International Journal of Plant and Environment
Vol. 3, No. 2 (2017), 17-29

Dissecting Papaya Leaf Curl Disease (PLCD) Complex and Assessing its Potential for siRNA Based Targeting

Saurabh Verma and Sangeeta Saxena*

Department of Biotechnology, Babasaheb Bhimrao Ambedkar University, Lucknow-226024, INDIA

Publication Info

***Article history*:**
Received : 02.05.2017
Accepted : 24.07.2017
DOI : https://doi.org/10.18811/ijpen.v3i02.10433

***Key words*:**
Begomovirus
Genetic variability
Papaya leaf curl disease complex
Phylogenetic analysis
siRNA

Corresponding author:
Dr. Sangeeta Saxena
Tel.: +91- 9450645342
Email:
dr_sangeeta_saxena@yahoo.com

Abstract

Carica papaya L. (papaya) production is severely affected by leaf curl and related disease symptoms caused by various species of begomoviruses making it a complex disease. Papaya leaf curl disease (PLCD) complex has broad host range and high genetic variability, therefore, the study was conducted to dissect the genetic complexity of PLCD. Begomoviruses infect various host crops and weeds associated with cultivation of papaya. Intermixing among their genomes and mutational forces drive molecular variability in the disease complex, which enables them to expand their host range. This molecular variability is important determinant when RNA interference based strategy is applied against these viruses. In this study, we have observed that the broad host range of PLCD complex viral components is due to high frequency of recombination events that took place at different time scales. The rise of recombinants led to host variability resulting in similar symptomatic infection over new crops and weeds in nearby fields. Therefore, we have taken into account molecular variability and recombination regions of the viral genome as the targets for providing resistance against these viruses, thus, preventing serious implications of similar diseases spreading over to new hosts.

1. Introduction

Papaya fruit is used for various medicinal and commercial purposes apart from being an edible delicacy and an important source of daily nutrition like minerals, vitamins and papain. Due to its commercial importance, it is cultivated at large scale in almost all parts of India. However, in last decade, papaya farming has suffered a huge setback due to various pathogens infecting papaya crops in every season, especially the rainy season. Fungus, anthracnose, leaf curl, crumpled growth and early fruit fall are some of the problems related to rainy season (Roden and Ingle, 2009; Singh *et al.,* 2012; Varun *et al.,* 2017). Among all problems, the leaf curl disease of papaya has been a major adversity faced by papaya plantations causing a deep dip in papaya productivity (Saxena and Verma, 2016). Upward or downward leaf curl, dark green coloration, vein swelling, leaf mottling, crumpling and crippled plant development (Saxena *et al.,* 2016) characterize the leaf curl disease symptoms in papaya caused by a plant virus infection. This disease is endemic and causes huge losses to papaya farmers in India, especially in the northern and central regions of Uttar Pradesh, Haryana, Delhi and Rajasthan. Apart from papaya leaf curl, crops in India such as cotton, tomato, okra, chili and squash are highly susceptible to leaf curl disease caused by various other begomoviruses (Varma and Malathi, 2003; Borah and Dasgupta, 2012).

Begomoviruses are single stranded DNA viruses, characterized by the geminate shaped structure consisting of a capsid and viral genome. Their genome reported to be containing DNA-A, DNA-B, beta (DNA-β) and alpha satellite molecules. The genetic composition of these viruses depends upon their host plants and vector biotypes i.e. *Bemisia tabaci* (Gennadius) also known as whitefly (Brown *et al.,* 1995; Gutierrez, 2000; David, 2003; King *et al.,* 2011). Monopartite begomoviruses consist of DNA-A and associated satellites, whereas, bipartite consist of DNA-A and DNA-B. The common genetic component, DNA-A, of begomoviruses harbor structural and multi-functional protein encoding open reading frames (ORFs). All genetic components have been found to be associated with a 200bp long, highly conserved region composed of repetitive elements, promoter region, origin of replication and a hairpin loop (Harrison, 1985). Satellite molecules are incomplete defective genetic components which may or may not be functional inside the host crops. In addition, DNA-β has been reported to be associated with disease severity in several cases (Zhou, 2013). DNA-B harbors ORFs responsible for the movement of viral DNA from cytoplasm to nucleus, nucleus to cytoplasm and from one cell to another cell, hence also known as movement component of bipartite begomoviruses. Together, these ORFs encode proteins that take part in structure formation, replication initiation, transcriptional activation, replication

enhancement, host immune suppression and viral movement (Fondong, 2013). Hence, begomoviruses are complete parasites of plants, forming a very tight virus-vector-host complex (Hanley-Bowdoin *et al.,* 2013; Saxena and Verma, 2016).

The PLCD (Nadeem *et al.,* 1997; Saxena *et al.,* 1998a,b,c) is reported to be a complex system due to different begomoviral species involved in causing this symptom. The begomovirus isolates belonging to Papaya leaf curl virus (PaLCuV), Papaya leaf curl China virus (PaLCuCNV), Papaya leaf curl Guangdong virus (PaLCuGdV), Chili leaf curl virus (ChiLCV), Tomato leaf curl virus (ToLCV), Pedilanthus leaf curl virus (PeLCuV), Papaya leaf crumple virus (PapayaLCV) and beta (β) satellites associated with other monopartite begomoviruses causes severe leaf deformation symptoms in plants (Fig. 1). Among all PLCD causing begomoviruses, the Indian species form the components of largest disease complex infecting papaya and associated crops and weeds (Hallan *et al.,* 1998a,b; Raj *et al.,* 2008) (Table 1). There is an urgent need to develop assessment parameters to propose an effective and sustainable resistance strategy against PLCD complex as it is caused by more than one type of begomovirus species (Singh-Pant *et al.,* 2012).

In this study, several parameters such as conservation, sequence diversity and the occurrence of recombination events have been studied to assess the level of complexity of PLCD. The complexity will decide if this disease complex has potential for a broad spectrum strategy that could target most of the components of PLCD complex. Therefore, this study on PLCD complex is necessary for assessing the potential for siRNA based targeting (Saxena *et al.,* 2011, 2013; Ghosal *et al.,* 2012) of PLCD complex to develop PLCD resistant papaya crops.

2. Materials and Methods

2.1. Taxonomic analysis and sequence retrieval of PLCD complex

The taxonomic diversity of the leaf curl disease complex in papaya was investigated using Taxonomy browser at NCBI (https://www.ncbi.nlm.nih.gov/taxonomy). The virus species were manually screened for their presence in India and infection host as *Carica papaya* and associated weeds and crop plants. The

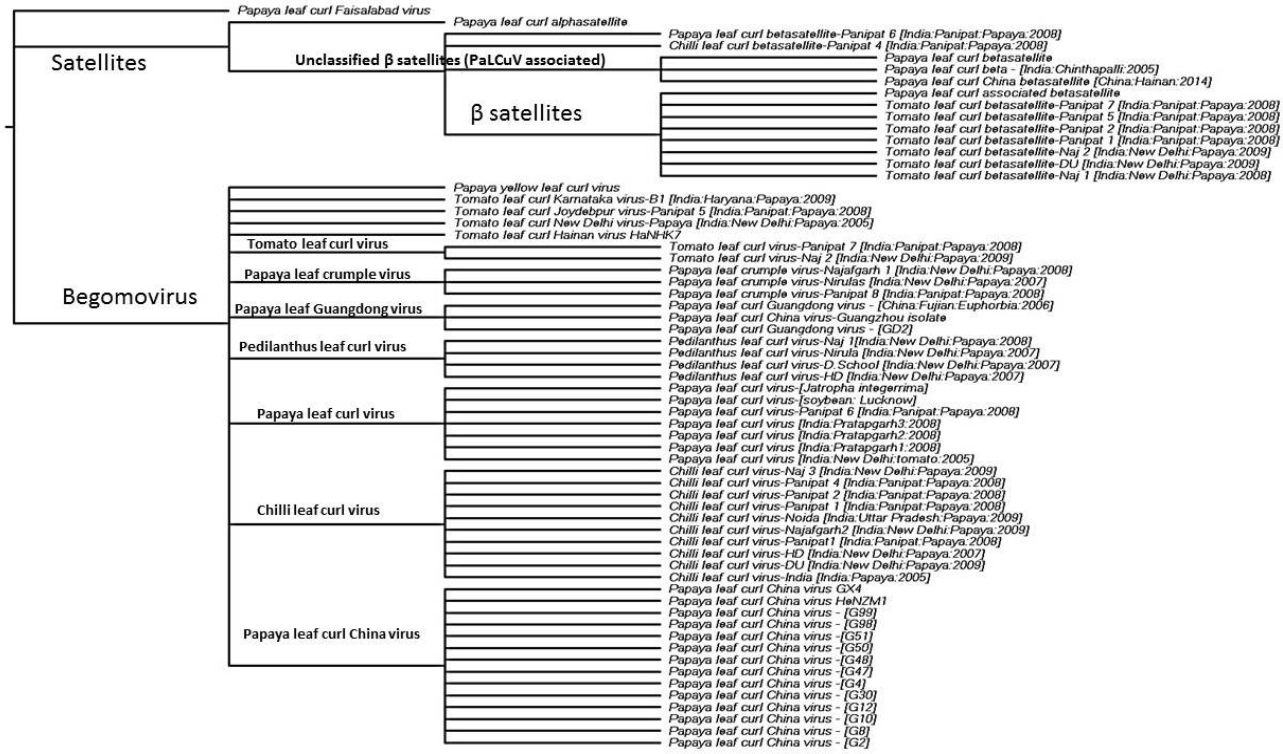

Fig. 1: Schematic representation of the Papaya leaf curl disease (PLCD) complex. The PLCD complex consists of Begomoviruses, incomplete defective satellite DNA molecules and unclassified viruses and satellite DNA molecules. The PLCD complex contains diverse strains exhibiting leaf curl and crumpling symptoms in papaya and associated crops and weeds

Table 1: List of DNA-A sequences of Indian PLCD complex begomoviruses causing leaf curl disease in papaya and associated crops and weeds

Accession ID	Virus	Annotation	Host	Size	AV2	AV1	AC3	AC2	AC1	AC4	AC5	AC6	AV3
Y15934.1	Papaya leaf curl virus	PLCD-1	*Carica papaya*	2746	150-506	310-1080	1077-1481	1222-1626	1529-2614	2200-2457	301-618		
NC_004147.1	Papaya leaf curl virus	PLCD-2	*Carica papaya*	2746	150-506	310-1080	1077-1481	1222-1626	1529-2614	2200-2457	301-618		
HM143914.1	Papaya leaf curl virus segment DNA A	PLCD-3	*Nicotiana glutinosa*	2746	150-506	310-1080	1077-1481	1222-1626	1529-2614	2200-2457			
JN135233.1	Papaya leaf curl virus	PLCD-4	*Amaranthus cruentus* L.	2746	150-506	310-1083	1077-1481	1222-1626	1529-2614	2200-2457			
JQ954859.1	Papaya leaf curl virus	PLCD-5	*Aster alpinus* L.	2746	150-506	310-1080	1077-1481	1222-1626	1529-2614	2200-2457			
KM525657.1	Papaya leaf curl virus	PLCD-6	*Croton bonplandianus* L.	2746	150-506	310-1080	1077-1481	1222-1626	1529-2614	2200-2457			
KU376493.1	Papaya leaf curl virus isolate CN2	PLCD-7	*Solanum lycopersicum*	2746	150-506	310-1080	1077-1481	1222-1626	1529-2614	2200-2457			
KX302713.1	Papaya leaf curl virus isolate Wellington clone WB2&5	PLCD-8	*Carica papaya*	2763	149-511	315-1085	1082-1270	1267-1623	1535-2674	2400-2621	729-980	1227-1487	
KY800906.1	Papaya leaf curl virus isolate India/New Delhi/ Papaya/ 2016	PLCD-9	*Carica papaya*	2763	147-503	307-1077	1074-1478	1219-1623	1529-2611	2161-2460			
KY026597.1	Papaya leaf curl virus clone Rad38	PLCD-10	*Raphanus sativus*	2745	143-499	303-1073	1070-1474	1215-1619	1522-2607	2193-2450			
KY026598.1	Papaya leaf curl virus clone Rad07	PLCD-11	*Raphanus sativus*	2746	143-499	303-1073	1070-1474	1215-1619	1522-2607	2193-2450			
GU136803.1	Chilli leaf curl virus isolate India: Amritsar: Papaya: 2009	PLCD-12	*Carica papaya*	2763	147-512	307-1077	1074-1478	1219-1623	1526-2460	2161-2460			
JN558352.1	Cotton leaf curl Multan virus isolate	PLCD-13	*Carica papaya*	2725	117-482	277-1047	1050-1454	1147-1599	1496-2581	2128-2430			
DQ629103.1	Papaya leaf curl virus [India:New Delhi: tomato: 2005]	PLCD-14	*Carica papaya*	2765	147-503	307-1077	1074-1478	1219-1623	1526-2611	2161-2450			
KX302707.1	Cotton leaf curl Burewala virus isolate Guntur clone LK_2N	PLCD-15	*Carica papaya*	2758	131-487	291-1061	1058-1462	1294-1503	1504-2595	2241-2681			
GQ200446.1	Papaya leaf curl virus [India:Pratapgarh1:2008] clone SHLD-NIFL-Pra-01	PLCD-16	*Crotalaria juncea*	2738	146-502	306-1076	1060-1464	1205-1609	1512-2597	2183-2500			

Accession ID	Virus	Annot-ation	Host	Size	AV2	AV1	AC3	AC2	AC1	AC4	AC5	AC6	AV3
GQ200447.1	Papaya leaf curl virus [India:Pratapgarh2:2008] clone SHLD-NIFL-Pra-02	PLCD-17	*Crotalaria juncea*	2738	146-502	306-1076	1060-1464	1205-1609	1512-2597	2183-2500			
GQ200448.1	Papaya leaf curl virus [India:Pratapgarh2:2008] clone SHLD-NIFL-Pra-03	PLCD-18	*Crotalaria juncea*	2738	146-502	306-1076	1060-1464	1205-1609	1512-2597	2183-2500			
DQ989325.1	Tomato leaf curl New Delhi virus-Papaya [India:New Delhi: Papaya: 2005]	PLCD-19	*Carica papaya*	2735	127-465	287-1057	1199-1603	1054-1464	1506-2597	2264-2440	317-802		48-434
DQ989326.1	Chilli leaf curl virus-India [India: Papaya: 2005]	PLCD-20	*Carica papaya*	2764	148-504	308-1078	1075-1479	1220-1624	1527-2612	2162-2455	338-823		
HM140364.1	Chilli leaf curl virus-DU [India: New Delhi : Papaya: 2009]	PLCD-21	*Carica papaya*	2763	147-503	307-1077	1074-1478	1219-1623	1526-2611	2161-2460			
HM140365.1	Chilli leaf curl virus-HD [India: New Delhi: Papaya: 2007]	PLCD-22	*Carica papaya*	2763	147-503	307-1077	1074-1478	1219-1623	1526-2611	2161-2460			
HM140366.1	Chilli leaf curl virus-Panipat1 [India: Panipat: Papaya: 2008]	PLCD-23	*Carica papaya*	2761	145-510	305-1075	1072-1476	1217-1621	1524-2609	2159-2458			
HM140370.1	Chilli leaf curl virus-Najafgarh2 [India:New Delhi: Papaya: 2009]	PLCD-24	*Carica papaya*	2763	147-503	307-1077	1074-1478	1219-1623	1526-2611	2161-2460			
HM140371.1	Chilli leaf curl virus-Noida [India: Uttar Pradesh: Papaya :2009]	PLCD-25	*Carica papaya*	2762	146-502	306-1076	1073-1477	1218-1622	1525-2610	2160-2459			
HM140367.1	Papaya leaf crumple virus-Panipat 8 [India: Panipat: Papaya: 2008]	PLCD-26	*Carica papaya*	2736	120-485	280-1050	1047-1451	1192-1599	1523-2587	2128-2430	258-548		
NC_014707.1	Papaya leaf crumple virus-Panipat 8 [India: Panipat: Papaya: 2008]	PLCD-27	*Carica papaya*	2736	120-485	280-1050	1047-1451	1192-1599	1523-2587	2128-2430	258-548		
HM140368.1	Papaya leaf crumple virus-Nirulas [India:New Delhi: Papaya: 2007]	PLCD-28	*Carica papaya*	2736	120-485	280-1050	1047-1451	1192-1599	1523-2587	2128-2430	258-548		

Accession ID	Virus	Annot-ation	Host	Size	AV2	AV1	AC3	AC2	AC1	AC4	AC5	AC6	AV3
HM140369.1	Papaya leaf crumple virus-Najafgarh 1 [India:New Delhi: Papaya: 2008]	PLCD-29	*Carica papaya*	2736	120-485	280-1050	1047-1451	1192-1599	1523-2587	2128-2430	258-548		
KJ028210.1	Papaya leaf crumple virus clone Moh7	PLCD-30	*Solanum nigrum*	2736	120-485	280-1050	1047-1451	1192-1599	1523-2587	2128-2430	258-548		
KM359408.1	Papaya leaf crumple virus isolate A-87	PLCD-31	*Andrographis paniculata*	2737	121-459	281-1051	1048-1452	1193-1600	1524-2588	2129-2431			
KR052159.1	Papaya leaf crumple virus isolate Mohali	PLCD-32	*Carica papaya*	2736	120-458	280-1050	1047-1451	1192-1599	1523-2587	2128-430			
JN807765.2	Papaya leaf curl virus-[soybean: Lucknow]	PLCD-33	*Glycine max*	2746	150-506	310-1080	1077-1481	1222-1626	1529-2614	2200-2457			
KX353622.2	Papaya yellow leaf curl virus isolate DP2	PLCD-34	*Carica papaya*	2759	145-501	305-1075	1072-1476	1217-1621	1524-2606	2159-2452			
KP725055.1	Tomato leaf curl virus isolate C1	PLCD-35	*Carica papaya*	2757	144-491	304-1074	1071-1475	1216-1620	1532-2608	2158-2451			

AV1=Coat protein; AV2=variable function; AC1=replication initiation protein; AC2=transcription activator protein; AC3=replication enhancer protein; AC4=host RNA machinery suppressor; AC5, AC6 and AV3=uncharacterized open reading frames of begomoviruses of PLCD complex.

details were transformed into a schematic representation using the available tree tool. The manually screened virus species were used to retrieve nucleotide sequence information of complete DNA-A genome using Nucleotide browser at NCBI (https://www.ncbi.nlm.nih.gov/nuccore). All sequences were transferred as a FASTA format into a text file. The ORF information was retrieved using Sequence information tool for every nucleotide sequence.

2.2. Phylogenetic analysis of PLCD complex

Phylogenetic analysis was performed using two algorithms i.e. Bayesian and Maximum likelihood. Bayesian analysis was performed using TOPALi v2.5 software (www.topali.org/). The input file for bayesian phylogenetic analysis was prepared by performing a multiple sequence alignment using CLUSTALW software (www.ebi.ac.uk/Tools/msa/clustalw2/). The parameters selection was based upon prediction of suitable model by Model Selection tool and final analysis was run using MrBayes version 3.1.1 tool (http://mrbayes.sourceforge.net/) available in TOPALi v2.5. The maximum likelihood algorithm was used using the same input file as used previously. PhyML-aLRT version 2.4.5 tool (www.atgc-montpellier.fr/phyml/) available in TOPALi v2.5 software was used for inferring phylogenetic association of PLCD complex viral DNA-A components. The inferred trees were prepared for illustration using FigTree v1.4.3 (http://tree.bio.ed.ac.uk/ software/figtree/).

2.3. Pairwise similarity and analysis

To construct a similarity matrix based on pairwise alignment of DNA-A, the pairwise distance estimation of DNA-A sequences of PLCD complex was carried out using Sequence demarcation tool (SDT) version 1.2 (web.cbio.uct.ac.za/SDT). The aligned viral components showing 90-100% sequence similarity were grouped together and assumed to be isolates of same species.

2.4. Recombination analysis

The CLUSTALW aligned file was used as input file for recombination analysis using Recombination detection program (RDT) v4.94 software (http://web.cbio.uct.ac.za/~darren/rdp.html). The recombination events and breakpoint distribution analysis was conducted using various tools available in Options toolbar of RDT software with default parameters as stated in RDP manual. The recombination events were screened manually according to the procedure listed in RDP manual. The recombination events were selected and analyzed manually for breakpoint boundary and parental contribution to assess recombinant viral component.

3. Results and Discussion

3.1. Components of PLCD complex

The PLCD complex consists of 13 begomovirus species associated with β satellites (Fig. 1). The begomovirus complex components are distributed mainly in Central Asia i.e. China, India, Pakistan and parts of Bangladesh. PaLCuCNV, PaLCuGnV and Tomato leaf curl Hainan virus (ToLCuHaV) are mainly reported in China infecting papaya, tomato and euphorbia. The Chinese components are associated with Papaya leaf curl China betasatellites. There are 18 different virus isolates comprising the Chinese PLCD complex components. The Pakistan PLCD complex consists of Papaya leaf curl Faisalabad virus, an unclassified virus that has been excluded from begomovirus. The Indian PLCD complex shows the most diverse nature in terms of number of species reported to cause leaf curl symptom in papaya. Nine begomovirus species i.e. ChiLCV, PaLCuV, PeLCuV, PapayaLCV, ToLCV, ToLCNDV, PaYLCuV etc. have been observed to infect papaya with broad host range. The Indian PLCD complex consists of 28 isolates belonging to the above mentioned begomoviruses. Overall, the PLCD complex consists of 47 different isolates identified to cause leaf curl disease in papaya and associated weeds and crops.

3.2. High molecular diversity of PLCD complex

The phylogenetic analysis was performed for manually screened isolates of Indian PLCD complex. Total thirty five isolates were chosen for phylogenetic analysis. The CLUSTALW output was used for Bayesian inference with MrBayes v.3.1.1 (Huelsenbeck and Ronquist, 2001) using General time reversible with gamma (GTR-G) substitution model (Larget and Simon, 1999) with two runs for 2,50,000 generations, sample frequency at 10 and burn-in ratio at 42% i.e. first 1,05,000 generations were excluded from the inference. The same set of PLCD complex DNA-A sequences were analyzed for maximum likelihood analysis using PhyML-aLRT v2.4.5 (Guindon and Gascuel, 2003) and HKY-G (Hasegawa, Kishino and Yano, 1985-Gamma) as a substitution model (Hasegawa *et al.,* 1985). The above phylogenetic analysis produced likelihood values that were very similar to each other i.e. 26743.85 and 26732.18 for MrBayes and PhyML respectively. The close lying likelihood scores means that the two trees can be interpreted in similar manner without much divergence. Therefore, the results and interpretation for MrBayes tree will be same as PhyML-aLRT tree in all future discussions.

The complexity of Indian PLCD complex is evident from the posterior distribution of model GTR-G parameters obtained using Markov Chain Monte Carlo (MCMC) methods in MrBayes inference (Larget and Simon, 1999; Ronquist and Huelsenbeck, 2003; Gamerman and Lopes, 2006) (Fig. 2). The use of GTR-G substitution model resulted in convergence of potential scale reduction factor (PSRF) to 1.0 for all the parameters, therefore, the analysis carried out using above mentioned models and default parameters was true and the bayesian inference could be accepted for the Indian PLCD complex DNA-A dataset. The PSRF convergence with range from 1.0 to 1.2 also suggest that the cladogram represents a clade credibility tree, whose branches lie in the region of 95% or more accumulated posterior probability. Therefore, the branch lengths represent the measured expected substitution per site that signifies the phylogenetic relatedness among all clades and sub-clades of the tree.

The MrBayes and PhyML tree resulted in a mid-rooted tree (Fig. 2 and 3) which was divided into two major clades and an out-group i.e. PLCD19 (Tomato leaf curl New Delhi virus). Other out-groups lying in the two major clades include PLCD13, PLCD14, PLCD34, PLCD35 and PLCD31 representing CLCuMuV, PaLCuV [India:New Delhi:tomato:2005], PaYLCuV [DP2], ToLCuV [C1], PapayaLCV [A-87] respectively. The first major clade consists of Papaya leaf crumple virus isolates within branch lengths 0.1 to 0.17 signifying low expected substitution rates per site. Despite having narrow branch length, notably, this virus species used only *A. paniculata* as its reservoir host, which is an annual weed widely found in south India, whereas tomato and papaya as a primary host in north India. This indicates that this is a new virus that has been introduced in northern regions via anthropogenic activities. During its course for survival, the Papaya leaf crumple virus isolates have adapted tomato and papaya as its primary host and undergone parallel evolution with Papaya leaf curl virus group in the northern regions. Its recombination analysis can provide much more details about its origin and propagation pattern.

The second major clade is further subdivided into two minor sub-clades i.e. ChiLCV and PaLCuV. The ChiLCV clade consists of ChiLCV and CLCuBuV [LK_2N] whereas PaLCuV clade clustered the papaya and associated crop infecting PaLCuV isolates according to their hosts. The ChiLCV clade has a narrow branch length range i.e. 0.24 - 0.29 signifying low substitution rate per site. Therefore, whole ChiLCV clade shows close genetic relatedness and host singularity as evident from *C. papaya* being the only host crop infected by members of this clade. The pattern suggests that these clade viruses have evolved to infect papaya crops lying nearby

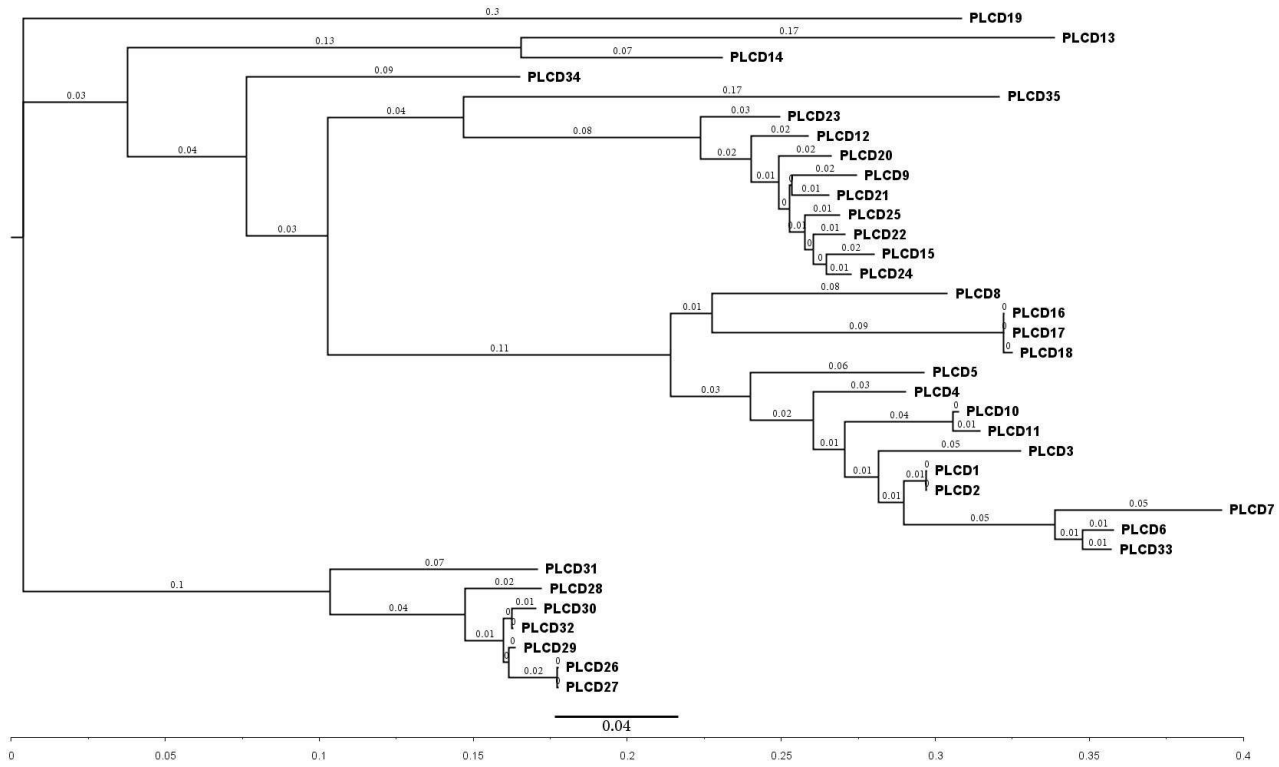

Fig. 2: MrBayes tree analysis. Phylogenetic tree based on the complete DNA-A sequences of the begomoviruses detected in *C. papaya* and additional sequences from associated crops and weeds infecting begomoviruses causing leaf curl disease. The tree was constructed by Bayesian inference using the GTR-G nucleotide substitution model available in MrBayes v3.1.1 method in TOPALi v2.5. Numbers at the branches indicate estimated clade credibility score inferred from statistical analysis of posterior probabilities assessed for the substitution model

chili fields in north Indian region and they are under very low mutational pressure to evolve and expand their host range. The PaLCuV clade on the other hand, has a very broad range of branch length i.e. 0.23-0.39 signifying high substitution rate per site; showing expansive genetic relation and diverse host range. Analysis also suggests that they might have evolved independently on diverse hosts in the geographically different regions. The out-group species are sparingly related to the clade species and have acquired the ability to infect papaya and associated host crops and weeds.

3.3. Species diversity identification by pairwise sequence comparison

The International Council for Taxonomy of Viruses (ICTV) has decided to solve the species nomenclature issues by using pairwise sequence identity criterion as a standard methodology for species demarcation. Earlier cutoff i.e., >89% sequence similarity was elevated to >91% and use of SDT tool was made mandatory for naming of new viral species (Brown *et al.,* 2015). Therefore, this tool was employed to study if the taxonomic species associated with Indian PLCD complex are really as diverse as indicated by the phylogenetic analysis.

The pairwise matrix (Fig. 4) obtained by using SDT v1.4 shows a similarity matrix characteristic of a complex association as the minimum and maximum similarity obtained was >75% and 100%. The species with highest similarity group i.e. >90% to 100% were grouped together using inbuilt function of the software. The PLCD complex thus got distributed into 11 different groups (Table 2). The groups labeled as Papaya leaf curl virus group, Chili leaf curl virus group, Papaya leaf crumple virus group and Papaya leaf curl virus (Pratapgarh) consist of 10, 9, 6 and 3 isolates respectively. Among all 35 PLCD complex isolates, 28 belonged to above mentioned groups and rest of the DNA-A sequences were phylogenetically isolated as individual species associated with Indian PLCD complex. Though, this grouping based on pairwise similarity is useful in inferring a good correlation with phylogenetic analysis done using MrBayes and PhyML-aLRT algorithms, still, it would be too premature to declare them similar or different species due to a significant overall similarity score in the range of >70%

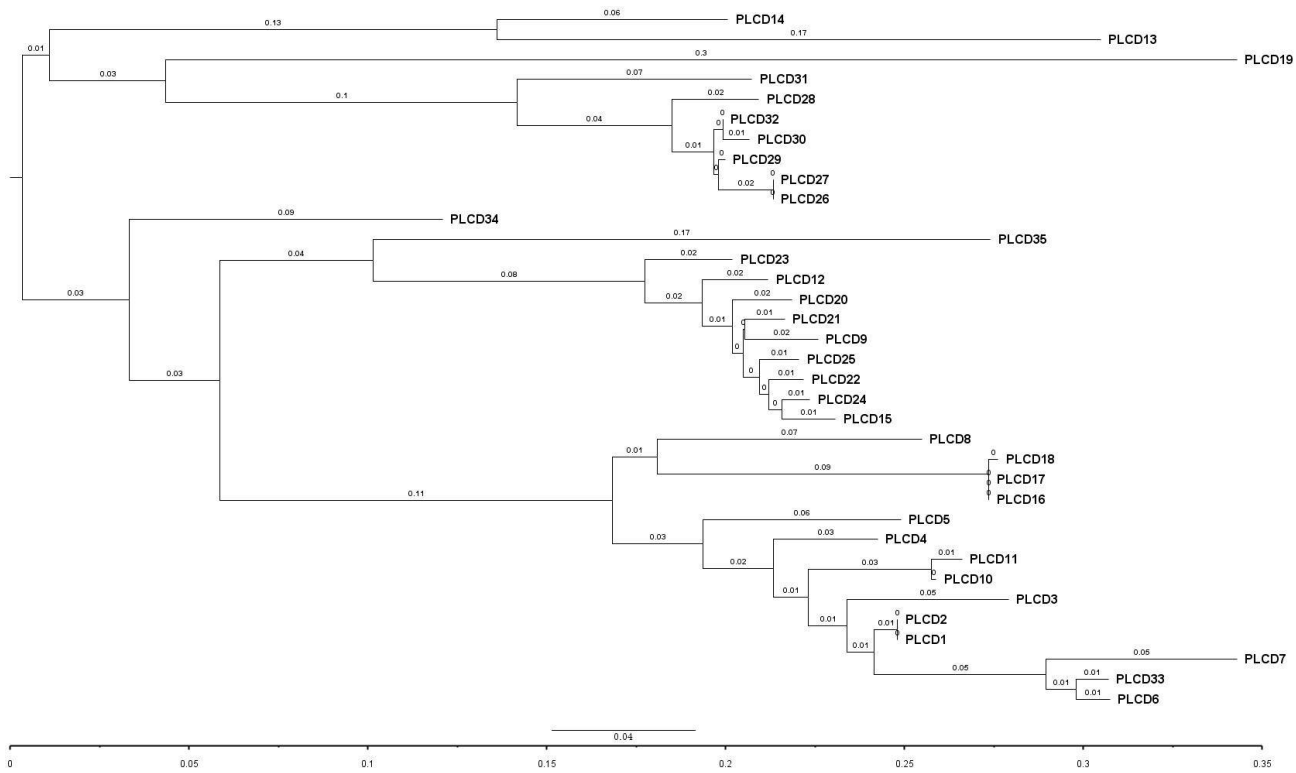

Fig. 3: PhyML tree analysis. A maximum likelihood tree based upon complete DNA-A sequences of the begomoviruses detected in *C. papaya* and additional sequences from associated crops and weeds infecting begomoviruses causing leaf curl disease. The tree was constructed by using PhyML-aLRT v2.4.5 tool in TOPALi v2.5 and the HKY-G nucleotide substitution model. Numbers at the branches indicate estimated substitution rate per nucleotide position

to 80%. Therefore, it was important to study the recombination pattern of the DNA-A sequences involved in the Indian PLCD complex to solve the complex nature of PLCD group of begomoviruses.

Pairwise sequence demarcation was carried out using SDT v1.2 program. The groups have been distributed according to the >90 to 100% sequence similarity with each other.

3.4. Recombination pattern of Indian PLCD complex

Recombination is an important molecular event that leads to translocation, transversion, addition, deletion or frame-shift mutations in a genome (Lima *et al.,* 2017). Due to the rolling circle replication mechanism in begomoviruses, these organisms have evolved to survive inside their hosts through acquisition of useful regions from other viruses or the host themselves, so that they can successfully evade host mediated RNA interference based immune response (van der Walt *et al.,* 2009). In the above analysis, it was observed that the Indian PLCD complex is well defined with major contribution made by PaLCuV, ChiLCV, PapayaLCV and discreet reports of other species infecting papaya and associated crops and weeds. To study if recombination is responsible for occurrence of such complex virus-host interaction; the DNA-A sequences of Indian PLCD complex were subjected to recombination analysis using RDP v4.95 (Martin and Rybicki, 2000) using available recombination analysis tools (Fig. 5) and default parameters as recommended by RDP4 manual (Smith, 1992; Padidam *et al.,* 1999; Gibbs *et al.,* 2000; Posada and Carandall, 2001; Martin *et al.,* 2005; Boni *et al.,* 2007). The analysis ended up predicting 66 recombination events, which further reduced to 12 events after careful manual screening as per the RDP4 manual (Martin *et al.,* 2015) (Table 3).

According to the manually screened recombination events results, event no. 10 i.e. PLCD9 and PLCD31 (inferred closest to the unknown minor parent) seems to be the major contributor of fragments giving a chance for origin of a new recombinant species e.g., PLCD8 and thirteen others i.e. PLCD1, PLCD2, PLCD3, PLCD4, PLCD5, PLCD6, PLCD7, PLCD10,

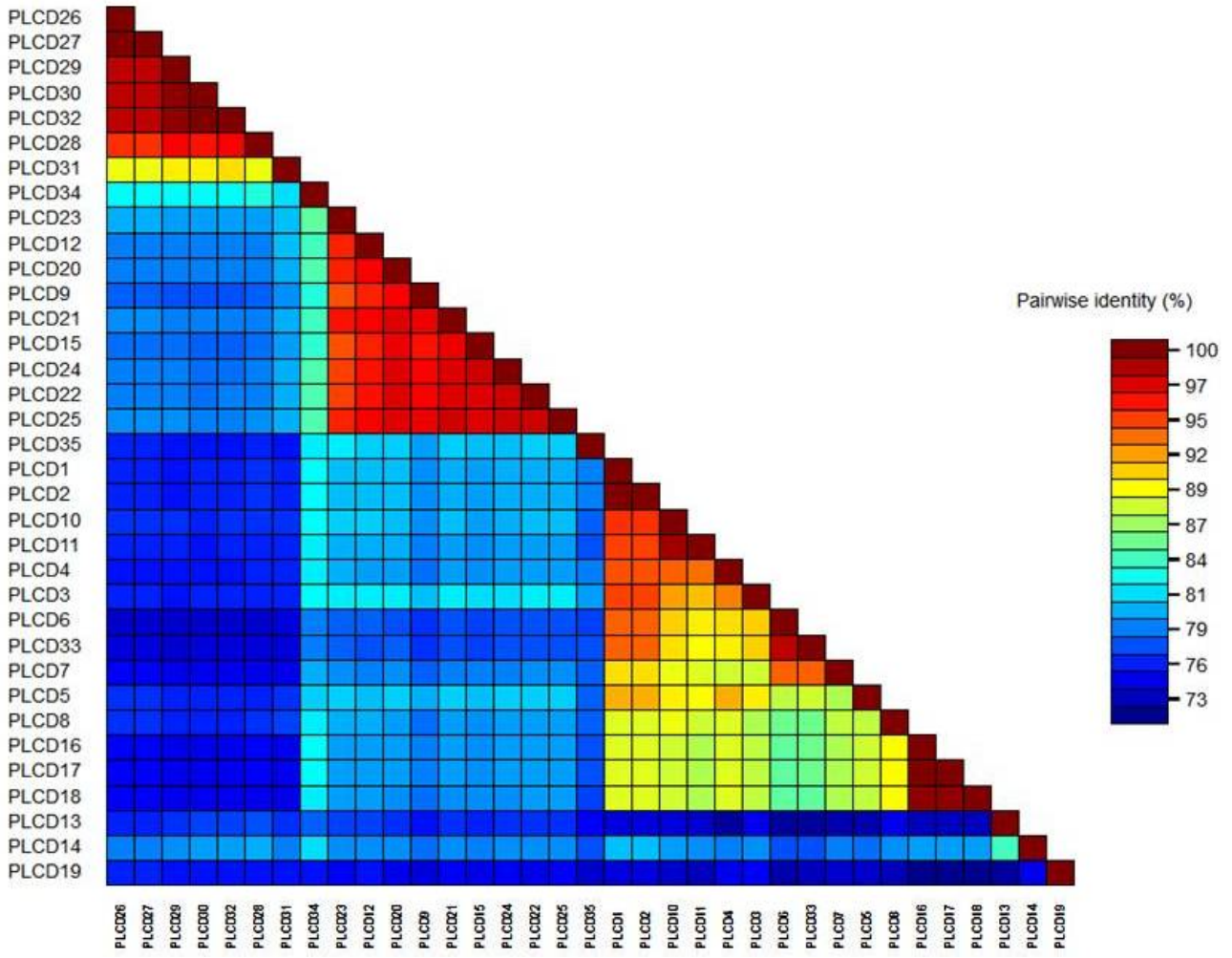

Fig. 4: Estimation of pairwise sequence similarity using Sequence Demarcation Tool v1.2. The matrix output indicates the diversity of begomoviruses infecting papaya and associated crops and weeds. The color key indicates highest and least similarity with brown and blue color respectively

Table 2: Pair wise sequence similarity among begomovirus components in PLCD complex

Group No.	No. of species	PLCD complex	Virus Taxons included in group
1	6	PLCD26; PLCD27; PLCD28; PLCD29; PLCD30; PLCD32	Papaya leaf crumple virus
2	1	PLCD31	Papaya leaf crumple virus (A-87)
3	1	PLCD34	Papaya yellow leaf curl virus (DP2)
4	9	PLCD9; PLCD12; PLCD21 PLCD22; PLCD23;PLCD24; PLCD25	Papaya leaf curl virus (New Delhi/2016) Chili leaf curl virus (Amritsar/2009)
5	1	PLCD35	Tomato leaf curl virus (C1)
6	10	PLCD1; PLCD2; PLCD3; PLCD4; PLCD5; PLCD6; PLCD7; PLCD10; PLCD11; PLCD33	Papaya leaf curl virus
7	1	PLCD8	Papaya leaf curl virus (WB2&5)
8	3	PLCD16; PLCD17; PLCD18	Papaya leaf curl virus (Pratapgarh)
9	1	PLCD13	Cotton leaf curl Multan virus
10	1	PLCD14	Papaya leaf curl virus (New Delhi/2005)
11	1	PLCD19	Tomato leaf curl New Delhi virus

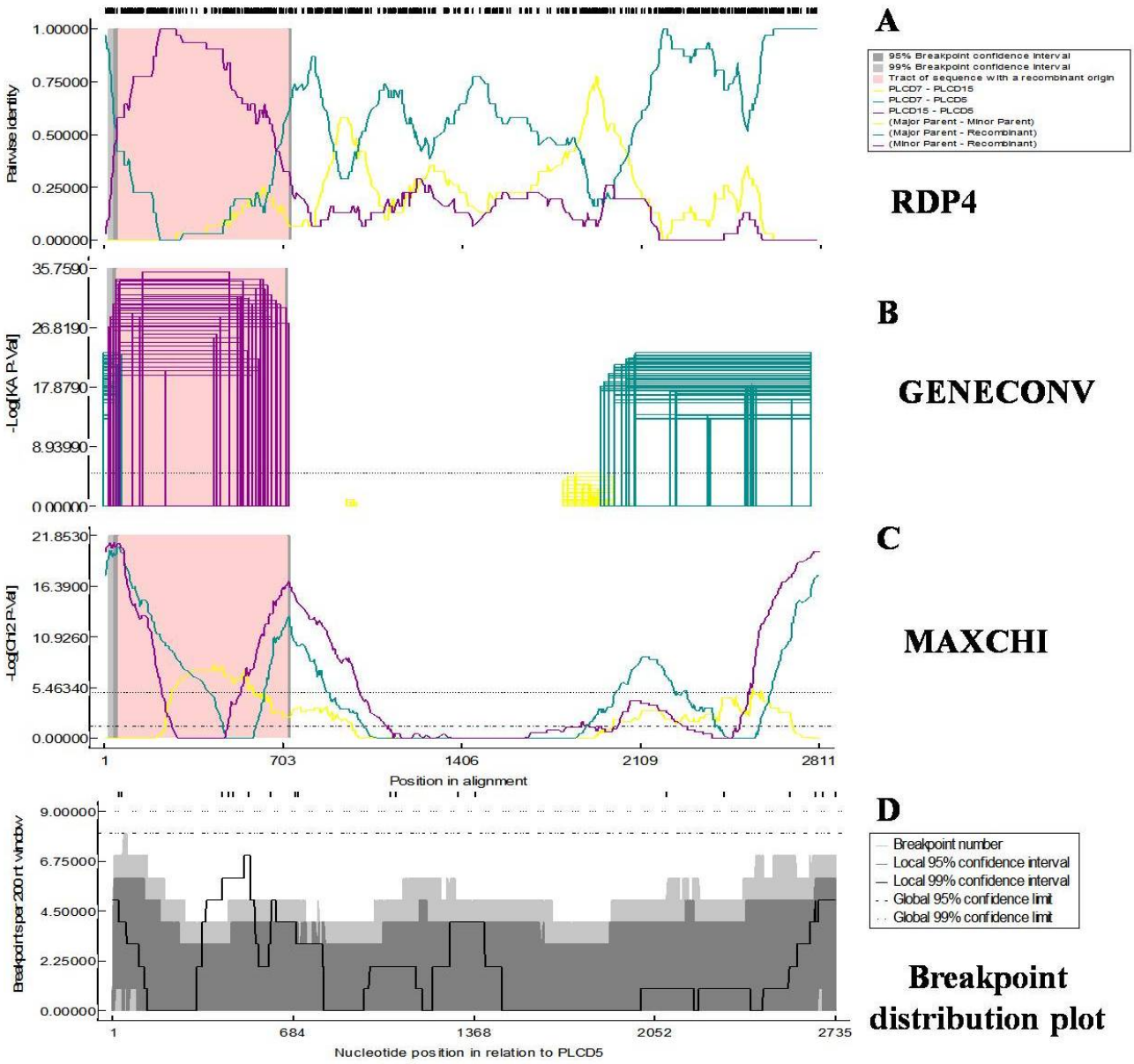

Fig. 5: Recombination analysis showing results of event number 1 using Recombination Detection Program v4.95 (RDP v4.95). The pink region indicates the occurrence of the recombination event with intersection points in the case of RDP and overlapping peaks in case of MAXCHI on the plot as the recombination starting and ending breakpoints. The grey region at the ends represents 99% and 95% confidence interval of prediction of breakpoints as mentioned in the key. A. RDP output; B. GENECONV output; C. MAXCHI output and D. Recombination breakpoint distribution plot: The light and dark grey regions in the plot represent the 99% and 95% confidence limit respectively for a 200nt scanning window. The breakpoint distribution was assessed with PLCD5 DNA-A sequence as a reference sequence just for qualitative purpose

PLCD16, PLCD17, PLCD18, PLCD23 and PLCD33. Thus, a Chili leaf curl virus and Papaya leaf crumple virus infecting papaya and *A. paniculata* contributed significantly to the Papaya leaf curl virus sub-clade, which partially explains the clades affinity to infect weeds and associated crops. The event no. 5, PLCD2 and PLCD28 (inferred closest to the unknown minor parent) give rise to PLCD33, PLCD6 and PLCD7 (already contains fragments from event no. 10) resulting into a *G. max*, Croton and *S. lycopersicum* infecting Papaya leaf curl virus isolates. Similarly, PLCD5 became a weed infecting recombinant after acquiring fragments

Table 3: Recombination analysis of Indian PLCD complex using RDPv4.95 program

Event No.	Recombinant	Major parent	Minor Parent	Frequency of recombination event	Detection Method						
					R	G	B	M	C	S	T
1	PLCD5	PLCD7	PLCD15	1	+	+	+	+	+	+	+
2	PLCD14	PLCD13	PLCD8	1	+	+	+	+	+	+	+
3	PLCD35	PLCD8*	PLCD3	1	+	+	+	+	+	+	+
4	PLCD7	PLCD10*	PLCD6	1	+	+	+	+	+	-	+
5	PLCD33	PLCD2	PLCD28*	3	+	+	+	+	+	+	+
6	PLCD13	PLCD2*	PLCD28	2	+	+	+	+	+	+	+
7	PLCD19	PLCD29	PLCD8	1	+	+	+	+	+	+	+
8	PLCD26	PLCD2	PLCD7	2	+	+	-	+	+	-	+
9	PLCD3	PLCD4	PLCD28*	1	+	+	+	+	+	+	+
10	PLCD8	PLCD9	PLCD31*	14	+	+	+	+	+	+	+
11	PLCD23	PLCD1	PLCD28	1	+	+	+	+	+	+	+
12	PLCD34	PLCD28	PLCD12	1	+	+	+	+	+	+	-
Total=12	Total frequency = 29										

*Unknown parent detected in recombination analysis but the specified parent was the closest inferred sequence in this analysis.

The analysis was performed against DNA-A sequences of 35 begomoviruses using default parameters and following algorithms. R=RDP; G=GENCONV; B=BOOTSCAN; M=MAXCHI; C=CHIMAERA; S=SISCAN; T=3SEQ

through event no. 1 i.e. Papaya leaf curl virus infecting *S. lycopersicum* and Chili leaf curl virus infecting papaya. Overall, the Indian PLCD complex is a repertoire of recombinants as evident from the recombination analysis and as the level of recombination complexity increases, the host range of recombinant virus also expands.

4. Conclusion

The PLCD complex (Fig. 1) comprises of genetically diverse group of begomoviruses as observed from their distinct phylogenetic patterns, pairwise sequence similarity and recombination analysis in this study. The diverse range of hosts mentioned in the study could be well attributed to putative recombination events between begomovirus species and strains infecting different host crops or weeds. Hence, in case of high genetic variability, the symptoms should only be used as a parameter for estimation of severity index while studying disease impact. The overall, >75% sequence similarity among begomoviruses of PLCD complex (Fig. 4) in this study signifies evidence of some sequence conservation that could be used to design siRNA based strategy.

The high frequency recombination fragments have prospects to be targeted for siRNA based strategy until and unless the target regions lie within the borders of a recombination hotspots i.e. a region most amenable to undergo complete transposition under normal recombination circumstances. Hence, the recombination hotspots despite being genetically mobile fragments can be employed as regions contributing to present and future siRNA strategies. Similar study can be employed to assess disease complexes such as Cassava mosaic disease (Pita *et al.,* 2001), Cotton leaf curl disease (Farooq *et al*., 2011) and Tomato leaf curl disease (Moriones *et al.,* 2017) complex widespread in central and south-east Asia, Africa and American continents. Therefore, the recombination patterns and sequence conservation trends of similar disease complexes need careful investigation to develop an efficient and sustainable strategy against begomovirus disease complexes in future.

Acknowledgment

SS and SV acknowledge Babasaheb Bhimrao Ambedkar University, Lucknow for providing the necessary infrastructure required to carry out this research. SV acknowledges Babasaheb Bhimrao Ambedkar University, Lucknow for fellowship.

References

Boni, M.F., Posada, D. and Feldman, M.W. 2007. An exact nonparametric method for inferring mosaic structure in sequence triplets. *Genetics* **176**:1035-1047.

Borah, B.K. and Dasgupta, I. 2012. Begomovirus research in India: A critical appraisal and the way ahead. *Journal of Biosciences* **37**:791-806.

Brown, J.K., Frohlich, D.R. and Rosell, R.C. 1995. The sweetpotato or silverleaf whiteflies: Biotypes of *Bemisia tabaci* or a species complex? *Annual Reviews of Entomology* **40**:511-534.

Brown, J.K., Zerbini, F.M., Navas-Castillo, J., Moriones, E., Ramos-Sobrinho, R., Silva, J.C.F. and Fiallo-Olive, E. 2015. Revision of begomovirus taxonomy based on pairwise sequence comparisons. *Archives of Virology* **160**:1593-1619.

David, R.J. 2003. Plant viruses transmitted by whiteflies. *European Journal of Plant Pathology* **109**:195-219.

Farooq, A., Farooq, J., Mahmood, A., Shakeel, A., Rehman, A., Batool, A., Riaz, M., Shahid, M.T.H. and Mehboob, S. 2011. An overview of cotton leaf curl virus disease (CLCuD) a serious threat to cotton productivity. *Australian Journal of Crop Science* **5**:1823-1831.

Fondong, V.N. 2013. Geminivirus protein structure and function. *Molecular Plant Pathology*. **14**:635-649.

Gamerman, D. and H. F. Lopes. 2006. MarkovChain Monte Carlo: Stochastic Simulation for Bayesian Inference. 2nd ed. Chapman and Hall, London, pp. 342.

Ghosal, S., Das, S., and Chakrabarti, J. 2012. Chapter 11: Computational approaches for designing efficient and specific siRNAs. In: Pérez-Sánchez, H. (Ed.) *Bioinformatics*, InTech open, Croatia-EUROPEAN UNION, pp. 261-276.

Gibbs, M.J., Armstrong, J.S. and Gibbs, A.J. 2000. Sister-Scanning: a Monte Carlo procedure for assessing signals in recombinant sequences. *Bioinformatics* **16**:573-582.

Gutierrez, C. 2000. Geminiviruses and the plant cell cycle. *Plant Molecular Biology* **43**:763-772.

Guindon, S. and Gascuel, O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic Biology* **52**:696-704.

Harrison, B.D. 1985. Advances in geminivirus research. *Annual Reviews of Phytopathology* **23**:55-82.

Hallan, V., Saxena, S. and Singh, B.P. 1998a. Ageratum, Croton and Malvastrum harbor geminiviruses: evidence through PCR amplification. *World Journal of Microbiology and Biotechnology* **14**:931-932.

Hallan, V., Saxena, S. and Singh, B.P. 1998b. Yellow net of *Triumfetta* is caused by a Geminivirus: A first report. *Plant Disease* **82**:127-127.

Hanley-Bowdoin, L., Bejarano, E.R., Robertson, D. and Mansoor, S. 2013. Geminiviruses: masters at redirecting and reprogramming plant processes. *Nature Reviews in Microbiology* **11**:777-788.

Hasegawa, M., Kishino, H. and Yano, T. 1985. Dating of human-ape splitting by a molecular clock of mitochondrial DNA. *Journal of Molecular Evolution* **22**:160-174.

Huelsenbeck, J.P., Ronquist, F. 2001. MrBayes: Bayesian inference of phylogeny. *Biometrics* **17**:754-755.

King, A.M.Q., Adams, M., Eric, B.C. and Lefkowitz, E.J. 2011. Virus Taxonomy: Ninth Report of the International Committee on Taxonomy of Viruses. Elsevier. San Diego, CA, pp. 1339.

Larget, B. and Simon, D. 1999. Markov chain Monte Carlo algorithms for the Bayesian analysis of phylogenetic trees. *Molecular Biology and Evolution* **16**:750-759.

Lima, A.T.M., Silva, J.C.F., Silva, F.N., Castillo-Urquiza, G.P., Silva, F.F. and Seah, Y.M. 2017. The diversification of begomovirus populations is predominantly driven by mutational dynamics. *Virus Evolution* **3**:1-14.

Martin, D. and Rybicki, E. 2000. RDP: detection of recombination amongst aligned sequences. *Bioinformatics* **16**:562-563.

Martin, D.P., Posada, D., Crandall, K.A. and Williamson, C. 2005. A modified bootscan algorithm for automated identification of recombinant sequences and recombination breakpoints. *AIDS Research and Human Retroviruses* **21**:98-102.

Martin, D.P., Murrell, B., Golden, M., Khoosal, A. and Muhire, B. 2015. RDP4: Detection and analysis of recombination patterns in virus genomes. *Virus Evolution* **1**:1-5.

Moriones, E., Praveen, S. and Chakraborty, S. 2017. Tomato Leaf Curl New Delhi Virus: An emerging virus complex threatening vegetable and fiber crops. *Viruses* **9**:e264.

Nadeem, A., Mehmood, T., Tahir, M., Khalid, S. and Xion, Z.1997. First report of Papaya Leaf Curl Disease in Pakistan. *Plant Disease* **81**:1333-1333.

Padidam, M., Sawyer, S. and Fauquet, C.M. 1999. Possible emergence of new geminiviruses by frequent recombination. *Virology* **265**:218-225.

Pita, J.S., Fondong, V.N., Sangare, A., Ogwal, S. and Fauquet, C.M. 2001. Recombination, pseudorecombination and synergism of geminiviruses are determinant keys to the epidemic of severe cassava mosaic disease in Uganda. *Journal of General Virology* **82**:655-665.

Posada, D. and Crandall, K.A. 2001. Evaluation of methods for detecting recombination from DNA sequences: Computer simulations. *Proceedings of National Academy of Science USA* **98**:13757-13762.

Raj, S.K., Snehi, S.K., Khan, M.S., Singh, R. and Khan, A.A. 2008. Molecular evidence for association of Tomato leaf curl New Delhi virus with leaf curl disease of papaya (*Carica papaya L.*) in India. *Australasian Plant Disease Notes* **3**:152-155.

Roden, L.C. and Ingle, R.A. 2009. Lights, Rhythms, Infection: The role of light and the circadian clock in determining the outcome of plant-pathogen interactions. *Plant Cell* **21**:2546-2552.

Ronquist, F. and Huelsenbeck, J.P. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**:1572-1574.

Saxena, S., Hallan, V., Singh, B.P. and Sane, P.V. 1998a. Leaf Curl Disease of *Carica papaya* from India may be caused by a bipartite geminivirus. *Plant Disease* **82**:126-126.

Saxena, S., Hallan, V., Singh, B.P. and Sane, P.V. 1998b. Nucleotide sequence and inter-geminiviral homologies of the DNA-A of papaya leaf curl Geminivirus from India. *Biochemistry and Molecular Biology International* **45**:101-113.

Saxena, S., Hallan, V., Singh, B.P. and Sane, P.V. 1998c. Evidence from nucleic acid hybridization test for a Geminivirus infection causing leaf curl disease of papaya in India.

Indian Journal of Experimental Biology **36**:229-232.

Saxena, S., Kesarwani, R.K., Singh, V. and Singh, S. 2013. Designing of putative siRNA against geminiviral suppressors of RNAi to develop Geminivirus resistant Papaya crop. *International Journal of Bioinformatics Research and Applications* **9**:3-12.

Saxena, S., Singh, N., Ranade, S.A. and Babu, G.S. 2011. Strategy for a generic resistance to Geminivirus infecting papaya and tomato through *in-silico* siRNA search. *Virus Genes* **43**:409-434.

Saxena, S., Singh, V.K. and Verma, S. 2016. PCR mediated detection of sex and PaLCuV infection in papaya: A review. *Journal of Applied Horticulture* **18**:80-84.

Saxena, S. and Verma, S. 2016. Chapter 20: Harnessing the genetic variability in Plant-Virus-Vector complex interaction in Begomovirus family to prevent viral diseases. In: Sobti, R.C., Mishra, S. and Jaiswal, K. (Eds.) *Recent Advances in Applied Biosciences*, Narendra Publishing House, New Delhi, pp. 173-183.

Singh, P., Mishra, A.K. and Tripathi, N.N. 2012. Assessment of mycoflora associated with postharvest losses of papaya fruits. *Journal of Agricultural Technology* **8**:961-968.

Singh-Pant, P., Pant, P. and Mazumdar-Leighton, S. 2012. Spatial and temporal diversity of begomoviral complexes in papayas with leaf curl disease. *Archives of Virology* **157**:1217-1232.

Smith, J.M. 1992. Analyzing the mosaic structure of genes. *Journal of Molecular Evolution* **34**:126-129.

Varma, A. and Malathi, V.G. 2003. Emerging geminivirus problems: A serious threat to crop production. *Annals of Applied Biology* **142**:145-64.

Varun, P., Ranade, S.A. and Saxena, S. 2017. A molecular insight into papaya leaf curl—a severe viral disease. *Protoplasma* doi: 10.1007/s00709-017-1126-8.

van der Walt, E., Rybicki, E.P., Varsani, A., Polston, J.E., Billharz, R., Donaldson, L., Monjane, A.L. and Martin, D.P. 2009. Rapid host adaptation by extensive recombination. *Journal of General Virology* **90**:734-746.

Zhou, X. 2013. Advances in understanding Begomovirus Satellites. *Annual Reviews of Phytopathology* **51**:357-381.