAGTT TAATAGGTTT CCGTTCTGGG TGTCATTTAA AGATACCTCT GATTCTGAGC
        1030        1040        1050        1060        1070        1080
   CAAGTGGCCG GTACAGCAAC GTATCTAAGA ATGCTCTCAT AGTTTATTAT TTGGCTTTGT
        1090        1100        1110         V1 stop    1130        1140
   GACACAAATG TAACTGCAGA TTCGTATGTA AAGTACGATT TGAACTACAT TGGATAAATA
        1150        1160        1170        1180        1190        1200
   AAATTATATT TGCTTTGCAT TTGAAGATAC ATTACTCCCT TCCATACATA GTTCAACTGT
     C1 stop   1220        1230        1240        1250        1260
   TTTTTTAATA ATATTAATTA CATCGTCATT TACTTTGTTG CTTCCTGCAA CAACCTCCGA
        1270        1280        1290        1300        1310        1320
   CGCCGAAGGG CCTGGATCGA GAGTTGCATC TTGTAGCTGA TGCAAATGTC TATATGGGTA
        1330        1340        1350        1360        1370        1380
   TTCCTCGTTC AATTGGTTCA CGACTGTGCT TGCGGAAGCC CAAGTATGCC CTTGTGGAAT
        1390        1400        1410        1420        1430        1440
   GGCATCTGTA GTGTATCTGG AAGACTGTGA CCTCATGTTT GATAGGGCTT GGACTGGTTT
        1450        1460        1470        1480        1490        1500
   CCTTGATACC TTGGATTGAG GCACATGCCA GAAGTCTATG TGGGTCATGT TATACGCCTT
        1510        1520        1530        1540        1550        1560
   TGACAGTATT TCAATCTTGG GGGATTTAAA TGTTATCTCC GAGGACTGTT TTGCAGATGA
        1570        1580        1590        1600        1610        1620
   CATCTTTAAT TTTCCTTGCA TCCTGCAGAA GTGAACTCCA TTTACCACGT TAGTGTCGTC
        1630        1640        1650        1660        1670        1680
   CACTCGGTAC AGTACTCTCC ATGGATTTGG GTCTTTGGGG GAGAAATAGG AAGAGGAGTA
        1690        1700        1710        1720        1730        1740
   GTAGTGTATG TTGCAGTTGC ACCCTATGGG TATAGTGAAC TCAGCCTGCT TCGAGTCACC
        1750        1760        1770        1780        1790        1800
   TTCGTGTAAC CTTGTGTCGT GCATTTCTAT GACCACATGA CCAGTTGCAT TTACAGGGAC
        1810        1820        1830        1840        1850        1860
   TTGGTTCCTG TATTGGAGGA TGACGTGGTC AATTCTGAGA CACTTTCCCA AAAGCATCGA
        1870        1880        1890        1900        1910        C1 start
   CAATTTTTGA TCGACCGTGG AGGGAAACAA CAGCGTAACT TCTGTTTTTT CATTTGACAG
        1930        1940        1950        1960        1970        1980
   CTGAAATTCA GTCCTATCCG ACGTCCGTAT ATGCAATATT GTTAGTTCTG GACTCCATTA
        1990        2000        2010        2020        2030        2040
   ATTGTTGATG GCCAATAAAA TAATGATAAT TCCATGTTTT TATAGACTTT TGAAGAGGCT
        2050        2060        2070        2080        2090        2100
   GTCACATTAG TGGAATGTGA CTAGTCGAGT TTTTGCATTG AAAATGTACT ATCTCAAAGG
        2110        2120        2130        2140        2150        2160
   ACCAATGAGA TACGTCCACG TATGTGCTAG ACCTGACAAG AGATAAGGTT AAATACTTGA
        2170        2180        2190        2200        2210        2220
   ATGTGGTTCA GCCACGTCGC CAATATGATA ATGGAAAACA GCTGTGTCCA CTAGCAGGTC
        2230        2240        2250        2260        2270        2280
   TTGTTGTTTT AAGTTGTTAA ATATAGCATT TTGAGATATT TGTCCAATAT TAATAATATA
        2290        2300        2310        2320        2330        2340
   ATTAAAAATT AAAGGAAGTA ATAGAAGTTA TGTAAGTGGT ACTTTAGTAC CACAATTTAA
        2350        2360        2370        2380        2390        2400
   ACCCAAATTG AGAAGATTCG AGCCTCAATT AAAGAACGAT TTGTAAATCG TTATGAGTTA
        2410        2420        2430        2440        2450        2460
   TCGTGAAATG GACAACTTAG ACCAATTGTT AACTGATCCA ACAATCTAAT TACGAGAAAA
        2470        2480        2490        2500        2510        2520
   CAGAAAACAC ATACTTGAAC ATCTTGAGAA AACCCGCGCG CAGCGGCTGT GTGCCGAAAT
        2530        2540        2550        2560        2570        2580
   ATATTTAGAA TTCCAGAACT GTTAATCGCT AATTATCATA TCTAATTTTA AAGTGAAGTA
        2590        2600        2610        2620        2630        2640
   TCGTAATTCA GTGTCATGCA AATGGTATGT ATCTCCTGCA GCTAACATAA TTAAAGTTAA
        2650        2660        2670        2680        2690        2700
   CAACATTAAT GCAATATAAT. TATTGCATAT TTACAGAAAA TACTAAAAAT ACATGTTTAG
        2710        2720        2730
   AGAGAGAAAG TCTAGAGAGA AGGAAGAGCG ATTGACT 3'
```

Figure 1 Nucleotide sequence of TYLCV-THL component A and B. The veral strand sequence is given for both components. The first base at 5′ end of the 200 base-common region is designed as nucleotide 1

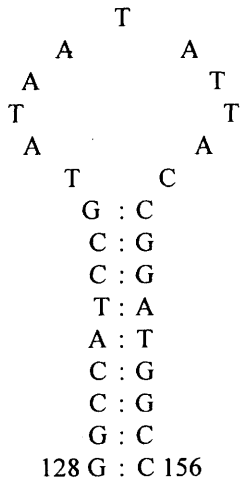

Figure 2 Nucleotide sequence of 29 base hairpin structure within TYLCV-THL component A and B.

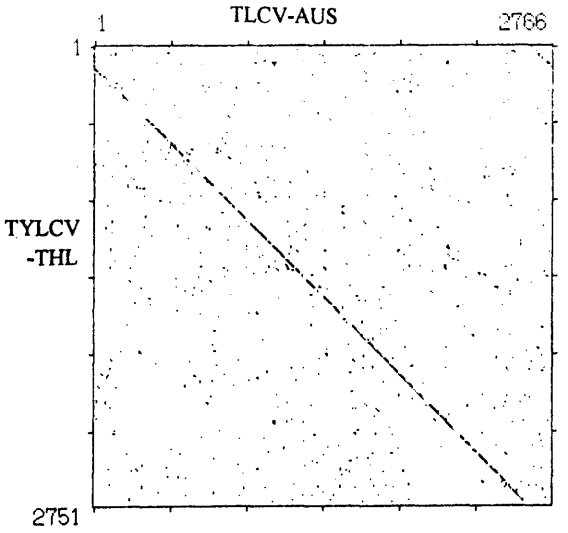

Figure 3 Computer plot of a comparison of the nucleotide sequences of TYLCV-THL component A and TLCV-AUS genomes (Check size :15, Match base no.: 11).

Table 1 Open reading frame in TYLCV-THL DNA, and size of putative translated products.

ORF*	Frame	Nucleotide		Amino acid	Protein M.W.
		Start	Stop		(Da)
Component A					
V1	1	433	1201	256	29,698
V2	3	327	609	94	10,730
C1	3	2743	1656	361	40,847
C2	2	1755	1347	136	15,567
C3	2	1604	1202	134	15,906
C4	1	2589	2292	99	11,342
C5	1	741	408	111	12,409
Component B					
V1	1	347	1061	238	26,555
V2	2	371	698	121	12,238
C1	1	1857	1146	237	26,680

*V and C denote viral strand and complementary strand ORFs ,respectively.

by using Seq. Ed computer software. The nucleotide sequences of TYLCV-DNA components A and B were compared with other reported geminivirus DNA sequences using the DNASIS program (Hitachi, Co. Ltd.).

RESULTS AND DISCUSSION

Nucleotide sequence of TYLCV-THL genomes

The sequence of both viral genomes in pTYLCA and pTYLCB are shown in Figure 1. The component A comprises 2751 nucleotides and component B comprises 2737 nucleotides. The first base of the 5′ end of the intergenic non-coding region (Lazarowitz, 1987) was designed as nucleotide 1. This region occupied 326 nucleotides, in which a conserved 29 base-hairpin structure was detected (Figure 2). The results of Harr Plot construction showed that TYLCV-THL component A is most similar to the component A-DNA of tomato leaf curl virus -Australia isolate (TLCV-AUS) (Dry *et al.*, 1993) and tomato yellow leaf curl virus-Israel isolate (TYLCV-ISR) (Navot *et*

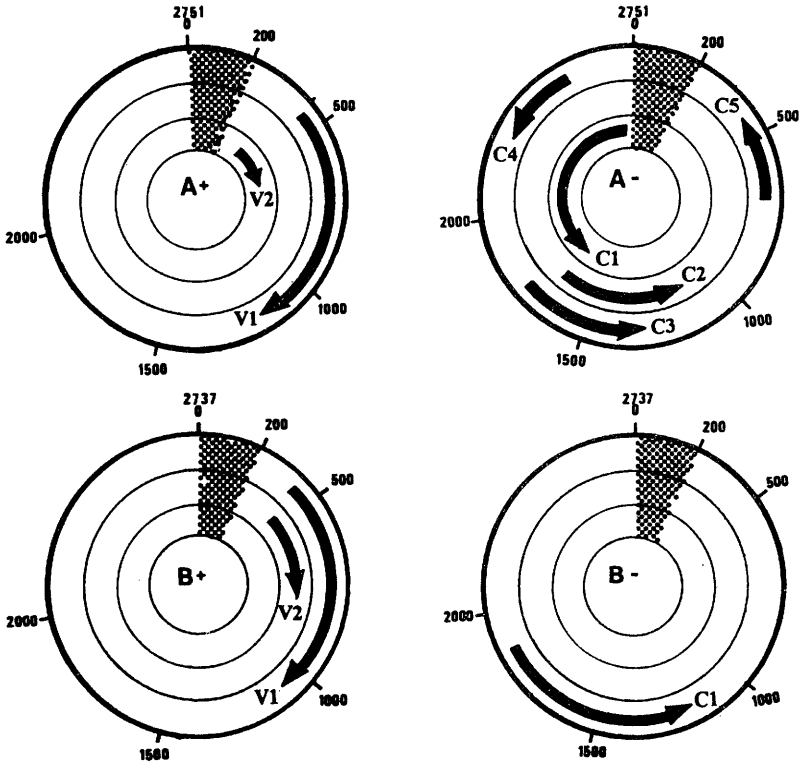

Figure 4 Potential coding regions in TYLCV-THL component A and B. All ORFs start with an ATG codon and coding for proteins with molecular weight more than 10,000 Da are shown. The common region is shaded.

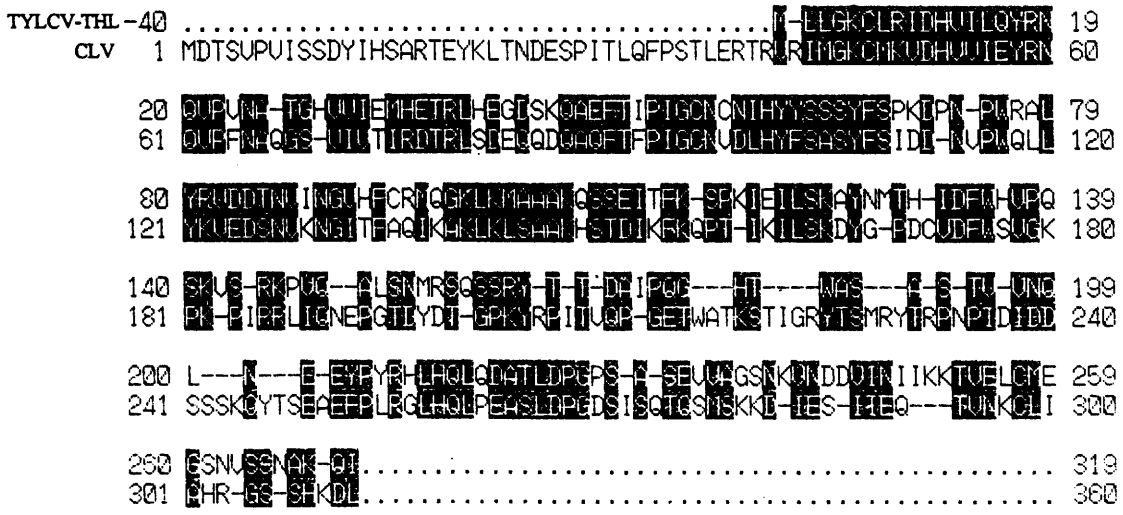

Figure 5 Amino acid homology between translated products coded for by the complementary strand ORFs of component B of TYLCV-THL and CLV DNA 2.

Table 2 Number of amino acids and sequence homology percentage between putative polypeptide products of TYLCV-Thailand isolate, TYLCV-THL, and selected whitefly-transmitted geminiviruses. TLCV-AUS: tomato leaf curl virus-Australia isolate, TYLCV-ISR : tomato yellow leaf curl virus-Israel isolate, TGMV: tomato golden mosaic virus, SqLCV: squash leaf curl virus, MYMV:mungbean yellow mosaic virus, CLV:cassava latent virus, BGMV: bean golden mosaic virus, AbMV:Abutilon mosaic virus.

ORF of TYLCV-THL	% Amino acid homology (no.of amino acid)							
(no.of amino acid)	TLCV -AUS	TYLCV -ISR	TGMV	SqLCV	MYMV	CLV	BGMV	AbMV
Component A								
V1(256)	87(256)	84(260)	85(247)	85(251)	84(244)	84(258)	84(261)	82(241)
V2(94)	72(115)	84(116)	-	46(98)	-	86(113)	43(114)	-
C1(361)	89(362)	88(357)	39(156)	67(316)	83(361)	82(358)	41(158)	39(115)
C2(136)	79(135)	82(134)	65(132)	65(134)	65(134)	76(134)	66(132)	79(135)
C3(134)	74(135)	75(135)	69(129)	65(131)	48(99)	70(135)	59(172)	63(129)
Component B								
V1(238)	-	-	52(256)	52(283)	52(256)	51(256)	53(256)	55(257)
C1(237)	-	-	84(184)	64(293)	59(298)	66(298)	64(293)	61(292)

al., 1991) with 74% and 69% total nucleotide similarity, respectively (Figure 3), despite that both latter geminiviruses had been reported to possess only one component which correspond to the component A of TYLCV-THL.

Alignment of the sequence of TYLCV-THL component B with those of bipartite geminiviruses showed that the B component of TYLCV-THL and tomato golden mosaic virus (TGMV) (Hamilton *et al.*,1984) as well as squash leaf curl virus (SqLCV-e) (Lazarowitz and Lazdins,1991) were about 55% nucleotide homology. In addition, the conserved hairpin loop structure of component B of the latter showed only two bases differences from that of TYLCV-THL as following ;

GGCCATCCGTA-TAATATTACCGGATGGCC.
C-A

Coding region and putative translation products

The viral strands of both components A and B of TYLCV-THL have two open reading frames (ORFs) with coding potential for proteins of molecular weight more than 10,000 daltons (Da). All ORFs are described in Table 1 and their organization along the genomes are depicted in Figure 4.

Direct comparison of individual ORFs of TYLCV-THL with those of other mono-and bi-partite geminivirus genomes showed a striking high degree of homology to TLCV-AUS and TYLCV-ISR, especially an ORF V1 which is the viral coat protein (Table 2). The only exception is that an ORF V2 of 10,730 Da of TYLCV-THL is more closely related (86% homology) to an ORF V2 of 13,101 Da of cassava latent virus (CLV) (Stanley and Gay,1983) rather than those of tomato-infected geminiviruses.

The presence of four ORFs in the complementary strand of component A is very similar to those observed for whitefly transmitted bipartite geminiviruses except for bean golden mosaic virus, BGMV (Howarth *et al.*,1985), mungbean yellow mosaic virus, MYMV (Ikegami,1988), and TGMV. The largest ORF (C1) encoded for protein molecular weight of 40,789 Da which located on the complementary strand of the A component showed 89% and 88% amino acid homology with C1 ORF of TLCV-AUS and TYLCV-ISR, respectively. This C1 polypeptide is believed to be a protein involving the replication of each viral genomic components (Elmer *et al.*, 1988).

Based on the above comparison, both on the nucleic acid sequences and the predicted amino acid sequences, TYLCV-THL showed greatest homology with TLCV-AUS and TYLCV-ISR when compared their component A sequences.

The component B of TYLCV-THL has two ORFs in the viral strand and one ORF in the complementary strand. Unlike component A ,the viral strand ORF showed comparatively low amino acid homology with those of other geminiviruses, ranging from 51-55% for CLV, SqLCV, MYMV, TGMV, BGMV and Abutilon mosaic virus, AbMV (Frischmuth *et al.*, 1990), descendantly. It is also very interesting that the ORF of complementary strand exibited high amino acid homology with those of CLV (66%), SqLCV (64%), BGMV (64%) and AbMV (61%). A significant homology is found with TGMV (84%), although this translated TGMV peptide (184 amino acids) is shorter than TYLCV-THL (237 amino acids). Extensive amino acid homology between translated products coded for by C1 of TYLCV-THL and CLV-DNA2 is shown in Figure 5.

Considering the biological properties, especially host range, with other tomato infected geminiviruses according to the data previously published (Thongrit *et al.*,1986; Mansour and Al-Musa,1992; Dry *et al.*, 1993), TYLCV-THL is more similar to TLCV-AUS than TYLCV-ISR. Unlike TYLCV-ISR, TYLCV-THL could not infect legume and cotton plants. As it is proposed that the component B is needed for systemic spread(Elmer *et al.*,1988), it will be very useful to study gene function of the component B of TYLCV-THL and the difference between this bipartite tomato-infecting geminivirus and other monopartite ones which infected tomato and some other hosts as well.

ACKNOWLEDGEMENT

This work was financially supported by STDB. The authors sincerely thank Prof. Dr. M. Ikegami for his excellent advice on computer analysis of DNA sequences. An automated DNA sequencer is the special donation of KU-JAPAN Phase II Project.

LITERATURE CITED

Attathom, S., P. Chiemsombat, T. Sutabutra and R. Pongpanitanond. 1990. Characterization of nucleic acid of tomato yellow leaf curl virus. Kasetsart Journal (Nat. Sci. Suppl.) 24:1-5.

Chiemsombat, P., W. Kositratana, S. Attathom, T. Sutabutra and N. Sae-Ung. 1990. DNA Probe

and nucleic acid hybridization for plant virus detection. Kasetsart Journal (Nat. Sci. Suppl.) 24:12-16.

Dry, I.B., E.R. Justin, R.K. Leslie, P.M. Mullineaux and M.A. Rezaian. 1993. Nucleotide sequence and genome organization of tomato leaf curl geminivirus. J.Gen.Virol. 74:147-151.

Elmer, J.S., L.Brand, G.Sunter, W.E. Gardiner, D.M.Bisaro and R.N.Beachy.1988. Genetic analysis of tomato golden mosaic virus II. The product of the AL1 coding sequences required for replication. Nucleic Acids Res. 16:7043-7060.

Frischmuth, T., G. Zimmat and H. Jeske. 1990. The nucleotide sequence of Abutilon mosaic virus reveals prokaryotic as well as eukaryotic features. Virology 178:461-468.

Hamilton, W.D.O., V.E.Stein, R.H.A. Coutts,and K.W.Buck. 1984. Complete nucleotide sequence of the infectious cloned DNA components of tomato golden mosaic virus : potential coding regions and regulatory sequences. EMBO J. 3(9):2197-2205.

Howarth, A.J., J. Caton, M. Bossert and R.M. Goodman. 1985. Nucleotide sequence of bean golden mosaic virus and a model for gene regulation in geminiviruses. Proc. Natl. Acad. Sci. U.S.A. 82:3572-3576.

Ikegami, M. 1988. Comparative studies between amino acid sequences of potential gene products of mungbean yellow mosaic virus and other whitefly transmitted geminiviruses. J. Agric. Sci. Tokyo Nogyo Daigaku 33(2):181-189.

Kheyr-Pour, A., M.Bendamane, V. Matzeit, G.P. Accotto, S. Crespi and B. Gronenborn. 1991. Tomato yellow leaf curl virus from Sardinia is a whitefly-transmitted monopartite gemini - virus. Nucleic Acids Res.19(24):6763-6769.

Lazarowitz, S.G. 1987. The molecular characterization of geminiviruses. Plant. Mol. Biol. Rptr. 4(4): 177-192.

Lazarowitz, S.G. and I.B.Lazdins. 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.

Mansour, A. and A. Al-Musa.1992. Tomato yellow leaf curl virus : host range and virus-vector relations. Plant Pathol. 41:122-125.

Navot, N., E.Pichersky, M.Zeidan, D.Zamir, and H. Czosnek. 1991. Tomato yellow leaf curl virus: A whitefly-transmitted geminivirus with a single genomic component. Virology 185:151-161.

Rochester, D. E., W. Kositratana and R.N. Beachy. 1990. Systemic movement and symptom production following agroinoculation with a single DNA of tomato yellow leaf curl geminivirus (Thailand). Virology 178:520-526.

Sanger, F., S. Nicklen and A. R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA. 74:5463-5467.

Stanley, J. and M.R. Gay. 1983. Nucleotide sequence of cassava latent virus DNA. Nature (London) 301:260-262.

Thongrit, D. , S.Attathom and T.Sutabutra. 1986. Tomato yellow leaf curl disease in Thailand. *In* Plant virus diseases of horticultural crops in the tropics and subtropics. FFTC Book Series No.33. Taipei,Taiwan. pp.60-63.