ISSN 2348-6201

In Silico Characterization of Ageratum Enation Virus, Ageratum Leaf Curl Betasatellite and Marigold Leaf Curl Alphasatellite Infecting an Important Ornamental Plant Marigold (Tagetes Patula) in Indian Subcontinent.

Avinash Marwal[1], Rajneesh Prajapat[2], R K Gaur[3]
[1] Department of Science, Faculty of Arts, Science and Commerce, Mody Institute of Technology and Science
marwal_avinash@yahoo.co.in
[2]Department of Science, Faculty of Arts, Science and Commerce, Mody Institute of Technology and Science
prajapat.rajneesh@gmail.com
[3]Department of Science, Faculty of Arts, Science and Commerce, Mody Institute of Technology and Science
gaurrajarshi@hotmail.com
Avinash Marwal and Rajneesh Prajapat have contributed equally to the present research

ABSTRACT

Ornamental plants act as an alternate host of begomoviruses and its associated satellite molecules in the absence of main crop. Marigold (*Tagetes patula*) an important ornamental plant, widely cultivated in India were found infected with three begomovirus components i.e. *Ageratum enation virus* (AEV: KC589699), *Ageratum leaf curl betasatellite* (ALCB: KC589700) and *Marigold leaf curl alphasatellite* (MLCuA: KC206078). This study supported evidence that virus associated with Marigold was studied through *In Silico* analysis for in depth study. The phylogenetic studies revealed closeness to other virus prominent in other neighbouring countries and can be replicated by other begomovirus species which increases the possibility of recombination and reassortment events.

Indexing terms/Keywords

Ornamental, Marigold, Begomovirus components, *In Silico*.

Academic Discipline And Sub-Disciplines

Plant Computational Biology.

Subject Classification

Virology and Bioinformatics.

Type (Method/approach)

Research Article on *In Silico* Characterization.

INTRODUCTION

Begomovirus are an outsized varied family of plant viruses [1] which infects an expansive assortment of plants such as ornamentals, weeds and crops and causes a noteworthy loss to Agriculture and Horticulture worldwide [2]. Ornamental plants are extensively scattered worldwide and have high environmental adaptability. Ornamentals are considered as a foundation of new viruses and reservoirs of unidentified economically imperative viruses but are often neglected during diversity study [3]. Many scientific reports have demonstrated that ornamental plants serve as reservoir or alternative hosts for begomovirus survival [4] and spread in the absence of the main crops [5]. Thus, there is a pressing need for additional information on the diversity and distribution of begomovirus in ornamental plants.

Marigold (*Tagetes patula*) species belonging to family *Asteraceae* are most common in plant kingdom, used for cosmetic preparation, medicines and most importantly as ornamentals [6]. It is found in different colors and different fragrance. Yellow color is most common. Flowers are mainly used for all these purpose by the extraction process. The leaves are reported to be effective against piles, kidney problem, muscular pain, ulcers and wounds. The flower is useful in fever, epileptic fits (Ayurveda), astringent, carminative, stomachic, scabies and liver complaints and is also employed in disease of eyes [7]. It shows different pharmacological activities like anti-bacterial activity, hepatoprotective activity, insecticidal activity, mosquitocidal activity, nematicidal activity, antioxidant and analgesic activity [8].

We have earlier molecularly characterized three begomovirus components infecting Marigold plant i.e. *Ageratum enation virus* (AEV: KC589699), *Ageratum leaf curl betasatellite* (ALCB: KC589700) and *Marigold leaf curl alphasatellite* (MLCuA: KC206078). All the three components were successfully cloned and sequenced in our earlier report [9]. Here we are presenting and highlighting the In Silico characterization of the three components for in depth study.

MATERIALS AND METHODS

Identifying Conserved Domain and Genome organization

For better understanding of genomic organization of begomovirus NCBI Sequence Viewer (http://www.ncbi.nlm.nih.gov/projects/sviewer/) was used which displays the graphical presentation of Nucleotide and Protein sequence. NCBI Genome Workbench is an integrated application for viewing and analyzing sequence data. With Genome Workbench, you can view data in publically available sequence databases at NCBI, and mix this data with your own private data. Genome Workbench can display sequence data in many ways, including graphical sequence views, various alignment views, phylogenetic tree views, and tabular views of data. It can also align your private data to data in public databases, display your data in the context of public data, and retrieve BLAST results.

Genome Workbench is built on the NCBI C++ ToolKit and uses cross-platform APIs for graphics. It runs on your local machine, and is available for Windows 2000/XP, Linux, MacOS X, and various flavors of UNIX. NCBI Genome Workbench is an integrated application for viewing and analyzing sequence data. Genome Workbench was developed entirely in-house at NCBI and makes use of the NCBI C++ ToolKit. The C++ ToolKit provides a convenient and flexible cross-platform API for managing system internals, database connections, network sockets, and the NCBI data model. In addition, the C++ ToolKit provides the Object Manager, which abstracts handling of sequences and sequence-related objects.

Phylogenetic Analysis

The sequence of begomovirus genomic components was aligned in Clustal-W sequence alignment program [10] using IUB matrix for DNA alignments in the Molecular Evolutionary Genetics Analysis Program (MEGA) version 4.0 [11]. The MEGA 4.0 software includes a unique facility to generate captions, written in figure legend format, in order to provide natural language descriptions of the models and methods used in the analyses. This facility aims to promote a better understanding of the underlying assumptions used in analyses, and of the results generated. Another feature is the Maximum Composite Likelihood (MCL) method for estimating evolutionary distances between all pairs of sequences simultaneously, with and without incorporating rate variation among sites and substitution pattern heterogeneities among lineages.

The FASTA format of NCBI reported begomovirus isolates and its homologous sequences were retrieved and subjected to Neighbor-Joining (NJ) construction. MEGA4.0 detects and analyzes the phylogenetic relationships among a set of aligned DNA sequences. Neighbor-Joining (NJ) analysis was carried out using Maximum Composite Likelihood model with uniform rates among the sites, the 500 bootstraps replicates were used to evaluate the significance of generated tree.

This MCL method also can be used to estimate transition / transversion bias and nucleotide substitution pattern without knowledge of the phylogenetic tree. This new version is a native 32-bit Windows application with multi-threading and multi-user supports, and it is also available to run in a Linux desktop environment (via the Wine compatibility layer) and on Intel-based Macintosh computers under the Parallels program [12].

Construction of Sequence Matrix and Entropy Plot

Pairwise BLAST is used to calculate similarity, but its limitations are that only two sequences may be analyzed at one time and percent similarity/identity is based on local alignment – not global alignment. Nucleotide sequence and amino acid identities between begomovirus infecting ornamental plants and selected begomoviruses were analyzed by MatGAT

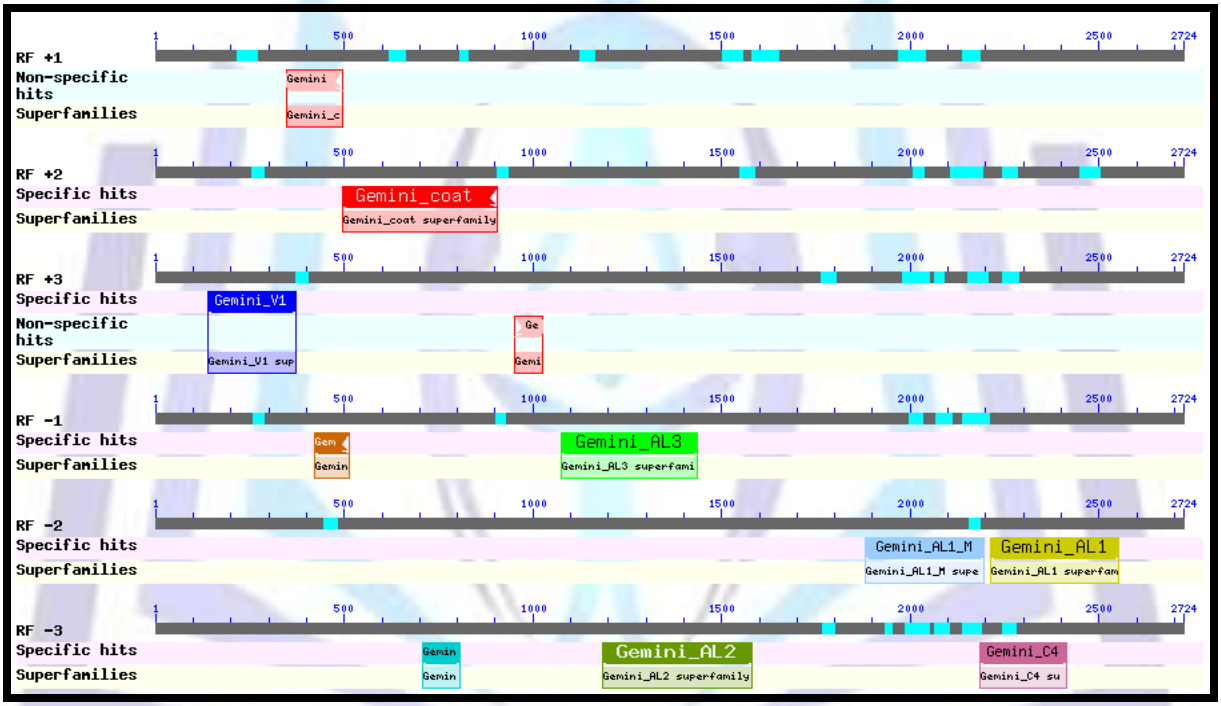
software version 2.01. MatGAT (Matrix Global Alignment Tool) is a simple, easy to use similarity/identity matrix generator that calculates the similarity and identity between every pair of sequences in a given data set without requiring pre-alignment of the data. The program performs a series of pairwise alignments using the Myers and Miller global alignment algorithm, calculates similarity and identity, and then places the results in a distance matrix [13].

Entropy power explains the sequence mutates by the function of entropy plot using BIOEDIT. If there are peaks above value 1 it confirms high variable regions. Such variations are responsible for recombination. So entropy power can be used to screen the variable region in the begomovirus genome [14].

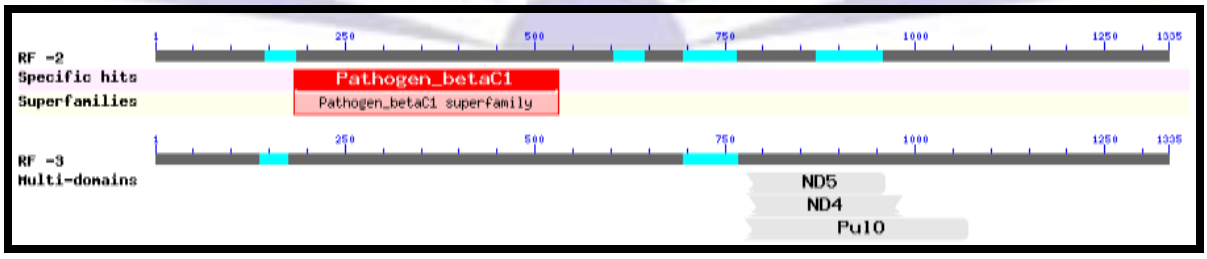
RESULTS AND DISCUSSION

Conserved domains were initially described as stable or autonomously folding units of protein structure, inspired by first results from the experimental characterization of protein three-dimensional (3D) structure. This definition of protein domains tends to coincide remarkably often with what has emerged from systematic analyses of sequence data—the characterization of protein domains as units of molecular evolution [15].

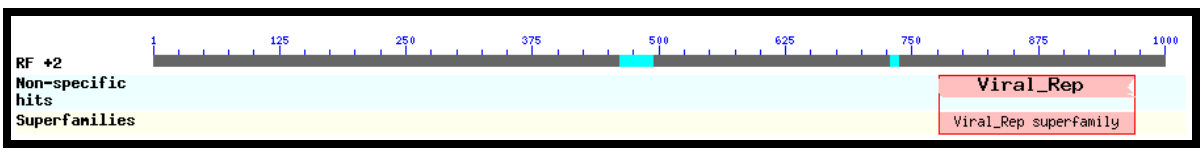
Therefore Geminivirus Conserved Domains of various proteins were detected while analyzing the sequences obtained from infected Marigold plants (Figure 1) consisting of *Ageratum enation virus* (AEV: KC589699), *Ageratum leaf curl betasatellite* (ALCB: KC589700) and *Marigold leaf curl alphasatellite* (MLCuA: KC206078).



(a)



(b)



(c)

Figure: 1. Conserved Domains identified in (a) *Ageratum enation virus* (AEV: KC589699), (b) *Ageratum leaf curl betasatellite* (ALCB: KC589700) and (c) *Marigold leaf curl alphasatellite* (MLCuA: KC206078) infecting Marigold. Well characterized Conserved domain annotation on a nucleotide sequence. Shown here is the view generated by the CD-Search tool, using pre-calculated alignment information. The view is divided into graphical summary detailing the individual matches. The query sequence coordinates are indicated on a gray bar in the graphical summary. 'Specific hits' to NCBI-curated domain models are positioned in a separate area below the query sequence, with corresponding balloons rendered in saturated colors. The extent of the best-scoring hit for a region on the query also determines the annotation with the corresponding conserved domain 'Superfamily'. 'Superfamilies' are positioned in the area below the 'Specific hits', and together these are enclosed in boxes to indicate superfamily membership of the NCBI-curated models. 'Non-specific hits' and 'Superfamily' balloons are rendered in pastel colors, with each superfamily being assigned a separate color. Only the best-ranked non-overlapping multi-domain models are shown. Sites are mapped from the highest ranked model only, and they are colored according to their source.

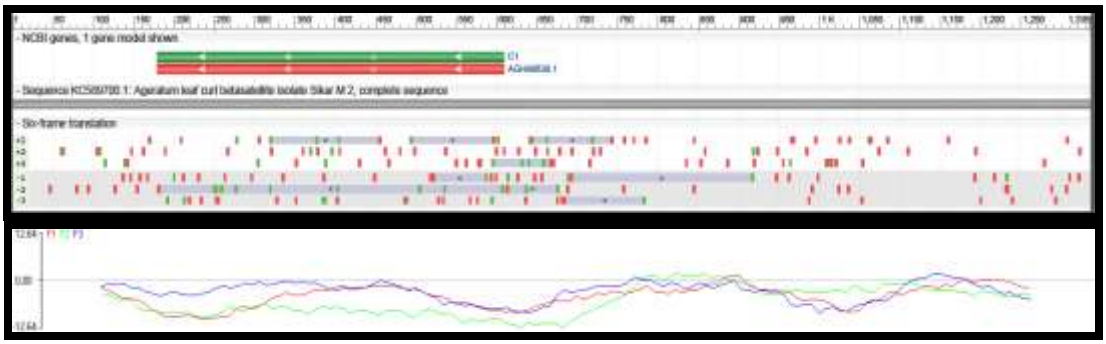
A total of 11 Conserved Domains was detected in six frames of the nucleotide sequence of *Ageratum enation virus* (AEV: KC589699). This information helps in proposing the position of genes in the begomovirus which are explained later. The *Ageratum leaf curl betasatellite* (ALCB: KC589700) sequenced from Marigold plants have a putative conserved domain of the geminivirus Pathogen_beta_C1 family which encodes for symptoms determining protein, responsible for typical begomovirus disease host plants. The betasatellite is a circular single stranded DNA and has a putative beta C1genes on the complementary- sense strand [16].

The Conserved Domain of Viral_Rep_superfamily was found in the *Marigold leaf curl alphasatellite* (MLCuA: KC206078) component, which are satellite-like circular ssDNA molecule. They encode a single gene, a rolling circle replication initiator protein and are capable of autonomous replication in plant cells. Closely related to the replication associated protein encoding components of nanoviruses, from which they are believed to have evolved, they require a helper begomovirus for movement within and between plants [17, 18]. Alphasatellite components are phenotypically silent; playing no part in the symptoms of the complex and their precise function remains unclear [19].

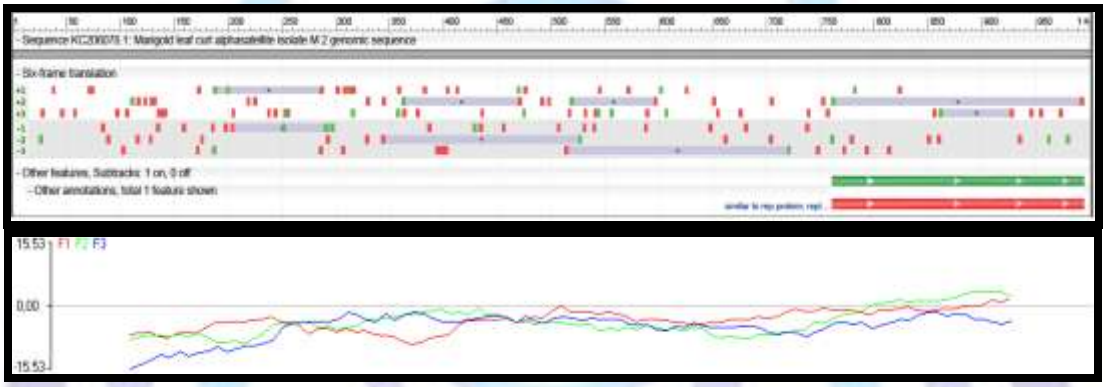
The majority of protein domain models collected in databases such as Pfam (Protein family). Conserved Domain models curated by NCBI often carry annotation of functional sites. These are recorded as co-ordinates on the MSA (Multiple Sequence Analysis) and resulting position-specific score matrices, and are mapped to protein query sequences via the CD-Search service [20]. The annotation of sequences with the location of domains is a common practice in the analysis of sequence data. The identification of a conserved domain footprint may be the only clue towards cellular or molecular function of a protein, as it indicates local or partial similarity to other proteins, some of which may have been characterized experimentally [21].

Based on the above observed Conserved Domains it becomes easy to identify the possible predicted ORFs in the begomovirus as well as in their satellites molecules infecting Marigold plants. The DNA-A of isolate M 2 consisted of 2,724 nucleotides (GenBank accession no. KC589699). The DNA-A (Figure 2a) contains six predicted open reading frames (ORFs AV1, AV2, AC1, AC2, AC3 and AC4) that are conserved among begomoviral DNA-A [22].

(a)



(b)



(c)

Figure: 2. Graphical sequence view of the *Ageratum enation virus* (AEV: KC589699), *Ageratum leaf curl betasatellite* (ALCB: KC589700) and *Marigold leaf curl alphasatellite* (MLCuA: KC206078), depicting the positions of ORF in the genome. Below the sequence is give all the possible six reading frames and predicted ORFs consisting of start codons, highlighted in green and the stop codons are highlighted in red, out of which the best possible one was selected after identifying the Conserved Domains. It further demonstrates the graphical representation of codon usage in three reading frames. Codon usage in first reading frame is coloured red; subsequently codon usage in reding frame second and third is highlighted as green and blue respectively. Codon usage reflects the probability that a sequence is coding in a given reading frame. Codon preference takes only the uneven use of synonymous codons into account. Codon usage is calculated as the log-likelihood for a sequence to be coding, based a codon usage table which contains the frequencies of codons in coding regions of a species.

The detailed results of individual ORF are depicted in Table 1 along with their predicted highest amino acid sequence identity during NCBI database search. An intergenic region (IR) between the AV2 gene and the AC1 gene contain the controlling elements for virus replication and transcription, including a stem-loop structure with an 11-nucleotide stem and an 11-nucleotide loop containing the conserved nonanucleotide sequence TAATATTAC in the loop. This conserved sequence contains the nick site for the initiation of viral strand DNA rolling-circle replication [23].

The IR sequence analysis of the begomovirus infecting Marigold in India with related isolates from the database sequences were carried out. The conserved region for TAATATTAC is boxed in blue. The so-called "variable region" is boxed in purple. The conserved region is boxed in orange and the iterons are boxed in brown. Whereas the TATA box is highlighted in the red box. Virus sequence from Marigold is written at the top (Figure 3). The accession numbers of the sequences from GenBank are indicated on the left.

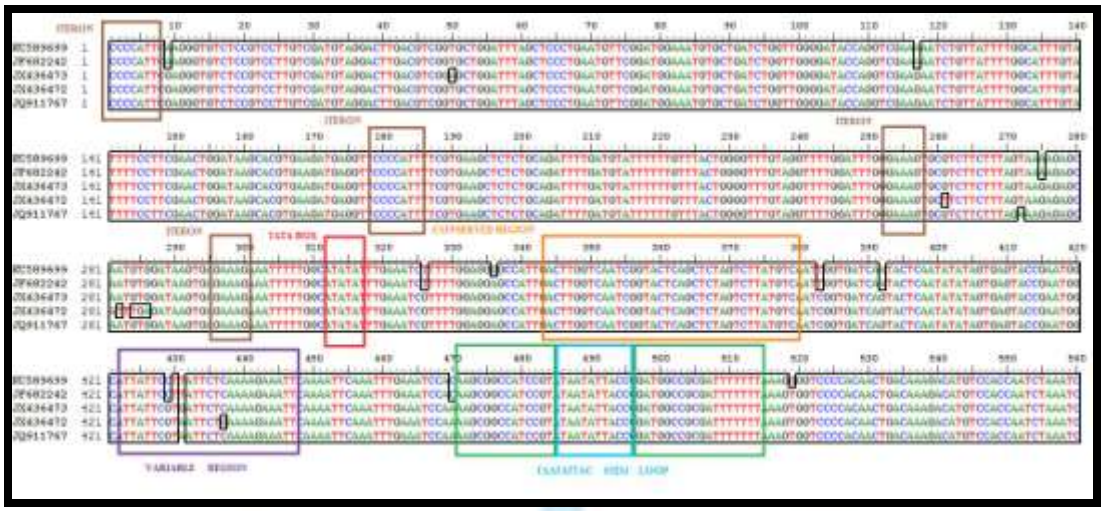


Figure: 3. Alignment of IR region nucleotide sequences of begomovirus isolated from Marigold. The sequences used here are as follows: KC589699 (*Ageratum enation virus*), JF682242 (*Ageratum enation virus*), JX436473 (*Ageratum enation virus*), JX436472 (*Ageratum enation virus*) and JQ911767 (*Ageratum enation virus*).

Betasatellites require begomoviruses for replication, encapsidation, insect transmission, and movement in plants [24]. Alphasatellites (formerly DNA1) are circular; single stranded DNA molecules associated with begomovirus/ betasatellite complexes. Consequently, alphasatellites are capable of self-replication in host plants, but require helper begomoviruses for movement in plants as well as insect transmission [25]. DNA-β associated with Marigold consisted of 1,335 nucleotides (GenBank accession no. KC589700). It contains a satellite-conserved region of 119 nucleotides (nt) in the 1,231 nt to 14 nt region. Moreover there is an adenine (A)-rich region of 281 nt in the 692 nt to 972 nt region (57% A). An ORF C1 located on the complementary-sense strand encoding a protein of 142 amino acids (Figure 2b).

Table: 1. Positions and coding capacity of predicted genes for the genome of begomovirus and its satellite molecules isolated from Marigold, and their highest amino acid sequence identities.

Components	Description	ORFs	Strand	Frame	Start codon (nucleotide coordinates)	Stop codon (nucleotide coordinates)	Predicted size (no. of amino acids)	Predicted molecular weight (kDa)	Predicted highest amino acid identities (%)
DNA-A	Pre coat protein	AV2	Sense strand	3rd frame (+)	138	551	137	15.86	*Ageratum enation virus* (AFQ32266) 100 %
	Coat protein	AV1	Sense strand	2nd frame (+)	485	964	159	18.29	*Ageratum enation virus* (AFB69494) 99%
	Replication enhancer protein	AC3	Complement strand	1st frame (–)	913	1440	175	20.67	*Ageratum enation virus* (ACG60167) 97%
	Transcriptional activator protein	AC2	Complement strand	3rd frame (–)	1181	1585	134	15.05	*Ageratum enation virus* (ACG60168) 99 %
	Replication associated protein	AC1	Complement strand	2nd frame (–)	1488	2573	361	40.85	*Ageratum enation virus* (AFQ32269) 98 %
	C4 Protein	AC4	Complement strand	3rd frame (–)	2159	2416	85	9.71	*Ageratum enation virus* (AFQ32270) 99 %
DNA-β	Symptoms inducing protein	C 1	Complement strand	2nd frame (–)	180	608	142	16.42	*Ageratum yellow leaf curl betasatellite* (ADR79368) 100 %
DNA-α	Similar to rep protein	rep	Sense strand	2nd frame (+)	761	994	77	8.76	alphasatellite (CCI72000) 93 %

The complete nucleotide sequence of the alphasatellite isolated from Marigold was determined, having GenBank accession no. KC206078. A-rich regions are maintained by all alphasatellites immediately downstream of the Rep gene as reported for other alphasatellites. This A-rich region is approximately 153-169 nts long with an A-content of between 52.3-58.4 % [26]. This molecule has an arrangement typical of alphasatellites, containing a single ORF in the virion sense which encodes a protein similarly to the replication-associated protein (Rep; a rolling-circle replication initiator protein) of alphasatellite. The predicted Rep shows the highest levels of amino acid sequence identity (93%) to the Rep of alphasatellite (GenBank accession no. CCI72000) reported from Palampur, India (Figure 2c).

The phylogenetics is the study of the evolutionary relationships among different species which have common ancestor. These relationships are shown in the form of phylogenetic trees composed of branches which indicate the descendents and nodes which represent the most recent common ancestors. Phylogenetic approaches have been used on the sequences of DNA and proteins to determine the ancestral relationships of living organisms in the form of tree of life [27].

These sequence-based approaches have also been used to study the relationships among different viruses and considered more reliable compared to other non sequence-based approaches. Further the multigene phylogenetic analysis is more reliable than single gene analysis, improves the resolution of the tree and thus been used to study the relationships of different organisms [28].

Pairwise BLAST is used to calculate similarity, but its limitations are that only two sequences may be analyzed at one time and percent similarity/identity is based on local alignment – not global alignment. MegAlign, which comes with the DNASTAR package (DNASTAR, Inc.), also generates similarity matrices, but it is quite expensive and not available as a stand-alone product. MatGAT (Matrix Global Alignment Tool) is a simple, easy to use similarity/identity matrix generator that calculates the similarity and identity between every pair of sequences in a given data set without requiring pre-alignment of the data. The program performs a series of pairwise alignments using the Myers and Miller global alignment algorithm, calculates similarity and identity, and then places the results in a distance matrix [13].

Phylogenetics can be defined as the systematic study of the relationships between organisms that leads to a taxonomical classification based on how closely they are related in terms of evolutionary differences. Based on the close sequence identity and the length of the sequences, begomovirus isolated from Marigold and sequences of various begomoviruses were downloaded from Gen Bank with the accession numbers provided by the FASTA output and were fed into Molecular Evolutionary Genetics Analysis Program (MEGA) version 4.0.

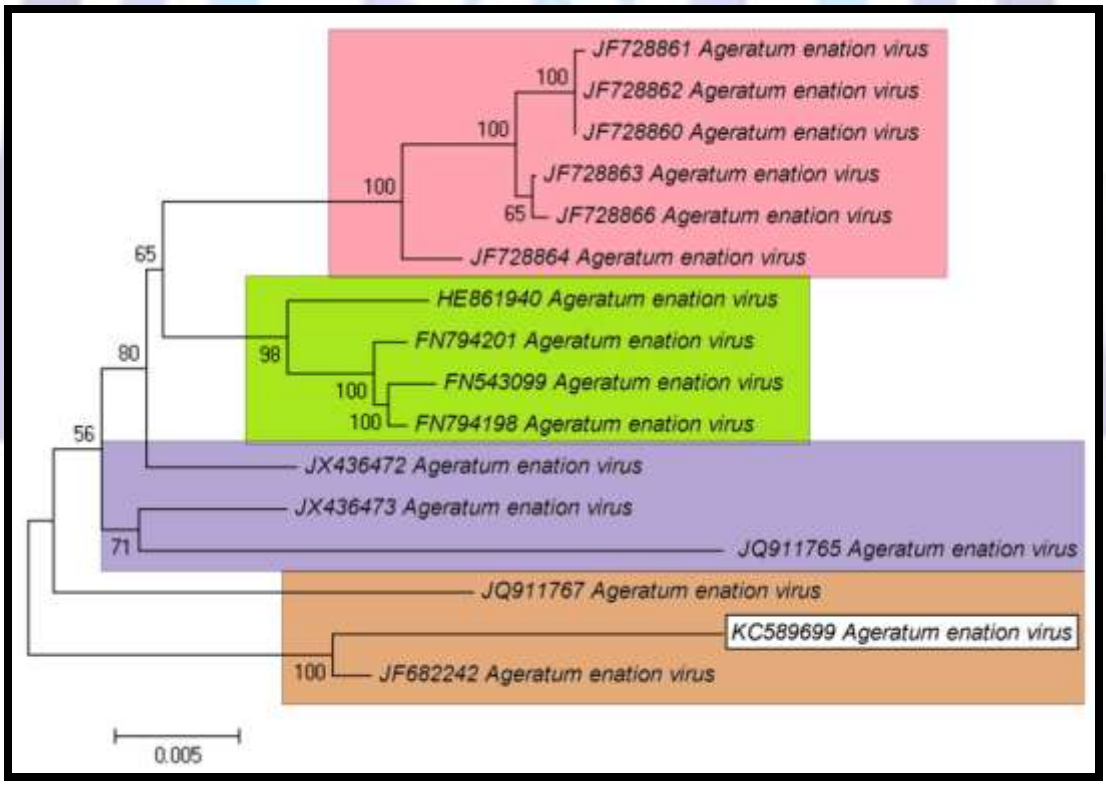


Figure: 4. Phylogenetic dendrograms based on alignments of selected begomovirus genome (DNA-A genomic). Begomovirus sequences used for comparison were *Ageratum enation virus* (AEV). Vertical branches are arbitrary, horizontal branches are proportional to calculated mutation distances. Values at nodes indicate percentage bootstrap values (1000 replicates). The different begomovirus used in the construction of phylogenetic tree are depicted in the figure.

Phylogenetic analysis based on the complete DNA-A sequences of Marigold begomovirus and other selected complete DNA-A sequences indicates that isolate M 2 cluster with the isolates of *Ageratum enation virus* from India. The DNA-A sequences of the M 2 isolate had high nucleotide sequence identity (93-97%) with other begomoviruses reported from India (JF728861, FN794201, JQ911767, FN543099, and JX436472), Pakistan (AM701770, AM261836, AM698011) and Nepal (AJ437618), which are all isolates of *Ageratum enation virus* (Figure 4).

One of the most important analyses that can be employed in phylogenetics is the pairwise determination of similarity or identity between DNA or protein sequences. The percent identity is the calculated percentage of how two sequences compare at a base-to-base or residue-to-residue level [29]. The percent similarity identity matrix is a more strict calculation where sequence gaps and mismatches are included in the evaluation and scored using a more complex formula and a comparison look-up table.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1. KC589699																
2. JF728861	94.5															
3. FN543099	94.4	97.0														
4. JF728862	94.5	100.0	97.1													
5. JF728860	94.5	100.0	97.1	100.0												
6. FN794201	94.5	97.1	99.6	97.1	97.1											
7. JF728863	94.6	99.7	97.3	99.7	99.7	97.4										
8. JF728866	94.6	99.6	97.4	99.6	99.6	97.5	99.9									
9. FN794198	94.5	97.1	99.7	97.2	97.2	99.7	97.5	97.5								
10. JQ911767	94.7	96.1	97.3	96.2	96.2	97.4	96.4	96.5	97.5							
11. JX436472	95.1	97.5	98.3	97.5	97.5	98.4	97.7	97.8	98.4	97.6						
12. JX436473	95.5	97.2	97.9	97.3	97.3	98.0	97.3	97.3	98.0	97.5	98.8					
13. JF682242	97.3	96.5	96.7	96.5	96.5	96.8	96.6	96.6	96.8	97.1	97.5	98.0				
14. HE861940	91.9	94.4	95.9	94.4	94.4	96.0	94.5	94.5	96.0	94.0	95.3	95.4	94.2			
15. JQ911765	93.8	96.0	95.8	96.1	96.1	95.9	96.0	96.0	95.9	95.4	96.6	97.0	96.1	93.6		
16. JF728864	94.1	99.1	97.7	99.1	99.1	98.0	99.2	99.2	97.8	96.6	98.0	97.5	96.3	95.0	95.7	

Figure: 5. *Ageratum enation virus* (AEV: KC589699) DNA-A nucleotide sequence identities which are expressed in percentage for the begomovirus associated with leaf curl disease of Marigold in India and selected previously characterized begomoviruses. The begomoviruses name corresponds to the accession numbers given in the analyses are the same as used for Phylogenetic analysis.

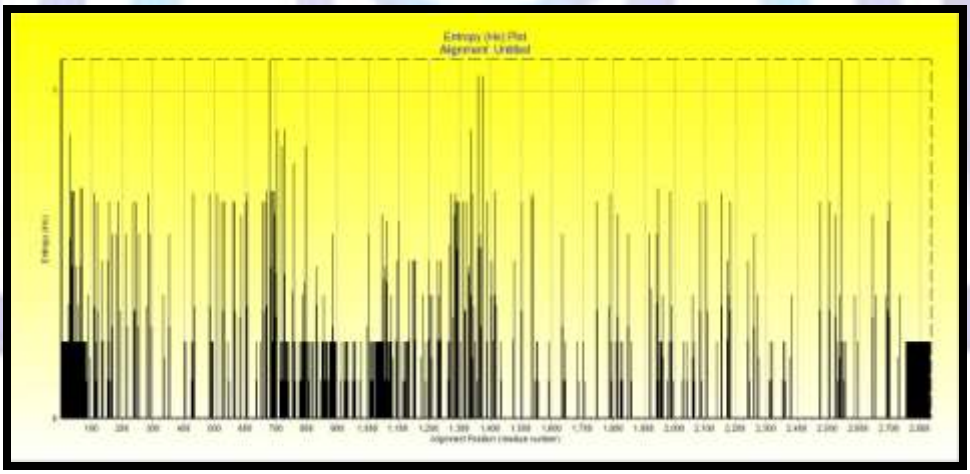


Figure: 6. The Entropy power result generated from the CLUSTAL-W alignment of DNA-A sequences of *Ageratum enation virus* (AEV: KC589699) found in association with leaf curl disease of Marigold.

Therefore as per the Myers and Miller global alignment algorithm the isolate M 2 showed similar result (Figure 5) with highest nucleotide identity of 97.3 % with *Ageratum enation virus* (JF682242) reported from India in association of vein enation disease on *Amaranthus hypochondriacus* and lowest nucleotide identity of 91.9 % with *Ageratum enation virus* (HE861940). This isolate is also reported from India and causing disease symptoms in *Glycine max*. The DNA-A sequence of *Ageratum enation virus* (AEV: KC589699) was subjected to multiple sequence alignment by using CLUSTAL-W and resulted in an Entropy plot. In the Entropy power plot hardly two or three regions are scored above 1, thus the Entropy plot shows negligible amount of variable regions in the sequences (Figure 6). This is very much in relation to the nucleotide sequence identity matrix of the aligned sequences, in which the begomovirus infecting Marigold showed almost similar amount of identity with other begomovirus genome retrieved from NCBI i.e. the range of highest and lowest nucleotide identity is quite less (92 - 97 %).

The use of DNA and amino acid sequences to estimate evolutionary history is denominated molecular phylogenetics [30]. Thus, molecular characteristics are used to classify organisms placing them on a map of evolutionary relationships known as the phylogenetic tree that shows the probable evolution of various organisms. Phylogenetic analysis based on complete DNA-β sequences demonstrates that it cluster with *Ageratum yellow leaf curl betasatellite* (HQ407397) from India (Figure 7).

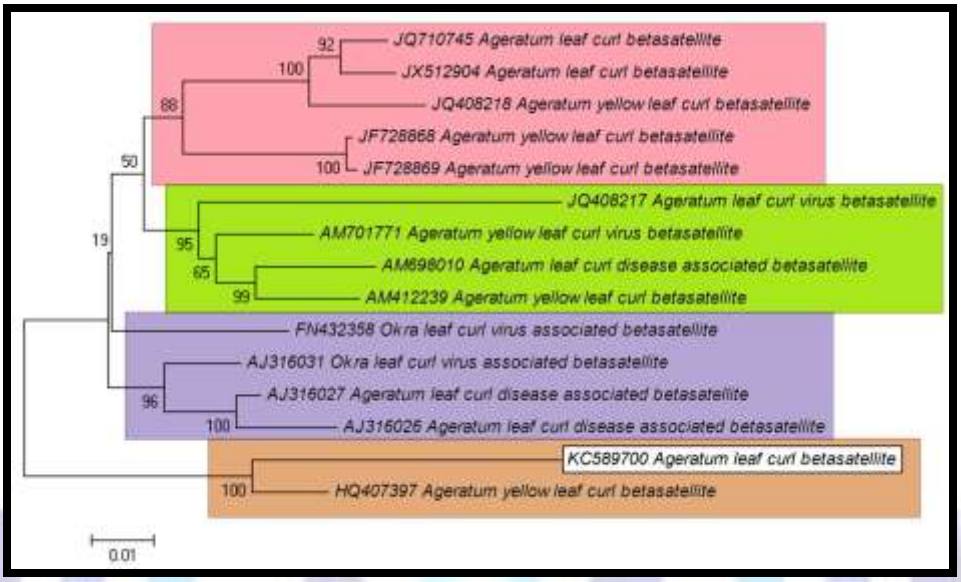


Figure: 7. The Neighbor-Joining tree based on complete nucleotide sequence of *Ageratum leaf curl betasatellite* (ALCB: KC589700) infecting Marigold in India and other betasatellite sequences available in GenBank. Bootstrap values at major nodes are indicated. Horizontal distances are proportional to the genetic distance between isolates and vertical distances are arbitrary. Scale bar indicates the proportion of sites changing along each branch.

Further nucleotide sequence comparison with other betasatellites showed highest sequence identity ranging from 92 to 96% reported from India (JQ408218, JF728868, JQ710745, AM412239) and Pakistan (AM701771, AM698010) which are isolates of *Ageratum leaf curl betasatellite* and *Ageratum yellow leaf curl betasatellite*. The Global alignment results suggests that the *Ageratum leaf curl betasatellite* (ALCB: KC589700) infecting Marigold had 93.2 % nucleotide sequence identity with *Ageratum leaf curl betasatellite* (HQ407397) infecting wild sunflower in India. It further showed lowest sequence identity with *Ageratum yellow leaf curl betasatellite* (JQ408218) causing leaf curl disease in *Ageratum sp.* in India (Figure 8).

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1. KC589700															
2. HQ407397	93.2														
3. FN432358	86.5	90.6													
4. AJ316027	87.1	91.1	93.6												
5. AJ316031	86.3	90.7	94.7	96.2											
6. AJ316026	85.5	89.4	91.5	96.7	94.0										
7. AM701771	84.9	89.0	90.6	93.2	92.3	90.9									
8. AM698010	84.5	88.8	89.6	92.3	91.1	90.1	96.0								
9. JQ408218	79.2	81.8	83.2	83.5	83.1	82.6	85.8	84.7							
10. JF728868	84.2	87.6	91.5	90.6	89.8	88.3	93.8	92.9	86.7						
11. JQ710745	83.6	86.9	89.8	89.4	89.3	87.4	92.5	91.0	89.4	93.0					
12. JF728869	83.9	87.3	91.0	90.2	90.2	88.0	93.4	92.7	85.9	98.5	92.6				
13. JX512904	79.6	81.9	83.4	83.7	83.0	82.9	86.5	85.2	95.1	87.5	91.3	86.6			
14. AM412239	84.0	87.7	88.9	91.8	91.6	89.4	95.9	95.8	84.5	93.1	91.3	94.1	85.1		
15. JQ408217	77.0	80.1	81.4	82.5	81.7	81.6	86.5	86.2	89.2	84.8	84.2	84.8	90.0	86.0	

Figure: 8. *Ageratum leaf curl betasatellite* (ALCB: KC589700) nucleotide sequence identities which are expressed in percentage associated with leaf curl disease of Marigold in India and selected previously characterized begomoviruses. The begomoviruses name corresponds to the accession numbers given in the analyses are the same as used for Phylogenetic analysis.

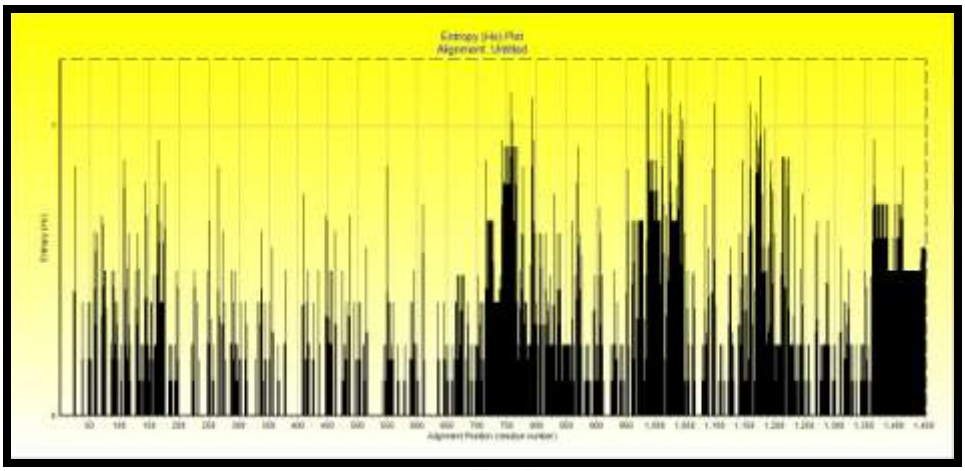


Figure: 9. The Entropy power result generated from the CLUSTAL-W alignment of DNA-β sequences of *Ageratum leaf curl betasatellite* (ALCB: KC589700) found in association with leaf curl disease of Marigold.

Similarly the betasatellite sequence of *Ageratum leaf curl betasatellite* (ALCB: KC589700) was also subjected to multiple sequence alignment by using CLUSTAL-W and resulted in an Entropy plot. In the Entropy power plot some regions are scored above 1 and some are almost near to 1, thus the Entropy plot shows fewer amounts of variable regions in the sequences (Figure 9). This is very much in relation to the nucleotide sequence identity matrix of the aligned sequences, in which the betasatellite infecting Marigold showed little variation of identity with other begomovirus genome retrieved from NCBI. The dark shaded region at the bottom of the plot depicts the miss matches in the alignment.

The complete nucleotide sequence of the alphasatellite isolated from Marigold was determined, having GenBank accession no. KC206078 and was most closely related to alphasatellite rep gene for truncated replication associated protein, clone UK7 [Pakistan: Faisalabad: 2010] (AM930246), with 93% nucleotide sequence identity. In phylogenetic comparisons the Marigold alphasatellite forms a cluster with various alphasatellites (Figure 10) reported from Pakistan (AM884370, AM930244), AM930248, AM930245) and India (HE861941, JX512905, FN794199, AJ512959). This molecule has an arrangement typical of alphasatellites, containing a single ORF in the virion sense which encodes a protein similarly to the replication-associated protein (Rep; a rolling-circle replication initiator protein) of alphasatellite.

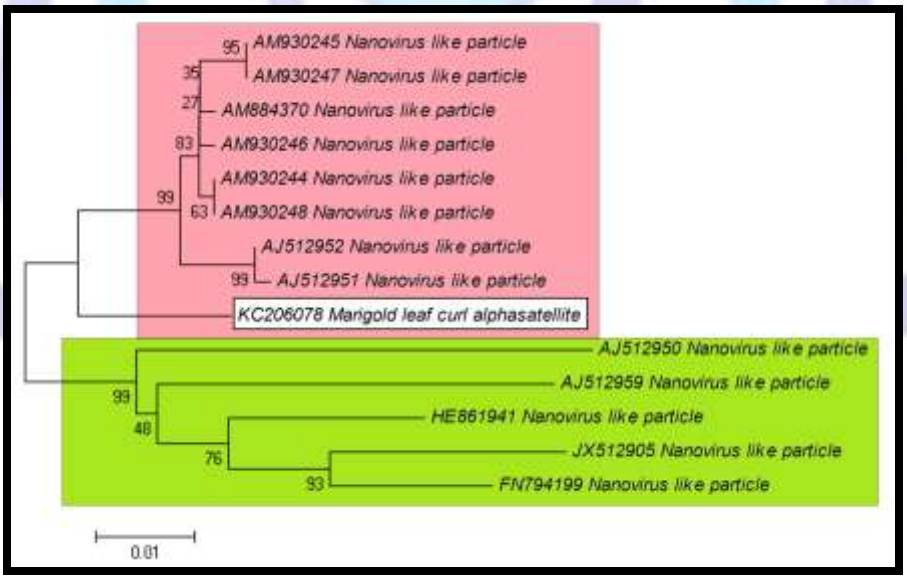


Figure: 10. Neighbor-Joining tree based on the complete sequence of *Marigold leaf curl alphasatellite* (MLCuA: KC206078) isolated from Marigold and other alphasatellite sequences available in GenBank. The different alphasatellite used in the construction of phylogenetic tree are depicted in the figure. Vertical branches are arbitrary, horizontal branches are proportional to calculated mutation distances. Values at nodes indicate percentage bootstrap values (1000 replicates).

The predicted Rep shows the highest levels of amino acid sequence identity (93%) to the Rep of alphasatellite (GenBank accession no. CCI72000) reported from Palampur, India. The ubiquitous association of alphasatellites with begomovirus / betasatellite complexes indicates that alphasatellites may play an important role in the occurrence, diffusion, and epidemiology of begomovirus / betasatellite complexes. The Global alignment reveals that the alphasatellite associated

with Marigold showed 40 - 42 % nucleotide sequence identity with all the isolates of nanovirus reported from India and Pakistan (Figure 11).

Marigold is a widely used ornamental plant in India was found to be infected with begomovirus and its associated satellites under natural conditions. It is likely that the whitefly vector is carrying the components (begomovirus, DNA-β and DNA-α). Thus, this identification represents the possibility of a serious threat to other economically important ornamental and crop plants and there is a need for a more comprehensive study to identify possible further begomoviruses infection in the country and to assess the contribution each makes to losses with a view to devising control strategies.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1. KC206078														
2. AM930246	42.1													
3. AM884370	42.2	99.6												
4. AM930244	42.3	99.6	99.7											
5. AM930248	42.3	99.6	99.7	100.0										
6. AM930245	42.2	99.6	99.6	99.6	99.6									
7. AM930247	42.2	99.6	99.6	99.6	99.6	100.0								
8. AJ512952	40.3	97.2	97.2	97.3	97.3	97.2	97.2							
9. AJ512951	40.4	97.2	97.2	97.3	97.3	97.2	97.2	99.3						
10. HE861941	41.5	91.9	92.0	92.0	92.0	91.9	91.9	92.0	92.0					
11. JX512905	41.5	91.4	91.5	91.5	91.5	91.4	91.4	91.4	91.4	92.2				
12. FN794199	41.5	92.7	92.7	92.7	92.7	92.7	92.7	92.5	92.5	95.1	94.8			
13. AJ512950	42.1	92.7	92.7	92.8	92.8	92.6	92.6	91.9	91.9	91.4	91.2	91.4		
14. AJ512959	42.2	90.6	90.6	90.7	90.7	90.5	90.5	90.4	90.5	91.0	91.8	92.8	91.9	

Figure: 11. *Marigold leaf curl alphasatellite* (MLCuA: KC206078) nucleotide sequence identities which are expressed in percentage associated with leaf curl disease of Marigold in India and selected previously characterized begomoviruses. The begomoviruses name corresponds to the accession numbers given in the analyses are the same as used for Phylogenetic analysis.

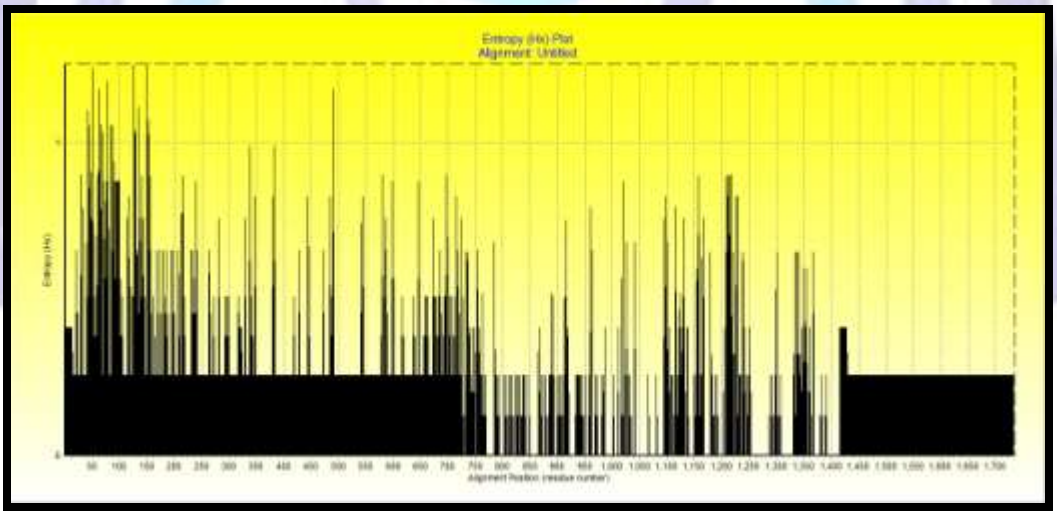


Figure: 12. The Entropy power result generated from the CLUSTAL-W alignment of DNA-α sequences of *Marigold leaf curl alphasatellite* (MLCuA: KC206078) found in association with leaf curl disease of Marigold.

The alphasatellite sequence of *Marigold leaf curl alphasatellite* (MLCuA: KC206078) was subjected to multiple sequence alignment by using CLUSTAL-W and resulted in an Entropy plot. In the Entropy power plot some regions are scored above 1 and some are almost near to 1, thus the Entropy plot shows fewer amounts of variable regions in the sequences (Figure 12). This is very much in relation to the nucleotide sequence identity matrix of the aligned sequences, in which the alphasatellite infecting Marigold showed almost similar amount of identity with other alphasatellite genome retrieved from NCBI i.e. the range of highest and lowest nucleotide identity is negligible (40 - 42 %). The alignment score is quite low; this proves from the energy plot, where the black box in the bottom reveals non-alignment in this portion of sequences.

CONCLUSION

In conclusion, we have established the *In Silico* characterization of *Ageratum enation virus* (AEV: KC589699), *Ageratum leaf curl betasatellite* (ALCB: KC589700) and *Marigold leaf curl alphasatellite* (MLCuA: KC206078) infecting Marigold plants in India. This study supported evidence that virus associated with Marigold was studied through bioinformatics analysis. The phylogenetic studies revealed closeness to other virus prominent in other neighbouring countries and can be replicated by other begomovirus species which increases the possibility of recombination and reassortment events. This could lead to evolution of new recombinant viruses or begomovirus complexes with different biological properties. The genetic diversity and evaluation of entropy power were viewed against the phylogenetic background which highlights the variable region where possibility of recombination increases. Our results provide much new information on these topics.

ACKNOWLEDGMENTS

The authors would like to acknowledge a vote of thanks to Department of Biotechnology (DBT project No. BT/PR13129/GBD/27/197/2009) and Department of Science and Technology (DST project no SR/FT/LS-042/2009), India for their financial support.

REFERENCES

[1] Mansoor, S., Briddon, R. W., Zafar, Y., and Stanley, J. 2003. Geminivirus disease complexes: An emerging threat. Trends Plant Sci 8: 128-134.

[2] Lima, A. T., Sobrinho, R. R., González-Aguilera, J., Rocha, C. S., Silva, S. J., Xavier, C. A., Silva, F. N., Duffy, S., and Zerbini, F.M. 2013. Synonymous site variation due to recombination explains higher genetic variability in begomovirus populations infecting non-cultivated hosts. J Gen Virol 94: 418–31.

[3] Urbino, C., Gutiérrez, S., Antolik, A., Bouazza, N., Doumayrou, J., Granier, M., Martin, D. P., and Peterschmitt, M. 2013. Within-host dynamics of the emergence of *tomato yellow leaf curl virus* recombinants. PLoS ONE. 2013; 8: e58375. doi:10.1371/journal.pone.0058375.

[4] Raj, S. K., Khan, M. S., Snehi, S. K., Kumar, S., and Khan, A. A. 2007. Natural occurrence of a Begomovirus on *Dimorphotheca sinuate* in India. Australasian Plant Disease Notes 2: 25–26.

[5] Ilyas, M., Nawaz, K., Shafiq, M., Haider, M. S., and Shahid, A. A. 2013. Complete nucleotide sequences of two begomoviruses infecting Madagascar periwinkle (*Catharanthus roseus*) from Pakistan. Arch Virol 158: 505–10.

[6] Perez Gutierrez, R. M., Luha, H. H., and Garrido, S. H. 2006. Antioxidant activity of *Tagetes erecta* essential oil. Journal of Chilean Chemical society, 51: 883-886.

[7] Marcia, M., Marques, M., Selene, M., Morais Icaro, G., Vieira, P., Mariano, G., Ana, R., Silva, A., De Almeida, R. R., and Guedes, M. I. 2011. Larvicidal activity of *Tagetes erecta* against *Aedes aegypti*. Journal of the American Mosquito Control Association, 27: 156-158.

[8] Belcher, J. V., and Hussey, R. H. 1977. Influence of *Tagetes patula* and *Arachis hypogaea* on *Meloidogyne incognita*. Plant Disease Reporter, 61: 525-528.

[9] Marwal, A., Sahu, A. K., Choudhary, D. K., and Gaur R. K. 2013. Complete nucleotide sequence of a begomovirus associated with satellites molecules infecting a new host *Tagetes patula* in India. Virus Genes. 47 (1): 194 – 198.

[10] Thompson, J. D., Higgins, D. G., and Gibson T. J. 1994. CLUSTAL-W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. Nucleic Acids Research, 22: 4673-4680.

[11] Kumar, S., Dudley, J., Nei, M., and Tamura, K. 2008. MEGA: A biologist-centric software for evolutionary analysis of DNA and protein sequences. Briefings in Bioinformatics, 9: 299-306.

[12] Tamura, K., Dudley, J., Nei, M., and Kumar, S. 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Molecular Biology and Evolution, 24: 1596-1599.

[13] Campanella, J. J., Bitincka, L., and Smalley, J. 2003. MatGAT: An application that generates similarity/identity matrices using protein or DNA sequences. BMC Bioinformatics 4:29.

[14] Sofy, A. R., Soliman, A. M., Mousa, A. A., and El-Dougdoug, K. A. 2012. Molecular characterization and Bioinformatics Analysis of Viriod isolate associated with Cirtus Gummy Bark Disease in Egypt. International Journal of Virology, 8: 133-150.

[15] Bauer, A. M., Zheng, C., Chitsaz, F., Derbyshire, M. K., Geer, L. Y., Geer, R. C., Gonzales, N. R., Gwadz, M., Hurwitz, D. I., Lanczycki, C. J., Lu, F., Lu, S., Marchler, G. H., Song, J. S., Thanki, N., Yamashita, R. A., Zhang, D., and Bryant, S. H. 2013. CDD: conserved domains and protein three-dimensional structure. Nucleic Acids Research, 41: 348-352.

[16] Namrata, J., Saritha, R. K., Datta, D., Singh, M., Dubey, R. S., Rai, A. B., and Rai, M. 2010. Molecular Characterization of *Tomato leaf curl Palampur virus* and *Pepper leaf curl betasatellite* Naturally Infecting Pumpkin (*Cucurbita moschata*) in India. Indian Journal of Virology, 21: 128–132.

[17] Mansoor, S., Khan, S. H., Bashir, A., Saeed, M., Zafar, Y., Malik, K. A, Briddon, R., Stanley, J., and Markham, P. G. 1999. Identification of a novel circular single-stranded DNA associated with cotton leaf curl disease in Pakistan. Virology, 259: 190-199.

[18] Saunders, K., and Stanley, J. 1999. A nanovirus-like DNA component associated with yellow vein disease of *Ageratum conyzoides*: Evidence for interfamilial recombination between plant DNA viruses. Virology, 264: 142-152.

[19] Tiendrebeogo, F., Lefeuvre, P., Hoareau, M., Villemot, J., Konate, G., et al., 2010. Molecular diversity of *Cotton leaf curl Gezira virus* isolates and their satellite DNAs associated with okra leaf curl disease in Burkina Faso. Virology Journal, 7: 48-48.

[20] Finn, R. D., Mistry, J., Tate, J., Coggill, P., Heger, A., Pollington, J. E., Gavin, O. L., Gunasekaran, P., Ceric, G., Forslund, K. et al. (2010). The Pfam protein families database. Nucleic Acids Research, 38: D211-D222.

[21] Bauer, A. M., Lu, S., Anderson, J. B., Chitsaz, F., Derbyshire, M. K., Scott, C. D., Fong, J. H., Geer, L. Y., Geer, R. C., Gonzales, N. R., Gwadz, M., Hurwitz, D. I., Jackson, J. D., Ke, Z., Lanczycki, C. J., Lu, F., Marchler, G. H., Mullokandov, M., Omelchenko, M. V., Robertson, C. L., Song, J. S., Thanki, N., Yamashita, R. A., Zhang, D., Zhang N., Zheng, C., and Bryant, S. H. 2011. CDD: a Conserved Domain Database for the functional annotation of proteins. Nucleic Acids Research, 39: 225-229.

[22] He, Z. F., Mao, M. J., Yu, H., Li, H. P., and Chen X. 2009. Molecular characterization of a distinct begomovirus infecting *Allamanda cathartica* in Guangdong, China. Achieves of Virology, 54: 1199-202.

[23] Ndunguru, J., Legg, J. P., Aveling, T. A. S., Thompson, G., Fauquet, C. M. 2005. Molecular biodiversity of cassava begomoviruses in Tanzania: evolution of cassava geminiviruses in Africa and evidence for East Africa being a center of diversity of cassava geminiviruses. Virology Journal, 2: 21.

[24] Mansoor, S., Briddon, R. W., Zafar, Y., and Stanley, J. 2003. Geminivirus disease complexes: An emerging threat. Trends in Plant Science, 8: 128-134.

[25] Yan, X., Peijun, W., Pei, L., Huanran, G., and Zhou, X. 2010. Characterization of alphasatellites associated with monopartite begomovirus / betasatellite complexes in Yunnan, China. Virology Journal, 7: 178.

[26] Marwal, A., Sahu, A., Prajapat, R., and Gaur, R. K. 2012. First report of association of begomovirus with the leaf curl disease of a common weed, Datura inoxia. Indian J Virol 23: 3–84.

[27] Forterre, P., and Philippe, H. 1999. Where is the root of the universal tree of life? BioEssays, 21: 871-879.

[28] Kurtzman, C. P., and Robnett, C. J. 2003. Phylogenetic relationships among yeasts of the 'Saccharomyces complex' determined from multigene sequence analyses. FEMS, Yeast Research, 3: 417-32.

[29] Pearson, W. R., and Lipman, D. J. 1988. Improved Tools for Biological Sequence Comparison. Proceedings of the National Academy of Science, 85: 2444-2448.

[30] Archer, J. P. 2008. Phylogenetic trees. In "The diversity of HIV-1". PhD Thesis (pp. 75-80. The University of Manchester, Manchester.

Author' biography with Photo

Avinash Marwal

Avinash Marwal has submitted the thesis and completed his PhD from a DBT funded project (DBT project No. BT/PR13129/GBD/27/197/2009), under the supervision of Dr R. K. Gaur, Principal Investigator. He has 35 international publications in reputed journals and has command over various molecular, biotechnological and bioinformatics tools and techniques. His core interests are Homology Modeling, Molecular Docking, Biological Database development, *In Silico* drug designing, Plant Molecular Biology and Development of Transgenic Plants.

Dr Rajneesh Prajapat

Dr. Rajneesh Prajapat has completed his under the supervision of Dr R K Gaur on the project entitled "In Silico molecular characterization of Begomovirus in Rajasthan, India. He graduated from M.D.S University, Ajmer, Rajasthan (India) in 2003 and did M.Sc Biotechnology from Universiry of Rajasthan, Jaipur, Rajasthan (India) in 2005 and obtains M.Phil in Environmental Biotechnology from Department of Environmental studies, M.D.S University, Ajmer, Rajasthan (India) in 2006. He has 30 publications and published 3 books on Bioinformatics and Virology.

ISSN 2348-6201

Dr Rajarshi Kumar Gaur

Dr. Rajarshi Kumar Gaur is presently working as Head and Associate Professor, Department of Science, Mody Institute of Technology and Science (Deemed University), Lakshmangarh, Sikar, Rajasthan. He did his Ph.D on molecular characterization of sugarcane viruses of India. He had partially characterized three sugarcane virus viz., sugarcane mosaic virus, sugarcane streak mosaic virus and sugarcane yellow luteovirus. He received MASHAV fellowship in 2004 of Israel government for his post doctoral studies and joined The Volcani Centre, Israel and then shifted to Ben Gurion University, Negev, Israel. In 2007 he received the Visiting Scientist Fellowship from Swedish Institute Fellowship, Sweden for one year to work in the The Umeå University, Umeå, Sweden. He is also a recipient of ICGEB, Italy Post Doctoral fellowship in 2008. He worked on development of marker-free transgenic plant against cucumber viruses. He has made significant contributions on sugarcane viruses and published 80 national/international papers and presented near about 50 papers in the national and international conferences. He was awarded as Fellow of International Society of Biotechnology, Fellow of Madhaswami Educational Trust and Fellow of International Consortium of Contemporary Biologist. He has also visited Thailand, New Zealand, London and Italy for the sake of attending the conference/workshop. Recently, he received two projects on begomovirus from Department of Biotechnology, Government of India and Department of Science and Technology, India, New Delhi. Presently, he is working on the characterization of geminiviruses and RNAi technology.