bioRxiv preprint doi: https://doi.org/10.1101/2021.09.27.462025; this version posted September 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Evolutionary dynamics of *Begomoviruses* causing Papaya leaf curl disease in India

Aarshi Srivasatava[1$], Vineeta Pandey[1$], Anurag Kumar Sahu[2], Dinesh Yadav[1] and R.K. Gaur[1*]

1. Department of Biotechnology, D.D.U. Gorakhpur University, Gorakhpur-273009, Uttar Pradesh, India
2. International Centre for Genetic Engineering & Biotechnology, Aruna Asaf Ali Marg, New Delhi, Delhi - 110067, India

[$]**Equal contribution**

***Corrsponding author: gaurrajarshi@hotmail.com**

Abstract

The genus *begomovirus* represents a group of multipartite viruses that significantly damages many agricultural crops, including papaya and drastically influence the overall production. Understanding the genetic variations, mutations and recombination of the *begomovirus* population infecting papaya has several important implications in alleviating substantial losses, mainly in developing countries, including India. In an attempt for a bioinformatics driven study of diversity and variability of papaya leaf curl disease in India, a total of thirty-two (32) DNA-A and sixteen (16) betasatellite sequences were retrieved from GenBank. An uneven distribution of evolutionary divergence has been observed across the branch length, which triggered the estimated recombinational event. Interestingly, a maximum of the *begomoviruses* were found to be intra-species recombinants. Further genetic variability, selection pressure, and substitution rate acting on the population were estimated and found to be high enough to support the evolution of geminiviruses. Genetic divergence composition in all *begomovirus* datasets revealed predominance of nucleotide diversity driven by mutation. The analysis indicates that even though a significant fraction of the genetic variations might be due to recombination but, it was constantly lower than the mutation rate. Thus, the diversification of the *begomovirus* population is principally impelled by mutational dynamics.

Keywords: *Begomovirus*, Papaya leaf curl disease, Genetic variability, Phylogenetic analysis, Recombination

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.27.462025; this version posted September 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Introduction

The genus *Begomovirus* belongs to the family *Geminiviridae* consisting of 424 species, 2020 release Virus Taxonomy by the International Committee on Taxonomy of Viruses (ICTV) (Zerbini et al. 2017). A closed circular single stranded DNA encapsidated in a quasi-isometric non-enveloped twinned particle of ~ 2.8 kb size constitutes a virus genome (Stanley et al. 2005). The native New World *begomoviruses* have only bipartite genomes (DNA-A and DNA-B components) whereas, *begomoviruses* of the Old World constitute both monopartite and bipartite genomes (lack DNA-B and have DNA-A homolog) (Brown et al. 2012). Although in bipartite *begomovirus* DNA-A can replicate independently but they require DNA-B for nuclear localization. The uniqueness of monopartite is the presence of small ssDNA satellite molecules which greatly enhance the virus virulence. These molecules are alphasatellite (~1.4 kb), betasatellite (~1.3 kb) and a newly reported deltasatellite of approximately size of 0.7 kb (Brown et al. 2012; Lozano et al. 2016; Fiallo-Olivé et al. 2016). These *begomoviruses* are transmitted by whiteflies of the *Bemisia tabaci* cryptic species which has been adapted and co-evolved with the *begomovirus* genome (Brown et al. 2012; Marwal et al. 2021). The *B. tabaci* is considered as the second most important vector because of its semi-persistent mode and behavioural manipulation for transmitting plant viruses (Moreno-Delafuente et al. 2013). The genomic DNA of the monopartite genome and DNA-A of the bipartite genome have corresponding genome organization consisting of six open reading frames (AV2, AV1, AC3, AC2, AC1, AC4), whereas DNA-B consists of two open reading frames (BV1, BC1) which also contribute to intra as well as inter–cellular movement of viral particles within the host. Thus, together they cause systemic infection and develop typical symptoms (Nawaz-ul-Rehman et al. 2009; Hanley – Bowdoin et al. 2013). *Begomoviruses* commonly induce severe symptoms in their hosts, including golden mosaic, yellow mosaic and leaf curl. Devastating pathogens include members of species such as *Cotton leaf curl Multan virus*, *African cassava mosaic virus*, *Chilli leaf curl virus*, *Bean golden mosaic virus* and *Tomato yellow leaf curl virus.* A vast range of initial symptoms, such as green– yellow mottle/mosaic, leaf curling, interveinal yellowing, vein swelling, and yellow spots are found associated with dicotyledonous plants infected with *begomoviruses*. In India, papaya leaf curl disease was first identified by Thomas and Krishnaswamy in 1939, which was confirmed by Saxena and colleagues that the disease is caused by geminivirus, i.e., *Papaya leaf curl* virus (Saxena et al. 1998). The infection rate of disease depends on the severity of symptoms, type of *begomovirus*, vector and associated helper viruses, and their

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.27.462025; this version posted September 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

severity rate increases with the rise in the whitefly population. The symptoms like reduction in leaf size, leaf curling, vein thickening, yellow mosaic, interveinal chlorosis, and stunted growth of plants with small distorted fruits or no fruits can be found consociated with papaya (Shahid et al. 2013; Varun et al. 2017). However, stunting may be seen at a severe stage of infection, which may eventually cause the death of infected plants.

Diseases associated with *begomoviruses* are found to be a threat to global papaya production. *Carica Papaya*, a major tropical, sweet, large and herbaceous food crop, belongs to the order Brassicales and the family *Caricaceae* is cultivated throughout India, and valued for its medicinal and nutritional benefits (Yadav et al. 2016). The market demand for tropical fruits is rapidly increasing for papaya, and now ranking third due to its benefits and importance in agribusiness (Evans 2012). Despite being an important wide-ranging topical fruit grown throughout the year in many portions of the country, papaya yields are degraded due to infection by a number of viral diseases which are particularly important (Usha 1980; Nehra et al. 2019). Fruit is significantly important due to early bearing and space conserving variety and is considered as a rich source of fibres, minerals, and antioxidant nutrients. In addition, it is a source of the digestive enzyme Papain, which is found as an important industrial ingredient in meat tenderizing, cosmetics, beauty products, brewing and pharmaceuticals (Vij and Prashar 2015; Urgessa et al. 2019). Above all, papaya acts as a metabolic activator, detoxifier, homeostasis maintainer, and cell rejuvenator. The medicinal and nutritional value of Papaya aids human health (Adiaha and Adiaha 2017). Furthermore, its local and commercial cultivation is difficult to achieve its full potential (FAOSTAT 2019) due to the incidence and emergence of a large number of plant-infecting viruses resulting in significant crop loss (Nascimento et al. 2010). Majorly, leaf curl symptoms caused by *papaya leaf curl virus* are found associated with infected papaya plants from different regions, which initiates serious production losses and can act as potential inducer for transmission of viruses *via* vector whitefly (Guo et al. 2015). The substantial diversity of *begomoviruses* affecting papaya in India includes several reports such as *papaya leaf curl virus* (PaLCuV) found infecting papaya for the first time described in 1939 in India. Since then, *papaya leaf curl virus* has been found as an important repressor for papaya cultivation. In spite of being economically important tropical fruit, little attention has been paid to assess the genetic diversity at the molecular level for *papaya leaf curl virus* causing leaf curl disease in papaya (Fougat et al. 2015).

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.27.462025; this version posted September 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

This article provides an insight into genomic components with its satellite molecules associated with the *begomovirus* causing leaf curl disease, its diversity and evolution patterns using bioinformatics approach.

Materials and methods

Sequence retrieval and Sequence alignment

Genome sequences of *begomoviruses* infecting papaya reported till June 2021 were retrieved from GenBank (www.ncbi.nlm.nih.gov) and their distributions along with its sub-viral satellite across India were depicted. (Table 1; Fig1). Three sequence datasets for whole genome sequences of 32 DNA-A segment, its six ORFs and 16 associated betasatellite were prepared for further analysis. Only the sequences containing distinct nanomer "TAATATT" were considered for the present studies. Each specific datasets were aligned through multiple sequence alignments algorithm with Clustal W, using MEGA X software (Kumar et al. 2018).

Phylogenetic analysis

Evolutionary divergence helps to observe sequence diversity among the virus isolates. The wide range of distribution helps in understanding the viruses' origin, dispersion, development and etiology of diseases. Based on lowest BIC value, the best nucleotide substitution model were chosen using model test ultimately constructed a phylogenetic tree utilizing 1,000 bootstrap replicates for maximum likelihood (ML) algorithm embed in MEGA X software (Kumar et al. 2018). Different transversion and transitional substitutions rate together with transition/transversion bias (R) were also calculated using MEGA X program. (Kumar et al. 2018).

Detection of Recombination

The aligned nucleotide sequences of reference *begomovirus* were used to analyse the evidence of recombination by utilizing recombination detection program (v.4.2). Different algorithms of RDP 4.2 software such as RDP, BOOTSCAN, CHIMAERA, GENECONV, MAXCHI, 3SEQ and SISCAN (Martin et al. 2015) were used to detect a potential recombination event by predicting foremost and rearmost breakpoints, parental isolates and origin of viruses. For analysis, towards 0.05 highest acceptable Bonferroni corrected p-value

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.27.462025; this version posted September 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

and default detection thresholds datasets were subjected. To escape false-positive results, out of seven at least three algorithms were regarded appropriate to detect recombination events.

Population structure and Coalescent analysis

To investigate the genetic variability and diversity in the virus population several parameters embed in DnaSP (v. 6.12) software used (Rozas et al. 2017). The aligned sequence datasets were analysed for, total number of segregating sites (s), the average number of nucleotide differences between sequences (k), total number of mutations (η), nucleotide diversity (π), additionally, Watterson's estimate of the population mutation rate based on the total number of segregation sites ($\theta - w$) and the total number of mutations ($\theta - \eta$) were also calculated along with a number of haplotypes (h), and haplotype diversity (Hd) (Lima et al. 2017).

Furthermore, an implemented method in DnaSP (v.6.12.) software i.e., Neutrality test, are executed to calculate the hypothesis of selection pressure occurring in population. Therefore, sequence datasets separated by different geographical location are tested by employing tests such as Tajima's *D*, Fu and Li's *D** for identifying the difference between the total number of mutations and the number of singletons, and Fu and Li's *F** for identifying the difference between the average number of nucleotide differences between paired sequences and the number of singletons (Rozas et al. 2017).

Consequently, the mean substitution rate and mutational bias in the sequence datasets of *begomoviruses* were determined by using the Bayesian Markov Chain Monte Carlo (MCMC) parameter of BEAST (v.1.10) (Suchard et al. 2018). Coalescent constant demographic models and Best-fit molecular clock were detected employing BEAST and resultant were operated under Tracer program (v 1.5) to achieve effective sample size (Rambaut et al. 2018). MCMC chain used 10% burn-in value with 10^7 run lengths to provide 95% highest posterior density (HPD) interval for determining statistical uncertainty.

Results

Since 1990s, humans adapting to modernisation and India's changing climatic condition have been found promoting factors in the emergence and evolution of many *begomovirus* diseases. Frequent recombination and mutation such as nucleotide substitution are apparent as topological incongruence in evolutionary events which depend upon type of virus and host plant (Varun and Saxena 2018). The occurrence of leaf curl disease in papaya was observed

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.27.462025; this version posted September 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

which allowed to device a molecular perceptive for understanding the DNA polymorphism of *begomoviruses*.

Phylogenetic analysis and Detection of Substitution bias

Using MEGA X program, the evolutionary history for sequence datasets DNA-A and associated betasatellite were calculated using the ML Tree based on the best fit nucleotide substitution model i.e., (TN93+G) for DNA-A, its ORFs and (TN93+G+I) for betasatellite computing with CLUSTAL W pairwise alignment. The four distant lineages were observed with 1000 bootstrap support and grouped as (ChiLCV I a & b), (PaLCuV II), (PaYLCuV III), (PaLCrV IV) (Fig2 a). The lineages comprise the isolates collected from different smaller and larger geographic locations (Fig1). The branch length among population suggests the level of differentiation within them. In addition, well-defined clusters are observed in case of betasatellite depicting four clades distinct grouped as (ToLCB I), (ChiLCB II), (ToLCB III), (CroYVMB IV) (Fig2 b). Since, longer branches are reported to be associated with well supported recombination events though short branches also support the recombination (Lima et al. 2017). Accordingly, the mean branch lengths found in the present study might be contributing to recombinational events. Nevertheless, the results also indicate that *begomoviruses* infecting *Carica papaya* were not restricted to any solitary geographical region of India.

Moreover, different rate of transitional, transversion substitutions and transition/transversion bias (R) were estimated for the *begomoviruses* causing papaya leaf curl disease (Table 2). The rate of transitional substitutions ranged from 9.77 to 15.6, and transversional substitutions ranged from 5.41 to 7.71 while transition/transversion bias (R) of 0.97 was observed for DNA-A. Further, its six ORF's showed variable values based on gene nature for different substitution rate. The transitional rate was maximum for C4 gene while Rep gene showed minimum value. Similarly, maximum transversional substitutions rate was observed for REn gene while pre-CP gene showed minimum value. The highest and lowest transition/transversion bias (R) was observed for TrAP and C4 gene respectively. Similarly, for betasatellite the range obtained for transitional and transversional substitutions was 8.49 to 14.62 and 5.64 to 9.26 respectively while, the transition/transversion bias (R) was 0.82. However, the estimation might support the contribution of mutation for nucleotide polymorphisms in a population.

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.27.462025; this version posted September 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Detection of Recombination

Further, the tree- like phylogenetic divergence obtained for sequence datasets directed us to detect the occurrence of non-tree-like evolution within populations to explain the potential of recombination events in aligned sequences. For analysis, different parameters were used for determining the shared overlapping intra- and inter-specific recombination events distributed throughout the genome with different parental combinations. Thirty-two putative recombination events were observed for (DNA-A) datasets and sixteen putative recombination breakpoints were identified for betasatellites (Table 3a). However, intra-specific recombination among different reference sequences were predominately observed in AC1, AV1 and AC2 rich genome region while AV2, AC3, C4 genome region showed minimum effect. This substantiate the possibility of putative recombination breakpoints among ORF's of theses DNAs (Table 3b). Moreover, recombination events distributed predominantly in the βC1 genome region of betasatellites supports the prevalence of recombination that is involved in virus movement by suppressing host antiviral silencing gene (Kumar et al. 2015). Further, relevant recombination events were obtained by selecting at least three or more methods which minimize incompetent outcomes. Thus, significant amounts of genetic variation were supported by maximum putative recombinational events among sequence datasets.

Population Structure and Coalescent analysis

An evolutionary scenario signifies the role of nucleotide substitution along with recombination in gaining genetic variation and evolution among *begomoviruses* (Mishra et al. 2020). The mean substitution rate among sequences datasets of DNA-A were 2.80×10^{-4} sub site^{-1} year^{-1} (DNA-A, 95% highest posterior density (HPD) interval ranging from 3.541×10^{-5} to 9.747×10^{-4}), which is comparatively higher when compared with the range of nucleotide substitution rate of RNA viruses thus, suggesting that though geminivirus evolve at analogous rate as many RNA viruses reported so far (Jenkins et al. 2002; Duffy and Holmes 2009; Kumar et al. 2015) but importantly the high substitution frequency detected here shows short term mutational phenomenon acting on population rather than long term substitution rate. To justify above, high rate of nucleotide substitutions was also detected superficial in the three gene datasets i.e., pre-CP, REn, Rep (Table 4). In addition, 5.14×10^{-5} sub site^{-1} year^{-1} (β, 95% HPD interval ranging from 7.157×10^{-7} to $1.56 \times 10^{-4)}$ substitution rate was detected for betasatellites. Conversely, relaxed molecular clock is used as prior to get the suitable

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.27.462025; this version posted September 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

value of mean substitution rate and the detected high substitution value are questionable to be caused by strong positive selection (Duffy and Holmes 2008, 2009).

Since, selection pressure acting for genetic variation was also effectuated by codon degeneracy, therefore mutational selection pressure within three nucleotide codon positions i.e., CP1, CP2, CP3 respectively, were appraised and found higher at codon position C3 for DNA-A and among ORFs highest codon degeneracy were found in CP gene at codon position C3 as compared to other genes. Similarly, for betasatellites chief mutation rate were found at codon position C3 (Table 4).

Moreover, to determine the degree of genetic variability, demography structure analysis was estimated (Table 5). However, the analysis revealed for DNA-A, number of polymorphic sites (s) were 1737 with 2720 number of mutation (η) having nucleotide diversity (π) of (π=0.2) which were found greater on comparison with CLCuMulV (RV-1) having a low degree of genetic variability ($\pi > 0.08$) (Mishra et al. 2020). Similarly, variability was seen in the ORFs of the DNA- A, the significant contributor of high degree of genetic divergence was verified by the non-randomness of nucleotide variability throughout genome region which were performed effectively by C4, Rep and REn gene depicting highest nucleotide diversity (π) value (Table 5). Simultaneously, for betasatellite a number of polymorphic sites (s) were 834 with 1482 number of mutation(η) having nucleotide diversity(π) of (π=0.3) were largely high on comparison with ChiLCB (MM-2) ($\pi > 0.06$) (Mishra et al. 2020). Therefore, the estimation suggests diverseness among populations. Furthermore, the number of haplotype (h) and haplotype diversity (Hd) was also estimated for the sequence datasets, which decipher the total number of haplotype (h) distribution for DNA-A was 32 and its haplotype diversity was detected equal to 1(1.000). Similarly, among its ORFs the haplotype diversity (Hd) were found close to 1 for gene pre-CP, REn, Rep, C4 and equal to 1 for gene CP and TrAP which support the relative contribution of genes in DNA polymorphism. Simultaneously, for betasatellite the total number of haplotype (h) and its diversity were detected 17 and equal to 1(1.000) respectively. Therefore, the estimation suggests the uniqueness within population and subpopulation (Table 5).

Neutrality tests were used to assess and understand the demographic selection acting on genetic population of *begomoviruses* and associated satellite molecules. Tajima's D was chosen for evaluation criteria which statistically reflect the negative Tajima's D value for DNA-A, its ORFs and associated betasatellites. Predominately, the statistically significant

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.27.462025; this version posted September 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

values were negative, which indicates the large proportion of genetic segregation might be there within sequence datasets which were unique to individual sequences. Simultaneously, the other parameters such as Fu & Li's *D* and Fu & Li's *F* tests of population statistics were also evaluated resulting in negative values for DNA-A, its ORFs and betasatellites, indicating reiterating of purifying selection and population expansion which might be due to the inherent diversity. Nevertheless, the combination of Tajima's D, Fu & Li's *D,* and Fu & Li's *F* negative values for DNA-A, its ORFs and associated satellite population revealed the conserved nature of the gene. Such evidences of nucleotide diversity might be expected when a selective sweep succeed the expansion of population and when most observable segregation functioning at the nucleotide level in a population are momentary and are eventually withdrawn by purifying selection (Table 6).

Discussion

India shares a large portion of the population which depends on small-area agricultural farming for their subsistence and income. A wide variety of diseases and their infection rate has been seen causing devastating effects both on crop yield and human persistence. Undoubtedly, the cultivation practices and presence of tropical climate conditions in the Indian subcontinent aids to the prevalence of a large number of plant viruses here, *Papaya leaf curl virus* (PaLCuV) found causing papaya leaf curl disease (PaLCuD), which affects plant growth, fruit size, quality and quantity, slowing its yield (Shahid et al. 2013; Varun et al. 2017) thus accelerating the spread of viral diseases. Additionally, climate change, adaptability and fast distribution of vectors and viruses is of major concern for the agriculture sector as they are greatly contributing to the Indian economy.

Across India, a full-length sequence of reported *begomoviruses* infecting *Carica papaya* were collected from NCBI and arranged into three specific datasets containing sequence of DNA-A, its ORFs and associated betasatellites. The recurrent occurrence of recombination and nucleotide substitution alike to RNA viruses are mostly attributed factors to contribute high genetic variability among *begomovirus* populations which may significantly step up their evolution by expanding the combinations of pre-existing nucleotide segregation created by mutation (Duffy and Holmes 2008, 2009). Accordingly, recombination and mutation are often stated as the chief contributors to genetic variability which is the subject matter of investigation in the present study by the aid of molecular and computational efficacy.

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.27.462025; this version posted September 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

The results revealed partitioning-based diversification, recombination together with mean substitution and purifying selection, as major contributor of observed levels of genetic variability across *begomovirus* genomes. The phylogeny-based partitioning method qualitatively estimated from the aligned whole-genome sequences and satellite molecules showing mean branch length were found useful in quantifying the effect of recombination event. To exclude any others biases that strengthen the significant differences between the degrees of intra and inter-specific variability, we checked for different rate of transition and transversion substation rate and transition /transversion bis(R) which supports the contribution of mutation for nucleotide polymorphisms in a population (Mishra et al. 2020).

Previous studies have revealed that recombination happens at high frequencies in *begomovirus* populations, which uses a conserved feature i.e., rolling circle mechanism for replicating their genomes and make it mechanistically recombination-prone thus, generating recombination breakpoints in a non-random location (Martin et al. 2011). Recombination rates are threatened for plant viruses. Our experimental analysis detected a recombination event in genomic region of datasets, showing variable parents which results in uneven distribution of recombination breakpoints supporting genetic diversity. However, the statistically measurable recombination rate seems to be lower than the mutation rate in sequence datasets. Even in such case recombinational event act actively but consequently, the mutational dynamics were still the leading forces in shaping the standing genetic variability. In other words, the relative contribution is potentially better than that estimated from our studies.

The substantial aspects towards the population genetics are possibly accommodated by mutation along with recombination, neutral selection, genetic drift and gene flow, which acts significantly to shape the genetic structure of populations. Further, we used coalescent Bayesian skyline model, strict and relaxed clock log normal and found relaxed clock as a prior to explore genetic diversity. Additionally, it is important to address the key issue that refers to the uneven distribution of the genetic variation across *begomovirus* genomes. In this context, the combination and pattern of various factors are responsible for affecting genetic variability through distribution of polymorphisms in non-random manner in the genomic regions of *begomoviruses* of datasets (Mishra et al. 2020).

Nevertheless, the standing genetic variability in all *begomovirus* populations was dominated by mutation, since for all datasets nucleotide diversity(π) were found high, suggesting the diverseness. Neutrality methods are used for finding positive selection to validate the significance of positive or diversifying selection in shaping the uneven levels of genetic

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.27.462025; this version posted September 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

variability across genome (Lima et al. 2017). Based on analysis negative values for given datasets indicates the reiterating of purifying selection and population expansion. Moreover, not any sporadic cases of positively selected sites were spotted. Therefore, our results clearly eliminate positive selection as a major contributor for the neutral selection in population. By some mean purifying selections are found responsible in accelerating the genetic variability in specific regions of genes.

Although the number of sequence data for DNA-A and betasatellites are small, and for alphasatellite the database is insufficient, the results of the present study possibly provide meaningful basic information that contributed greatly to diversification of *begomoviruses* particularly causing papaya leaf curl diseases and acknowledged the evolutionary potential, principally in the context of recombination, adaptation, genetic diversity, emergence and evolution to novel *begomovirus* and host type.

Contributions

AS and VP performed experiments and wrote the manuscript, AKS helped in data collection. RKG and DY guided the design of the whole test scheme. All authors have read and approved the final manuscript.

Ethical Declaration

Competing Interest: Author's proclaims no conflict of interest.

Funding: This study established no external funding.

References

Adiaha MS, Adiaha MS (2017) Effect of nutritional, medicinal and pharmacological properties of papaya (*Carica papaya* L.) to human development: a review. World Sci News 2:238–249.

Brown JK, Fauquet CM, Briddon RW, Zerbini M, Moriones E, Navas Castillo J (2012) Geminiviridae. In: King AMQ, Adams MJ, Carstens EB, Lefkowitz EJ (ed) In Virus Taxonomy- Ninth Report of the International Committee on Taxonomy of Viruses. Elsevier Academic Press., San Diego, CA: pp 351–373

Duffy S, Holmes EC (2008) Phylogenetic evidence for rapid rates of molecular evolution in the single-stranded DNA *begomovirus* Tomato yellow leaf curl virus. J Virol 82:957–965. https://doi.org/10.1128/JVI.01929-07

Duffy S, Holmes EC (2009) Validation of High Rates of Nucleotide Substitution in Geminiviruses: Phylogenetic Evidence from East African Cassava Mosaic Viruses. J of Gen Virol 90:1539–47. https://doi.org/10.1099/vir.0.009266-0

Evans EA, Ballen FH (2012) An Overview of Global Papaya Production, Trade, and

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.27.462025; this version posted September 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Consumption. FE913.

FAOSTAT (2019) Food and Agriculture organization of the United Nation Database. http://www.fao.org/faostat/en/#data/QC. Accessed 26 June 2021.

Fiallo-Olivé E, Tovar R, Navas-Castillo J (2016) Deciphering the biology of deltasatellites from the New World: maintenance by New World *begomoviruses* and whitefly transmission. New Phytol 212:680–692. https://doi.org/10.1111/nph.14071

Fougat R S, Purohit AR, Kumar S, Parekh MJ, Kumar M (2015) SSR based genetic diversity in *Abelmoschus* species. Indian J Agr Sci 85:1223–1228. https://doi.org/10.1007/s13562-016-0378-2

Guo T, Guo Q, Cui X, Liu Y, Hu J, Liu S (2015) Comparison of transmission of papaya leaf curl China virus among four cryptic species of the whitefly *Bemisia tabaci* complex. Sci Rep 25:15432. https://doi.org/10.1038/srep15432

Hanley-Bowdoin L, Bejarano ER, Robertson D, Mansoor S (2013) Geminiviruses: masters at redirecting and reprogramming plant processes. Nat Rev Microbiol 11:777–788. https://doi.org/10.1038/nrmicro3117

Jenkins GM, Rambaut A, Pybus OG, Holmes EC (2002) Rates of molecular evolution in RNA viruses: a quantitative phylogenetic analysis. J Mol Evol 54:156–165. https://doi.org/10.1007/s00239-001-0064-3

Kumar RV, Singh AK, Singh AK, Yadav T, Basu S, Kushwaha N, Chattopadhyay B, Chakraborty S (2015) Complexity of *begomovirus* and betasatellite populations associated with chilli leaf curl disease in India. J Gen Virol 96(10):3143-3158. https://doi.org/10.1099/jgv.0.000254

Kumar S, Stecher G, Li M, Knyaz C, Tamura K (2018) MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. Mol Biol Evol 35:1547–1549. https://doi.org/10.1093/molbev/msy096

Lima ATM, Silva JCF, Silva FN, Castillo-Urquiza GP, Silva FF, Seah YM (2017) The diversification of *begomovirus* populations is predominantly driven by mutational dynamics. Virus Evol 3:vex005. https://doi.org/10.1093/ve/vex005

Lozano G, Trenado HP, Fiallo-Olivé E, Chirinos D, Geraud-Pouey F, Briddon RW (2016) Characterization of non – coding DNA satellites associated with sweepoviruses (genus *Begomovirus*, Geminiviridae) – definition of a distinct class of *begomovirus* – associated satellite. Frontiers Microbiology 7:162. https://doi.org/10.3389/fmicb.2016.00162

Martin DP, Lefeuvre P, Varsani A et al (2011) Complex Recombination Patterns Arising During Geminivirus Coinfections Preserve and Demarcate Biologically Important Intra-Genome Interaction Networks. PloS Pathogens 7:e1002203. https://doi.org/10.1371/journal.ppat.1002203

Martin DP, Murrell B, Golden M, Khoosal A, Muhire B (2015) RDP4: Detection and analysis of recombination patterns in virus genomes. Virus Evol 1:vev003. https://doi.org/10.1093/ve/vev003

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.27.462025; this version posted September 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Marwal A, Nehra C, Verma RK, Mishra M, Srivastava D, Choudhary P, Gaur RK (2021) First report of papaya leaf curl virus and its associated papaya leaf curl betasatellite infecting *Catharanthus roseus* plants in India. The Journal of Horticultural Science Biotechnology https://doi.org/10.1080/14620316/2021/1912646

Mishra M, Verma RK, Marwal A, Sharma P, Gaur RK (2020) Biology and Interaction of the Natural Occurrence of Distinct Monopartite *Begomoviruses* Associated with Satellites in *Capsicum annum* from India. Frontiers Microbiology 11:512957. https://doi.org/10.3389/fmicb.2020.512957

Moreno-Delafuente A, Garzo E, Moreno A, Fereres A (2013) A plant virus manipulates the behavior of its whitefly vector to enhance its transmission efficiency and spread. PloS one 8(4):e61543. https://doi.org/10.1371/journal.pone.0061543

Nascimento AKQ, Lima JAA, Nascimento ALL, Beserra EAJr (2010) Biological, physical, and molecular properties of a *Papaya lethal yellowing virus* isolate. Plant Disease 94(10):1206-1212. https://doi.org/10.1094/PDIS-11-09-0733

Nawaz-ul-Rehman MS, Fauquet CM (2009) Evolution of geminiviruses and their satellites. FEBS letters 583(12):1825-1832. https://doi.org/10.1016/j.febslet.2009.05.045

Nehra C, Marwal A, Verma RK, Mishra M, Sharma P, Gaur RK (2019) *Papaya yellow Leaf curl virus*: A newly identified *Begomovirus* infecting *Carica papaya* L. from the Indian Subcontinent. The Journal of Horticultural Science and Biotechnology. https://doi.org/10.1080/14620316.2019.1570827

Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA (2018) Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7. Syst Biol 67(5):901-904. https://doi.org/10.1093/sysbio/syy032

Rozas J, Ferrer-Mata A, Sánchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE (2017) DnaSP 6: DNA sequence polymorphism analysis of large data sets. Mol Biol Evol 34:3299–3302. https://doi.org/10.1093/molbev/msx248

Saxena S, Hallan V, Singh BP, Sane PV (1998) 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. https://doi.org/10.1080/15216549800202472

Shahid MS, Yoshida S, Khatri-Chhetri GB, Briddon RW, Natsuaki KT (2013) Complete nucleotide sequence of a monopartite *Begomovirus* and associated satellites infecting *Carica papaya* in Nepal. Virus Genes 46(3):581-4. https://doi.org/10.1007/s11262-013-0888-0

Stanley J, Bisaro DM, Briddon RW, Brown JK, Fauquet CM, Harrison BD, Rybicki EP, Stenger DC (2005) Geminiviridae. In: Fauquet CM, Mayo MA, Maniloff J, Desselberger U, Ball LA (ed) VIIIth Report of the International Committee on Taxonomy of Viruses. Virus Taxonomy. Elsevier/Academic Press, London, pp 1163-1169

Suchard MA, Lemey P, Baele G, Ayres DL, Drummond AJ, Rambaut A (2018) Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. Virus Evol 4(1):vey016. https://doi.org/10.1093/ve/vey016

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.27.462025; this version posted September 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Urgessa OE, Itana DD, Raga OT (2019) Extraction of Papain from Papaya *(Carica Papaya L.)* Fruit Latex and Its Application in Transforming Tannery Raw Trimming. Ethiopian Journal of Science and Sustainable Development 6(2):22-32. https://doi.org/10.20372/ejssdastu:v6.i2.2019.92

Usha R (1980) Characterization, diagnosis & management of plantviruses. In: Rao GP, Kumar PL, Holguı´n-Pen˜a RL (ed) Studium Press, Houston, pp 387–392

Varun P, Ranade SA, Saxena S (2017) A molecular insight into papaya leaf curl – a severe viral disease. Protoplasma 254 (6):2055-2070. https://doi.org/10.1007/s00709-017-1126-8

Varun P, Saxena S (2018) Association of tomato leaf curl Gujarat virus and tomato leaf curl Bangladesh betasatellite on Papaya showing typical leaf curl symptoms in North India. 3Biotech 8:243. https://doi.org/10.1007/s13205-018-1254-7

Vij T, Prashar Y (2015) A review on medicinal properties of *Carica papaya* Linn. Asian Pacific Journal of Tropical Disease 5(1):1–6. https://doi.org/10.1016/S2222-1808(14)60617-4

Yadav J, Yadav S, Mishra S (2015) Experimental evidences showing nutritional and medicinal property of *Carica papaya* plant. International journal of science and research 6(12):21-25. https://doi.org/10.21275/ART20178569

Zerbini, F.M., Briddon, R.W., Idris, A., Martin, D.P., Moriones, E., Navas-Castillo, J., Rivera-Bustamante, R., Roumagnac, P., Varsani, A., 2017. ICTV report consortium. ICTV virus taxonomy profile: geminiviridae. J. Gen. Virol. 98 (2), 131–133. https://doi.org/10.1099/jgv.0.000738.

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.27.462025; this version posted September 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Legend for Figures and Tables

Fig1. Distribution of distinct *begomoviruses* isolates and associated satellites in India causing leaf curl disease of *Carica papaya* (Follow Table 1).

Fig 2 . Maximum-likelihood phylogeny-based partitioning tree associated with papaya leaf curl disease in India aligned using CLUSTAL W within MEGA v.10. (a) DNA-A; (b) Betasatellites.

Table 1. Features of *begomoviruses* causing leaf curl disease in papaya are identified in India.

Table 2. Substitution rate for *begomoviruses* and betasatellites associated with PaLCuD in India.

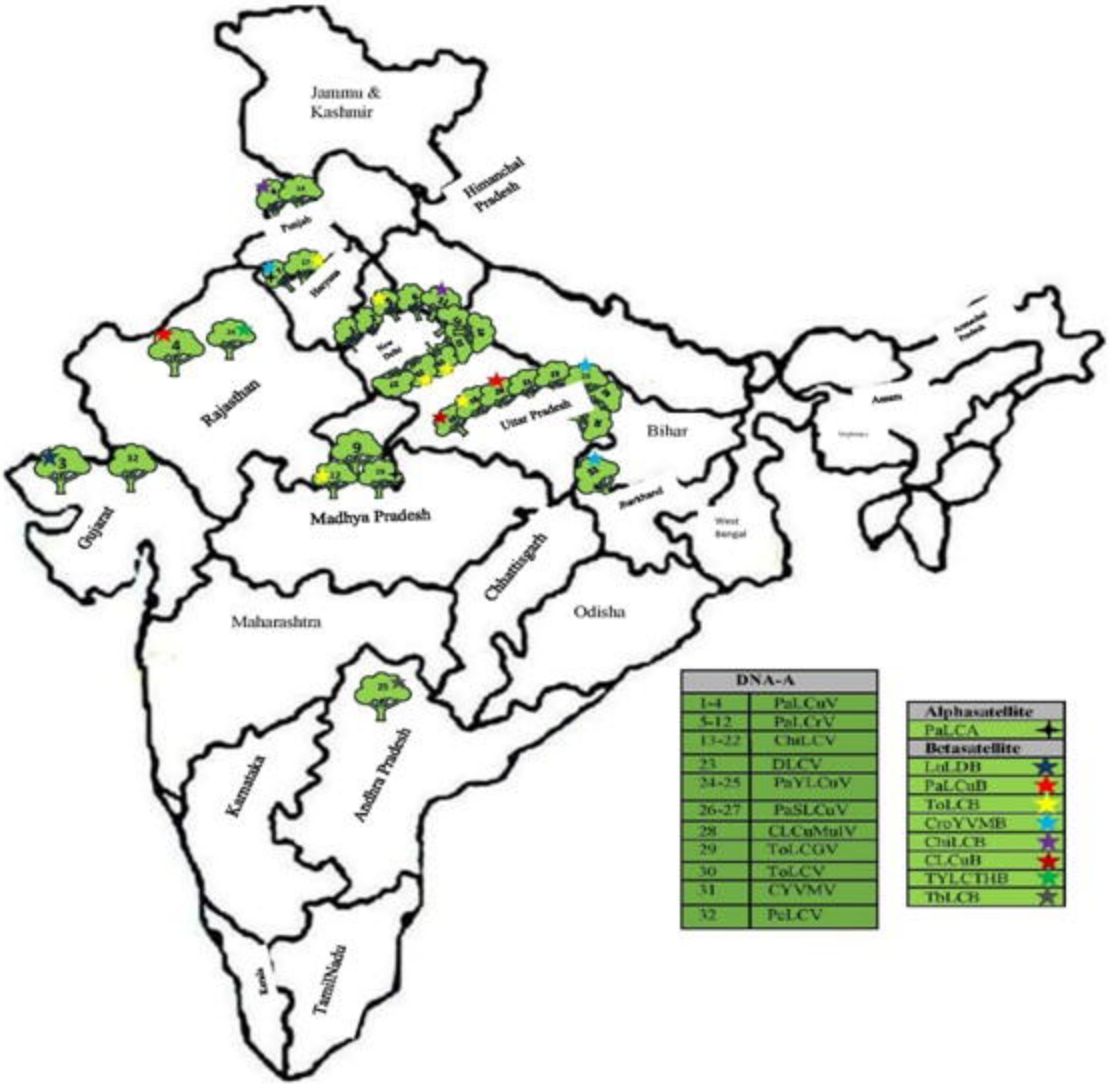
Table 3 (a) Putative recombination events detected among *begomoviruses* associated with papaya leaf curl disease, based on provided datasets from India; DNA-A and associated betasatellite.

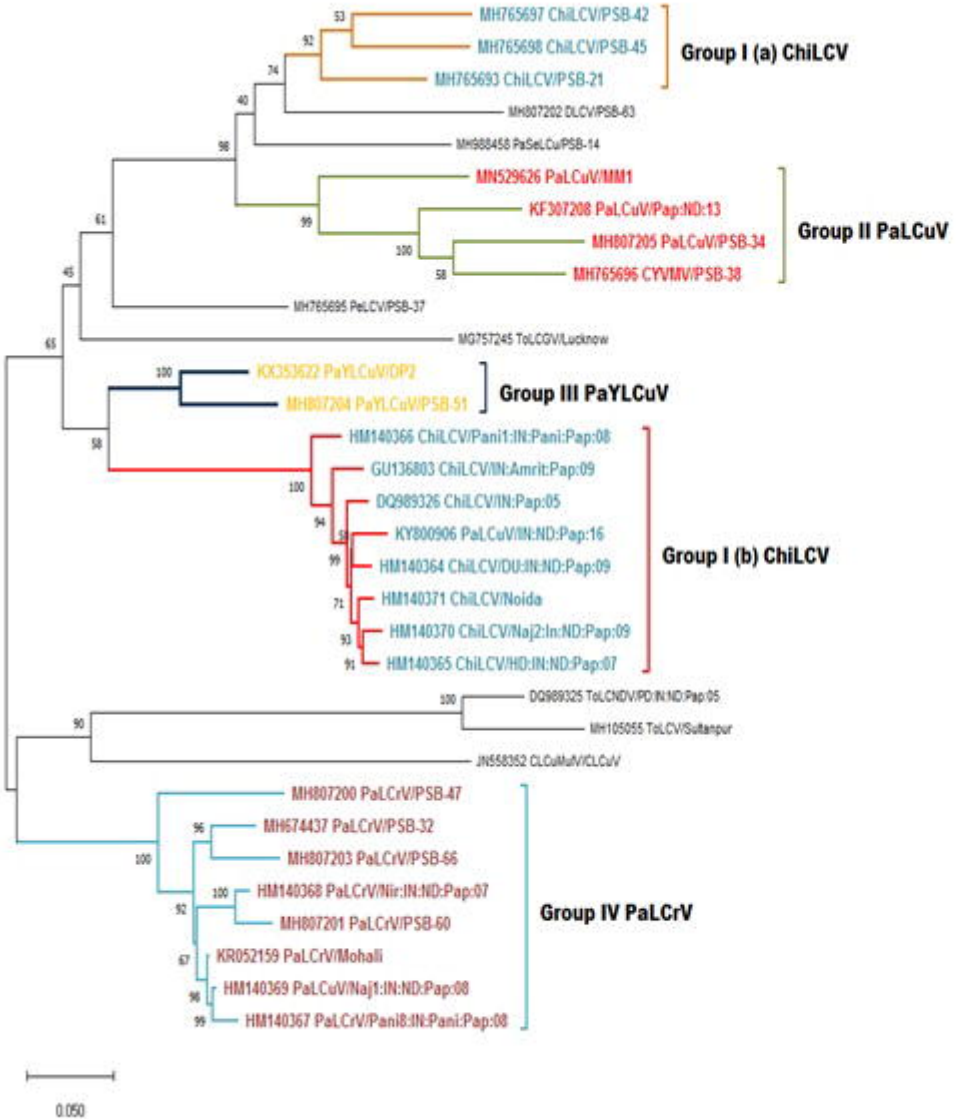
Table 3 (b) Putative recombination events detected among *begomoviruses* associated with papaya leaf curl disease, based on provided datasets from India; six genes/ORF's of DNA-A.

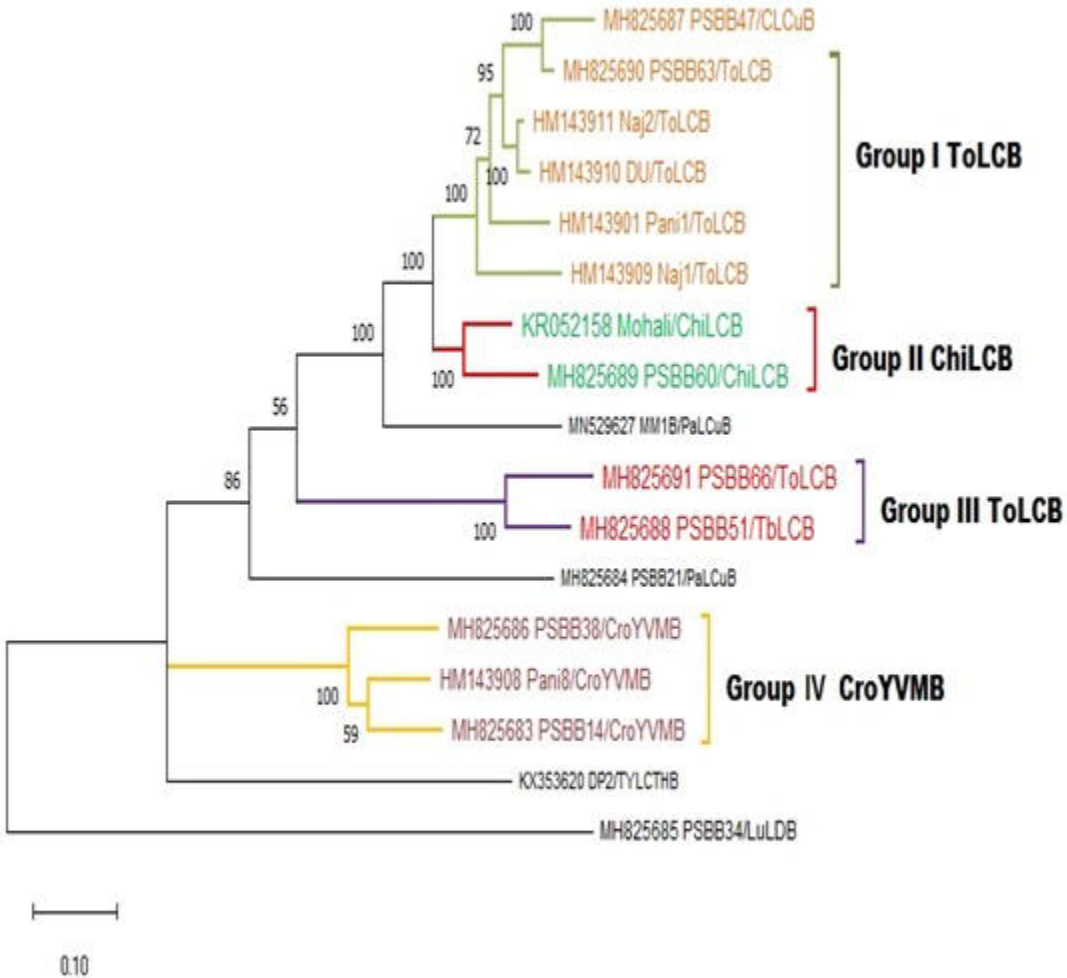
b Mean substitution and codon degeneracy rates for PaLCuD associated *begomoviruses*.

Table 5. Genetic diversity of *begomoviruses* and betasatellites associated with PaLCuD in India.

Table 6. Different neutrality tests for the datasets obtained from identified *begomoviruses* causing PaLCuD in India.


Jammu & Kashmir
Himachal Pradesh
Punjab
Haryana
Rajasthan
Uttar Pradesh
Bihar
Jharkhand
West Bengal
Madhya Pradesh
Gujarat
Chhattisgarh
Odisha
Maharashtra
Karnataka
Andhra Pradesh
Tamil Nadu
Kerala
Assam
Meghalaya
DNA-A
1-4 PaLCuV
5-12 PaLCrV
13-22 ChiLCV
23 DLCV
24-25 PaYLCuV
26-27 PaSLCuV
28 CLCuMuIV
29 ToLCGV
30 ToLCV
31 CYVMV
32 PeLCV
Alphasatellite
PaLCA
Betasatellite
LuLDB
PaLCuB
ToLCB
CroYVMB
ChiLCB
CLCuB
TYLCTHB
ThLCB


MH765697 ChiLCV/PSB-42
MH765698 ChiLCV/PSB-45
MH765693 ChiLCV/PSB-21
Group I (a) ChiLCV
MH807202 DLCV/PSB-63
MH988458 PaSeLCu/PSB-14
MN529626 PaLCuV/MM1
KF307208 PaLCuV/Pap:ND:13
MH807205 PaLCuV/PSB-34
MH765696 CYVMV/PSB-38
Group II PaLCuV
MH765695 PeLCV/PSB-37
MG757245 ToLCGV/Lucknow
KX353622 PaYLCuV/DP2
MH807204 PaYLCuV/PSB-51
Group III PaYLCuV
HM140366 ChiLCV/Pani1:IN:Pani:Pap:08
GU136803 ChiLCV/IN:Amrit:Pap:09
DQ989326 ChiLCV/IN:Pap:05
KY800906 PaLCuV/IN:ND:Pap:16
HM140364 ChiLCV/DU:IN:ND:Pap:09
HM140371 ChiLCV/Noida
HM140370 ChiLCV/Naj2:In:ND:Pap:09
HM140365 ChiLCV/HD:IN:ND:Pap:07
Group I (b) ChiLCV
DQ989325 ToLCNDV/PD/IN:ND:Pap:05
MH105055 ToLCV/Sultanpur
JN558352 CLCuMuV/CLCuV
MH807200 PaLCrV/PSB-47
MH674437 PaLCrV/PSB-32
MH807203 PaLCrV/PSB-66
HM140368 PaLCrV/Nir:IN:ND:Pap:07
MH807201 PaLCrV/PSB-60
KR052159 PaLCrV/Mohali
HM140369 PaLCuV/Naj1:IN:ND:Pap:08
HM140367 PaLCrV/Pani8:IN:Pani:Pap:08
Group IV PaLCrV
0.050


MH825687 PSBB47/CLCuB
MH825690 PSBB63/ToLCB
HM143911 Naj2/ToLCB
HM143910 DU/ToLCB
HM143901 Pani1/ToLCB
HM143909 Naj1/ToLCB
Group I ToLCB
KR052158 Mohali/ChiLCB
MH825689 PSBB60/ChiLCB
Group II ChiLCB
MN529627 MM1B/PaLCuB
MH825691 PSBB66/ToLCB
MH825688 PSBB51/TbLCB
Group III ToLCB
MH825684 PSBB21/PaLCuB
MH825686 PSBB38/CroYVMB
HM143908 Pani8/CroYVMB
MH825683 PSBB14/CroYVMB
Group IV CroYVMB
KX353620 DP2/TYLCTHB
MH825685 PSBB34/LuLDB
100
95
72
100
100
100
100
56
86
100
100
100
59
0.10