ISSN NO. 2320-5407

Journal homepage: http://www.journalijar.com

INTERNATIONAL JOURNAL OF ADVANCED RESEARCH

Short Communication

Estimating GC Profile and Restriction Map analysis of three begomovirus components infecting an ornamental plant Marigold (*Tagetes patula*) in India

Avinash Marwal[#], Rajneesh Prajapat[#] and R.K. Gaur*

Department of Science, Faculty of Arts, Science and Commerce, Mody Institute of Technology and Science, Lakshmangarh, Sikar - 332311, Rajasthan, India.

[#] These Authors contributed equal to this work

Manuscript Info

Manuscript History:

Received: 11 November 2013
Final Accepted: 25 November 2013
Published Online: January 2014

Key words:
Begomovirus, betasatellite, alphasatellite, GC Profile, Restriction Map.

Corresponding Author:
gaurrajarshi@hotmail.com

Abstract

Begomovirus is one of the largest group of *Bemisia tabaci* transmitted plant viruses (family *Geminiviridae*) containing single-stranded circular DNA that encapsulated in geminate particles and prevalent in the tropical and subtropical regions of the world. Sometimes begomovirus has been found associated with their other genomic components i.e betasatellite and alphasatellite. This study supported evidence that three begomovirus components associated with Marigold was studied through computational techniques for in depth analysis using GC Profile tool and Restriction map analysis. The relationships between the G+C content and other genomic features, such as distributions of genes, are analyzed in a perceivable manner.

Introduction

Begomovirus are an outsized varied family of plant viruses (Mansoor et al., 2003) which infects an expansive assortment of plants such as ornamentals, weeds and crops and causes a noteworthy loss to Agriculture and Horticulture worldwide (Lima et al., 2013). Ornamental plants are extensively scattered worldwide and have high environmental adaptability (Marwal et al., 2012). Ornamentals are considered as a foundation of new viruses and reservoirs of unidentified economically imperative viruses but are often neglected during diversity study (Marwal et al., 2013a), (Urbino et al., 2013). Many scientific reports have demonstrated that ornamental plants serve as reservoir or alternative hosts for begomovirus survival (Raj et al., 2207) and spread in the absence of the main crops (Ilyas et al., 2013). Thus, there is a pressing need for additional information on the diversity and distribution of begomovirus in ornamental plants.

Horizontal gene transfer is recognized as a major force for microbial evolution, as it leads to 'evolution in quantum leaps' (Koonin et al., 2001). Genomic islands are formerly mobile genetic elements that have been acquired by the core genomes via horizontal gene transfer (Groisman and Ochman, 1996). They often consist of DNA regions that differ from the core genome in their G + C content and codon usage. Depending on the functions they encode, genomic islands can be classified further as pathogenicity islands, metabolic islands, secretion islands, resistance islands and symbiosis islands (Hentschel and Hacker, 2001), (Hacker and Carniel, 2001).

In the present study the previously identified begomovirus and its components are subjected to computational analysis for their in depth study using two online available softwares i.e GC profile and NEBcutter.

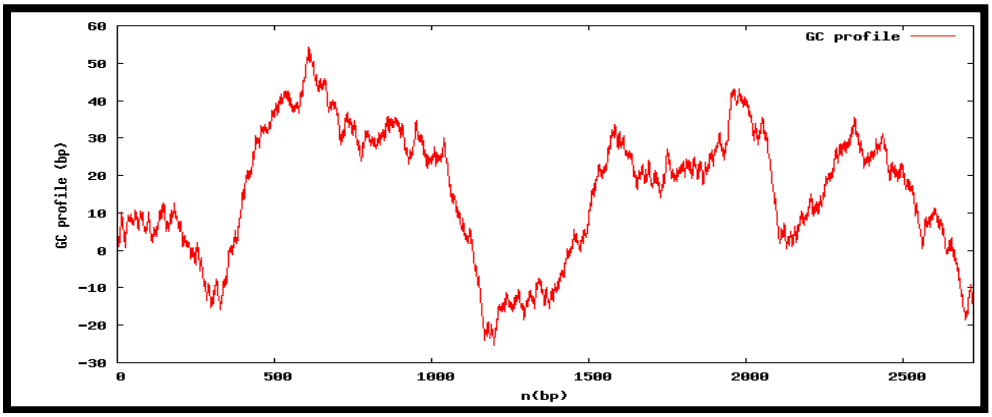
Results and Discussion

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 (Marwal et al., 2013b). In order to understand the evolution, structure and function of genomes, it is

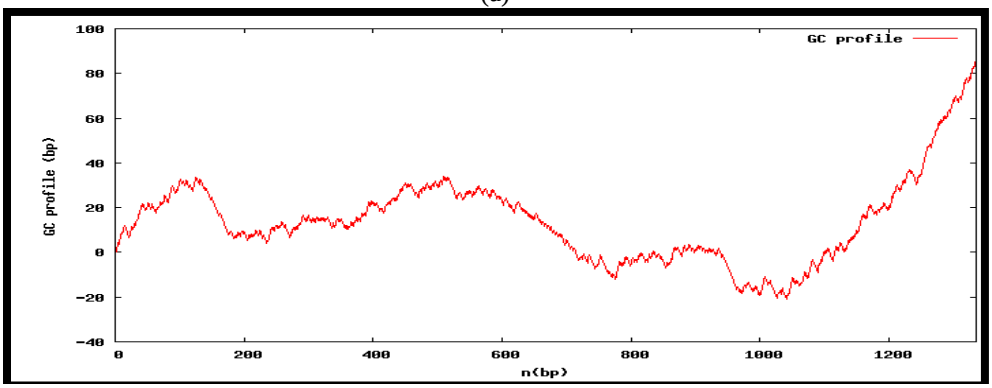
important to know the general compositional features of DNA sequences. Based on the quadratic divergence, a new segmentation algorithm to partition a given genome or DNA sequence into compositionally distinct domains has been put forward (Gao and Zhang, 2006). With the aid of the technique of cumulative GC profile, the distribution of segmentation points can be displayed intuitively. GC-Profile provides a quantitative and qualitative view of genome organization (Zhang et al., 2005).

GC-Profile implements a new segmentation algorithm based on the quadratic divergence, and integrates a windowless method for the G + C content computation, known as the cumulative GC profile. The integration of cumulative GC profile with the coordinates of segmentation points leads to a clear graphical representation of the G + C content variation along a genome or chromosome and enables to establish the relationships between the G + C content and other genomic features (Gao and Zhang, 2006).

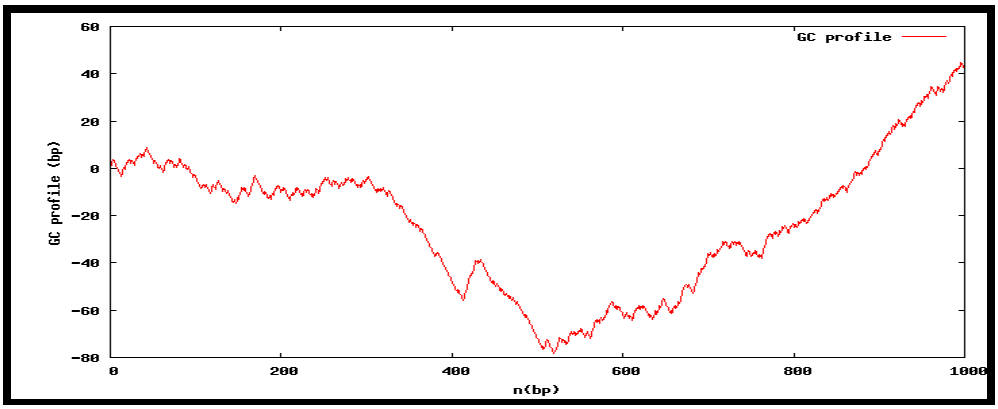
As shown in Figure 1a, there is a desert like region in an up and down distribution, which was calculated in 2.7 kb long genomic sequence of *Ageratum enation virus* (AEV: KC589699). It is also shown that the obtained segmentation points have clear biological implications. Three segmented boundaries were detected. There is an abrupt decrease of the density of GC profile at the first boundary below zero ranging from 250 to 350 bp. Second boundary has a poorest negative region (G + C) between 1100 to 1500 bp and the third boundary is observed at 2600 to 2700 bp. The overall GC content of the DNA-A sequence was calculated as 42.03 %.



(a)



(b)



(c)

Figure: 1. GC profile of (a) *Ageratum enation virus* (AEV: KC589699), (b) *Ageratum leaf curl betasatellite* (ALCB: KC589700) and (c) *Marigold leaf curl alphasatellite* (MLCuA: KC206078) infecting Marigold.

Whereas in the case of *Ageratum leaf curl betasatellite* (ALCB: KC589700) only a single stretch of low GC profile was observed from 650 to 1150 bp (Figure 1b). The negative cumulative GC profile for the genomic islands is distinct from that of the rest of the genome, in that the genomic islands have relatively low GC content, as reflected by abrupt drops in the negative cumulative GC profile at the regions of the genomic islands identified. The abrupt drop in the negative cumulative GC profile indicates that there are clear boundaries between the genomic islands and the surrounding regions. The overall GC content of the betasatellite (DNA-β) sequence was calculated as 38.50 %.

In case of *Marigold leaf curl alphasatellite* (MLCuA: KC206078) there is a sharp decrease in the GC profile below zero from position 350 to 800 bp (Figure 1c). The overall GC content of the alphasatellite (DNA-α) sequence was calculated as 39.20 %. This large-scale variation in base composition affects the coding sequences and seems to reflect a fundamental level of genome organization. This organization shows marked variation in a number of important biological properties, including gene density, patterns of codon usage, gene length, replication timing and the rate of recombination.

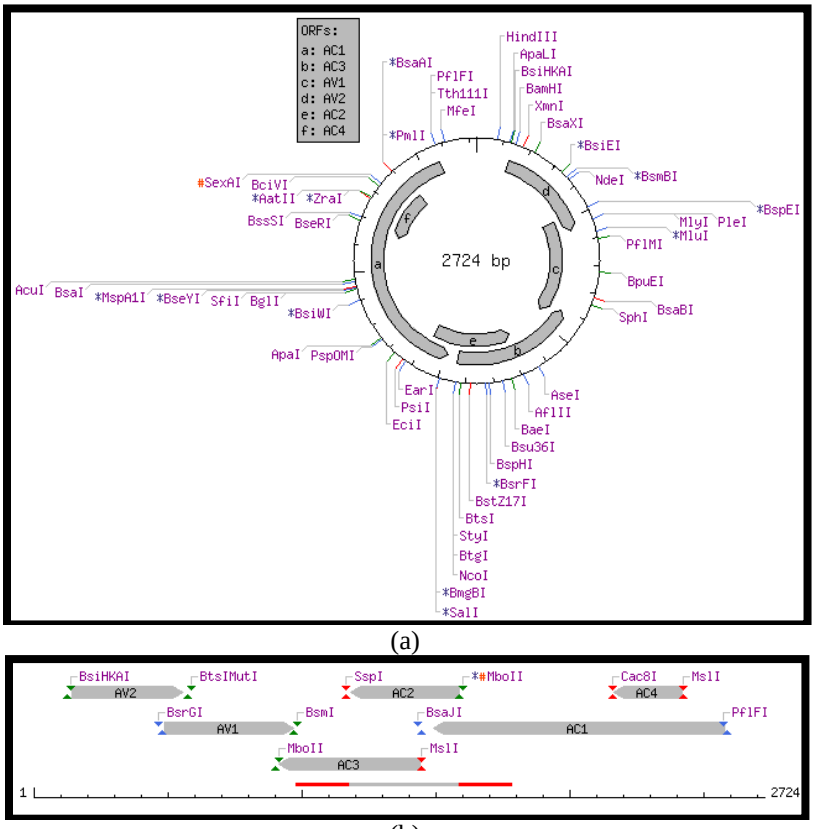
Until the advent of the polymerase chain reaction (PCR), restriction enzymes provided the most convenient way to manipulate individual genes and move them from one vector to another. For a while, it seemed that the ability of PCR to permit precise amplification of individual stretches of DNA might render the use of restriction enzymes obsolete (Pingoud and Jeltsch, 2001). However, they merely found new utility by then serving as diagnostic reagents to show that DNA constructs had been made correctly. They still provide one of the cheapest and most convenient ways to characterize DNA constructs (Roberts et al., 2003). NEBcutter, version 1.0, is a program available via a web server (http://tools.neb.com/NEBcutter) that will accept an input DNA sequence and produce a comprehensive report of the restriction enzymes that will cleave the sequence. It produces a variety of outputs including restriction enzyme maps and theoretical digests (Vincze et al., 2003).

The *Ageratum enation virus* (AEV: KC589699), *Ageratum leaf curl betasatellite* (ALCB: KC589700) and *Marigold leaf curl alphasatellite* (MLCuA: KC206078) sequences was retrieved from NCBI as a GenBank file via its accession number. Size of the ORFs to be displayed and the set of restriction enzymes to be used were selected. The program calculates the positions of all restriction enzyme sites and finds the ORFs in the sequence (Table 1). It then displays a schematic diagram of the sequence, the long ORFs, based on the rules described that all restriction enzymes that cut it just once. The initial display also shows the enzymes that can be used in a complete digest to excise each ORF that is displayed.

Components	Description	ORFs	Strand	Frame	Start codon	Stop codon
DNA-A	Pre coat protein	AV2	Sense	3rd frame (+)	138	551
	Coat protein	AV1	Sense	2nd frame (+)	485	964
	Replication enhancer protein	AC3	Complement	1st frame (–)	913	1440

	Transcriptional activator protein		AC2	Complement	3rd frame (–)	1181	1585
	Replication associated protein		AC1	Complement	2nd frame (–)	1488	2573
	C4 Protein		AC4	Complement	3rd frame (–)	2159	2416
DNA-β	Symptoms inducing protein		C 1	Complement	2nd frame (–)	180	608
DNA-α	Similar to rep protein		rep	Sense strand	2nd frame (+)	761	994

Table: 1. Positions and coding capacity of predicted genes for the genome of begomovirus and its satellite molecules isolated from Marigold

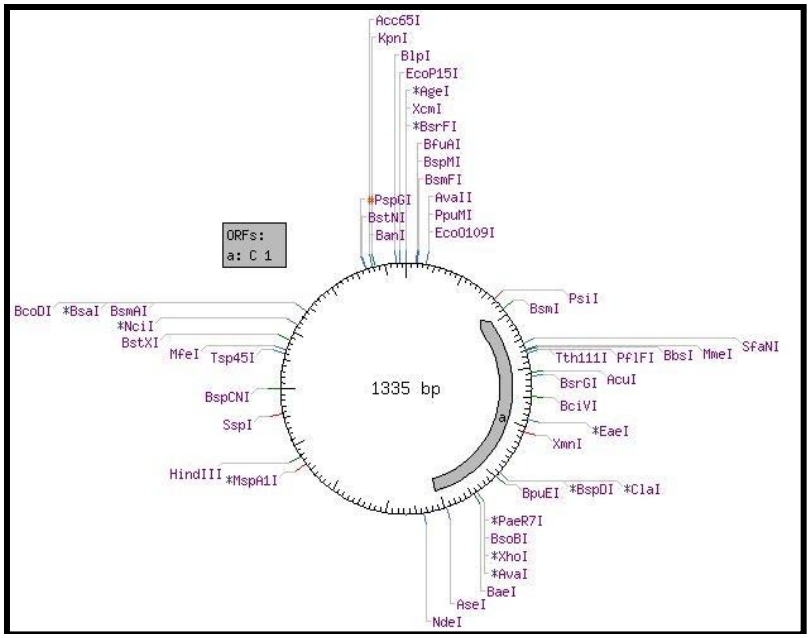


(a)

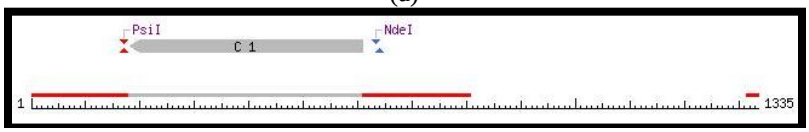
(b)

Figure: 2. (a) Circular representation of the restriction map of *Ageratum enation virus* (AEV: KC589699) DNA-A component. In this display only enzymes that cleave the sequence once are shown. (b) The linear display of a digest of the *Ageratum enation virus* (AEV: KC589699).

The six main ORFs in *Ageratum enation virus* (AEV: KC589699) i.e. AV2, AV1, AC3, AC2, AC1 and AC4 (Figure 5.34a) are flanked by sites BsiHKAI and BisIMutI; BsrGI and BsmI; MboII and Ms1I; SspI and MboII; BsaJI and Pf1FI; Cac8I and Ms1I respectively that could be used to excise them from the complete begomovirus sequence (Figure 2b). These are the closest enzymes at 5` end and at 3` end of all the ORFs.

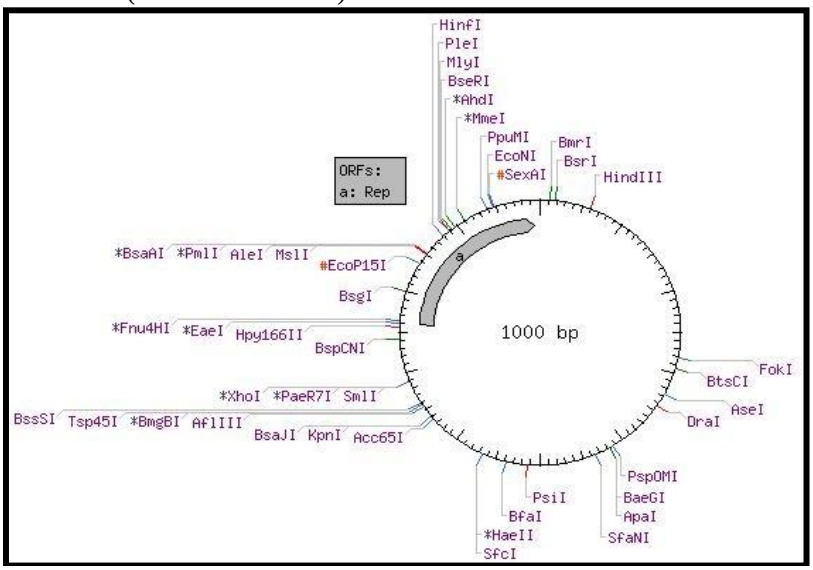


(a)

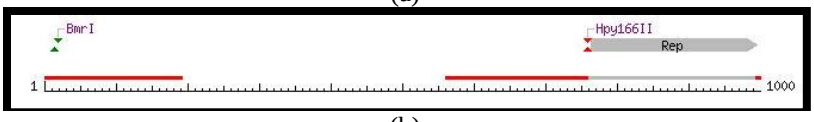


(b)

Figure: 3. (a) Circular representation of the restriction map of *Ageratum leaf curl betasatellite* (ALCB: KC589700). In this display only enzymes that cleave the sequence once are shown. (b) The linear display of a digest of the *Ageratum leaf curl betasatellite* (ALCB: KC589700).



(a)



(b)

Figure: 4. (a) Circular representation of the restriction map of *Marigold leaf curl alphasatellite* (MLCuA: KC206078) infecting Marigold. In this display only enzymes that cleave the sequence once are shown. (b) The linear display of a digest of the *Marigold leaf curl alphasatellite* (MLCuA: KC206078).

In *Ageratum leaf curl betasatellite* (ALCB: KC589700) the single ORF; C1 (Figure 3a) is flanked by sites PsiI and NdeI (Figure 3b). In case of *Marigold leaf curl alphasatellite* (MLCuA: KC206078) the single ORF Rep (Figure 4 a) is flanked by BarI and Hpy166II (Figure 4b) that could be used to excise them from the complete sequence for inserting into vector. Restriction enzymes have found use in analyzing these genomes of using restriction fragment length polymorphisms (RFLPs) as physical markers or by directly detecting the presence of single nucleotide polymorphisms (SNPs).

Conclusion

Based on the obtained results, the relationships between the G+C content and other genomic features, such as distributions of genes, can be analyzed in a perceivable manner. It shows that GC-Profile would be an appropriate starting point for analyzing the isochore structure of higher eukaryotic genomes, and an intuitive tool for identifying genomic islands in prokaryotic genomes. Restriction mapping is a very useful technique when used for determining the orientation of an insert in a cloning vector, by mapping the position of an off-center restriction site in the insert. Restriction maps were represented in a linear or circular fashion. The tool used analyzed the viral DNA sequence and find the large, non-overlapping open reading frames using the *E.coli* genetic code and the sites for all Type II and commercially available Type III restriction enzymes that cut the sequence just once.

References

1. Gao, F. and Zhang, C. (2006). GC-Profile: a web-based tool for visualizing and analyzing the variation of GC content in genomic sequences. Nucleic Acids Res., 34, W686–W691.
2. Groisman, E.A. and Ochman, H. (1996). Pathogenicity islands: bacterial evolution in quantum leaps. Cell, 87: 791–794.
3. Hacker, J. and Carniel, E. (2001). Ecological fitness, genomic islands and bacterial pathogenicity: a Darwinian view of the evolution of microbes. EMBO Rep, 2: 376–381.
4. Hentschel, U. and Hacker, J. (2001). Pathogenicity islands: the tip of the iceberg. Microbes Infection, 3; 545–548.
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. Koonin, E.V., Makarova, K.S. and Aravind, L. (2001). Horizontal gene transfer in prokaryotes: quantification and classification. Annual Review of Microbiology, 55: 709–742.
7. 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.
8. Mansoor, S., Briddon, R.W., Zafar, Y. and Stanley, J. (2003). Geminivirus disease complexes: An emerging threat. Trends Plant Sci 8: 128-134.
9. Marwal, A., Sahu, A., Sharma, P. and Gaur, R.K. (2013a). Molecular Characterizations of Two Begomoviruses Infecting *Vinca rosea* and *Raphanus sativus* in India. Virologica sinica 28: 53–56.
10. Marwal, A., Sahu, A. K., Choudhary, D.K. and Gaur R.K. (2013b). Complete nucleotide sequence of a begomovirus associated with satellites molecules infecting a new host *Tagetes patula* in India. Virus Genes. 47 (1): 194 – 198.
11. 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.
12. Pingoud, A. and Jeltsch, A. (2001). Structure and function of type II restriction endonucleases. Nucleic Acids Research, 29: 3705–3727.
13. 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.
14. Roberts, R.J., Vincze, T., Posfai, J and Macelis, (2003). REBASE—restriction enzymes and methyltransferases. Nucleic Acids Research, 31: 418–420.
15. 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.
16. Zhang, C.T., Wang, J. and Zhang, R. (2001). A novel method to calculate the G + C content of genomic DNA sequences. J. Biomol. Struct. Dyn. 19, 333–341.

17. Vincze, T., Posfai, J. and Roberts, R.J. (2003). NEBcutter: a program to cleave DNA with restriction enzymes. Nucleic Acids Research, 31: 3688-3691.