
Sequence and recombination analyses of the geminivirus replication initiator protein

T VADIVUKARASI¹, K R GIRISH² and R USHA^{2,*}

¹Center for Excellence in Bioinformatics, and ²Department of Plant Biotechnology, School of Biotechnology, Madurai Kamaraj University, Madurai 625 021, India

*Corresponding author (Fax, 91-452-2459105; Email, usha@mrna.tn.nic.in)

The sequence motifs present in the replication initiator protein (Rep) of geminiviruses have been compared with those present in all known rolling circle replication initiators. The predicted secondary structures of Rep representing each group of organisms have been compared and found to be conserved. Regions of recombination in the Rep gene and the adjoining 5' intergenic region (IR) of representative species of *Geminiviridae* have been identified using Recombination Detection Programs. The possible implications of such recombinations on the increasing host range of geminivirus infections are discussed.

[Vadivukarasi T, Girish K R and Usha R 2006 Sequence and recombination analyses of the geminivirus replication initiator protein; *J. Biosci.* 32 17–29]

1. Introduction

Geminiviruses are circular single-stranded DNA viruses, which replicate through rolling circle replication (RCR) mechanism. The initiation of RCR is by the replication initiator protein (Rep) encoded by the DNA replicons. The current analysis was carried out between replication initiator proteins of all groups of organisms undergoing RCR to understand the evolution of the geminivirus Rep. The geminivirus Rep is a multifunctional protein; i.e. it cleaves and ligates single-stranded DNA in the invariant sequence of the hairpin loop (Laufs *et al* 1995; Orozco and Hanley-Bowdoin 1996), hydrolyzes ATP (Desbiez *et al* 1995; Orozco *et al* 1997) and has a site-specific topoisomerase activity (Pant *et al* 2001). Rep interacts with itself (Orozco *et al* 1997), the viral replication enhancer protein AC3 (Settlage *et al* 1996), and a maize homologue of the cell cycle regulatory protein, retinoblastoma (Ach *et al* 1997). In addition Rep induces the expression of a host DNA synthesis protein, proliferating cell nuclear antigen (PCNA),

in non-dividing plant cells (Nagar *et al* 1995), possibly through interactions with the plant cell cycle machinery.

Recombination can provide selective advantage in the evolution of viruses within strains, species, genera and family (Keese and Gibbs 1993; Morse 1994; Gibbs *et al* 1995; Holland 1998). It has now been accepted that recombination contributed to the diversity of geminiviruses and therefore, to the emergence of new variants and species reported worldwide. For instance, cotton leaf curl disease in Pakistan became severe during the past decade, causing extensive damage to cotton production. In Trinidad and Tobago, a geminivirus disease on tomato observed in 1989 has re-emerged throughout the country. A new cassava mosaic virus has devastated cassava production in Uganda. Tomato production in Spain and Italy is severely constrained by *Tomato yellow leaf curl virus-Sardinia* (TYLCV). All of these viruses are recombinants (Umaharan *et al* 1998). In India, Kirthi *et al* (2002) have detected recombination between strains of *Tomato leaf curl virus* from Bangalore. Girish and Usha (2005) have analysed recombination events

Keywords. Geminivirus; motifs; recombination; Rep

Abbreviations used: RCR, Rolling circle replication; RDP, Recombination Detection Program; Rep, replication initiator protein; TYLCV, *Tomato yellow leaf curl virus-Sardinia*

between legume-infecting begomoviruses from South and Southeast Asia.

The present analysis has revealed the conservation of sequence motifs and secondary structures among the RCR initiator proteins, providing a possible clue to the evolution of the geminivirus Rep. Recombination detection analysis on the *Rep* gene and the intergenic region (IR) of a total of 101 members of *Geminiviridae* has shown the regions of recombination that could explain the expanding host range of these viruses.

2. Materials and methods

2.1 Sequence motif prediction

The replication initiator proteins from the following organisms were chosen for the study:

(i) Four genera of geminiviruses (Mastre, Curto, Begomo and Topocuviruses); (ii) circoviruses; (iii) nanoviruses; (iv) parvoviruses; (v) papillomaviruses; (vi) bacteriophages; (vii) bacterial plasmids (pT181, pLS1, pUB110, pSN2 groups); (viii) phytoplasma plasmids (pOYW and EcOYW1); (ix) *P.pulchra* (red algae) plasmid; (x) transposases of bacterial IS elements; (xi) putative product of *Helitrons* (eukaryotic DNA transposons).

The above sequences were collected by running NCBI BLAST-2.2.13 with a *P*-value of 0.005, using the Rep sequence of *Mungbean yellow mosaic India virus*, (soybean isolate from Madhya Pradesh; Girish and Usha 2005), as the query sequence. EMBnet T-COFFEE run was performed among all replication initiator proteins and also within each group to arrive at the motifs common for RCR initiator proteins.

2.1a Secondary structure prediction: The three-dimensional structures for the N-terminal domain (4-121 residues) of Rep from TYLCV (Campos-Olivas *et al* 2002) and the catalytic domain of the *Adeno Associated Virus* (AAV) type 5 Rep (1-197 residues) (Hickman *et al* 2002) are known already. These two sequences were aligned (using T-COFFEE) with the sequences corresponding to the N-terminal domains of Rep representing various groups of organisms in order to predict their secondary structures. The output obtained in multiple sequence file (MSF) format was used as input for Jpred.

2.1b Phylogenetic tree construction: PileUp and PAU Psearch from GCG (Devereux *et al* 1984) version 10.3 were used. Alignments were constructed using PileUp program and the output obtained was used as input for PAUPsearch program with the option of bootstrap analysis using neighbor-joining distance. The number of bootstrap replications was set to 1000. *TreeView* was used to view the phylogenetic tree.

2.1c Analysis of recombination: Recombination Detection Program (v 1.08): Recombination Detection Program (RDP)

(Martin and Rybicki 2000) and GENECONV (Padidam *et al* 1999) were used. Multiple sequence alignment was constructed using EBI ClustalW (Thompson *et al* 1994) and the output was given as input for RDP. Options set for analysis are as follows:

General options: highest acceptable *P*-value 5×10^{-5} , Bonferroni correction on and List the events detected by > 1 method.

RDP options: Internal reference only and window size of 10.

GENECONV options: Scan sequence triplets; Treat indel blocks as one polymorphism and G-scale (mismatch penalty) of 2.

3. Results

Motifs predicted using T-COFFEE alignments are shown in figure 1.

Rep protein of geminiviruses shows five motifs namely: (i) Motif I (FLTY); (ii) Motif II (HXHUUU, where U represents a hydrophobic amino acid residue); (iii) Motif III (YXXK, where X represents any amino acid); (iv) Walker A motif or P-loop (GXXXXGKS/T); and (v) Walker B motif (DD).

Motif I shows variation both within and between different groups. Motif II (HXH) and motif III (YXXK) are highly conserved. Rep from *Porphyra pulchra*, circoviruses and *Onion yellows phytoplasma* plasmids have all the five motifs shown by geminiviruses. Bacterial plasmids lack Walker A and Walker B motifs, which are present in the C-terminus of the Rep of geminiviruses. In differentiated cells, human papillomaviruses replicate in high copy number (>100 copies per cell) by rolling circle replication (Flores and Lambert 1997). However the Rep of papillomaviruses failed to show any of these motifs.

3.1 Phylogenetic analysis

To study the evolution of the geminivirus Rep, two sets of phylogenetic analysis were carried out: (i) Using the entire Rep sequence from representative organisms of all groups and (ii) using only the N-terminal region of the protein, since Rep from all known RCR plasmids contain only the RCR initiator domain, which is present in the N-terminal region of these proteins (Oshima *et al* 2001).

3.2 Phylogenetic analysis using the entire length of Rep

In this tree (figure 2) the Rep of *P. pulchra* (red algae) is clustered along with the geminivirus Rep. However, no recombination could be detected between the above Rep. Phages, circo- and nanoviruses Rep form a separate cluster.

Accession number	[organism]	Motif I (FLTY)	Motif II (HxHUUU)	Motif III (YxxK)	P loop (GxxxxGKS/T)	WalkerB (DD)
<u>Geminivirus</u>						
gi 13249671 ref	[TGMV]	: FLTYPQC	HLHVLI	YIDK	GDSRTGKT	IDDV
gi 16151557 emb	[MYMIV-Sb-MP]	: FLTYPKC	HLHVLL	YMEK	GDSRTGKT	IDDV
gi 30146816 ref	[BSCTV]	: FLTYPQC	HLHALI	YVSK	GDSRTGKT	IDDV
gi 46402155 ref	[SpCTV]	: FLTYPQC	HLHALI	YVAK	GDSRTGKT	IDDV
gi 10257478 ref	[HrCTV]	: FLTYPQC	HLHCLI	YITK	GNSRTGKT	IDDV
gi 14794698 gb	[MSV]	: FLTYPHC	HLHALL	YILK	GPTRTGKS	VDDI
gi 68299253 emb	[WDV]	: FLTYPQC	HLHVLV	YITK	GPTRTGKT	IDDI
gi 8571400 gb A	[SSEV-Naga]	: FLTYPKC	HIHALA	YVLK	-----	----
gi 1419333 emb	[TPCTV]	: FLTYPNC	HLHVLI	YVDK	GESRTGKT	IDDV
		*****:	*:* *	*: *	*:*****	:**:
<u>Circovirus</u>						
gi 9630730 ref	[BFDV]	: FTLNNT	HLQGYF	YCSK	GPPGCGKS	LDDF
gi 39546118 gb	[MDC]	: FTINNPT	HLQGFL	YCSK	GPPGTGKS	MDDF
gi 12280942 ref	[PCV]	: FTLNNS	HLQGFA	YCSK	GPPGCGKS	LDDF
		:*:*:	**:	****	**** **	:***
<u>Nanovirus</u>						
gi 571496 gb AA	[SCSV]	: FTLN	HLQGFI	YAMK	GPAGNEGKS	
gi 4995172 emb	[FBNYV]	: FTLN	HIQGVI	YAQK	GPNGNEGKS	
gi 3798656 dbj	[MVDV]	: FTLN	HLQGFI	YAMK	GPKGGEKGS	
gi 19525669 gb	[BBTV]	: FTIN	HVQGYV	YCMK	GPNGGEGKT	
gi 9626659 ref	[CFDV]	: FTLN	HLQGFI	YCSK	GRDGDGKGS	
		****	*:** :	* *	* *.:**:	
<u>Eukaryotic plasmid</u>						
gi 7108457 gb	[<i>P.pulchra</i>]	: FLTYPQS	HFHVLL	YICK	GASGVGKT	FDDV
w.r.t begomovirus Rep		*****:.	*:***:	*: *	* * .***	***
<u>Onion yellows phytoplasma plasmid</u>						
gi 13434985 dbj	[pOYW]	: ----	HWHIYL	YMIH	GTSGSGKS	LDDL
gi 4115497 dbj	[EcoYW1]	: FLTY	HHHVFF	YVKK	GNSKSGKT	YDDI
			* *:::	*: :	* * :**:	**:
<u>Helitrons</u>						
gi 37533896 ref	[<i>O.sativa</i>]	: FLTM	HAHFLL	YLFK	GPGGTGKT	
gi 9294530 dbj	[<i>A.thaliana</i>]	: FITF	HAHILL	YLFK	GFGGTGKT	
gi 7331906 gb	[<i>C.elegans</i>]	: FLTF	HVHMLL	YLFK	GPGGSGKT	
		::	*.*:**	****	* **:*:**	
<u>Parvovirus</u>						
gi 55668312 gb	[PPV]	: FLNK	HCHVLL	YIEM	GPASTGKS	
gi 9627948 ref	[MPV]	: FLTK	HCHVLI	YIEM	GPASTGKS	
gi 9628650 ref	[GPV]	: LIPK	HLHCCI	YRSF	GPATGKT	
gi 86211074 gb	[HPV]	: LMKK	HIHVVI	YTLL	GPSTGKT	
gi 54646361 gb	[CPV]	: FLTK	HCHVLL	YIEM	GPASTGKS	
gi 51512245 gb	[AAV]	: LLPK	HMHVLV	YISF	GPATGKT	
gi 9626995 ref	[MMV]	: FLTK	HCHVLI	YIEM	GPASTGKS	
		:: *	* * :	* :	* . ***:	

Figure 1. For caption, see p. 20.

Figure 1. (Continued).

Plasmids**pE194/pLS1**

gi 132358 sp P1	[pLS1]	:	FLLYPES	HYHVLY	YDKA
gi 150185 gb AA	[pKMK1]	:	VITWPES	HYHFIL	YDER
gi 551918 gb AA	[pE194]	:	FVLYPES	HYHILV	YQKE
gi 511601 gb AA	[pCI414]	:	WIVYPES	HWHIII	YDKS
gi 3150198 emb	[pS194]	:	FLLYPDS	HYHVIY	YDKK
gi 7597481 ref	[pUB112]	:	FLLYPDS	HYHVIY	YDKK
		:	:*:*	*:*.:	*:.

pC194/pUB110 Plasmid

gi 9507322 ref	[pKAYM]	:	FLTL	HFHCLL	YSVK
gi 1075900 pir	[pBAA1]	:	FLTL	HFHVLI	YPVK
gi 97846 pir	[pC194]	:	FLTL	HMHVLV	YPVK
gi 282350 pir	[pBC1]	:	FLTL	HFHVLL	YPVK
gi 9507395 ref	[pBC16]	:	FLTL	HFHVLL	YPVK
gi 290013 gb	[pCA2.4]	:	FLTL	HFHCLL	YSVK
gi 580974 emb	[pCB101]	:	HLTL	HYHVAL	YMTK
gi 132376 sp	[pFTB14]	:	FLTL	HFHVLL	YPVK
gi 548724 sp	[pLAB1000]	:	FLTL	HMHVLL	YQVK
gi 132379 sp	[pLP1]	:	FLTL	HLHVLL	YEVK
gi 59800171 sp	[pUB110]	:	FLTL	HMHVLV	YPVK
gi 1049124 gb	[pTA1060]	:	FLTL	HFHVLI	YPVK
gi 151701 gb	[pTHT13]	:	FLTL	HFHVLL	YPVK
gi 32455781 ref	[pWC1]	:	FLTL	HIHVLL	YPVK
		:	.***	* * :	* .*

Phages

gi 138132 sp P0	[psiX174]	:		HFHAVH	YVAK
gi 11095672 gb	[S13]	:		HFHAVH	YVAK
gi 17981445 gb	[G4]	:		HFHAVH	YVAK
		:		*****	****

* Identical residues
 : very similar residues
 . Similar residues

The identities and similarities are from comparison of sequences within the group itself except for *P.pulchra*, which was compared with the begomovirus Rep.

Figure 1. Conserved sequence motifs in proteins mediating RCR initiation.

Helitrons (eukaryotic RC transposons) cluster together with prokaryotic RC IS elements. Rep of parvoviruses and plasmids fall into independent clusters.

separate single cluster. pUB110 Rep alone form a cluster. The third cluster is formed by *Helitron* putative proteins.

3.3 Analysis using N-terminal region of Rep

In this tree (figure 3) there are three main clusters. Interestingly, the Rep from EcOYW1, circo- and nanoviruses cluster with geminivirus Rep. Rep of *P. pulchra* forms a

3.4 Secondary structure prediction

Sequence identity between geminivirus Rep and RCR initiator proteins of all other groups is less than 25%. To identify the type of evolutionary relationship between these Rep with that of the geminivirus Rep, secondary structure

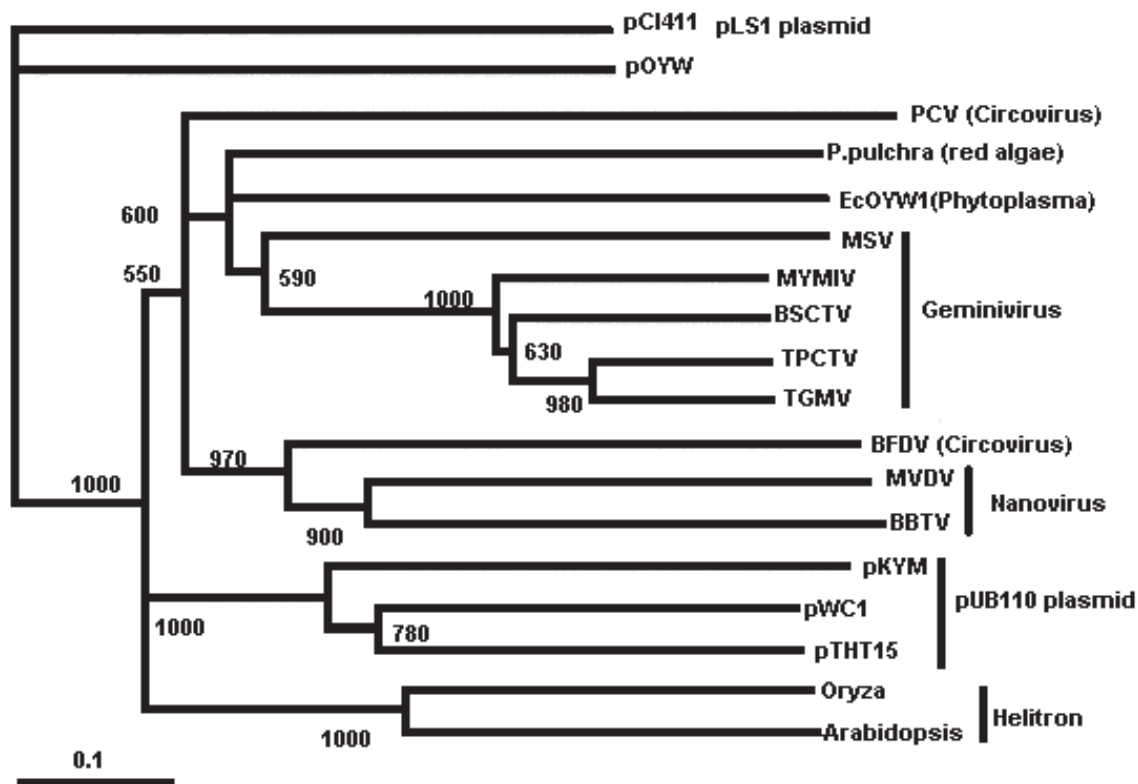


Figure 3. Phylogenetic tree constructed using only the N-terminal region of Rep protein.

Table 1. Order of helices and sheets predicted by Jpred for the Rep sequences

Rep source	Predicted secondary structure
TYLCV(geminivirus)	B-H-B-B-H
AAV (parvovirus)	B-H-B-B-B-H-B-B
BFDV (circovirus)	B-H-B-B-H-B-H
<i>Clostridium butyricum</i> (pUB110)	B-H-B-B-H-B-B
<i>P.pulchra</i> (red alga)	H-B-H-B-B-H
BBTV (nanovirus)	H-B-B-B-B-H-B-B-B
EcOYW1 (<i>Phytoplasma</i>)	H-B-H-H-B-B-B

B, beta sheet; H, alpha helices.

geminiviruses. Three motifs characterize the catalytic domain: I, (FLTYP); II, (H_xH); and III, (Y_{xxx}Y) or (Y_{xx}K). Motif III contains the active site tyrosine(s), motif II was postulated as a metal ion-binding site, and no function was ascribed to motif I (Orozco and Hanley-Bowdoin 1998). The C-terminal domain contains a NTP binding motif (G_{xxxx}GKT/S), specifying the phosphate-binding fold (P-loop). Alteration of this motif led to loss of

the ATPase and DNA helicase activities (Campos-Olivas *et al* 2002).

Motif I (FLTY), required for Rep/DNA binding and cleavage (Orozco and Hanley-Bowdoin 1998) shows high variation because Rep binding/cleavage sequences are not well conserved among different groups of organisms. The two His residues of motif II are involved in metal ion binding necessary for the activity of RCR initiator proteins (Orozco and Hanley-Bowdoin 1998), and its conservation among a wide range of RCR initiators suggests that they all are metal ion dependent proteins. The third motif, which has a tyrosine, is central to nick formation and shows conservation among a vast group of RCR initiators. Rep proteins from all known bacterial plasmids lack the helicase domain. Therefore the Walker A and B motifs are not present among the plasmid groups. *P. pulchra* plasmid replicase, *Phytoplasma* plasmid EcOYW1 initiator protein and the Rep of circoviruses show good conservation in all the five motifs with respect to the geminivirus Rep (figure 1), suggesting that the Rep proteins encoded by plasmids of *P. pulchra*, *Phytoplasma* and circoviruses are closely related to geminiviruses Rep. The probability of finding the motifs were found to be zero or very close to zero. The analysis was done with the MACAW program available at: <http://genamics.com/software/downloads/macawwin.exe>.

Table 2. Geminivirus sequences used for recombination detection.

Genus	<i>Mastrevirus</i>		
Species in the Genus		GenBank Accession Numbers	Abbreviations
1. Bean yellow dwarf virus - [South Africa:Mpumalanga:1994]		Y11023	(BeYDV-[ZA:Mpu:1994])
2. Chloris striate mosaic virus - [Australia]		M20021	(CSMV-[AU])
3. Digitaria streak virus - [Vanuatu]		M23022	(DSV-VU])
4. Maize streak virus - A[Nigeria1]		X01633	(MSV-A[NG1])
5. Miscanthus streak virus - [Japan:1991]		D01030	(MiSV-[JP91])
6. Panicum streak virus - Karino [South Africa:1989]		L39638	(PanSV-Kar[ZA:89])
7. Sugarcane streak virus - [South Africa:Natal]		M82918, S64567	(SSV-[ZA:Nat])
8. Sugarcane streak Egypt virus - [Egypt:Aswan]		AF039528	(SSEV-[EG:Asw])
9. Tobacco yellow dwarf virus [Australia]		M81103	(TYDV-[AU])
10. Wheat dwarf virus - [France:1989]		X82104	(WDV-[FR:89])
Genus	<i>Curtovirus</i>		
11. Beet mild curly top virus - [US:Worland]		U56975	(BMCTV-[US:Wor])
12. Beet severe curly top virus- [US:Cfh] (Beet curly top virus - CFH)		U02311	(BSCTV-[US:Cfh])
13. Beet curly top virus - [US:California:1985]		X04144	(BCTV-[US:Cal:85])
14. Horseradish curly top virus - [US:Salinas:1988]		U49907	(HrCTV-[US:Sal:88])
15. Spinach curly top virus - [US:Spinach 3:1996]		AY548948	(SpCTV-[US:Sp3:96])
Genus	<i>Topocuvirus</i>		
16. Tomato pseudo-curly top virus - [US:Florida:1994]		X84735	(TPCTV-[US:FL:94])
Genus	<i>Begomovirus</i>		
17. Abutilon mosaic virus - [Germany]		X15983	(AbMV-[DE])
18. African cassava mosaic virus - [Cameroon:1998]		AF112352	(ACMV-[CM:98])
19. Ageratum enation virus - [Nepal:2001]		AJ437618	(AEV-[NP:01])
20. Ageratum leaf curl virus - [China:Guangxi 52:2003]		AJ851005	(ALCuV-[CN:G52:03])
21. Ageratum yellow vein virus - [Singapore:1992]		X74516	(AYVV-[SG:92])
22. Bean calico mosaic virus - [Mexico:Sonora:1986]		AF110189	(BCaMV-[MX:Son:86])
23. Bean dwarf mosaic virus - [Colombia]		M88179	(BDMV-[CO])
24. Bean golden mosaic virus - [Brazil:Sao Paulo-Campinas1:1978]		M88686	(BGMV-[BR:SP-Ca1:78])
25. Bean golden yellow mosaic virus - [Cuba]		AJ544531	(BGYMV-[CU])
26. Bhendi yellow vein mosaic virus - [India: Madurai]		AF241479	(BYVMV-[IN:Mad])
27. Cabbage leaf curl virus - [US:Florida:1996]		U65529	(CabLCuV-[US:Flo:96])
28. Chayote yellow mosaic virus - [Nigeria:Ibadan]		AJ223191	(ChaYMV-[NG:Iba])
29. Chilli leaf curl virus - [Pakistan:Multan:1998]		AF336806	(ChiLCuV-[PK:Mul:98])
30. Chino del tomate virus - [Mexico:Sinaloa IC:1983]		AF101476	(CdTV-[MX:SinIC:83])
31. Corchorus yellow vein virus - [Vietnam:Ho Binh:2000]		AY727903	CYVV-[VN:Ho Binh:00]
32. Cotton leaf crumple virus - [Mexico:Sonora:1991]		AF480940	(CLCrV-[MX:Son:91])
33. Cotton leaf curl Rajasthan virus - [India:Sriganganagar]		AF363011	(CLCuRV-[IN:Sri])

Table 2. (Continued)

Species in the Genus	GenBank Accession Numbers	Abbreviations
34. Cowpea golden mosaic virus - [Nigeria:Nsukka:1990]	AF029217	(CPGMV-[NG:Nsu:90])
35. Croton yellow vein mosaic virus - [India]	AJ507777	(CYVMV-[IN])
36. Cucurbit leaf crumple virus - [US:Arizona:1991]	AF256200	(CuLCrV-[US:Ari:91])
37. Dicliptera yellow mottle virus - [US:Florida:1998]	AF139168	(DiYMoV-[US:Flo:98])
38. Dolichos yellow mosaic virus - [Bangladesh:Gazipur]	AY271891	(DoYMV-[BD:Gaz])
39. East African cassava mosaic Cameroon virus - [Cameroon:1998]	AF112354	(EACMCV-[CM:98])
40. Eupatorium yellow vein mosaic virus - [Japan:SOJ3:2000]	AJ438937	(EpYVMV-[JR:SOJ3:00])
41. Eupatorium yellow vein virus - [Japan:Kumamoto]	AB007990	(EpYVV-[JR:Kum])
42. Euphorbia leaf curl virus - [China:Guangxi 35:2002]	AJ558121	(EuLCV-[CN:G35:02])
43. Hollyhock leaf crumple virus - [Egypt:Cairo 1:1997]	AY036009	(HoLCrV-[EG:Cai1:97])
44. Honeysuckle yellow vein virus - [UK:Norwich 1:1999]	AJ542540	(HYVV-[UK:Nor1:99])
45. Honeysuckle yellow vein mosaic virus	AB020781	(HYVMV)
46. Horsegram yellow mosaic virus - [India:Coimbatore]	AJ627904	HgYMV-[IN:Coi]
47. Indian cassava mosaic virus - [India:Kerala 2:2002]	AJ575819	(ICMV-[IN:Ker2_02])
48. Lindernia anagallis yellow vein virus - [China:Hainan:2004]	AY795900	(LAYVV-[CN:Hai:04])
49. Ludwigia yellow vein China virus - [China:Guangxi 38:2003]	AJ965540	(LuYVCNV-[CN:G38:03])
50. Luffa yellow mosaic virus - [Vietnam]	AF509739	(LYMV-[VN])
51. Macroptilium mosaic Puerto Rico virus - [Puerto Rico:1990]	AY044133	(MaMPRV-[PR:90])
52. Macroptilium yellow mosaic virus - [Cuba]	AJ344452	(MaYMV-[CU])
53. Malvastrum leaf curl virus - [China:Guangxi 87:2004]	AJ971263	(MaLCV-[CN:G87:04])
54. Malvastrum yellow vein virus - [China:Yunnan 47:2001]	AJ457824	(MYVV-[CN:Y47:01])
55. Melon chlorotic leaf curl virus - [Costa Rica:Guanacaste:1998]	AY064391	(MCLCuV-[CR:Gua:98])
56. Mungbean yellow mosaic India virus - [India:Jabalpur:Soybean MP]	AJ416349	(MYMIV-[IN:Jab:SbTN])
57. Mungbean yellow mosaic virus - [India:Madurai:Soybean]	AJ421642	(MYMV-[IN:Mad:Sb])
58. Okra yellow mosaic Mexico virus - [Mexico:Mazatepec 3:2004]	DQ022611	(OYMMV-[MX:Maz3:04])
59. Okra yellow mottle Iguala virus - [Mexico:Iguala]	AY751753	(OYMoIV-[MX:Igu])
60. Okra yellow vein mosaic virus - [Pakistan:Faisalabad 201:1995]	AJ002451	(OYVMV-[PK:Fai201:95])
61. Papaya leaf curl virus - [India:Lucknow]	Y15934	(PaLCuV-[IN:Luc])
62. Pepper golden mosaic virus - [Costa Rica]	AF149227	(PepGMV-[CR])
63. Pepper huasteco yellow vein virus - [Mexico:Sinaloa:1988]	AY044162	(PHYVV-[MX:Sin:88])
64. Pepper leaf curl virus	AF134484	(PepLCV)
65. Pepper yellow vein Mali virus - [Mali]	AY502935	(PepYVMV-[ML])
66. Potato yellow mosaic virus - [Venezuela]	D00940	(PYMV-[VE])
67. Rhynchosia golden mosaic virus - [Honduras:Comayagua:1999]	AF239671	(RhGMV-[HN:Com:99])
68. Senecio yellow mosaic virus - [China:Guangxi 46:2005]	AJ876550	(SeYMV-[CN:G46:05])
69. Sida golden mosaic Honduras virus - [Honduras]	Y11097	(SiGMHV-[HN])
70. Sida golden yellow vein virus - [Cuba:Havana]	AJ577395	(SiGYVV-[CU:Hav])
71. Sida leaf curl virus - [China:Hainan 57]	AM050730	(SiLCuV-[CN:H57])
72. Sida micrantha mosaic virus - [Brazil:A2B2]	AJ557451	(SiMMV-[BR:A2B2])
73. Sida mottle virus - [Brazil:A1B3]	AJ557450	(SiMoV-[BR:A1B3])
74. Sida yellow mosaic virus - [Brazil:Minas Gerais-Vicosa2:1999]	AY090558	(SiYMV-[BR:MG-Vi2:99])

Table 2 (Continued)

Species in the Genus	GenBank Accession Numbers	Abbreviations
75. Sida yellow vein virus - [Honduras]	Y11099	(SiYVV-[HN])
76. South African cassava mosaic virus - [Madagascar:12]	AJ422132	(SACMV-[MG:12])
77. Soybean crinkle leaf virus - [Japan]	AB050781	(SbCLV-[JR])
78. Squash leaf curl virus - [US:Imperial Valley:1979]	M38183	(SLCV-[US:Imp:79])
79. Squash mild leaf curl virus -[US:Imperial Valley:1979 (Squash leaf curl virus - R)	AF421552	(SMLCV-[US:Imp:79])
80. Sri Lankan cassava mosaic virus - [Sri Lanka:Colombo:1998]	AJ314737	(SLCMV-[LK:Col:98])
81. Stachytarpheta leaf curl virus - [China:Hainan 6.1:2001]	AJ564742	(StaLCuV-[CN:Hn6.1:01])
82. Sweet potato leaf curl virus - [US:Louisiana:1994]	AF104036	(SPLCV -[US:Lou:94])
83. Tobacco curly shoot virus - [China:Yunnan 1:1999]	AF240675	(TbCSV-[CN:Y1:99])
84. Tobacco leaf curl Yunnan virus - [China:Yunnan 3:1999]	AF240674	(TbLCYNV-[CN:Y3:99])
85. Tomato chino La Paz virus - [Mexico:Baja El Carrizal:2002]	AY339619	(ToChLPV-[MX:BEC:02])
86. Tomato chlorotic mottle virus - [Brazil:Bahia-Seabra1:1996]	AF490004	(ToCMoV-[BR:BA-Se1:96])
87. Tomato curly stunt virus - [South Africa:Onderberg:1998]	AF261885	(ToCSV-[ZA:Ond:98])
88. Tomato golden mosaic virus - Yellow vein [Brazil]	K02029	(TGMV-YV[BR])
89. Tomato golden mottle virus - [Guatemala:R2:1994]	AF132852	(ToGMoV-[GT:R2:94])
90. Tomato leaf curl Bangalore virus - [India:Kolar]	AF428255	(ToLCBV-[IN:Kol])
91. Tomato mild yellow leaf curl Aragua virus - [Venezuela:10]	AY927277	(ToMYLCAV-[VE:10])
92. Tomato mosaic leaf curl virus - [Venezuela:Trujillo]	AY508991	(ToMLCV-[VE:Tru])
93. Tomato mosaic Havana virus - [Cuba:Quivicán]	Y14874	(ToMHV-[CU:Qui])
94. Tomato mottle virus - [US:Florida:1989]	L14460	(ToMoV-[US:Flo:89])
95. Tomato rugose mosaic virus - [Brazil:Uberlandia:1996]	AF291705	(ToRMV-[BR:Ube:96])
96. Tomato severe leaf curl virus - [Guatemala:Sansirisay:1996]	AF130415	(ToSLCV-[GT:San:96])
97. Tomato severe rugose virus - [Brazil:Minas Gerais:2000]	AY029750	(ToSRV-[BR:MG:00])
98. Tomato yellow leaf curl virus - [Israel:Rehovot:1986]	X15656	(TYLCV-[IL:Reo:86])
99. Tomato yellow margin leaf curl virus - [Venezuela:Merida 57]	AY508993	(TYMLCV-[VE:Mer57])
100. Watermelon chlorotic stunt virus - [Iran:1997]	AJ245652	(WmCSV-[IR:97])
101. Vernonia yellow vein virus	AM182232	AYVV

4.2 Phylogenetic analysis

Signatures of the Rep proteins of geminiviruses are found in the replicase of the dsDNA plasmids of the evolutionarily ancient red alga *Porphyra* (Moon and Goff 1996). *P. pulchra* plasmid replicase is placed along with geminivirus Rep in both the trees indicating that it is closely related to geminivirus Rep. Only in the tree constructed using N-terminal Rep sequence, EcOYW1 is placed along with geminivirus showing that its N-terminal domain is closely related to geminivirus Rep. It was clear from the alignment that EcOYW1 protein lacks sequences after Walker B motif at its C-terminal end. The fact that *Phytoplasma* replicates in both plant and insect (leaf

hopper) cells (Oshima *et al* 2001) may contribute to the close evolutionary relationship with geminiviruses. The phylogenetic tree shows that these proteins might have diverged from a common ancestor.

The product of Helitrons shows evolutionary relationship with prokaryotic IS elements. The identification of motifs among the Helitron-encoded products suggests that prokaryotic and eukaryotic RC elements arose from a common ancestral element (Feschotte and Wessler 2001).

The N-terminal sequences of nanovirus Rep are more closely related to the corresponding circovirus Rep sequences (Gibbs and Weiller 1999). Therefore circo and nanoviral Rep proteins are closely placed. High sequence variation between *Porcine circovirus* and *Beak and feather disease*

Table 3. A few examples of hosts common to both daughter and parental viruses.

Daughter	Host	Parent'	Host
<i>Hollyhock leaf crumple virus</i> [Egypt:Cairo]	<i>Malva parviflora</i> (E), <i>Nicotiana benthamiana</i> (E) <i>N. tabacum</i> (E)	<i>Chino del tomate virus</i> [Mexico:Sinaloa]	<i>Malva parviflora</i> , <i>Nicotiana spp.</i> (E)
<i>Ageratum yellow vein virus</i> [Singapore]	<i>Phaseolus vulgaris</i> (E)	<i>Bean golden mosaic virus</i> [Brazil:Sao Paulo-Campinas]	<i>Phaseolus vulgaris</i>
<i>Tobacco leaf curl Yunnan virus</i>	<i>Nicotiana spp.</i>	<i>Bean dwarf mosaic virus</i> [Colombia]	<i>Nicotiana benthamiana</i> (E), <i>N. tabacum</i> (E)
<i>Squash leaf curl virus</i> [US:Imperial Valley]	<i>Phaseolus vulgaris</i> (E)	<i>Macroptilium mosaic Puerto Rico virus</i>	<i>Phaseolus spp.</i>
<i>Rhynchosia golden mosaic virus</i> – [Mexico:Chiapas]	<i>Nicotiana spp.</i>	<i>Chino del tomate virus</i> [Mexico:Sinaloa]	<i>Nicotiana spp.</i> (E)

(E), experimental host.

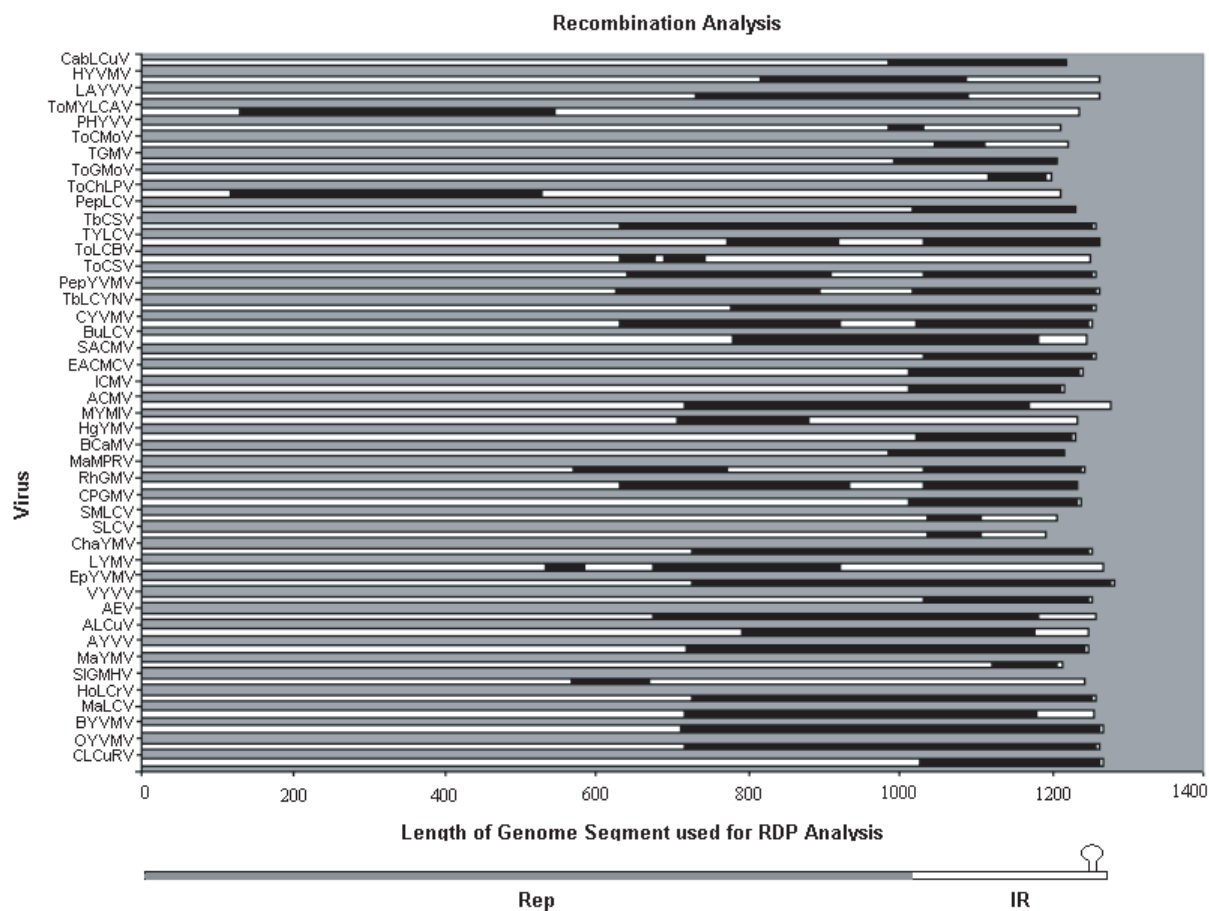
Source:

Gemini detective: A project maintained by the laboratory of Dr Judith Brown at the University of Arizona, USA.

URL: <http://gemini.biosci.arizona.edu/>

Plant Viruses Online: Descriptions and lists from the VIDE database.

URL: <http://image.fs.uidaho.edu/videl/refs.htm>

**Figure 4.** Summary of recombination events predicted by RDP and GENECONV.

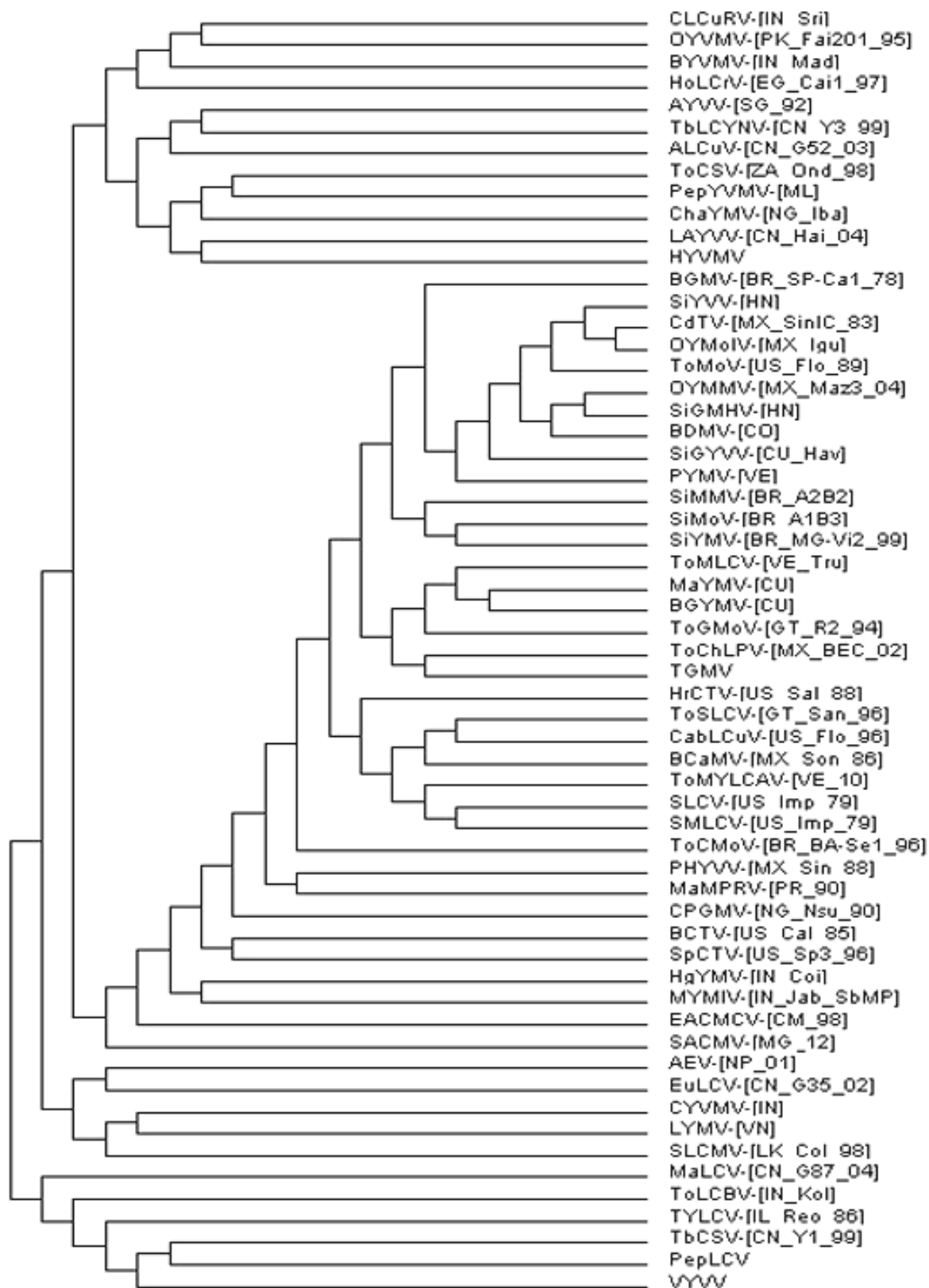


Figure 5. NJ tree constructed using ClustalW program for the recombinant regions predicted by both RDP and GENECONV.

virus is the reason for their separation as two independent clusters even though both are circoviruses (figure 3).

There is extensive homology between the N-terminal regions of pOYW replication initiator proteins with pLSI Rep (Oshima *et al* 2001) leading to their close proximity in both the trees.

The Rep proteins from pUB110 and pLSI plasmids were used as out-groups for the trees shown in figures 2 and 3 respectively. The resulting trees support the concept of Ilyina and Koonin (1992) that geminiviruses might have evolved from eubacterial replicons.

4.3 Secondary structure prediction

The secondary structures predicted for the Rep proteins of BFDV, *Clostridium*, pUB110 and *P. pulchra* are similar to that of the geminivirus Rep (table 1). BBTV and *Phytoplasma* plasmid Rep proteins show additional β strands and α helices. This might be due to insertion of regions during the course of evolution. Percentage identities between geminivirus Rep and all other groups of RCR initiator proteins are less than 25%. However they all carry out similar functions, shows sequence motif conservation and secondary structure conservation, as all of them are found to belong to the $\alpha + \beta$ class. Therefore it may be concluded that these structures have evolved through divergent evolution.

4.4 Recombination detection analysis

Recombination is very frequent in the evolution of geminiviruses and occurs between species and within and across genera (Padidam *et al* 1999). Among the predicted events in the present study, 58.4% occurs in the 5' end of IR (figure 4), the most variable part of begomovirus DNA-A (Sanz *et al* 2000). Out of this, more than 93% occurs along with the 5' end of *Rep* gene and thus the association of the 5' half of the IR and its cognate *Rep* gene is maintained. Such an association is probably required for virus variability because, a specific motif in the N-terminal portion of the Rep protein must recognize the iteron in the 5' half of the IR for the replication of the viral genomic DNA (Sanz *et al* 2000). The region of the Rep interacting with host proteins is located in the N-terminal domain. Therefore the recombination events predicted in the corresponding part of *Rep* gene has important implications for increasing the host specificity of geminiviruses.

There are many examples of distinct species of whitefly-transmitted geminiviruses with hosts in common (Zhou *et al* 1997). Sequences typical of viruses from non-malvaceous plants were found in viruses infecting malvaceous plants (Sanz *et al* 2000). Interestingly, while searching through natural and experimental hosts for geminiviruses it was

found that 14 daughters and the corresponding parents infect common hosts (table 3). With the emergence of the B biotype whitefly that can feed on hundreds of plant species, the host range of geminivirus has expanded dramatically (Padidam *et al* 1999).

Acknowledgements

The Department of Biotechnology, New Delhi is gratefully acknowledged for funding and fellowship (TV). Prof. S Krishnaswamy is thanked for helpful discussions. The computational facilities at the Center for Excellence in Bioinformatics, are acknowledged.

References

- Ach R A, Durfee T, Miller A B, Taranto P, Hanley-Bowdoin L, Zambriski P C and Gruissem W 1997 An alternatively spliced, multigene family in maize encodes retinoblastoma-related proteins which can interact with a plant D type cyclin and a geminivirus replication protein; *Mol. Cell. Biol.* **17** 5077–5086
- Campos-Olivas R, Louis J M, Clerot D, Gronenborn B and Gronenborn A M 2002 The structure of a replication initiator unites diverse aspects of nucleic acid metabolism; *Proc. Natl. Acad. Sci. USA* **99** 10310–10315
- Desbiez C, David C, Mettouchi A, Laufs J and Gronenborn B 1995 Rep protein of tomato yellow leaf curl geminivirus has an ATPase activity required for viral DNA replication; *Proc. Natl. Acad. Sci. USA* **92** 5640–5644
- Devereux J, Haeblerli P and Smithies O 1984 A comprehensive set of sequence analysis programs for the VAX; *Nucleic Acids Res.* **12** 387–395
- Feschotte C and Wessler S R 2001 Treasures in the attic: Rolling circle transposons discovered in eukaryotic genomes; *Proc. Natl. Acad. Sci. USA* **98** 8923–8924
- Flores E R and Lambert P F 1997 Evidence for a switch in the mode of human papillomavirus type 16 DNA replication during the viral life cycle; *J. Virol.* **10** 7167–7179
- Gibbs A, Calisher C H and Garcia-Arenal F 1995 *Molecular basis of viral evolution* (Cambridge: Cambridge University Press)
- Gibbs M J and Weiller G F 1999 Evidence that a plant virus switched hosts to infect a vertebrate and then recombined with a vertebrate-infecting virus; *Proc. Natl. Acad. Sci. USA* **96** 8022–8027
- Girish K R and Usha R 2005 Molecular characterization of two soybean-infecting begomoviruses from India and evidence for recombination among legume-infecting begomoviruses from South-East Asia; *Virus Res.* **108** 167–176
- Hickman A B, Ronning D R, Kotin R M and Dyda F 2002 Structural unity among viral origin binding proteins: crystal structure of the nuclease domain of adeno-associated virus Rep; *Mol. Cell* **10** 327–337
- Holland J D E 1998 Origin and Evolution of viruses; *Virus Genes* **16** 13–41
- Ilyina T V and Koonin E V 1992 Conserved sequence motifs in the initiator proteins for rolling circle DNA replication

- encoded by diverse replicons from eubacteria, eukaryotic and archaeobacteria; *Nucleic Acids Res.* **20** 3279–3285
- Keese P and Gibbs A 1993 Plant viruses: master explorers of evolutionary space; *Curr. Opin. Genet. Dev.* **3** 873–877
- Kirthi N, Maiya, S P, Murthy M R N and Savithri H S 2002 Evidence for recombination among the tomato leaf curl virus strains/species from Bangalore, India; *Arch. Virol.* **147** 255–272
- Laufs J, Schumacher S, Geisler N, Jupin I and Gronenborn B 1995 Identification of tyrosine of geminivirus rep protein; *FEBS Lett.* **377** 258–262
- Martin D and Rybicki E 2000 RDP: detection of recombination amongst aligned sequences; *Bioinformatics* **16** 562–563
- Moon D A and Goff L J 1996 Small plasmids in *Porphyra pulchra* are related to flowering plant geminiviruses; *J. Phycol.* **32** 33–34
- Morse S S 1994 *The evolutionary biology of viruses* (New York: Raven Press)
- Nagar S, Pederson T J, Carrick K, Hanley-Bowdoin L and Robertson D 1995 A geminivirus induces expression of a host DNA replication protein in terminally differentiated plant cells; *Plant Cell* **7** 705–719
- Orozco B M and Hanley-Bowdoin L 1996 A DNA structure is required for geminivirus origin function; *J. Virol.* **70** 148–158
- Orozco B M and Hanley-Bowdoin L 1998 Conserved sequence and structural motifs contribute to the DNA binding and cleavage activities of a geminivirus replication protein; *J. Biol. Chem.* **273** 24448–24456
- Orozco B M, Miller A B, Settlege S B and Hanley-Bowdoin L 1997 Functional domains of geminivirus replication protein; *J. Biol. Chem.* **272** 9840–9846
- Oshima K, Kakizawa S, Nishigawa H, Kuboyama T, Miyata S, Ugaki M and Namba S 2001 A plasmid of phytoplasma encodes a unique replication protein having both plasmid- and virus-like domains: clue to viral ancestry or result of virus/plasmid recombination?; *Virology* **285** 270–277
- Padidam M, Sawyer S and Fauquet C M 1999 Possible emergence of new geminiviruses by frequent recombination; *Virology* **265** 218–225
- Pant V, Gupta D, Choudhury N R, Malathi V G, Varma A and Mukherjee S K 2001 Molecular characterization of the Rep protein of the blackgram isolate of Indian mungbean yellow mosaic virus; *J. Gen. Virol.* **82** 2559–2567
- Sanz A I, Fraile A, Garcia-Arenal F, Zhou X, Robinson D J, Khalid S, Butt T and Harrison B D 2000 Multiple infection, recombination and genome relationships among begomovirus isolates found in cotton and other plants in Pakistan; *J. Gen. Virol.* **81** 1839–1849
- Settlege S B, Miller B and Hanley-Bowdoin L 1996 Interaction between geminivirus replication proteins; *J. Virol.* **70** 6790–6795
- Thompson J D, Higgins D G and Gibson T J 1994 CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice; *Nucleic Acids Res.* **22** 4673–4680
- Umaharan P, Padidam M, Phelps R H, Beachy R N and Fauquet C M 1998 Distribution and Diversity of Geminiviruses in Trinidad and Tobago; *Phytopathology* **88** 1262–1268
- Zhou X, Liu Y, Calvert L, Munoz C, Otim-Nape G W, Robinson D J and Harrison B D 1997 Evidence that DNA-A of a geminivirus associated with severe cassava mosaic disease in Uganda has arisen by interspecific recombination; *J. Gen. Virol.* **78** 2101–2111

ePublication: 18 September 2006