Virology 482 (2015) 41–50

Contents lists available at ScienceDirect

Virology

journal homepage: www.elsevier.com/locate/yviro

Potential linkage between compound microsatellites and recombination in geminiviruses: Evidence from comparative analysis

B. George [1], Ch Mashhood Alam [1], R. Vinoth Kumar, Prabu Gnanasekaran, S. Chakraborty *

Molecular Virology Laboratory, School of Life Sciences, Jawaharlal Nehru University, New Delhi 110067, India

ARTICLE INFO

Article history:
Received 14 November 2014
Returned to author for revisions
16 February 2015
Accepted 5 March 2015

Keywords:
IMEx
dMAX
cSSR%
Relative density
Relative abundance
Hotspot
Recombination
Geminivirus

ABSTRACT

The compound microsatellites consist of two or more individual microsatellites, originate from mutation or imperfection in simple repeat sequences. The reports on systematic analysis of the occurrence, size and density of compound microsatellite (cSSR) types are very rare. Our study indicates that cSSRs are clustered at specific regions in the begomovirus genomes. cSSRs were overrepresented in majority of begomovirus genomes indicating that they might have some functional significance. Further, non-random distribution pattern of cSSR in begomovirus genomes was significantly correlated with the recombination breakpoint positions in the genome. The analysis of cSSR regions in the viral genome indicates the presence of stem loop (hairpin) secondary structure. The significance of these findings in biology of geminiviruses is discussed based on our present understanding of recombination and repetitive DNA. To our knowledge, this is the first analysis suggesting the possible association between recombination and microsatellites in any viral genome.

Introduction

Geminiviridae, the second largest family of plant viruses, is characterized by twinned icosahedral particles with monopartite or bipartite single stranded DNA (ssDNA) genome replicating via dsDNA intermediate in the host nucleus (Hanley-Bowdoin et al., 2013). Depending on the insect vector and genomic characteristics, geminiviruses are grouped into seven genera: *Becurtovirus*, *Begomovirus*, *Curtovirus*, *Eragrovirus*, *Mastrevirus*, *Topocuvirus* and *Turncurtovirus* (Adams et al., 2013). Their genomes are characterized by divergent transcriptional units and a 5′ intergenic region containing origin of rolling circle replication (RCR). For bipartite viruses, the two genomic components share a common region containing the origin of replication and regulatory regions for bidirectional transcription. The functions of geminivirus coded genes have been exhaustively studied (Hanley-Bowdoin et al., 2013; Gilbertson et al., 2003; Lazarowitz and Beachy, 1999; Patil and Fauquet, 2009).

Begomoviruses (transmitted by whiteflies, *Bemisia tabaci* Genn.) constitute the largest genus as evident from the number of species available in genome databases and are considered as one of the major constraints of crop production. Based on genome organization, they are further classified as either monopartite (DNA A – like component) or bipartite (DNA A and DNA B components) encoding 5–7 proteins necessary for viral replication, movement and pathogenesis (Fauquet and Stanley, 2003; Fauquet et al., 2008). In addition, monopartite begomoviruses infecting both crops as well as weeds are often associated with satellite and satellite-like molecules named as betasatellites and alphasatellites, respectively (Briddon et al., 2003; Saunders et al., 2008; Chattopadhyay et al., 2008; Kumar et al., 2012; Singh et al., 2012). Recently, association of satellite molecules has also been reported with bipartite begomoviruses (Romay et al., 2010; Sivalingam and Varma, 2012).

In the evolution of ssDNA viruses, the rate of recombination and nucleotide substitution are closer to that in RNA viruses (Drake 1993; Martin et al., 2011). Besides rolling circle replication (RCR), geminiviruses also employ recombination dependent replication (RDR) strategy for their multiplication (Jeske et al., 2001; Preiss and Jeske, 2003). It is known that recombination plays a vital role in the evolution of begomoviruses (Duffy and Holmes, 2008, 2009; Zhou et al., 1997; Monci et al., 2002; García-Andrés et al., 2006). Several studies have also indicated that high frequency of recombination prevails within recombination hotspots in begomoviruses (Ndunguru et al., 2005; Fauquet et al., 2005; García-Andrés et al., 2007).

* Corresponding author. Tel.: +91 11 26704153 (work).
E-mail address: supriyachakrasls@yahoo.com (S. Chakraborty).
[1] These authors contributed equally.

http://dx.doi.org/10.1016/j.virol.2015.03.003
0042-6822/

The simple sequence repeats (SSRs, also called as microsatellites and minisatellites) are tandem repetitions of relatively short motifs of DNA. The microsatellites may be considered as either simple or compound depending on the constituent sequences. Simple microsatellites are ubiquitously present in both prokaryotic (Mrazek et al., 2007) and eukaryotic (Toth et al., 2000) genomes, but are relatively rare in viruses; especially in those with small genomes such as *human immunodeficiency virus* Type 1, potyviruses and geminiviruses (Chen et al., 2009; Zhao et al., 2011; George et al., 2012). The compound microsatellites (cSSR) are composed of two or more microsatellites located adjacent to each other. Some of these cSSRs are functionally important in regulating gene expression in yeast (Struhl, 1985) and higher metazoans including human (Curi et al., 2005; Borrmann et al., 2003). Importance of SSRs in meiotic recombination (Schultes and Szostak, 1991; Gendrel et al., 2000; Kirkpatrick et al., 1999; Treco and Arnheim, 1986), evolution of species (Bowcock et al., 1994), genome mapping (Hong et al., 2010), differentiation of viral strains (Deback et al., 2009), studying population genetics and linkage association (Rosenberg et al., 2002; Abdurakhmonov et al., 2005) have been documented. Several examples of simple sequence repeats have been found in menovirus (Duke et al., 1990; Hahn and Palemberg, 1995), vesicular stomatitis virus (Barr et al., 1997), hepatitis C virus (Yamada et al., 1996) and human respiratory syncytial virus (Garcia-Barreno et al., 1994). Changes in length of tri- and hexanucleotide repeats at the hemagglutinin cleavage site in avian influenza virus have been associated with increased virulence (Perdue et al., 1997). However, the significance of cSSR is not available for any viruses.

Jeffreys et al. (1998) reported the minisatellite-associated recombination hotspot in humans. Recently, the association of SSRs with hotspots of meiotic recombination in yeast has been reported (Bagshaw et al., 2008). It is known that certain DNA sequence such as dinucleotide repeats [d(CG)n and d(TG)n] can form left-handed Z-DNA with hotspot activity (Stringer, 1985) as recombination enzymes have preference for those sequences (Blaho and Wells, 1987; Kmiec and Holloman, 1984, 1986). Available literature suggests that repeat sequences might have a role in regulating recombination in both prokaryotes and eukaryotes (Schultes and Szostak, 1991; Gendrel et al., 2000; Kirkpatrick et al., 1999; Treco and Arnheim, 1986; Murphy and Stringer, 1986). In order to understand the control of location of recombination in geminiviruses, it would be relevant to study the association of microsatellite repeats with the recombination hotspots. Specifically, the abundance, distribution and variation of cSSRs would give an insight into the mutational changes and evolution of repeat sequences.

Results

Divergence in number, relative density and relative abundance of compound microsatellites in geminivirus genomes

Scanning of geminivirus genomes revealed the presence of 0–4 cSSRs in each (Table S3). Analysis of relative density and relative abundance indicated that 50 out of 181 begomovirus DNA-A/DNA-A like did not possess imperfect microsatellite (cSSR, Fig. 1 and Table S3). On an average, one cSSR was present in each genome, suggesting a lower abundance of cSSRs in geminiviral genomes. Analysis of null distribution of cSSR motifs in begomovirus genomes indicated that their occurrence was significantly higher than expected by chance (Table S3). Maximum percentage of cSSR (27.27%) was observed in *Tomato yellow margin leaf curl virus* followed by *Dicliptera yellow mottle Cuba virus*, *Euphorbia leaf curl Guangxi virus* (EuLGxCV) and *Tomato yellow leaf curl Sardinia virus*, in which, 25% of SSRs could be classified as cSSRs. The maximum relative density of cSSRs (30.2 bp/kb) was observed in EuLGxCV genome. The relative abundance of cSSR varied among all selected genomes from 0.00/kb (in 50 DNA A/DNA-A like which lack cSSR) to 1.46/kb in EuLGxCV (Table S3). The presence of cSSR could be ascertained in all curtovirus species analyzed except in *Pepper yellow dwarf virus* genome, whereas in mastrevirus, 5 out of 11 genomes were devoid of cSSR. The relative density and abundance of cSSR in curtovirus were 8.3 and 0.54 and for mastrevirus it was 5.2 and 0.24, respectively. The relative density and abundance of cSSR in the only available topocuvirus species were 11.5 and 0.69, respectively (Table S3). The range of relative abundance and density of cSSR vis-a-vis various species are shown in Fig. 1.

Effect of dMAX on the occurrence of compound microsatellites

To determine the impact of dMAX, we selected five distinct species of geminiviruses with or without cSSRs. The percentage of individual SSR being part of a cSSR (cSSRs %) was introduced as a measure to check cSSR variability with increase in dMAX (Fig. 2). It is noteworthy that the dMAX value can only be set between 0 and 50 for IMEx. A linear increase of cSSR frequency with dMAX can be expected if SSRs are randomly distributed. Interestingly, we did not observe a linear increase, suggesting a non-random distribution of

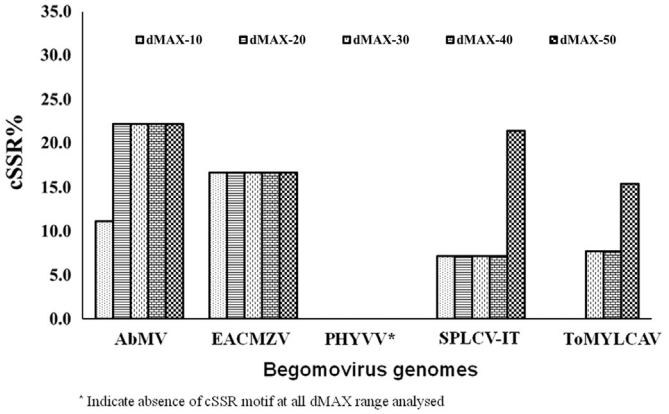

Fig. 2. Effect on cSSR% (percentage of individual microsatellites being part of a compound microsatellite) on varying the dMAX in six different species of begomovirus.

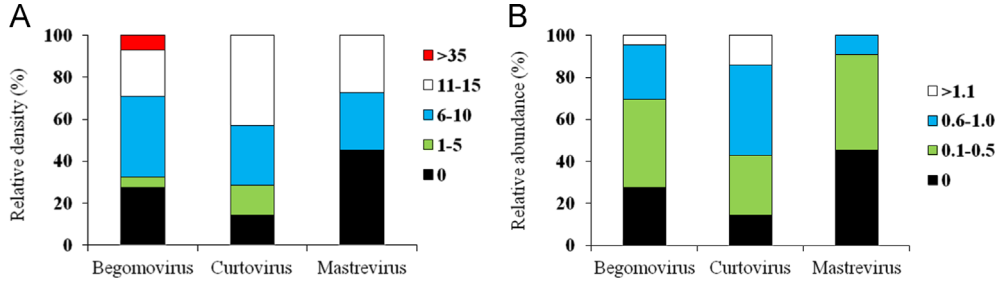

Fig. 1. Relative density (A) and relative abundance density (B) of imperfect microsatellites in geminivirus genomes. Relative density is defined as the total length (bp) contributed by each microsatellite per kb of sequence analyzed whereas relative abundance depicts the number of microsatellites present per kb of the genome.