OPEN ACCESS Journal Name

ISSN
DOI:

Research Article
A Pathogenic Isolate of Monopartite PepYLCV DNA A-like Genome Differs Significantly in C1 Gene and CR Sequence, but not in their other Genes

[1]Jamsari Jamsari, [1]Istino Ferita, [2]Ade Noverta, [1]Eko Dharma Husada, [3]Friedrich W. Herberg, [4]Wolfgang Nellen and [1]Lily Syukriani

[1]Department of Molecular Breeding, Agricultural Faculty University of Andalas, Campus Limau Manis, 25136 Padang, West Sumatra, Indonesia
[2]Department of Agroecotechnology, Agricultural Faculty University of Andalas, Campus Dharmasraya, West Sumatra, Indonesia
[3]Department of Biochemistry, Kassel University, Heinrich-Plett Str. 40, 34132, Building AVZ I, 2 Stock, Kassel, Germany
[4]Department of Genetics, Kassel University, Heinrich-Plett Str. 40, 34132 Kassel, Germany

Abstract

Background and Objective: In order to elucidate pathogenic determining factor of pepper yellow leaf curl virus isolated from a lowland chili pepper cultivation field in West Sumatra, Indonesia, it was conducted whole genome analysis via primer walking strategy. **Materials and Methods:** The whole DNA A-like genome sequence of PepYLCIpsV was elucidated via five steps of primer walking, started with universal primers PAR1c715 and PAL1v1978 as the start point. **Results:** Whole genome comparison analysis identified only one single base insertion/deletion event located in the common region from a previously described isolate collected from the upland location (PpepYLCItdV). The BLAST comparison on the nucleotide level showed 90.6% maximal homology with the existing DNA A-like genome from many pepper yellow leaf curl viruses deposited in public database so far. Further detail comparison each of six Open Reading Frames (ORFs) between PepYLCIpsV and PpepYLCItdV indicated that V1 and V2 displayed 94-95% homology, respectively, while C2 and C3 had homology in range of 99 and 97%, respectively. Interestingly C1 and C4 showed homology only 79 and 68%, respectively and the Common Region (CR) shared only 74% similarity. **Conclusion:** The C1 and CR could be the determining factors for the pathogenicity, therefore, characterizing of these two regions in the population could be used for management and controlling of the virus. The pathogenic isolate PepYLCIpsV (KT809345) appears to be derived by recombination from two isolates originating from different regions and hosts. Its existence seems to be more ancient than its currently mild dominant counterpart PepYLCItdV (KT809346).

Key words: *Capsicum annuum*, pathogenicity factor, replicase protein, common region

Received: Accepted: Published:

Citation: Jamsari Jamsari, Istino Ferita, Ade Noverta, Eko Dharma Husada, Friedrich W. Herberg, Wolfgang Nellen and Lily Syukriani, 2016. A pathogenic isolate of monopartite PepYLCV DNA A-like genome differs significantly in C1 gene and CR sequence, but not in their other genes.

Corresponding Author: Jamsari Jamsari, Department of Molecular Breeding, Faculty of Agricultural, University of Andalas, Campus Limau Manis, 25136 Padang, West Sumatra, Indonesia Tel: +62-751-72776

Competing Interest: The authors have declared that no competing interest exists.

Data Availability: All relevant data are within the paper and its supporting information files.

INTRODUCTION

Pepper yellow leaf curl virus (PepYLCV) is the causing agent of pepper yellow leaf curl disease in *Capsicum annuum* species. The virus belongs to the family of Geminiviridae, grouped into mono and bipartite-genome classification. The bipartite genome group is characterized by the presence of genom A and genom B which are almost similar in size but contain different genes accessories. On the other hand, the monopartite group is characterized by having only a DNA A-like genome accompanied with a β-satellite component[1,2].

Six genes namely V1/AV1, V2/AV2, C1/AC1, C2/AC2, C3/AC3 and C4/AC4 have been identified from DNA A-like/DNA-A genome of monopartite/bipartite *Begomoviruses*[3]. The DNA-B genome in bipartite *Begomovirus* normally contains only small number of genes enhancing the role of genes located on the DNA-A genome, which is important for systemic symptoms and cell to cell movement of virus[4]. The role of the β-satellite component in monopartite group is believed to be involved in pathogenicity[5] and symptom expression as well as enhancement of the accumulation of helper virus in host plants[6].

West Sumatra is one of the central chilli productions in Indonesia, in which diet culture places chili pepper fruit as one of the important components in their spices. Previous studies by Jamsari and Pedri[2] identified the existence of monopartite PepYLCV in West Sumatra designated as PepYLCItdV. Although, also bipartite PepYLCIV has previously been reported in Indonesia by Sakata *et al.*[7]. Artificial infection via whitefly mediated inoculation of PepYLCItdV showed apparent symptoms just after 20 Days After Inoculation (DAI). On the other hand, PepYLCIpsV isolate showed symptoms after only 8 DAI. Hence, the PepYLCIpsV isolate was denominated pathogenic. In order to further develop of defense strategy against geminivirus for pepper chilli species it was performed a detailed analysis of its sequence characteristic. Here it was reported on genome characteristics of PepYLCIpsV, a pathogenic strain compared to PepYLCItdV and some PepYLCIV isolates from Indonesia. It was also reported on recombination analysis inferred from our isolate. The information presented here is of importance regarding pathogenicity factors in PepYLCV.

MATERIALS AND METHOD

Plant material, inoculums source and artificial inoculation: Plant materials for artificial inoculation were prepared by soaking chilli seeds in sterile water. Three seedlings each were transferred into polybags for 2 weeks to ensure the viability of plants. Watering, fertilizing and other maintaining aspects were performed according to standard procedures. All experimental units were kept in screenhouse, protected from insects potentially interfering with the experiment. All sample plants were maintained until 3 weeks after soaking. Inoculum sources were plants showing symptoms of PepYLCV infection. They were collected from an infected field in Pesisir Selatan and Tanah Datar Regency and further designated as PepYLCIpsV and PepYLCItdV, respectively. Two grams of fresh leaves were used for DNA isolation and successive PCR analysis of PepYLCIpsV for genome characteristic. The inoculated plants were incubated in the isolated screenhouse and maintained until they grew normally. Three days after plant incubation, five *Bemisia tabacci* were released and incubated together with the inoculums plants in the screenhouse. Three healthy plants were used for artificial infection. Virus infection was done by putting the three healthy plants in the surrounding of inoculated sources. Observation of symptom development was started after three DAI.

PCR technique and virus isolate typing: All standard molecular analysis procedures, such as DNA isolation, electrophoresis, PCR cocktail and sequencing procedures were done as previously described by Jamsari and Pedri[2]. Verification of PepYLCV from inoculated plant was done by PCR with universal primer pair[8] PAR1c715 and PAL1v1978. Further identification of virus genotype from each inoculums was done by using specific gene primer AC1-PSS14/TD21-WS-F and AC1-PSS14-WS-R for PepYLCIpsV strain while identification of virus PepYLCItdV strain was done by using gene specific primer pair AC1-PSS14/TD211-WS-F and AC-TD21-WS-R (Table 1).

Primer walking: The whole DNA A-like genome sequence of PepYLCIpsV was elucidated via five steps of primer walking. All primers used in this study are listed in Table 1. The universal primers PAR1c715 and PAL1v1978 were used as the start point for the primer walk. Extension of the genome coverage was done in both sense and antisense direction until whole the genome sequence was covered.

Bioinformatic analysis: Sequence analysis including editing was performed by means of BioEdit[9] version 7.2.5. Alignment for homology search at nucleotide as well as amino acid level was done with BLAST analysis at NCBI database[10]. Multi alignment at nucleotide level was done by MUSCLE available

Table 1: Primers, sequences and their expected PCR product length used for primer walking step

Primer ID	Sequence 5' - 3'	Length (Nucleotide)	Length of PCR product (bp)	Reference
PAR1c715	GATTTCTGCAGTTDATRTTYTCRTCCATCCA	31	1600	Rojas *et al.*[8]
PAL1v1978	GCATCTGCAGGCCCACATYGTCTTYCCNGT	30		
PAR 1c Compl	TGGATGGAYGARAAYATHAACTGCAGAAATC	31	1600	Present study
PAL1v1978-Compl	ACNGGRAAGACRATGTGGGCCTGCAGATGC	30		
AC1-PSS14/TD21-WS-F	GCAGTCTAAGTCAATACGTCT	21	1108	Present study
AC1-PSS14-WS-R	TGACCGAGGCATGTTTGACTC	20		
AC2-PSS14-WS-F	CATTGTTCTTAAATACTCTTA	21	423	Present study
AC2-PSS14-WS-R	ACGAAGGTCGCATTTTTTAG	20		
AC3-PSS14-WS-R	CTGTCATGGATATGGATTT	19	431	Present study
AC3-TD21/PSS14-WS-F	AAATCCATATCCATGACAG	20		
AC4-PSS14-WS-F	CGAACCTGCGTTCAAGGCTT	20	276	Present study
AC4-PSS14-WS-R	TGCAGAGACCCATCAGAATG	20		
AV1-TD21/PSS14-WS-F	GTAAAAATTATGCCGAAGCGT	21	794	Present study
AV1-TD21/PSS14-WS-R	AATGTTAACAAATTAATAAAGC	21		
AV2-PSS14-WS-F	CCACTATCTAAAGATGTGGGA	21	376	Present study
AV2-PSS14-WS-R	AAGGACCAGTGATGTTCCACG	21		

at:http://www.ebi.ac.uk/Tools/msa/muscle[11] and output data were saved ina ClustalW format. Multi alignment analysis at amino acid level was performed with clustal omega[12] available at (http://www.ebi.ac.uk/Tools/msa-/clustalo/). The ClustalW format was then converted into FASTA format on the web-based server program available at http://sequence conversion.bugaco.com/converter-/biology/sequences/). The fasta format data was subsequently used for analysis of recombination detection program (RDP)[13] version 4 beta 56 involving the sub packages GENECONV method[14], Bootscan/recscan method[15], MaxChi method[16], Chimaera method[17]: SiScan method[18] and 3Seq method[19]. Estimation of divergence time and tree time of strain emergence was analyzed by MEGA[20], version 6.0.

RESULTS AND DISCUSSION

Pathogenicity test: Eight days after artificial infection, plant samples treated with PepYLCIpsV strain already showed infection symptoms, while plants treated with PepYLCItdV showed symptoms only after twenty days of inoculation. Infection symptoms characterized by yellowish curling and stunted veins started from the shoot and young leaves (Fig. 1a, b).The successful artificial infection was further proved by the presence of expected geminivirus DNA fragment as verified by PCR technique by using universal primer PAL1v1978/PAR1c715 (Fig. 1c). Based on this result it was denominated PepYLCIpsV as a pathogenic isolate and PepYLCItdV as a mild isolate.

Genome comparison between PepYLCIpsV and PepYLCItdV: Five step chromosome walking was performed to elucidate the whole genome of DNA A-like sequence of PepYLCIpsV strain. After editing the sequence data, 2748 bases of sequence was fixed and confirmed (Fig. 2a). The length of total genome is one base shorter than its counterpart PepYLCItdV strain that was reported previously[2]. The whole genome sequence of PepYLCItdV has been deposited in the NCBI with accession number KT809346.

The additional base found in PepYLCItdV was caused by a deletion event that occurred in the Common Region (CR). The complete sequence of DNA A-like genome of PepYLCIpsV is deposited in NCBI database with accession number KT809345. The BLAST analysis in nucleotide level resulted in only 90.6% maximal homology with the existing DNA-A genome from most of pepper yellow leaf curl viruses deposited in public database for instance with AB267834 (91%), while PepYLCItdV showed up to 95% with AB267838, a PepYLCV isolated from *Ageratum* sp. in Indonesia[7]. Therefore, PepYLCIpsV could be classified as a novel PepYLCV strain. This is based on the fact, that its sequence homology is lower than 92% which is used as a threshold value for virus classification described by the International Committee on Taxonomy of Viruses (ICTV). However, the isolate could not be considered as new *Begomovirus* species, since its identity is more than 89% compared with other existing *Begomoviruses* deposited in GenBank so far[21].

Specific nucleotide sequence comparison of whole DNA-A like genome between PepYLCIpsV and PepYLCItdV in nucleotide level exhibited only 85% homology. Further gene annotation successfully identified six ORFs i.e., V1, V2, C1, C2, C3 and C4 which are similar to the previously published monopartite PepYLCV strain, PepYLCItdV[2]. All six ORFs fitted with common genes found in the DNA-A genome of bipartite *Begomoviruses* from old world hemisphere[21,22] as well as in DNA A-like genome of monopartite *Begomoviruses* which are considered to originate from new world hemisphere[23].

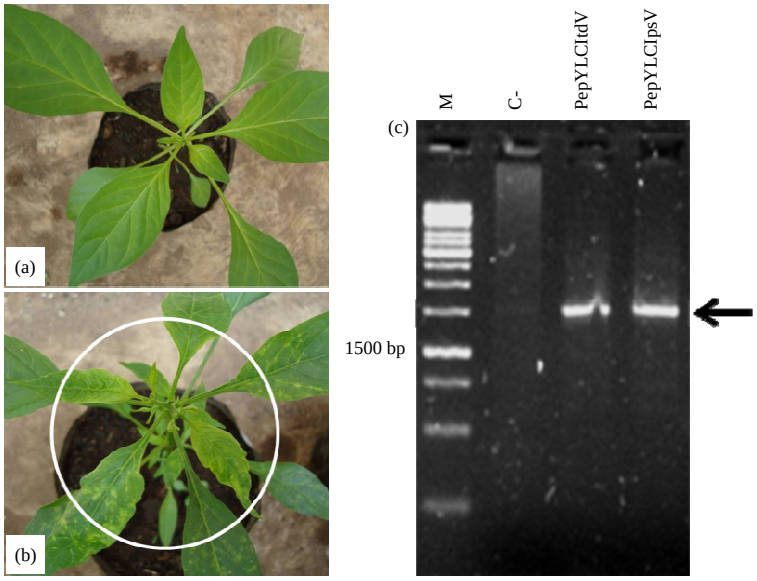

Fig. 1(a-c): Symptom shown by sample plants at 8 DAI and PCR verification, (a) Plant sample inoculated with PepYLCItdV-21 isolate, (b) Plant sample inoculated with PepYLCIpsV-14 isolate and (c) PCR verification of PepYLCV DNA from both infected with PepYLCItdV and PepYLCIpsV using universal primer PAR1c715 and PAL1v1978, exhibited a single fragment of about1600 bp in size, C: Negative control, M: 1 kb ladder

Detailed analyses were performed to compare each of all 6 ORFs found in PepYLCIpsV with the counterpart PepYLCItdV and some monopartite PepYLCV gene sequences existing in the NCBI database on the nucleotide as well as amino acid level. Four ORFs i.e., V1, V2, C2 and C3 showed homology more than 96%. The V1 and V2, gene showed homology up to 94-95%, respectively, while C2 and C3 showed homology in range of 99 and 97%, respectively on nucleotide sequence. Interestingly C1 and C4 showed homology only 79 and 68%, respectively. The DNA sequence of Common Region (CR) of PepYLCIpsV isolate consisted of 434 nucleotides in length, while TD-21 had 1 additional nucleotide. Both isolates shared only 74% similarity in that region.

Comparison in amino acid level of all six ORFs resulted 85% in average. The four ORFs V1, V2, C2 and C3 have similarity ranges from 96-99%. The highest similarity (99%) was displayedby V2, a gene encoding putative virus Coat Protein (CP), while the lowest similarity (97%) showed by V1, which is believed to be the gene responsible for the precoat protein. Linear result with that nucleotide comparison, C1 and C4 showed similarity only 77 and 41%, respectively.

In order to focus this study, it was delimited further comparison analysis for only two regions, C1/Rep and CR region. Both two regions are well known mainly to be involved in geminivirus replication.

Characteristic of common region of PepYLCIpsV: Common Region (CR), also known as intergenic region (IR/LIR) is believed to have an essential role in virus replication. Its sequence is conserved among all geminivirus members and characterized by an invariant motif TAATATTAC. Detailed analysis showed that the conserved 5′ region of almost all geminiviruses contains a replication origin characterized by a rep-binding site and hairpin structure[24,25].

Previous studies established that the hairpin structure is essential for replication[26], whereas its sequence characteristics determine the replication efficiency[26]. Moreover, sequences for nicking/joining activity[27] and direct repeat elements[28] are located in this region. Based on this, we compared our isolates by aligning this region using clustal omega[12]. The comparison along complete IR sequence showed only 64% homology. The low homology is due to sequence variation by base substitutions rather than insertion/deletion (indel) events. Both sequences differ in only one nucleotide in length i.e., 274 nucleotides for PepYLCIpsV and 275 nucleotides for PepYLCItdV. The low homology (51%) was apparently associated with the first 150 nucleotides 5′ upstream of the IR sequence where the three functional regions are located. The other 124 nucleotides in 3′ downstream exhibited 83% identity.

The stem-loop structure of both isolates of PepYLCIpsV and PepYLCItdV contains identical loop elements including

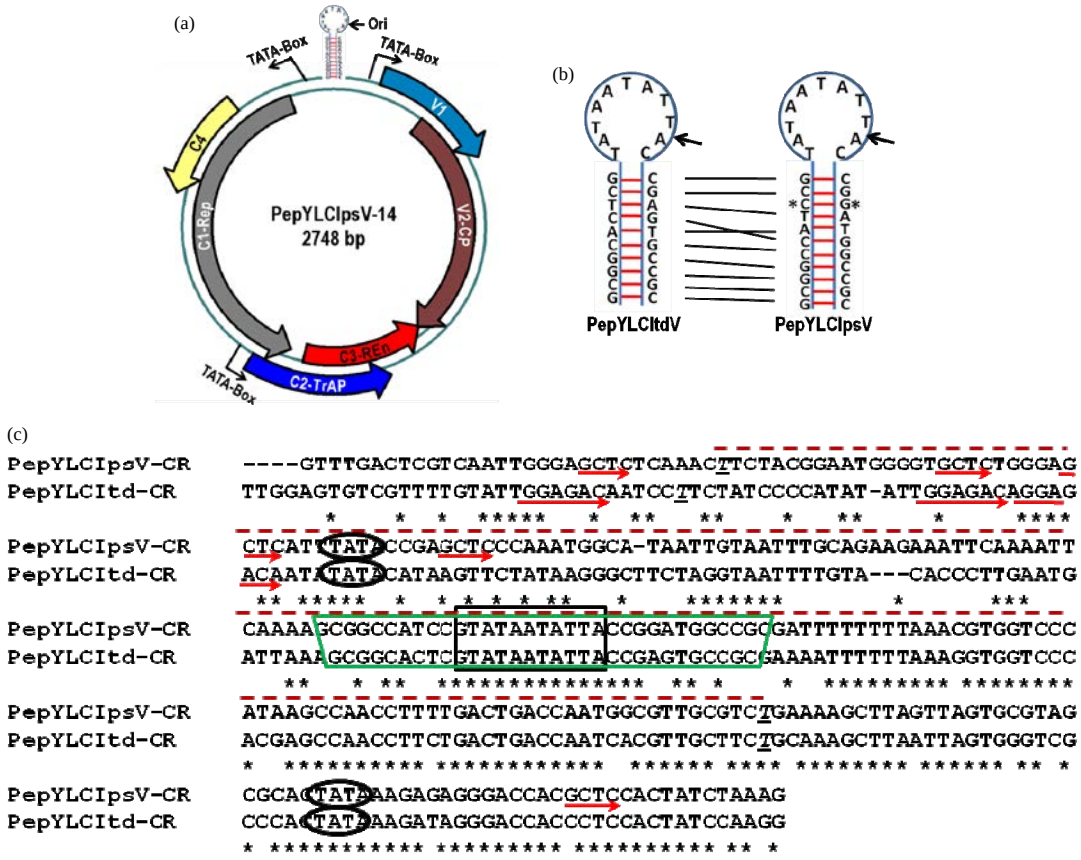

Fig. 2(a-c): Genome organization of PepYLCIpsV isolate 14 and comparison of the Intergenic Region (IR) between isolate PepYLCIpsV and PepYLCItdV, (a) All six ORFs and the origin of replication (Ori) in the stem loop structure are depicted according to their transcription orientation. Some putative promoters (TATA-box) and their orientations are also depicted, (b) Stem-loop structure, showed additional C-G bond in PepYLCIpsV and frame shift event involving A-T, C-G bonds. Nicking position between the last A with C is designated with arrowhead and (c) All functional sequences located in the intergenic region, involving: Direct repeat element (iteron) (underlined with arrowhead), TATA-box (oval), stem-loop sequences (trapezoid) and invariant sequences (rectangle) are depicted

11 nucleotides TATAATATTAC but their stem sequences are different by insertion of a C-G bond between the T-A and the C-G position in PepYLCIpsV. Additionally, a frame shift event occurred between the A-T and C-G bond (Fig. 2b). A TATA box in the 5′ upstream of stem-loop could be identified in both sequences, however, the direct repeat element (iteron) was located in different position and showed different motif patterns. The PepYLCIpsV isolate has a GCTC motif iteron with 5 repetitions. Interestingly the iteron motif in PepYLCIpsV is simpler, containing only 4 nucleotides "GCTC" and is located more dispersed. Three iterons are located in the rep-binding element region spanning between nucleotides 31-213. The other two iterons are located outside of the rep-binding region. Two of iterons located in the rep-binding element region are located 3 nucleotides upstream from the first TATA-box and are located close to each other but separated by 4 nucleotides GGGA as spacer (GCTCTGGGAGCTC) (Fig. 2c). On the other hand the PepYLCItdV isolate contains only three iterons characterized with longer repeat motif GGAGACA. Only two iterons are directly located in the rep-binding element and are located in a similar position 3 nucleotides upstream from the first TATA box (Fig. 2b). Considering the number of iterons and their distribution in the rep-binding element, we speculate that those characteristics might associate with the pathogenicity phenotype of PepYLCIpsV. Our hypothesis is based on the affinity binding probability theory, where simple nucleotide sequence motifs will bind more easily than complex motifs[29]. This will be further enhanced with an increased number of the motif. In turn, an increase the binding capability between rep-protein and CR

and thus improve virus replication efficiency is expected. However, our hypothesis still needs empirical testing. What appears to be clear is that differences in iterons are reflected indifferent pathogenic phenotypes.

Previous studies demonstrated that iterons serve for high affinity binding of the rep-protein. They have thus a crucial function in virus replication[30]. Using two strains of tomato leaf curl virus[31] demonstrated that mutations in the sequence of the repeat motifs or alteration in the arrangement of the motifs affect the ability of rep-protein to bind the DNA sequence and causes reduced accumulation of viral DNA in protoplasts. Their finding suggested that binding of rep-protein to its cognate iterons is an essential step in virus replication. Furthermore, a difference in sequence of two base pairs in the binding site of two the ToLCV-Nde strains was sufficient to affect DNA binding of the corresponding rep-protein and replication of the virus DNA in protoplasts. Based on this result we assume that this specific character of common region indeed contributes to the pathogenic phenotype of our PepYLCIpsV isolate.

Characteristic of C1/Rep gene: Nucleotide comparison of C1 gene sequence between PepYLCIpsV and PepYLCItdV showed only 79% homology, while amino acid sequence comparison revealed only 77% homology. The relative evolutionary distance between PepYLCIpsV and PepYLCItdV is largely determined by the N-terminus containing all functional domains published so far. Further detailed amino acid comparison exhibited only 60% homology of 200 amino acid in the left side of N-terminal rep-protein, while the remaining 162 amino acid in C-terminus showed relative higher homology (97%).

Nine functional domains, described previously could be identified in both isolates. Specificity determinant motif (SPD) can be found in the position 3-11 of amino acid sequence. They shared only 11% similarity, which is contributed by F-7 (Phe residue at amino acid 7). Londono *et al.*[32] concluded from tertiary structure modeling of different rep-proteins that SPD regions interact to form a small beta-sheet element that has been proposed to be critical for high-affinity DNA-binding of rep. The change of Asp10 to Asn coupled with point mutation of the 3rd and 10th nucleotide of the 13-mer binding sites in the common region could increase replication level of a mild *Begomovirus* strain[33]. Our pathogenic isolate PepYLCIpsV contains Lys 10 instead of Gln 10 in their C1 genes. Whether this plays role in pathogenicity of PepYLCIpsV requires further experiments.

Three conserved motifs, namely M-I, M-II and M-III[34], which are highly conserved among geminiviruses are also located in similar position. A slight variation was seen in M-III, whereas M-I and M-II showed 100% identity. The M-III motif characterized by a conserved sequence DVKXYXXKD, which is typical for geminiviruses infecting dicotyledonous, showed only 73% identity. Twenty seven percent of variation in M-III motif of PepYLCIpsV and PepYLCItdV is partly due to changes in amino acid 102 and 104 the flanking tyrosin Y-103 known to have catalytic activity[35]. Tyrosin Y-103 is present in both isolates. The M-III motif in both isolates was localized downstream of two amino acids of α-2 motif containing amino acid sequence SSSDVKSYIE in PepYLCIpsV and SSSDVKAYMD in PepYLCItdV.

The Geminivirus Rep Sequence (GRS) motif located at amino acid 152-176. Both isolates shared only 65% identity in amino acid level. The feature of a GRS was believed to be a conserved, essential motif characteristic of an ancient lineage of rolling-circle initiators and supports the idea, that geminiviruses may have evolved from plasmids associated with phytoplasma or algae[36]. However, its role in the RCR machinery seems not to be directly linked to replication efficiency. The putative sequence, predicted as oligomerization site in rep-protein[26] is located between position 132-178. The region covers a 21 amino acid stretch as the oligomerization major region. Interestingly, their amino acid sequences share only 52% identity.

Other conserved motif i.e., ATP/GTP binding domain with walker A and walker B motifs[37], P-loop element, RGG and RxL motif elements could be detected with 100% sequence identity (Fig. 3). Based on their identical sequences, the last three functional domains appear not to be involved in the pathogenicity phenotype of PepYLCIpsV isolate.

Origin of PepYLCIpsV isolate: The pathogenicity character of PepYLCIpsV as a monopartite strain could provide novel insights into pathogenicity factors of *Begomovirus*, particularly of PepYLCV. Better understanding of this aspect might provide valuable information for development of defense mechanism systems in plants, particularly in chilli pepper. We therefore performed clustering analysis of whole genome DNA sequence of PepYLCIpsV against complete DNA-A sequence of 34 *Begomoviruses* isolated from Indonesia, Thailand, Pakistan, India and Bangladesh together with β-satellite sequence isolated from PepYLCItdV.

As predicted, all *Begomoviruses* collected from Indonesia and Thailand clustered into one group, while isolates from

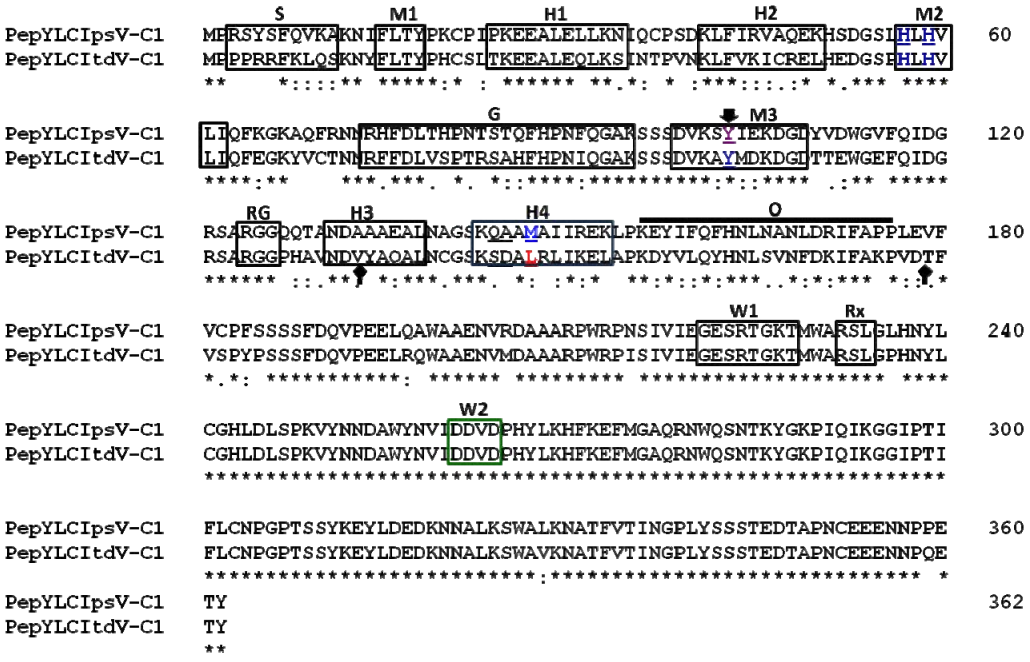

Fig. 3: Sequence alignment of rep-protein amino acid sequence of PepYLCIpsV and PepYLCItdV, S: Specifity determinant (SPD), MI-III: Motif I, II and III, H1-4: α-helices, G: GRS motif ⬇: Tyrosine residue (Y), cleavage site, RG: RGG motif, O: Oligomerization major region, W1-2: ATP/GTP binding domain, P-loop, Rx: RxL motif, ⬆: Start and end position of oligomerization region

India, Pakistan and Bangladesh were clustered together into a different group (Fig. 4a). The Indonesia and Thailand group consisted of 13 isolates, which are 7 of pepper yellow leaf curl viruses and 6 of tomato leaf curl viruses infecting tomatoe, eggplant and chilli pepper. The clustering result is in accordance with spatial distribution, where the 7 isolates belong to the West side of Indonesia (Sumatra and Java) and the rest of 6 isolates belong to the North side of Indonesia (Sulawesi and Thailand). Moreover, the Indonesia and Thailand groups can be sub divided into three sub groups containing the FJ sub group (FJ237618 and FJ237620, tomato leaf curl Sulawesi virus isolates), KF (KF446667 and KF446673, tomato yellow leaf curl kanchanaburi virus isolate), AF (AF511529 and AF511530, tomato yellow leaf curl kanchanaburi virus-[Thailand Kan1]), AB (AB267834, AB267836, AB267838), DQ (DQ083764, DQ083765) and KT (PepYLCIpsV and PepYLCItdV). The last sub group contains 7 isolates, which are originated from Java (AB and DQ) and Sumatra.

Interestingly, among *Begomoviruses* of sub group three (AB267834, AB267836, AB267838, DQ083764, DQ083765, PepYLCIpsV[KT809345] and PepYLCItdV[KT809-346]), this strain PepYLCIpsV is separated from PepYLCItdV, even though both isolates originated from a similar region (West Sumatra) and similar host (*Capsicum annuum*). The β-sequence which is isolated from PepYLCItdV, clusters into a very distant position with all sequences compared, indicating that the beta sequence of PepYLCItdV has evolved from a different origin of the cognate DNA-A like genome. Our finding is consistent with the previously published β-satellite sequences which commonly diverged from their cognate DNA-A like genome[38]. Hence, for the most cases this result is in accordance with the previous recombination hypothesis that the PepYLCItdV strain might have evolved from any recombination process of the existing isolates in the established population.

Further detailed clustering analysis from each gene and domain of 13 selected *Begomoviruses* in the second main group (Indonesia and Thailand) showed that only two domains (C1 also involving C4 gene and CR region) that could separate PepYLCIpsV from other six *Begomoviruses* originated from Sumatra and Java (Fig. 4b, c), while the other genes (C2, C3, V1 and V2) clustered all *Begomoviruses* together in similar group.

In order to identify recombination event and to infer the origin of PepYLCIpsV as pathogenic isolate, we performed recombination detection program analysis. For this purpose, we run RDP version 4 beta 56 with all 13 *Begomoviruses* used for clustering analysis with the complete sequence of their DNA-A like genomes. Eight unique events with

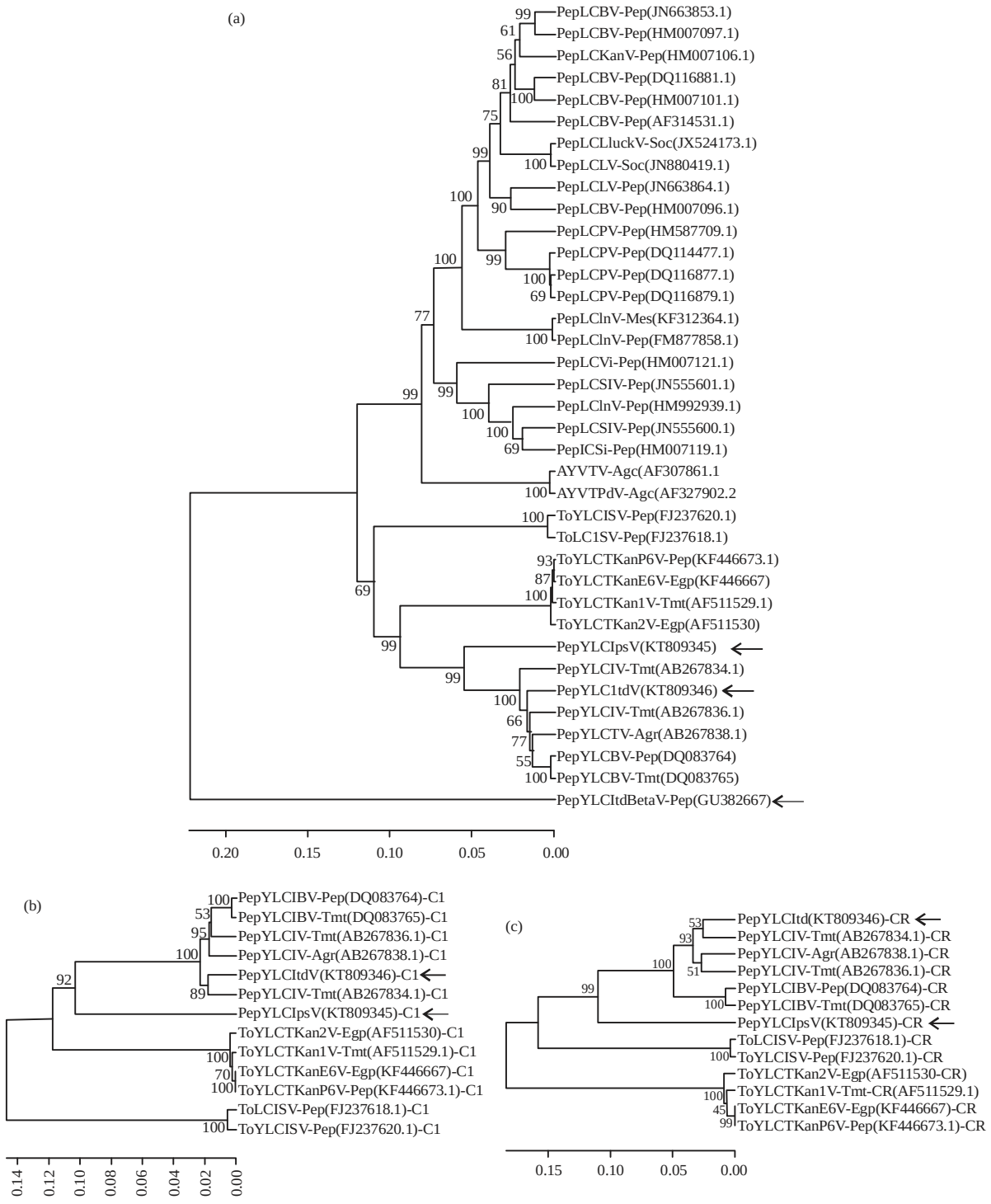

Fig. 4(a-c): Clustering analysis of whole sequence of DNA-A genome (for bipartite) and DNA-A like genome (for monopartite) and specific domain of *Begomoviruses* isolated from different main countries in Asia, (a) Clustering of whole genome and (b, c) Clustering of Common Region (CR) and rep-gene (C1), respectively

29 recombination signals could be detected by three (RDP, GENECONV and MaxChi) sub package programs integrated in the RDP 4 beta 5.6. The three algorithms showed 4-5 unique events and 5-6 recombination signals among 13 isolates analyzed with average p-value ranged between 7.558×10^{-13} to 1.646×10^{-44}.

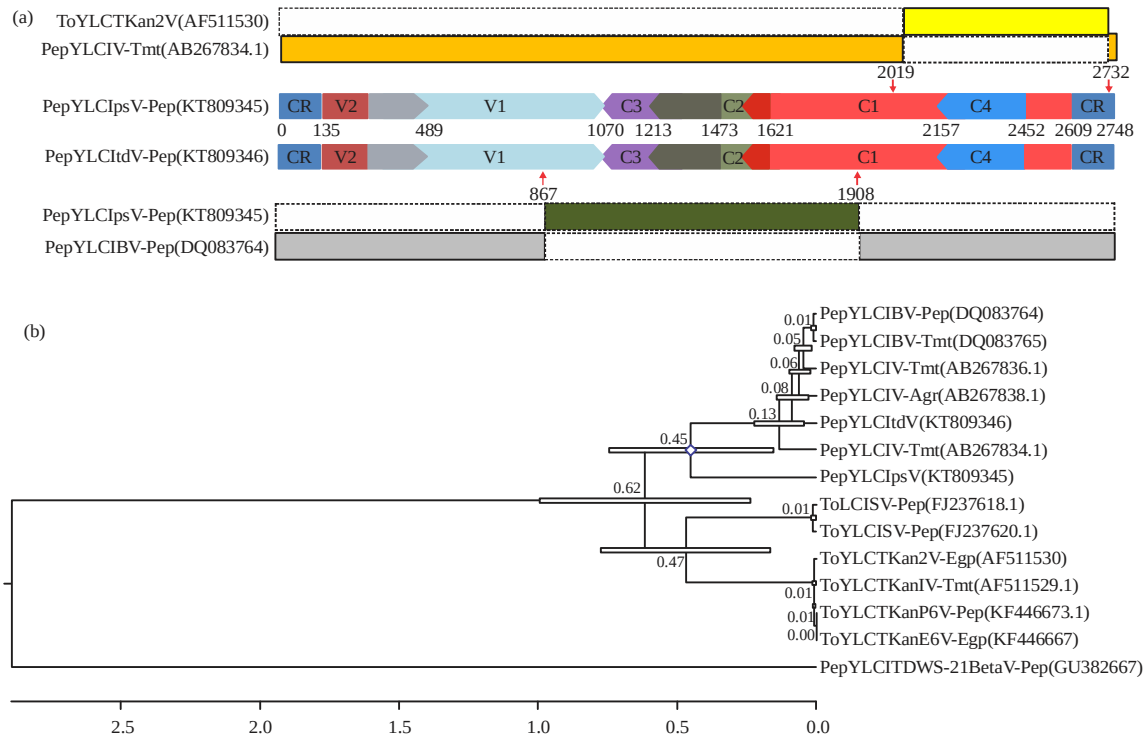

Fig. 5(a-b): Diagram representing (a) Breakpoint positions in both the recombinant PepYLCIpsV[KT809346] and the pathogenic PepYLCItdV[KT809346] isolates and (b) Divergence time of 13 isolates including β-sequence [GU382667] isolated from PepYLCItdV[KT809346] isolate. Begin and end of breakpoint positions are depicted with numberred arrowhead. Numbers in between of both isolates represented nucleotide position in each sequence

Setting the PepYLCIpsV as recombinant we detected that PepYLCIV-Tmt[AB267834.1] could be inferred as minor parent, while ToYLCVTKan2V-Egp[AF511530] which was isolated from Kanchanaburi, Thailandwas inferred as major parent (Fig. 5). Interestingly, RDP analysis also showed that PepYLCItdV[KT809346] was indicated as descendant of PepYLCIpsV[KT809345] (Fig. 5) and PepYLCIBV-Pep[DQ083764]. The later isolate was foundin chilli-pepper from Bogor (Java).

Breakpoint position in the whole genome of PepYLCIpsV[KT809345] started at base 2732 and ended at base 2019. Both positions were estimated with 99% of confidence. The break point positions covered from the N-terminal region of C1/rep gene until-7 from the upstream position of stem-loop structure. The low homology in nucleotide sequences between PepYLCIpsV and PepYLCItdV was caused by the intraspecies recombination involving pepper yellow leaf curl virus from tomatoe plant (PepYLCIV-Tmt[AB267834.1]) and tomatoe yellow leaf curl virus from eggplant (ToYLCVTKan2V-Egp[AF511530]). Their relative far distance in species relationship contributed not only to form a divergent species or strain indicating by low homology in nucleotide sequencebut also phenotypically shows a more pathogenic character.

Clustering analysis of six ORFs/genes from both PepYLCItdV and PepYLCIpsV showed that only the C1/rep and CR-domain could separate both isolates into different clusters (Fig. 4b, c). The data are consistent with the clustering results from above whole genome analysis, meaning that both isolates are divergent in both two regions. Based on this data we were interested to analyze these domains for their evolution time.

Time tree analysis from MEGA6 software based on maximum likelihood with 1000 bootstrap replication was shown that PepYLCIpsV[KT809345] existed 0.15 time unit earlier than the isolate PepYLCItdV[KT809346]. Its existence was similar with its ancestral PepYLCIV-Tmt[AB267834.1] and ToYLCVTKan2V-Egp[AF511530]. Mixed infection together between PepYLCIV-Tmt[AB267834.1] and ToYLCVTKan2V-Egp[AF511530] at 0.15 unit time might have produced the new recombinant isolate PepYLCIpsV[KT809345] that uses chilli pepper as a major host. In this case the newly produced recombinant isolate is more pathogenic. On the other hand, PepYLCItdV[KT809346] is milder and emerged later than its

ancestor. However, it has become more dominant in the region. Our field observation from 2009, indicated that population of strain monopartite PepYLCItdV[KT809346] in West Sumatera was dominant over PepYLCIpsV[KT809345][2]. The PepYLCIpsV[KT809345] strain could only be found in Pesisir Selatan Regency (lowland) but could not be found in other regions in West Sumatra.

CONCLUSION

The DNAA-like genome of a naturally pathogenic isolate PepYLCIpsV[KT809345] showed different genome structure particularly in C1/rep domain and CR domain with the currently published PepYLCV. Based on our characterization we conclude that pathogenic determining factors are located on the C1/rep and CR-domain. The recombination detection program indicated that the pathogenic isolate PepYLCIpsV[KT809345] might be derived from ancestors of PepYLCIV-Tmt[AB267834.1] and ToYLCVTKan2V-Egp [AF511530], while the milder isolate PepYLCItdV[KT809346] was a descendant from pathogenic PepYLCIpsV[KT809345] and PepYLCIBV-Pep[DQ083764].

SIGNIFICANCE STATEMENTS

Better understanding to agressivity and pathogenicity determining factors of Pepper Yellow Leaf Curl Virus (PepYLCV) could help researcher to elucidate a mechanism that in turn could enable to develop any defence mechanism of the host. In the submitted manuscript, we describe a detailed genome analysis of a pathogenic PepYLCV isolate from West Sumatra which is known as one of the centre of chilli-pepper in indonesia. According to our knowledge so far, informations described in the submitted manuscript is the first publication concerning pathogenic determining factor especially from PepYLCV infecting pepper-chilli. Therefore, the informations described here are novel and should complement the available information regarding the pathogenic determining factor in common *Begomoviruses*.

ACKNOWLEDGMENTS

We kindly thank the Directorate General of Higher Education for supporting this research via International Research Collaboration and Publication Grant fiscal year of 2015, contract no: 030/SP2H/PL/DITLITABMAS/I/2015. Part of data in this publication was also supported by Indonesian Agency for Agricultural Research and Development, Ministry of Agriculture.

REFERENCES

1. Briddon, R.W., J.K. Brown, E. Moriones, J. Stanley, M. Zerbini, X. Zhou and C.M. Fauquet, 2008. Recommendations for the classification and nomenclature of the DNA-β satellites of begomoviruses. Arch. Virol., 153: 763-781.
2. Jamsari, J. and J. Pedri, 2013. Complete nucleotide sequence of DNA A-like genome and DNA-β of monopartite pepper yellow leaf curl virus, a dominant *Begomovirus* infecting *Capsicum annuum* in West Sumatera Indonesia. Asian J. Plant Pathol., 7: 1-14.
3. Galvao, R.M., A.C. Mariano, D.F. Luz, P.F. Alfenas and E.C. Andrade *et al*., 2003. A naturally occurring recombinant DNA-A of a typical bipartite *Begomovirus* does not require the cognate DNA-B to infect *Nicotiana benthamiana* systemically. J. Gen. Virol., 84: 715-726.
4. Noueiry, A.O., W.J. Lucas and R.L. Gilbertson, 1994. Two proteins of a plant DNA virus coordinate nuclear and plasmodesmal transport. Cell, 76: 925-932.
5. Saeed, M., S.A.K. Behjatnia, S. Mansoor, Y. Zafar, S. Hasnain and M.A. Rezaian, 2005. A single complementary-sense transcript of a geminiviral DNA β satellite is determinant of pathogenicity. Mol. Plant Microbe Interact., 18: 7-14.
6. Briddon, R.W., S. Mansoor, I.D. Bedford, M.S. Pinner and K. Saunders *et al*., 2001. Identification of DNA components required for induction of cotton leaf curl disease. Virology, 285: 234-243.
7. Sakata, J.J., Y. Shibuya, P. Sharma and M. Ikegami, 2008. Strains of a new bipartite *Begomovirus*, pepper yellow leaf curl Indonesia virus, in leaf-curl-diseased tomato and yellow-vein-diseased ageratum in Indonesia. Arch. Virol., 153: 2307-2313.
8. Rojas, M.R., R.L. Gilbertson, D.R. Russell and D.P. Maxwell, 1993. Use of degenerate primers in the polymerase chain reaction to detect whitefly-transmitted geminiviruses. Plant Dis., 77: 340-347.
9. Hall, T.A., 1999. BioEdit: A user-friendly biological sequence alignment editor and analysis program for windows 95/98/NT. Nucleic Acid Symp. Ser., 41: 95-98.
10. Altschul, S.F., W. Gish, W. Miller, E.W. Myers and D.J. Lipman, 1990. Basic local alignment search tool. J. Mol. Biol., 215: 403-410.
11. Edgar, R.C., 2004. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. Nucl. Acids Res., 32: 1792-1797.
12. Sievers, F., A. Wilm, D. Dineen, T.J. Gibson and K. Karplus *et al*., 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using clustal omega. Syst. Biol., 7: 539-544.
13. Martin, D. and E. Rybicki, 2000. RDP: Detection of recombination amongst aligned sequences. Bioinformatics, 16: 562-563.

14. Padidam, M., S. Sawyer and C.M. Fauquet, 1999. Possible emergence of new geminiviruses by frequent recombination. Virology, 265: 218-225.
15. Martin, D.P., D. Posada, K.A. Crandall and C. Williamson, 2005. A modified Bootscan algorithm for automated identification of recombinant sequences and recombination breakpoints. AIDS Res. Hum. Retroviruses, 21: 98-102.
16. Smith, J.M., 1992. Analyzing the mosaic structure of genes. J. Mol. Evol., 34: 126-129.
17. Posada, D. and K.A. Crandall, 2001. Evaluation of methods for detecting recombination from DNA sequences: Computer simulations. Proc. Nat. Acad. Sci. USA., 98: 13575-13762.
18. Gibbs, M.J., J.S. Armstrong and A.J. Gibbs, 2000. Sister-scanning: A Monte Carlo procedure for assessing signals in recombinant sequences. Bioinformatics, 16: 573-582.
19. Boni, M.F., D. Posada and M.W. Feldman, 2007. An exact nonparametric method for inferring mosaic structure in sequence triplets. Genetics, 176: 1035-1047.
20. Tamura, K., G. Stecher, D. Peterson, A. Filipski and S. Kumar, 2013. MEGA6: Molecular evolutionary genetics analysis version 6.0. Mol. Biol. Evol., 30: 2725-2729.
21. Huang, J.F., T. Jiang and X.P. Zhou, 2006. Molecular characterization of begomoviruses infecting *Ludwigia hyssopifolia*. J. Plant Pathol., 88: 83-88.
22. Melgarejo, T.A., T. Kon, M.R. Rojas, L. Paz-Carrasco, F.M. Zerbini and R.L. Gilbertson, 2013. Characterization of a new world monopartite *Begomovirus* causing leaf curl disease of tomato in Ecuador and Peru reveals a new direction in *Geminivirus* evolution. J. Virol., 87: 5397-5413.
23. Pandey, P., S. Mukhopadhya, A.R. Naqvi, S.K. Mukherjee, G.S. Shekhawat and N.R. Choudhury, 2010. Molecular characterization of two distinct monopartite begomoviruses infecting tomato in India. BMC Virol. J., Vol. 7. 10.1186/1743-422X-7-337.
24. Bisaro, D., 1996. Geminivirus DNA Replication. In: DNA Replication in Eukaryotic Cells, Depamphilis, M.L. (Ed.). Cold Spring Harbor Laboratory Press, New York USA., pp: 833-854.
25. Boukhatem, N., S. Jdaini, Y. Muhovsky, J.M. Jacquemin, C.L. del Rincone, M.J. Diez and A. Bouali, 2008. Molecular characterisation of Tomato yellow leaf curl virus Alm [Ma:Bk:02] in Morocco: Complete sequence and genome organisation. J. Plant Pathol., 90: 109-112.
26. Orozco, B.M. and L. Hanley-Bowdoin, 1996. A DNA structure is required for geminivirus replication origin function. J. Virol., 70: 148-158.
27. Steinfeldt, T., T. Finsterbusch and A. Mankertz, 2006. Demonstration of nicking/joining activity at the origin of DNA replication associated with the rep and rep proteins of porcine circovirus type 1. J. Virol., 80: 6225-6234.
28. Behjatnia, S.A.A., I.B. Dry and M.A. Rezaian, 1998. Identification of the replication-associated protein binding domain within the intergenic region of tomato leaf curl geminivirus. Nucl. Acid Res., 26: 925-931.
29. Levo, M. and E. Segal, 2014. In pursuit of design principles of regulatory sequences. Nat. Rev. Genet., 15: 453-468.
30. Chatterji, A., R.N. Beachy and C.M. Fauquet, 2001. Expression of the oligomerization domain of the replication-associated protein (Rep) of Tomato leaf curl New Delhi virus interferes with DNA accumulation of heterologous geminiviruses. J. Biol. Chem., 276: 25631-25638.
31. Chatterji, A., U. Chatterji, R.N. Beachy and C.M. Fauquet, 2000. Sequence parameters that determine specificity of binding of the replication-associated protein to its cognate site in two strains of Tomato leaf curl virus-New Delhi. Virology, 273: 341-350.
32. Londono, A., L. Riego-Ruiz and G.R. Arguello-Astorga, 2010. DNA-binding specificity determinants of replication proteins encoded by eukaryotic ssDNA viruses are adjacent to widely separated RCR conserved motifs. Arch. Virol., 115: 1033-1046.
33. Chatterji, A., M. Padidam, R.N. Beachy and C.M. Fauquet, 1999. Identification of replication specificity determinants in two strains of Tomato leaf curl virus from New Delhi. J. Virol., 73: 5481-5489.
34. Orozco, B.M. and L. Hanley-Bowdoin, 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.
35. Laufs, J., S. Schumacher, N. Geisler, I. Jupin and B. Gronenborn, 1995. Identification of the nicking tyrosine of geminivirus Rep protein. FEBS Lett., 377: 258-262.
36. Nash, T.E., M.B. Dallas, M.I. Reyes, G.K. Buhrman, J.T. Ascencio-Ibanez and L. Hanley-Bowdoin, 2011. Functional analysis of a novel motif conserved across geminivirus Rep proteins. J. Virol., 85: 1182-1192.
37. Desbiez, C., C. David, A. Mettouchi, J. Laufs and B. Gronenborn, 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.
38. Briddon, R.W., S.E. Bull, I. Amin, A.M. Idris and S. Mansoor *et al*., 2003. Diversity of DNA β, a satellite molecule associated with some monopartite *Begomoviruses*. Virology, 312: 106-121.