Molecular evolution and genetic analysis of Mesta yellow vein mosaic virus and associated betasatellites

Rasel Ahmed^{1*}, Rajnee Hasan², Md. Wali Ullah¹, Borhan Ahmed¹

¹Basic and Applied Research on Jute Project, Bangladesh Jute Research Institute, Dhaka, Bangladesh.

²Department of Biochemistry, University of Nebraska-Lincoln, Lincoln, NE, USA.

*** Correspondence:**

Rasel Ahmed

Basic and Applied Research on Jute Project

Bangladesh Jute Research Institute

Manik Mia Avenue, Dhaka-1207, Bangladesh.

Email: rasel@jutegenome.org, rahmed.bg@gmail.com

Abstract

Mesta yellow vein mosaic disease (MYVMD), one of the major diseases circulating mesta growing regions of Indian sub-continent, is responsible for serious yield loss in mesta crops. A complex of monopartite begomovirus, Mesta yellow vein mosaic virus (MYVMV) and associated betasatellite, is reported in several studies as the causal agent of MYVMD. However, all-inclusive molecular evolutionary analysis of so far available MYVMVs and associated betasatellites disseminating in this region is still lacking. In this study, by estimating and analyzing various indexes of population genetics and evolutionary parameters, we discussed the sources of genetic variations, population dynamics and different forces acting on the evolution of MYVMVs and associated betasatellites. The study finds recombination as a vital force in the evolution and diversification of begomovirus complexes in different geographic locations however, betasatellites were found to be exposed to more diverse recombination events compared to MYVMVs. Indian isolates are reported to have high frequency of polymorphism in this study which suggests a balancing selection or expansion occurring in Indian populations of begomoviruses. Higher degree of genetic differentiation and lower rate of gene flow calculated between the viral populations of Bangladesh and Pakistan is justified by the relatively far geographical distance between these two countries. Although the study detects overall purifying selection, the degrees of constraints acting on individual gene tested are found different. Coat protein (AV1) is estimated with very high nucleotide substitution rate which is very likely to result from the strongest purifying selection pressure ($d_N/d_S = 0.131$) calculated in this study on coat protein. The findings of this study on different evolutionary forces that shape the emergence and diversification of MYVMVs and associated betasatellites may provide directions towards future evolutionary trend analysis and development of comprehensive disease control strategies for begomoviruses.

Keywords: evolution; recombination; Begomovirus; betasatellite

Introduction

Mesta is one of the cheapest and important bast fiber producing crops cultivated largely in Indian sub-continent. For being biodegradable and renewable in characteristics, the bast fibers are drawing remarkable attention worldwide to address several environmental issues. Two species of mesta namely *Hibiscus cannabinus* (kenaf) and *Hibiscus sabdariffa* (roselle) of the malvaceous family are cultivated mainly for producing lingo-cellulosic bast fibers which are used to manufacture traditional packaging materials such as rope, sacking, canvas, cordage, carpet backing etc. Beside this, the leaves and the seeds of both species are being used as vegetable and rich source of oil respectively and for having lower proportion of unsaturated fatty acids, the oil has potential medicinal value [1,2]. Moreover, mesta, particularly kenaf (*Hibiscus cannabinus*) is now gaining global interest as a non-woody source of quality pulp for tree-free newspaper industry.

The cultivation of mesta crops is highly susceptible to a disease called mesta yellow vein mosaic disease (MYVMD) which is alarmingly affecting and spreading the large mesta growing area in the sub-continent. The MYVMD is characterized by a typical “yellow vein network” showed in the leaf veins followed by chlorosis of the leaves. The bast fiber yield of the diseased plants is seriously affected due to the stunted growth, complete chlorosis and defoliation of the leaves in the advanced stage of the disease [3,4]. The disease was documented first time in West Bengal state of India which is located adjacent to Bangladesh [5]. Several districts, particularly Kishoreganj in Bangladesh also came under sporadic attack of yellow mosaic few years ago and the affected areas are increasing in a yearly basis (<http://www.newagebd.net/article/18783/yellow-mosaic-virus-attacks-jute-crop>). A monopartite begomovirus, Mesta yellow vein mosaic virus (hereafter called MYVMV for all the Mesta yellow vein mosaic viruses isolated from mesta plants) and an isolate of Cotton leaf curl Multan betasatellite are reported to be associated with MYVMD in eastern India [6,7]. A similar disease of mesta was also found to be associated with a monopartite begomovirus and a betasatellite in the northern part of India [3]. Some mesta leaf samples from northern India showing yellow mosaic symptoms were carried out for diagnosis and the identified virus particles showed similarity with another reported begomovirus Mesta yellow vein mosaic Bahraich virus and an isolate of Ludwigia leaf distortion betasatellite [8].

Begomovirus (transmitted by whiteflies, *Bemisia tabaci*) is the largest genus under the family Geminiviridae and are widely distributed geographically. Begomoviruses have either monopartite (DNA-A-like component) or bipartite (DNA-A and DNA-B) genomes [9] containing circular, single-stranded DNA (ssDNA) of approximately 2.7 Kbp. Typically, DNA-A has six open reading frames (ORFs) of AV1(coat protein), AV2 (pre-coat protein), AC1 (replication initiation protein or rep protein), AC2 (transcriptional activator protein), AC3 (replication enhancer protein) and AC4 (AC4 protein). Among these 6 ORFs AV1 and AV2 are on the virion-sense strand and rest are on the complementary-sense strand. Monopartite begomoviruses are often co-existed with satellite termed as betasatellites [9]. These betasatellite molecules depend on the helper begomoviruses for their replication and are approximately 1.3 Kbp in size containing one major ORF (β C1) on complementary strand. The β C1 ORF has been characterized to be involved in virus

movement, suppressing host antiviral silencing as a pathogenicity determinant [10,11] and symptom induction [12]. In evolutionary aspects, nucleotide substitution and recombination rate in the ssDNA viruses are quite similar to that of RNA viruses [13]. Recombination is recognized as a vital process in the evolution of begomoviruses [14,15,16] and some studies in begomoviruses found high rate of recombination within recombination hotspots [17,18,19]. On the other hand, compound microsatellites (cSSRs) are reported to be potentially linked with the recombination of begomoviruses [20,21]. The simple sequence repeats or microsatellites (SSRs) are tandem repeats of short DNA sequences and considered as cSSRs when two or more closely located SSRs are separated by non-repeat sequences. The microsatellites have been extensively studied in viruses, prokaryotes, and eukaryotes [22,23,24] and found to be involved in gene regulation and transcription processes [25,26].

In this study, we collected kenaf leaf samples with yellow mosaic symptoms from different regions of Bangladesh and identified the associated begomovirus particles. To our knowledge, this is the first report from Bangladesh for identification and molecular characterization of MYVMD associated begomovirus complexes in kenaf. Several sporadic investigations were carried out to detect, isolate and characterize the causing viral agents of MYVMD in India [3,4,5,6,7,8,27,28,29]. However, accumulated molecular variability analysis of MYVMVs and associated betasatellites considering various evolutionary parameters are yet to be studied. Hence, with an aim to address this issue, we reported and discussed the sources of genetic variations, population dynamics and different forces acting on the evolution of MYVMVs and betasatellites. For these analyses, we considered all the available MYVMV and betasatellite isolates so far (isolated from mesta plants or associated with mesta plants) in GenBank including the new isolates from Bangladesh (from this study). Moreover, we determined the potential linkage between the recombination events and presence of cSSR motifs in the studied begomovirus complexes. In a practical context, our findings can be useful to future epidemiological and evolutionary trend analyses towards development of comprehensive disease control strategies for begomoviruses.

Materials and Methods

Sample Collection and Viral Sequences

Naturally infected kenaf leaf samples showing typical yellow mosaic symptoms were collected from 7 kenaf producing districts of Bangladesh namely Faridpur, Gazipur, Jashore, Kishoreganj, Manikganj, Rangpur and Tangail. At least three samples were collected from each location. All the infected leaf samples were transported in an ice box to the laboratory and preserved at -80°C freezer until DNA isolation. The genomic DNA was isolated following the method described by Ghosh et al. 2009 [30] with necessary modifications. The complete genomic DNA of begomoviruses were PCR (Polymerase Chain Reaction) amplified with a set of forward (5'-GAGGGTGAGCTCTCCGAAGGTTGTAG-3') and reverse (5'-AATTATGAGCTCTGGGAG-TGCGCAAG-3') primers using DreamTaq DNA Polymerase (Thermo Fisher Scientific, USA). The isolated genomic DNAs were also subjected to amplify full length betasatellites using forward (5'-AGTAAGGGTACCACTACGCTACGCAG-3') and reverse (5'-ATTAATGGTACCTACC-

CTCCCAGGGGTAC-3') primer pair. Both the amplified begomovirus genomes and betasatellites were individually cloned into the pGEM®-T Easy Vector Systems (Promega Corporation, USA). The colonies with desired clones were selected based on restriction patterns and three clones per isolate were sequenced bi-directionally for both begomovirus genomes and betasatellites using sanger sequencing facilities at Basic and Applied Research on Jute Project, Manik Mia Avenue, Dhaka, Bangladesh. The raw sequencing reads as chromatographs were assembled and edited with Geneious Prime software program.

Phylogenetic Analysis, Recombination and Microsatellite Detection

In addition to the begomovirus and betasatellite sequences reported in this study, 28 MYVMV (23 from India and 5 from Pakistan) and 38 betasatellite (28 from India and 10 from Pakistan) sequences were downloaded from the GenBank and the details are presented in Table S1. All the sequences of downloaded MYVMVs and betasatellites were either isolated from or associated with mesta plants. MUSCLE alignment program was used to align the sequences [31] and Sequence Demarcation Tool (SDT v1.2) [32] was utilized to obtain the pairwise identities between the sequences. MUSCLE generated alignment for both the sequences were used to generate phylogenetic trees in molecular evolutionary genetics analysis version X (MEGA-X) [33] using a maximum-likelihood method with 1000 bootstrap replicates. Possible recombination events were detected with seven methods namely RDP, GENECONV, Bootscan, MaxChi, Chimaera, SiScan, and 3Seq using RDP4 (Recombination Detection Program version 4) [34]. Recombination predictions were only considered when the events were confirmed by at least 4 out of the 7 methods. Imperfect Microsatellite Extractor Version 2.1 (G-IMEx) was used to identify the presence of microsatellite in the viral sequences along with position [35] using the 'Advance-Mode' parameters as described in Chen et al. 2012 [36].

Population Demography Analysis

Basic parameters of genetic variability and sequence diversity such as the number of variable sites (s), total number of mutations (η), nucleotide diversity per site (π), average number of nucleotide differences (k), the number of haplotype (h) and haplotype diversity (Hd) were calculated using DnaSP v6 [37]. To explore the demographic history of MYVMVs and associated betasatellites, several neutrality tests incorporated in DnaSP v6 such as Tajima's D, Fu and Li's D* and Fu and Li's F* were implemented. The Tajima's D test is based on the differences between the number of segregating sites and the average number of nucleotide differences whereas the Fu and Li's D* and Fu and Li's F* tests were based on the differences between the number of singletons with the total number of mutations and the average number of nucleotide differences between pairs of sequences respectively. Both the complete genome sequences and gene sequences (coat protein and rep protein of MYVMV and β C1 protein of betasatellite) were considered for all these analyses.

Estimation of Selection Pressure

The acting evolutionary constraints or selection pressure on the coding regions of the viral populations were calculated based on the ratio of the nonsynonymous substitution rate (d_N) to the synonymous substitution rate (d_S). The nucleotide sequences of coat protein and rep protein of MYVMV and β C1 protein of betasatellite were used to calculate d_N/d_S ratio using single-likelihood ancestor counting (SLAC) method available in the DataMonkey server [38]. Next, a series of methods such as fixed-effects likelihood (FEL), SLAC and fast unbiased Bayesian approximation (FUBAR) incorporated in the DataMonkey server were utilized to find the sites under selection. The only positively selected sites identified by at least two of these methods were considered as under diversifying selection.

Population Genetic Differentiation and Gene Flow Analysis

Permutation-based statistical tests K_s^* and Z^* [39] were considered to evaluate the extent of genetic differentiation among the viral populations. For measuring gene flow between the populations tested, F_{ST} and N_m values were taken into consideration. F_{ST} is calculated based on the mean number of differences between sequences sampled from the same subpopulation and the mean number of differences between sequences sampled from the two different subpopulations. N_m is based on the estimate of average divergence time of pairs of genes sampled from within a subpopulation and the average divergence time of genes sampled from different subpopulations [40]. All of these tests were calculated in DnaSP v6 considering the complete genome sequences and gene sequences of coat protein, rep protein and β C1 protein.

Estimation of Nucleotide Substitution Rates

The rates of nucleotide substitution per site were estimated with the Bayesian Markov Chain Monte Carlo (MCMC) method available in the BEAST 1.10.4 package [41]. The nucleotide sequences of coat protein, rep protein and β C1 protein were used as input for BEAST analysis. For all the data set, both strict and relaxed (uncorrelated lognormal) molecular clocks were utilized along with the demographic model of piece-wise Bayesian skyline plot. Moreover, two substitution models (HKY and GTR) were used to generate BEAST files for each clock options (strict and relaxed). The resultant log files from BEAST analysis were analyzed into Tracer v1.6 to perform model comparison analyses (AICM analysis) and the best-fitting models were determined according to the AICM values. To ensure convergence, sufficient length of MCMC chains were run where an initial 10 % of the MCMC chains discarded as burn-in. The rates of nucleotide substitution per site were estimated from the best-fitting models of each data set using Tracer v1.6 program.

Results

Viral Sequence Analysis

Complete monopartite begomovirus genomes were PCR amplified with MYVMV specific primer pair from all the infected kenaf leaf samples collected from 7 districts of Bangladesh. However, betasatellite specific primer pairs were able to detect betasatellite genome in all the samples except the leaves collected from Rangpur district. All of these 13 amplified PCR products were cloned and sequenced for further analyses. The assembled contigs were pairwise aligned with known begomovirus and betasatellite sequences. The assembled monopartite genome sequences shared maximum nucleotide identities of 98.80 to 99.20% with MYVMVs, GenBank accession number NC_009088 and EF373060. Therefore, all the isolates of begomovirus reported in this study are identified as MYVMV isolates according to the demarcation threshold of begomovirus species [42]. On the other hand, the begomovirus associated betasatellite sequences shared maximum nucleotide identities of 96.90 to 98.30% with MYVMV associated betasatellites, GenBank accession number NC_009903 and DQ298137. As a result, all the isolates of begomovirus associated betasatellites reported in this study are identified as MYVMV associated betasatellite isolates according to the ICTV guidelines for betasatellite demarcation [43]. The pairwise comparison analysis of MYVMVs and betasatellites were presented in Figure 1a and Figure 1c respectively. The total 13 sequences reported in this study were deposited, assigned with accession number (MZ448274- MZ448286) and released in the GenBank. The details of this isolates with accession numbers are provided in Table S2. In addition to 7 MYVMVs and 6 associated betasatellites reported in this study, 28 MYVMV and 38 betasatellite sequences were downloaded from the GenBank and considered for further analysis.

Phylogenetic Relationship

As a part of the investigation of the genetic basis of evolution in MYVMV and associated betasatellites, we studied overall sequence variations and generated phylogenetic trees. MYVMV tree (Figure 1b) produced two major branches. All the MYVMV isolates from Pakistan were clustered into clade I under branch A while the Indian isolates were grouped into both the branches. Branch B was solely consisted of Indian isolates and all the Bangladeshi isolates are gathered under the branch A. Five of the 7 Bangladeshi isolates formed an independent clade however, rest two did not fall under any specific group. The phylogenetic tree constructed based on betasatellites was also divided into two major branches (Figure 1d). Branch A was consisted of Bangladeshi, Indian and Pakistani isolates whereas only Indian isolates occupied the branch B. As like MYVMV tree, all the betasatellites from Pakistan were clustered under single clade, IV while betasatellites isolated from Bangladesh were separated into clade I with some Indian isolates.

Recombination Analysis

Detection of recombination events and putative parental sequences were carried out with all the studied isolates of MYVMV and betasatellite. A total of 15 unique recombination events were detected in MYVMV isolates by at least 4 number of methods (Table 1). The analysis detected 20 isolates out of 35 as recombinants and among them, 12 isolates were found to have multiple events of recombination (Table 1 and Table S3). Maximum 4 events were possessed by each of the 3 South Indian isolates FJ159269, FJ159270 and FJ159271. All the isolates from Pakistan were detected as recombinants where 3 of them possessed one and rest 2 were contained with two recombination events. No isolate from Bangladesh was found as recombinants however, 3 of them (MZ448274, MZ448277 & MZ448280) were recognized as the parental sequences of both Indian and Pakistani recombinants (Table 1 and Table S3). Both inter-species and intra-species recombination events were detected and one of the parental sequences was found unknown in case of 8 events (Table 1).

In betasatellites, 23 out of 44 isolates were found to be recombinants possessing 15 unique recombination events (Table 1 and Table S4). Multiple events of recombination were contained by 7 isolates and Indian isolate KT390489 possessed 3 recombination events which is the highest among the betasatellites (Table S4). MZ448282 is the only Bangladeshi isolate found to be recombinant having another Bangladeshi isolate (MZ448285) as the major parent. The number of non-recombinant isolates was 9 and 7 out of 28 Indian and 10 Pakistani isolates respectively. The recombination events were found inter-species for 10 events and intra-species for 5 events. However, the detection program was unable to precisely determine the putative major parents in 7 recombination events and the number was 4 in case of minor parents.

Association Between cSSR Motifs and Recombination Events

To determine the presence and position of cSSR motifs in the viral genomes, we performed sequence analysis using G-IMEx. The analysis detected cSSR motifs in all the genomes except one betasatellite isolate MG587933 from Pakistan. The number of cSSRs present per genome ranged from 1-4 for MYVMV and 1-3 for betasatellites (Figure S1a and Table S4). Two cSSRs were present in more than 60% isolates of both of the genomes and 4 cSSRs was only detected in KT390482, betasatellite from Uttar Pradesh, India. Most abundant types of cSSRs are (CG)₃-x1-(T)₇ & (AT)₃-x1-(T)₇ in MYVMVs while (GA)₃-x3-(TTG)₄-x2-(TG)₃ and (A)₉-x1-(A)₆ are the prevailing types in betasatellites (Figure S1b).

Next, the association between the presence of recombination events and the cSSR motifs were studied. To observe the possible link of their co-existence, we counted the isolates that possessed both the cSSRs and recombination events, any of them or none of them. We analyzed that the presence of both the recombination event and cSSR were noticed in 57.14% of MYVMVs and 50% of betasatellites (Figure 2a). On contrary, 42.86% MYVMVs and 47.73% betasatellites were

Table 1. Recombination analyses of MYVMVs and associated betasatellites

Isolates	Event	Recombinant	Major parent	Minor parent	p-value ¹	Detection methods ²
MYVMV	1	MN005041	MZ448280	Unknown	1.000×10^{-39}	RGBMCS3
	2	MH538339	MZ448280	Unknown	3.489×10^{-36}	RGBMCS3
	3	FJ159267/EU360303/ FJ159268/NC_010818	Unknown	MZ448280	2.372×10^{-35}	RGBMCS3
	4	FJ159271/ FJ159269/ FJ159270	FJ345399	FJ159262	6.125×10^{-30}	RGBMCS3
	5	FJ159269/ FJ159270/ FJ159271	Unknown	MZ448274	5.304×10^{-19}	RGBMCS3
	6	HE578897	MH628534	FJ159265	1.966×10^{-13}	RGBMCS3
	7	FJ159269/ FJ159265/ FJ159270FJ159271	Unknown	MZ448277	1.642×10^{-14}	RGS3
	8	JX181786/FN645922/ KJ462083/MN005041	FJ159262	FJ159267	4.355×10^{-07}	RMCS3
	9	FJ345398/ FJ345399/ FJ345400	MZ448274	KR013745	1.000×10^{-05}	RMCS3
	10	KJ462083/ FN645922/ JX181786/MN005041	FJ345398	FJ159267	1.288×10^{-08}	RGMCS3
	11	KR013745/MH628534	MZ448277	Unknown	9.888×10^{-11}	RMCS3
	12	KT948076	KR013745	Unknown	2.459×10^{-13}	RGMCS3
	13	KR013745/ KT948076	MH628534	FJ345399	1.471×10^{-03}	RGBM3
	15	FJ345399/ FJ345398/ FJ345400	Unknown	EF428256	4.566×10^{-03}	RGS3
	16	FJ159271/ FJ159269/ FJ159270	EU360303	FJ159266	7.885×10^{-04}	RGMCS3
Betasatellite	1	KT390482	Unknown	EF614161	5.941×10^{-28}	RGBMCS3
	2	KT390489/ KT390482	Unknown	NC_028130	2.680×10^{-9}	RGBMCS3
	3	FJ159274/ EF614158/ EF614159/ FJ159274	MZ448282	KT390489	2.935×10^{-14}	RGBMCS3
	4	EF614158/ EF614159	DQ298137	NC_028130	6.247×10^{-12}	RGBMCS3
	5	KT390489	Unknown	KT390482	1.051×10^{-13}	RGBMCS3
	6	NC_028130/ KT390486	EF614161	Unknown	1.016×10^{-10}	RBMCS3
	7	MK358823	MZ448286	EU557374	6.088×10^{-8}	RGM3
	8	KT390489	Unknown	EU880233	1.901×10^{-14}	RGBMCS
	9	LT716981	Unknown	EF614161	3.132×10^{-7}	RGMCS3
	10	FJ159275	FJ159274	Unknown	2.805×10^{-5}	RMC3
	11	KT390479/ EF614161/F614162/EU557374/ FJ159272/ FJ159273/NC_010569/ KT390486 /NC_028130	KT948077	FJ159274	3.430×10^{-8}	RGMCS
	13	MF683831	Unknown	FR772083	1.706×10^{-8}	RGBMCS3
	14	EU825205/ EU825206/EF620566	Unknown	MG587941	2.006×10^{-4}	RBMCS3
	15	MZ448282	MZ448285	Unknown	6.311×10^{-8}	RGMCS3
	20	MG587933	MG587941	Unknown	2.249×10^{-4}	RMC3

Note. ² The lowest p-value detected by one of the methods.

¹ R-RDP; G-GENCONV; B-BOTSCAN; M-MAXCHI; C-CHIMERA; S-SISCAN; 3-3SEQ.

found to have cSSR motifs lacking recombination events in their genomes. However, this case is not found that both the cSSR and recombination event are missing. Moreover, we attempted to detect the trend of this relationship in the putative parental sequences identified by recombination detection program. Results showed that 61.54% parental sequence of MYVMVs and 47.73% of betasatellites were consisted of both the cSSRs and recombination events (Figure 2b) whereas 38.46% and 50% parental sequences were possessed cSSRs without any recombination event in MYVMVs and betasatellites respectively. Nevertheless, no recombination events were detected in the isolates lacking cSSR motifs.

Figure 2. Association between recombination events and microsatellites detected in MYVMVs and betasatellites. **a**, The correlation between the cSSRs and recombination events detected in the viral genomes. ‘+’ and ‘-’ sign in the parenthesis indicate the presence or absence of cSSRs and recombination events detected in viral genomes. **b**, The correlation between the cSSRs and recombination events in the putative parental genomes identified in recombination detection software RDP4. ‘+’ and ‘-’ sign in the parenthesis indicate the presence or absence of cSSRs and recombination events detected in viral genomes. **c**, Physical proximity of the cSSRs and recombination breakpoints. Distances were calculated between each cSSR and the nearest recombination breakpoint in viral isolates which possess both cSSRs and recombination events.

Figure 3. Genome-wide distribution and location of cSSR motifs and recombination breakpoints in MYVMV and betasatellite genomes. Schematic representation of a linearized MYVMV and betasatellite genome is placed over the respective graphical representation with similar scaling. In Schematic representation horizontal bars represents the genes, A-rich regions and conserved regions where applied. Horizontal arrow represents the directions of respective segment while vertical arrow represents the position of cSSR motifs. **a**, The distribution of total cSSR motifs and recombination breakpoints detected in the MYVMVs with schematic representation. AV1, coat protein; AV2, pre-coat protein; AC1, replication initiation protein or rep protein; AC2, transcriptional activator protein; AC3, replication enhancer protein and AC4, AC4 protein. **b**, The distribution of total cSSR motifs and recombination breakpoints detected in the betasatellites with schematic representation. β C1, β C1 protein and SCR, satellite conserved region.

Further, to find the location of cSSR motifs and recombination events we mapped their position onto the linearized genomes of MYVMV and betasatellite. The results indicated that, the recombination break points were distributed throughout the length of the MYVMV and betasatellite genomes compared to the position of cSSR motifs, although the majority of the break points in betasatellites were gathered in the satellite conserved region (SCR) (Figure 3a and Figure 3b). The cSSRs were found to be mostly located in the N-terminal rep protein and intergenic region, and non-coding and A-rich region of MYVMVs and betasatellites respectively. Furthermore, we sought to explore the correlation between cSSR motifs and recombination events in terms of their physical proximity which was calculated by measuring the nucleotide distance between each cSSR and the nearest recombination breakpoint if both were present in respective genome. Interestingly the majority of the cSSR motifs were found to be located within 300 bases of the nearest recombination break points in both MYVMVs and betasatellites (Figure 2c).

Genetic Diversity and Population Dynamics

Nucleotide diversity was found high both in MYVMVs ($\pi = 0.09$) and betasatellites ($\pi = 0.15$) detected in DnaSP v6 software program (Table 2). The number of haplotypes were detected 33 & 40 along with 0.997 & 0.996 haplotype diversity in MYVMV and betasatellite genomes respectively. When considered geographically, the genetic variability was highest ($\pi = 0.10$ for MYVMVs and $\pi = 0.18$ for betasatellites) in Indian isolates. Pakistani isolates also possessed higher degree of variations whereas, very less nucleotide diversity ($\pi = 0.006$ for MYVMV and $\pi = 0.017$ for betasatellites) was observed in the isolates of Bangladesh. We also measured the genetic diversity in the genes coded for coat & rep protein of MYVMVs and β C1 protein of betasatellites. Similar trends of nucleotide diversity were found in gene analysis compared to whole genome analysis i.e. Indian isolates scored highest value of π while lowest value was achieved by Bangladeshi isolates (Table 2). The haplotypes number was found higher in coat protein compared to rep protein and the haplotype diversity was found 0.992, 1.000 & 0.976 for coat, rep and β C1 protein respectively.

To test the hypothesis of neutral selection between lineages and populations of MYVMV and betasatellite, several neutrality tests such as Tajima's D, Fu and Li's D* and Fu and Li's F* were considered. Evolutionary events such as population expansion, bottlenecks and selection are speculated by Tajima's D while Fu and Li's D* tests are associated with population demographic expansion. All the values of these tests were found negative except the Fu and Li's D* values in MYVMV populations when considered the complete genome sequences (Table 3). However, in country wise consideration, all the three values were calculated positive in MYVMV populations from India. Whether cumulative or country wise analyses were done, all the betasatellites produced negative values in all three indices in case of both whole genome and β C1 gene sequences (Table 3). The Fu and Li's D* and Fu and Li's F* values for coat and rep protein genes were estimated positive for Indian MYVMV populations in which values for coat protein genes were found statistically significant.

Table 2. Genetic diversity of MYVMV and associated betasatellite isolates in different populations

Sequences considered	Viral component	Regions	Sample size	Number of site	Genetic variability parameters					
					s	η	π	k	h	Hd
Whole genome sequence	MYVMV	BD	7	2753	49	51	0.006±0.001	17.76	7	1.000±0.076
		IND	23	2682	856	942	0.100±0.012	268.99	21	0.992±0.015
		PAK	5	2717	419	463	0.076±0.014	205.70	5	1.000±0.126
		All	35	2639	965	1182	0.091±0.011	238.84	33	0.997±0.008
	Betasatellite	BD	6	1334	59	60	0.017±0.002	22.73	6	1.000±0.096
		IND	28	1667	604	849	0.180±0.023	188.99	24	0.989±0.012
		PAK	10	1406	287	331	0.075±0.009	94.09	9	0.978±0.054
		All	44	1689	638	942	0.153±0.019	150.01	40	0.996±0.006
Gene sequence	AV1	BD	7	771	9	9	0.004±0.001	2.86	7	1.000±0.076
		IND	23	743	242	282	0.124±0.020	91.98	19	0.980±0.020
		PAK	5	771	92	99	0.057±0.012	44.10	5	1.000±0.126
		All	35	743	275	341	0.104±0.018	76.95	31	0.992±0.010
	AC1	BD	7	1089	22	23	0.007±0.002	8.14	7	1.000±0.076
		IND	23	1086	340	391	0.090±0.011	97.36	13	0.941±0.027
		PAK	5	1083	191	203	0.085±0.019	92.30	5	1.000±0.126
		All	35	1080	401	488	0.084±0.010	91.22	25	0.975±0.014
	β C1	BD	6	357	6	6	0.006±0.001	2.00	6	1.000±0.096
		IND	28	394	200	263	0.160±0.037	50.71	14	0.944±0.020
		PAK	10	357	44	46	0.035±0.004	12.47	9	0.978±0.054
		All	44	394	205	277	0.129±0.027	40.10	29	0.976±0.010

Note. s, Number of variable sites; η , Total number of mutations; π , Nucleotide diversity (per site); k, Average number of nucleotide differences; h, Number of Haplotypes and Hd, Haplotype diversity. AV1, coat protein; AC1, replication initiation protein or rep protein and β C1, β C1protein. BD, Bangladesh; IND, India and PAK, Pakistan.

Population Genetic Differentiation and Gene Flow

Independent statistical tests of population differentiation K_s^* and Z^* were performed using DnaSP v6 between geographical origins of isolates (Bangladesh, India and Pakistan) of MYVMV and betasatellite. The K_s^* and Z^* values were comparatively higher when analyzing the Indian isolates with Bangladeshi or Pakistani isolates while comparatively lower values were calculated when comparing Bangladeshi with Pakistani isolates (Table 4). The completely similar tendency was found whether full genomes or gene sequences were considered during the analyses. To find out the evidence of gene flow between those populations the number of effective migrants (N_m) values were calculated along with another genetic differentiation parameter F_{ST} . Considering both genomes and gene sequences, the F_{ST} values ranged from 0.18603 to 0.54216 and 0.20264 to 0.73537 for MYVMV and betasatellite isolates respectively (Table 4). Regarding gene flow, the number of effective migrants (N_m) values were calculated <1 for all the data set except the AC1 gene sequences when comparing isolates of Bangladesh with Indian populations. Taking all the value trends under consideration, higher genetic differentiation and lower genetic flow were noticed between the isolates of Bangladesh and Pakistan.

Table 3. Neutrality tests of MYVMV and associated betasatellite isolates in different populations

Sequences considered	Viral component	Regions	Sample size	Tajima's D	Fu and Li's D*	Fu and Li's F*
Whole genome sequence	MYVMV	BD	7	-0.84865	-0.67052	-0.78647
		IND	23	0.22007	0.97331	0.86378
		PAK	5	-0.56878	-0.36589	-0.44331
		All	35	-0.64225	0.30508	-0.03319
	Betasatellite	BD	6	-0.86664	-0.80082	-0.89404
		IND	28	-0.67317	-0.20628	-0.42727
		PAK	10	-0.97988	-0.81145	-0.96531
		All	44	-1.14060	-0.53817	-0.91052
Gene sequence	AV1	BD	7	-1.18108	-1.32889	-1.41888
		IND	23	0.82744	1.67898**	1.65640*
		PAK	5	-0.54691	-0.40299	-0.46942
		All	35	-0.26901	0.91907	0.59838
	AC1	BD	7	-0.74918	-0.60431	-0.70261
		IND	23	-0.32941	0.33409	0.14666
		PAK	5	-0.40234	-0.28493	-0.33599
		All	35	-0.87793	-0.20287	-0.52280
	βC1	BD	6	-1.36732	-1.39992	-1.48974
		IND	28	-0.97946	-0.38526	-0.68764
		PAK	10	-1.13759	-1.25278	-1.38224
		All	44	-1.36372	-0.55617	-1.02007

Note. AV1, coat protein; AC1, replication initiation protein or rep protein and βC1, βC1protein. BD, Bangladesh; IND, India and PAK, Pakistan; BD, Bangladesh; IND, India and PAK, Pakistan. * p<0.05, ** p<0.02

Table 4. Genetic differentiation and gene flow analyses of MYVMV and associated betasatellite isolates in different populations

Viral Component	Population	Genetic differentiation		Gene flow	
		Ks*	Z*	F _{ST}	Nm
MYVMV genome sequence	BD-IND	4.66688	4.87185	0.21595	0.91
	BD-PAK	3.68327	2.70111	0.54216	0.21
	IND-PAK	5.11954	4.77762	0.29490	0.60
AV1 Gene sequence	BD-IND	3.35434	4.83831	0.24454	0.77
	BD-PAK	2.19763	2.70969	0.53150	0.22
	IND-PAK	3.82111	4.81786	0.26398	0.70
AC1 Gene sequence	BD-IND	3.63016	4.94258	0.18603	1.09
	BD-PAK	2.83008	2.72351	0.52229	0.23
	IND-PAK	4.08841	4.76933	0.26358	0.70
Betasatellite genome sequence	BD-IND	4.64418	5.22571	0.25336	0.74
	BD-PAK	3.92163	3.05172	0.64430	0.14
	IND-PAK	4.65535	5.21305	0.26528	0.69
βC1 Gene sequence	BD-IND	3.05795	5.26652	0.20264	0.98
	BD-PAK	2.00942	3.04262	0.73537	0.09
	IND-PAK	3.13289	5.18609	0.27281	0.67

Note. AV1, coat protein; AC1, replication initiation protein or rep protein and βC1, βC1protein. BD, Bangladesh; IND, India and PAK, Pakistan.

Selection Pressure Analysis

To detect the natural selection dominating on the coat, rep and β C1 protein, the ratio between the non-synonymous and synonymous substitution frequencies (d_N/d_S) were calculated in Datamonkey web server. The d_N/d_S ratio of coat protein was 0.131 which was approximately half compared to rep protein indicating coat protein under strong constrained for diversification (Table 5). The d_N/d_S ratio of the β C1 protein of betasatellites was calculated 0.266 which was slightly higher than rep protein and twice of coat protein of MYVMVs. Selection constraints operating on specific sites of the analyzed proteins were identified with three independent methods. General results reflected the d_N/d_S ratios (overall negative selection) showing a large number of sites under purifying selection than positive selection (Table 5). Only one site was found under positive selection in coat and rep protein by two methods whereas, no site was detected under diversifying selection by more than one method in β C1 protein. Not a single site was selected by SLAC method under positive selection however, the number of sites under negative selection detected by SLAC were comparatively lower than the sites detected by FEL and FUBAR (Table 5).

Table 5. Selection pressure analyses in the selected genes of MYVMVs and associated betasatellites

Genes	d _N /d _S	Sites under positive selection ¹			Number of negatively selected sites ²
		By maximum likelihood method		By Hierarchical Bayesian Method FUBAR	
		FEL	SLAC		
AV1	0.131	29	None	29	107/59/106
AC1	0.231	8, 182, 347, 351	None	8	111/65/110
βC1	0.266	94	None	32	26/15/23

Note. AV1, coat protein; AC1, replication initiation protein or rep protein and β C1, β C1protein. ¹ The positively selected amino acid residues detected by at least two methods are given by bold letters. ² The number of negatively selected sites detected FEL, SLAC and FUBAR, respectively

Table 6. Estimation of evolutionary rates in the selected genes of MYVMVs and associated betasatellites

Parameters	Coat protein (AV1)	Rep protein (AC1)	β C1 protein (β C1)
Best-fit molecular clock model	Relaxed (lognormal)	Relaxed (lognormal)	Strict
Best-fit substitution model	GTR	GTR	GTR
No. of sequences	35	35	44
Sequence length (nt)	771	1101	394
Chain length (millions)	80	60	100
Mean substitution rate (nucleotides/site/year)	5.99×10^{-3}	2.59×10^{-4}	5.39×10^{-6}
HPD substitution rate (nucleotides/site/year)	1.34×10^{-3} to 1.11×10^{-2}	4.18×10^{-6} to 6.68×10^{-4}	7.15×10^{-23} to 3.40×10^{-5}

Note. HPD, 95% high probability density

Estimation of Evolutionary Rates

Two substitution models HKY and GTR along with both strict and relaxed molecular clocks were used to detect nucleotide substitution rate nevertheless, GTR was selected as best-fitting model in model comparison for all the data sets. Relaxed molecular clock model was found better over strict clock in coat and rep protein data sets while strict molecular clock model was analyzed best-fitting in β C1 protein (Table 6). Notably, Bayesian skyline was used as demographic model for each data set and uncorrelated lognormal parameter was used for relaxed molecular clock model. The mean rate of substitution for coat protein was found 5.99×10^{-3} (95% HPD 1.34×10^{-3} to 1.11×10^{-2}) nucleotide substitutions/site/year which is higher than the rep protein estimated mean rate of 2.59×10^{-4} with 95% HPD 4.18×10^{-6} to 6.68×10^{-4} nucleotide substitutions/site/year (Table 6). The β C1 protein has been estimated with 5.39×10^{-6} and 7.15×10^{-23} to 3.40×10^{-5} nucleotide substitutions/site/year as the mean rate and 95% HPD respectively pointing out the lower rate of substitution of betasatellite protein than two MYVMV proteins.

Discussion

Mesta plant, which has two cultivated species *Hibiscus cannabinus* (kenaf) and *Hibiscus sabdariffa* (roselle), beside its' traditional role as the producer of natural fiber is a promising alternative source of quality pulp for tree-free paper manufacturing industries addressing the climate changing issues. The MYVMD, characterized with a typical yellow vein mosaic symptom, is one of the major factors for hindering the ultimate yield of mesta plants [3,4]. MYVMV and several betasatellites were proved to be the causal agents of the disease [6,7,8]. The mesta crops cultivated in Bangladesh are also affected by this mosaic disease seriously however, no study was reported to detect and characterize the causal agents of the disease in Bangladesh. In this circumstances, beside the isolation and identification of the causing agents of MYVMD from the infected samples of Bangladesh, we attempted to perform detailed molecular and evolutionary study in available MYVMVs and betasatellites associated with mesta plants including new isolates from Bangladesh.

The begomovirus particles isolated from 7 leaves samples of Bangladeshi kenaf plants were identified as MYVMV according to the demarcation threshold of begomovirus species [42] and the betasatellites isolated from 6 of these samples were identified as MYVMV associated betasatellites according to the ICTV guidelines for betasatellite demarcation [43]. The betasatellite specific primer did not able to amplify any product from the sample collected from Rangpur districts of Bangladesh. It was evident that in the absence of a betasatellite monopartite begomovirus could evolve [44]. Surprisingly, the MYVMV isolated from Rangpur district (MZ448279) was not clustered with other Bangladeshi isolates under clade IV in the phylogenetic analyses rather it was placed independently in the tree (Figure 1b). The isolates from eastern part of India were grouped in clade III and VI under branch A while MYVMVs from North and South India were separated into clade VII and VIII under major branch B. Notably, all the Pakistani isolates disregarding MYVMVs or betasatellites were set exclusively under single clade (Figure 1b and Figure 1d). However, isolates of Bangladesh were gathered under clade I with some Indian

isolates in betasatellite tree. It can be speculated from the phylogenetic tree analyses that geographic location has had major influence in the evolution of MYVMVs and associated betasatellites. Although 3 Tomato leaf curl Joydebpur DNA betasatellites (EU880232, EU880233 and EU431115) along with Croton yellow vein mosaic betasatellite (KT390489) were isolated from *Hibiscus cannabinus*, they were distinctly separated in the betasatellite tree forming a major branch B (Figure 1d).

Recombination was proved to be a crucial factor in the evolution of viral species under Geminiviridae family [45,46]. In accordance, more than 50% MYVMV isolates were detected as recombinants in this study (Table 1 and Table S3). The majority of the recombination break points were located within the region of AV1/AV2 and AC1 genes (Figure 3a). The identification of AV1/AV2 and AC1 regions as the recombination hotspots is in agreement with some previous reports [20,21]. Recombination events 2,6,11 and 12 were solely detected in Pakistani isolates while the rest of the maximum events were distributed among more than one Indian isolates (Table 1). Interestingly no Bangladeshi isolates were found as recombinant nevertheless, 3 of them (MZ448274, MZ448277 & MZ448280) were recognized as the parental sequences of both Indian and Pakistani recombinants suggesting the ancient and conservative lineages of MYVMV in Bangladesh. In case of betasatellites also, more than 50% of isolates were found to be recombinants possessing 15 unique recombination events (Table 1 and Table S4). Event 9,13 and 20 were solely shared by Pakistani isolates however, recombinant LT716981 was detected to have an Indian isolate as minor parent. Recombination event 11 was shared by 9 Indian isolates which was maximum per event and interestingly a Pakistani isolate KT948077 was identified as the major parent of this event (Table 1). The overall results indicated more diversified recombination in betasatellites compared to MYVMVs.

In recent time, the cSSR motifs were reported to be potentially linked with the recombination of begomoviruses [20,21]. Consequently, we detected the presence and position of cSSR motifs in the studied data sets and tried to figure out their relation with recombination breakpoints. The analyses indicated that almost 100% of isolates that were detected as recombinants also found to have cSSR motifs in their genomes (Figure 2a). Similar results were found when considered the putative parents of the recombination events in MYVMVs and betasatellites (Figure 2b). Moreover, the presence of recombination events in the absence of cSSR motifs in particular isolates is very rare which postulated the possible link between their co-existence and these observations are consistent with the findings of Kumar et al 2017. The genome wide distribution of cSSRs and recombination breakpoints of MYVMVs and betasatellites (Figure 3a and Figure 3b) depicted their appearance throughout the length of the genome however, cSSR motifs were mostly confined to N-terminal rep protein and intergenic region, and non-coding and A-rich region of MYVMVs and betasatellites respectively. In addition to the genome wide mapping, we calculated the physical proximity between each cSSR motif and the nearest recombination breakpoint and excitingly observed that 72% motifs in MYVMVs and 58% in betasatellites were located within 300 bases of corresponding break points (Figure 2c). All of these results are in well

accordance with the previous studies of George et al. 2015 [21] and Kumar et al. 2015 [47] in begomoviruses. The possible reasons behind this association were speculated in some reports where these repetitive sequences were supposed to be an inducing agent triggering crossing-over or a source of signal for recombination [21,48,49] however, these ideas are yet to be validated experimentally.

Low genetic variability tends to be common in natural populations of plant viruses [50] however, in this study the overall genetic diversity was found very high in both whole genome and gene based study of MYVMVs and betasatellites with nucleotide diversity ranging from 0.084 to 0.153 (Table 2) when calculated for the whole population. These values were found much more higher compared to other begomoviruses [51,52]. In geographical consideration, the highest diversity was shown by Indian isolates ranging from 0.090 to 0.180 compared to Pakistan and Bangladesh. Bangladeshi isolates showed comparatively very low variability with nucleotide diversity ranging from 0.004 to 0.017. In congruence with the recombination analyses of this study, betasatellites were observed to maintain more genetic diversity among its' isolates compared to MYVMVs and this finding is supported by the fact that higher genetic variability may be linked with more frequent and diversified recombination [51,52].

Three statistical tests, Tajima's D, Fu and Li's D* and Fu and Li's F* suggested the molecular diversity patterns between lineages and populations. Apparently all the three values were found negative in the populations of Bangladesh and Pakistan (Table 3) indicating low frequency of polymorphism maintained in the populations of these two regions caused either by background selection, genetic hitchhiking or recent expansion. In contrast, Indian populations produced positive values in all three indices (only found negative Tajima's D in rep protein) whether considered whole genome or gene sequences of MYVMVs. Moreover, the Fu and Li's D* and Fu and Li's F* values of coat protein were found much higher and statistically significant indicating polymorphism with high frequency which suggested that balancing selection or expansion has occurred in Indian populations. Two independent statistical tests Ks* and Z* values were calculated to measure population differentiation between three regions. In all the data sets, Ks* and Z* values were found lower in Bangladesh-Pakistan compared to Bangladesh-India and later one produced nearly similar values of India-Pakistan data sets suggesting high genetic differentiation between the populations of Bangladesh and Pakistan (Table 4). To estimate the degree of gene flow between geographical origins, the number of effective migrants (Nm) values were estimated along with F_{ST} which is a genetic differentiation index correlated with Nm. Generally, lower gene migration was observed in the populations with high degree of differentiation [53]. In accordance with the previous results based on Ks* and Z*, higher value of differentiation (F_{ST}) was calculated with lower degree of gene flow (Nm) in all the Bangladesh-Pakistan data sets (Table 4). All these tests reinforce the finding that the populations of Bangladesh-Pakistan data set are distantly related with each other compared to Bangladesh-India and India-Pakistan. One possible explanation for this correlation could be the relatively far geographical distance between Bangladesh and Pakistan.

In this study, the d_N/d_S ratios of coat, and rep proteins of MYVMVs and β C1 protein of betasatellites were calculated <1 indicating a negative or purifying selection, albeit a different degree of selection pressure over each gene was observed (Table 5). The results suggested a stronger selection constraints operating on coat protein compared to rep protein which was exposed to comparatively weak selection pressure. In accordance with the overall purifying selections identified by d_N/d_S ratios, the number of individual sites under negative selection outnumbered the positively selected sites in all the data types and this result is supported by the earlier reports which emphasized the role of purifying selection on begomovirus evolution [14,46,52]. Notably, for having structural constraints and maintaining specific molecular interactions, many plant viruses were found to have purifying selection as the main evolutionary force [50]. No site under diversifying selection was detected by SLAC which is known to be a more conservative method to detect sites under selection [54]. The amino acid 8, only positively selected site in rep protein (Table 5), was part of the N-terminal region comprising of different conserved motifs required for virus replication [55]. The estimated nucleotide substitution rates were found very high in coat and rep protein compared to β C1 protein (Table 6). The observation of high substitution rate in begomovirus coat protein is consistent with some other studies [14,15,20]. We found the strongest purifying selection pressure ($d_N/d_S = 0.131$) in coat protein and its' high nucleotide substitution rate is very likely to result from this strong negative selection acting on coat protein. The possible reasons for achieving such a high nucleotide substitution rate by begomoviruses might include the usage of error prone DNA polymerases in virus replication [14,56] or spontaneous mutations.

Conclusions

This study reports the MYVMVs and betasatellites associated with MYVMD of kenaf plants in Bangladesh along with their molecular characterization. To our knowledge, this is the first report from Bangladesh on MYVMD associated begomovirus complex. This study also describes detailed molecular evolutionary analyses of available MYVMVs and associated betasatellites based on different parameters. The findings generated through our studies reinforces the fact that recombination plays an important role in the evolution and diversification of begomoviruses and betasatellites. A potential linkage is also demonstrated between the recombination events and presence of cSSR motifs. In addition to stating high degree of evolutionary rate and genetic variability found in MYVMVs and betasatellites, the study describes the relation between various index of population genetics addressing the geographic locations of the isolates. The findings of this study on different evolutionary forces that shape the emergence and diversification of MYVMVs and associated betasatellites may provide directions for future evolutionary study of different species of begomoviruses.

Author Contributions

Rasel Ahmed conceived and designed the study, performed the experiments, performed bioinformatics analyses and wrote the manuscript. Rasel Ahmed, Borhan Ahmed, Md. Wali Ullah and Rajnee Hasan contributed to data analyses and graphics representation. Borhan Ahmed, Md. Wali Ullah and Rajnee Hasan contributed to critical reviewing and editing the manuscript. Borhan Ahmed and Md. Wali Ullah supervised the study.

Conflicts of Interest

The authors declare no conflicts of interest.

References

1. Duke, J.A. Isthmian Ethnobotanical Dictionary. Scientific Publisher, Jodhpur, India, 1986, pp. 100–102.
2. Pal, D.C.; Jain, S.K. Tribal Medicine. Naya Prakash, Calcutta, India, 1998, pp. 149–150.
3. Ghosh, R.; Paul, S.; Roy, A.; Mir, J.I.; Ghosh, S.K.; Srivastava, R.K.; Yadav, U.S. Occurrence of begomovirus associated with yellow vein mosaic disease of kenaf (*Hibiscus cannabinus*) in north India. *Plant Health Progr.* 2007. doi:10.1094/PHP-2007-0508-01-RS
4. Chatterjee, A.; Roy, A.; Ghosh, S.K. Acquisition, transmission and host range of a begomovirus associated with yellow vein mosaic disease of mesta (*Hibiscus cannabinus* and *Hibiscus sabdariffa*). *Australas. Plant Pathol.* 2008, 37, 511–519.
5. Chatterjee, A.; Roy, A.; Padmalatha, K.V.; Malathi, V.G.; Ghosh, S.K. Occurrence of a Begomovirus with yellow vein mosaic disease of mesta (*Hibiscus cannabinus* and *Hibiscus sabdariffa*). *Aust Plant Pathol.* 2005, 34:609–610.
6. Chatterjee, A.; Ghosh, S.K. A new monopartite begomovirus isolated from *Hibiscus cannabinus* L. in India. *Arch. Virol.* 2007, 152, 2113–2118.
7. Chatterjee, A.; Ghosh, S.K. Association of a satellite DNA molecule with mesta yellow vein mosaic disease. *Virus Genes*, 2007b, 35, 835–844.
8. Das, S.; Ghosh, R.; Paul, S.; Roy, A.; Ghosh, S.K. Complete nucleotide sequence of a monopartite begomovirus associated with yellow vein mosaic disease of mesta from north India. *Arch. Virol.* 2008, 153, 1791–1796.
9. Brown, J.K.; Fauquet, C.M.; Briddon, R.W.; Zerbini, F.M.; Moriones, E.; Navas-Castillo, J. in *Virus Taxonomy 9th Report of the International Committee on Taxonomy of Viruses*, ed. by King, A.M.Q.; Adams, M.J.; Carstens, E.B.; Lefkowitz, E.J. (Elsevier Academic Press, London, 2012), pp. 351–373.
10. Li, F.; Huang, C.; Li, Z.; Zhou, X. Suppression of RNA silencing by a plant DNA virus satellite requires a host calmodulinlike protein to repress RDR6 expression. *PLoS Pathog.* 2014, 10, e1003921.

11. Yang, X.; Xie, Y.; Raja, P.; Li, S.; Wolf, J. N.; Shen, Q.; Bisaro, D. M.; Zhou, X. Suppression of methylation-mediated transcriptional gene silencing by bC1-SAHH protein interaction during geminivirusbetasatellite infection. *PLoS Pathog.* 2011, 7, e1002329.
12. Briddon, R. W.; Stanley, J. Subviral agents associated with plant single-stranded DNA viruses. *Virology*, 2006, 344, 198–210.
13. Martin, D.P.; Lemey, P.; Posada, D. Analysing recombination in nucleotide sequences. *Mol. Ecol. Resour.* 2011, 11, 943–955.
14. Duffy, S.; Holmes, E.C. Phylogenetic evidence for rapid rates of molecular evolution in the single-stranded DNA begomovirus Tomato yellow leaf curl virus (TYLCV). *J. Virol.* 2008, 82, 957–965.
15. Duffy, S.; Holmes, E.C. Validation of high rates of nucleotide substitution in geminiviruses: phylogenetic evidence from East African cassava mosaic viruses. *J. Gen. Virol.* 2009, 90, 1539–1547.
16. García-Andrés, S.; Monci, F.; Navas-Castillo, J.; Moriones, E. Begomovirus genetic diversity in the native plant reservoir *Solanum nigrum*: evidence for the presence of a new virus species of recombinant nature. *Virology*, 2006, 350 (2), 433–442.
17. Ndunguru, J.; Legg, J.P.; Aveling, T.A.; Thompson, G.; Fauquet, C.M. 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. *Virol. J.* 2005, 22, 2–21.
18. Fauquet, C.M.; Sawyer, S.; Idris, A.M.; Brown, J.K. Sequence analysis and classification of apparent recombinant begomoviruses infecting tomato in the Nile and Mediterranean basins. *Phytopathology*, 2005, 95, 549–555.
19. García-Andrés, S., Tomás, D.M., Sánchez-Campos, S., Navas-Castillo, J., Moriones, E., Frequent occurrence of recombinants in mixed infections of Tomato yellow leaf curl disease-associated begomoviruses. *Virology*, 2007, 365, 210–219.
20. Kumar, R.V., Prasanna, H.C., Singh, A.K. et al. Molecular genetic analysis and evolution of begomoviruses and betasatellites causing yellow mosaic disease of bhendi. *Virus Genes*, 2017, 53, 275–285. <https://doi.org/10.1007/s11262-016-1414-y>
21. George, B.; Alam, C.M.; Kumar, R.V.; Prabu, G.; Chakraborty, S. Potential linkage between compound microsatellites and recombination in geminiviruses: Evidence from comparative analysis. *Virology*, 2015 148, 41–50.
22. Valot, B.; Knapp, J.; Umhang, G.; Grenouillet, F.; Millon, L. Genomic characterization of EmsB microsatellite loci in *Echinococcus multilocularis*, *Infection, Genetics and Evolution*, 2015, Volume 32, Pages 338–341, <https://doi.org/10.1016/j.meegid.2015.03.040>.
23. Alam, C.M.; Singh, A.K.; Sharfuddin, C.; Ali, S. Genome-wide scan for analysis of simple and imperfect microsatellites in diverse carlaviruses, *Infection, Genetics and Evolution*, 2014, Volume 21, Pages 287–294, <https://doi.org/10.1016/j.meegid.2013.11.018>

24. Alam, C.M.; George, B.; Sharfuddin, C.; Jain, S.K.; Chakraborty, S. Occurrence and analysis of imperfect microsatellites in diverse potyvirus genomes. *Gene*, 2013, 521(2), 238–244. <https://doi.org/10.1016/j.gene.2013.02.045>
25. Mrazek, J.; Guo, X.; Shah, A. Simple sequence repeats in prokaryotic genomes. *Proc. Natl. Acad. Sci.* 2007, USA 104, 8472–8477.
26. Kashi, Y.; King, D.G. Simple sequence repeats as advantageous mutators in evolution. *Trends Genet.* 2006, 22(5):253-9. doi: 10.1016/j.tig.2006.03.005
27. Chatterjee, A.; Sinha, S.K.; Roy, A.; Sengupta, D.N.; Ghosh, S.K. Development of diagnostics for DNA A and DNA β of a Begomovirus associated with mesta yellow vein mosaic disease and detection of geminiviruses in mesta (*Hibiscus cannabinus* L. and *H. sabdariffa* L.) and some other plant species. *J. Phytopathol.* 2007, 155: 683-689.
28. Das, S.; Roy, A.; Ghosh, R.; Paul, S.; Acharyya, S.; Ghosh, S.K. Sequence variability and phylogenetic relationship of betasatellite isolates associated with yellow vein mosaic disease of mesta in India. *Virus Genes*, 2008, 37, 414–424.
29. Roy, A.; Acharyya, S.; Das, S.; Ghosh, R.; Paul, S.; Srivastava, R.K.; Ghosh, S.K. Distribution, epidemiology and molecular variability of the begomovirus complexes associated with yellow vein mosaic disease of mesta in India, *Virus Research*, 2009, 141(2), 237-246. <https://doi.org/10.1016/j.virusres.2008.11.022>.
30. Ghosh, R.; Paul, S.; Ghosh, S.K.; Roy, A. An improved method of DNA isolation suitable for PCR-based detection of begomoviruses from jute and other mucilaginous plants, *Journal of Virological Methods*, 2009, 159(1), 34-39. <https://doi.org/10.1016/j.jviromet.2009.02.020>
31. Edgar, R.C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, 2004, 32(5):1792-1797 DOI: 10.1093/nar/gkh340
32. Muhire, B.M.; Varsani, A.; Martin, D.P. SDT: A Virus Classification Tool Based on Pairwise Sequence Alignment and Identity Calculation. *PLoS ONE*, 2014, 9(9), e108277. doi:10.1371/journal.pone.0108277
33. Kumar, S.; Stecher, G.; Li, M.; Knyaz, C.; Tamura, K. MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* 2018, 35, 1547–1549.
34. Martin, D.P.; Murrell, B.; Golden, M.; Khoosal, A.; Muhire, B. RDP4: Detection and analysis of recombination patterns in virus genomes, *Virus Evolution*, 2015, 1(1), vev003, <https://doi.org/10.1093/ve/vev003>
35. Mudunuri, S.B.; Nagarajaram, H.A. IMEx: imperfect microsatellite extractor. *Bioinformatics*, 2007, 23, 1181–1187.
36. Chen, M.; Tan, Z.; Zeng, G.; Zhuotong, Z. Differential distribution of compound microsatellites in various human immunodeficiency virus type 1 complete genomes. *Infect. Genet. Evol.* 2012, 12, 1452–1457.
37. Rozas, J.; Ferrer-Mata, A.; Sánchez-DelBarrio, J.C.; Guirao-Rico, S.; Librado, P.; Ramos-Onsins, S.E.; Sánchez-Gracia, A. DnaSP 6: DNA Sequence Polymorphism Analysis of Large Datasets. *Mol. Biol. Evol.* 2017, 34, 3299-3302. DOI: 10.1093/molbev/msx248

38. Weaver, S.; Shank, S.D.; Spielman, S.J.; Li, M.; Muse, S.V.; Pond, S.L.K. Datamonkey 2.0: A Modern Web Application for Characterizing Selective and Other Evolutionary Processes, *Molecular Biology and Evolution*, 2018, 35(3), 773–777, <https://doi.org/10.1093/molbev/msx335>
39. Hudson, R.R.; Boos, D.D.; Kaplan, N.L. A statistical test for detecting geographic subdivision. *Mol Biol Evol.* 1992, 9, 138–151.
40. Hudson, R.R.; SLATKIN, M.; MADDISON, W.P. Estimation of levels of gene flow from DNA sequence data. *Genetics*, 1992, 132, 583-589.
41. Suchard, M.A.; Lemey, P.; Baele, G. Ayres DL, Drummond AJ, Rambaut A. Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evol.* 2018, 4(1), vey016. doi:10.1093/ve/vey016
42. Brown, J.K.; Zerbini, F.M.; Navas-Castillo, J.; Moriones, E.; Ramos-Sobrinho, R.; Silva, J.C.; Fiallo-Olivé, E.; Briddon, R.W.; Hernández-Zepeda, C.; Idris, A.; Malathi, V.G.; Martin, D.P.; Rivera-Bustamante, R.; Ueda, S.; Varsani, A. Revision of Begomovirus taxonomy based on pairwise sequence comparisons. *Arch Virol.* 2015, 60(6), 1593-619. doi: 10.1007/s00705-015-2398-y.
43. Briddon, R.W.; Brown, J.K.; Moriones, E.; Stanley, J.; Zerbini, M.; Zhou, X.; Fauquet, C.M. Recommendations for the classification and nomenclature of the DNA-beta satellites of begomoviruses. *Arch Virol.* 2008, 153(4), 763-81. doi: 10.1007/s00705-007-0013-6.
44. Melgarejo, T.A.; Kon, T.; Rojas, M.R.; Paz-Carrasco, L.; Zerbini, F.M.; Gilbertson, R.L. 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.* 2013, 87(10), 5397-5413. doi:10.1128/JVI.00234-13
45. Lefeuvre, P.; Moriones, E. Recombination as a motor of host switches and virus emergence: geminiviruses as case studies. *Curr Opin Virol.* 2015, 10, 14-9. doi: 10.1016/j.coviro.2014.12.005
46. Kraberger, S.; Kumari, S.G.; Hamed, A.A.; Gronenborn, B.; Thomas, J.E.; Sharman, M; Harkins, G.W.; Muhire, B.M.; Martin, D.P.; Varsani, A. Molecular diversity of Chickpea chlorotic dwarf virus in Sudan: high rates of intra-species recombination - a driving force in the emergence of new strains. *Infect Genet Evol.* 2015, 29, 203-15. doi: 10.1016/j.meegid.2014.11.024.
47. Kumar, R.V.; Singh, A.K.; Singh, A.K.; Yadav, T.; Basu, S.; Kushwaha, N.; Chattopadhyay, B.; Chakraborty, S. Complexity of begomovirus and betasatellite populations associated with chilli leaf curl disease in India. *J Gen Virol.* 2015, 96(10), 3143-3158. doi: 10.1099/jgv.0.000254
48. Jeffreys, J.; Murray, J.; Neumann, R. High-resolution mapping of crossovers in human sperm defines a minisatellite-associated recombination hotspot. *Mol. Cell*, 1998, 2, 267–273.
49. Brandström, M.; Bagshaw, A.T.; Gemmell, N.J.; Ellegren, H. The relationship between microsatellite polymorphism and recombination hot spots in the human genome. *Mol Biol Evol.* 2008, 25(12), 2579-87. doi: 10.1093/molbev/msn201

50. García-Arenal, F.; Fraile, A.; Malpica, J.M. Variability and genetic structure of plant virus populations. *Annu. Rev. Phytopathol.* 2001, 39, 157–186.
51. Sobrinho, R.R.; Xavier, C.A.D.; Pereira, H.M.B.; Lima, G.S.A.; Assunção, I.P.; Mizubuti, E.S.G.; Duffy, S.; Zerbini, F.M. Contrasting genetic structure between two begomoviruses infecting the same leguminous hosts. *J Gen Virol.* 2014, 95(11), 2540-2552. doi: 10.1099/vir.0.067009-0.
52. Rocha, C.S.; Castillo-Urquiza, G.P.; Lima, A.T.; Silva F.N.; Xavier, C.A.; Hora-Júnior, B.T.; Beserra-Júnior, J.E.; Malta, A.W.; Martin, D.P.; Varsani, A.; Alfenas-Zerbini, P.; Mizubuti, E.S.; Zerbini, F.M. Brazilian begomovirus populations are highly recombinant, rapidly evolving, and segregated based on geographical location. *J Virol.* 2013, 87(10), 5784-99. doi: 10.1128/JVI.00155-13.
53. Schmidt, A.M.; Jacklyn, P.; Korb, J. Isolated in an ocean of grass: Low levels of gene flow between termite subpopulations. *Mol. Ecol.* 2013, 22, 2096–2105.
54. Pond, S.L.; Frost, S.D.; Muse, S.V. HyPhy: Hypothesis testing using phylogenies. *Bioinformatics* 2005, 21, 676–679.
55. Hanley-Bowdoin, L.; Bejarano, E.R.; Robertson, D.; Mansoor, S. Geminiviruses: masters at redirecting and reprogramming plant processes. *Nat Rev Microbiol.* 2013, 11(11), 777-88. doi: 10.1038/nrmicro3117.
56. Richter, K.S.; Götz, M.; Winter, S.; Jeske, H. The contribution of translesion synthesis polymerases on geminiviral replication. *Virology*, 2016, 15(488), 137-148. doi: 10.1016/j.virol.2015.10.027.