Plant Omics Journal POJ

POJ 10(02):88-96 (2017) ISSN:1836-3644
doi: 10.21475/poj.10.02.17.pne391

Genomic sequence characterization of Begomovirus infecting soybean and molecular evolutionary genomics of Legume yellow mosaic viruses (LYMVs)

Shunmugiah V. Ramesh*[#1], Bhagat S. Chouhan[#1,2], Girish K. Gupta[1], Syed M. Husain[1], Suresh Chand[2]

[1]ICAR-Indian Institute of Soybean Research (ICAR-IISR), Indore 452 001, Madhya Pradesh, India
[2]School of Life Sciences, Devi Ahilya Vishwavidhyalaya (DAVV), Indore 452 001, Madhya Pradesh, India

***Corresponding author: ramesh.sv@icar.gov.in**
[#]Authors have contributed equally to this work

Abstract

Begomoviruses infecting legumes (family *Geminiviridae*) pose a serious threat to the cultivation of grain legumes. Eventhough legume yellow mosaic viruses (LYMVs) cause significant loss in yield of legumes, studies regarding evolutionary lineage analysis of LYMVs are very rare. Previously, we have shown that Mungbean yellow mosaic virus (MYMV) and Mungbean yellow mosaic India virus (MYMIV) are major begomoviruses causing yellow mosaic disease (YMD) of soybean in India. In this study, complete genome sequence of begomovirus causing yellow mosaic disease of soybean in Central Indian region was characterized. Furthermore, whole genome sequences of legume begomoviruses [DNA A (108 isolates) and DNA B (89 isolates)] were analyzed to infer genetic diversity, gene flow and evolutionary lineage using nucleotide sequence-based computational approaches. Analysis of nucleotide diversity disclosed that LYMV population as a whole is diverse compared to MYMV and MYMIV. Test of neutral evolution also reiterates the operation of purifying selection and population expansion of MYMV and MYMIV. However, LYMVs as a whole, show decrease in population size and act of balancing or neutral selection. Genetic differentiation studies reveal greater diversity between MYMV and MYMIV. Frequent gene flow was detected between Dolichos yellow mosaic virus (DoYMV), Rhynchosia yellow mosaic India virus (RhYMIV) and other LYMVs. Recombinant events have been detected among LYMV species suggesting frequent genetic exchanges. Molecular phylogeny also revealed distinctness of Old World begomoviruses as New World begomoviruses formed a separate basal cluster. Hence, it is concluded that genetic exchanges are recorded among the LYMVs, and implications of breaching this seclusion is also discussed.

Keywords: Begomovirus; Evolutionary genomics; Genetic diversity; Population genetics; Soybean.
Abbreviations: DoYMV_Dolichos yellow mosaic virus; HgYMV_Horsegram yellow mosaic virus; KuMV_Kudzu mosaic virus; LYMV_Legume yellow mosaic virus; MYMIV_Mungbean yellow mosaic India virus; MYMV_Mungbean yellow mosaic virus; RCA_Rolling circle amplification; RDP_Recombination detection program; RhYMIV_ Rhynchosia yellow mosaic India virus; RhYMV_Rhynchosia yellow mosaic virus; YMD_Yellow mosaic disease; YMV_Yellow mosaic virus

Introduction

Legume infecting begomoviruses belong to family *Geminiviridae* (Fauquet et al., 2008). Legume begomoviruses are transmitted by *Bemisia tabaci* and cause yellow mosaic disease (YMD) that severely hamper production potential of legumes including soybean (Varma and Malathi 2003). Economic loss caused due to the infection of yellow mosaic viruses in legumes has been estimated to the tune of 300 million US $ (Varma and Malathi 2003). YMD of legumes in tropical, South East Asia is caused due to the following species of Begomoviruses, *Mungbean yellow mosaic virus* (MYMV), *Mungbean yellow mosaic India virus* (MYMIV), *Dolichos yellow mosaic virus* (DoYMV) and *Horsegram yellow mosaic virus* (HgYMV), *Kudzu mosaic virus*(KuMV), *Rhynchosia yellow mosaic virus* (RhYMV), *Rhynchosia yellow mosaic India virus* (RhYMIV). These begomoviruses are collectively called as legume yellow mosaic viruses (LYMVs) (Fauquet and Stanley 2003; Qazi et al., 2007; Briddon et al., 2010). Further, legumes are also infected with other begomoviruses such as Soybean chlorotic blotch virus, Soybean mild mottle virus (Alabi et al., 2010), Bean chlorosis virus (BYCV) and Bean white chlorosis mosaic virus (BWCMV) (Fiallo et al., 2013).

Genomes of legume yellow mosaic viruses, have two ssDNA genomic components *ie*) DNA-A and DNA-B each approximately of 2750nts in length. Virus genome encodes proteins in virion-sense and complementary-sense strands (Qazi et al., 2007). DNA A encodes for coat protein (AV1) and pre-coat proteins (AV2) in its virion-sense and genes Rep (AC1-replication associated protein), TrAP (AC2-Transcriptional activator protein), REn (AC3-replication enhancer protein), AC4 and AC5 in its complementary sense strand. Nuclear shuttle protein (NSP) encoded by ORF BC1 is involved in intracellular transport of viral ssDNA whereas ORF BV1 encoded movement protein (MP) is implicated in cell to cell movement of the virus nucleoproteins (Briddon et al., 2010). Thus DNA-A encodes for proteins involved in encapsidation, replication and gene expression whereas, DNA-B encodes for the proteins involved in virus movement. Soybean supplies more than 25% of edible oil requirement of the country and hence, is a major oil seed crop that has great

export potential of de-oiled cake (DOC) (Annual Report DSR 2013-14). The sudden outbreak of YMD caused due to yellow mosaic virus in the central Indian region during crop season –(*kharif* 2015) was a serious setback to meet the demand for soy crop and its products (Soybean News, (2015). In this study, genome sequence of the yellow mosaic virus causing YMD in soybean in Central India has been characterized (Supplementary fig.S1). Despite the significant yield losses due to LYMVs, information regarding genetic diversity, population selection and evolutionary lineage analysis of YMVs are absent. In order to fill this knowledge gap, global analysis of genomic components of legume yellow mosaic viruses (LYMVs) was conducted to study their molecular evolutionary genomics.

Results

Annotating whole genome sequence features

Symptomatic soybean leaves were initially diagnosed for the presence of yellow mosaic virus. PCR based detection of partial AV1 gene (encoding coat protein) was carried out to confirm the presence of yellow mosaic virus infecting soybean. PCR amplification of the coat protein region with the primer set MYMIV F and YMV R yielded an amplicon of ~391bp indicating MYMIV infection in symptomatic plant leaves (Ramesh et al., 2016). Multiply primed Rolling circle amplification (RCA) resulted in amplification of viral genome as multiple copy, high molecular weight, concatamers of circular DNA. Restriction digestion analysis of RCA derived DNA, using selected endonucleases (*Bam*HI, *Eco*RI, *Hin*dIII, *Pst*I), revealed ~2.7 Kb fragments corresponding to unit length of either DNA A or DNA B. The restriction enzyme digestion of RCA DNA did not divulge any satellite DNA components of the virus. Sequence characterization and annotation showed DNA A genome of MYMIV infecting soybean in Central Indian region was 2750nts in length (GenBank Acc. no. KC852204). DNA A component has two ORFs in the viral sense (AV2 and AV1) coding for pre-coat protein and coat protein genes. Presence of AV2 ORF is a characteristic feature of Old World begomovirus (Stanley et al., 2005). In its complementary sense, DNA A genome has 5 ORFs. Among them the protein products encoded by ORFs AC1, AC2 and AC3 were annotated as replication-associated protein, transcriptional-activator protein (TrAP) and replication-enhancer protein (REn) respectively. Nucleotide blast analysis of the complete DNA A revealed that it exhibits 99% sequence identity to MYMIV isolates (EU523045 and DQ389153) infecting soybean and cowpea respectively.

DNA B genomic component of the virus was found to be 2671nts in length (GenBank Acc. no. KP828155). Analysis of DNA B genome revealed two ORFs one each in both the DNA strands. ORF BV1 in the viral strand encodes for nuclear shuttle protein (NSP) whereas ORF BC1 in the complementary sense strand encodes for movement protein (MP). Similarly, BLASTn search for homologous nucleotide sequences revealed 97% sequence identity to DNA B derived from MYMIV infecting soybean (EU523045) and 96% identity to MYMIV isolates infecting cowpea (AY939925) and kidney bean (KC019305).

Genetic polymorphism and molecular phylogenetic relationships

In order to study the nucleotide sequence diversity and to perform haplotype analysis complete genome sequences of LYMVs were obtained from GenBank database (including sequences from this study) and were employed as query in DnaSP software (Librado and Rosas 2009) (Supplementary Table 1 and Supplementary Table 2). Nucleotide diversity (π) analysis of complete DNA A genome revealed negligible diversity between MYMIV (π=0.03875) and MYMV (π= 0.03103) (Table 1). However, LYMV population as a whole revealed relatively high diversity values (π=0.15089). The higher nucleotide diversity of LYMV population could be attributed to relatively low diversity levels of RhYMIV (0.00024) and HgYMV (0.02074). Considering the absolute number of polymorphic sites present in viral genomes, MYMIV (*S* =624) showed high polymorphism than MYMV (S =222). However, considering number of virus isolates under analysis, DoYMV (*S* =309 for a sample size of N=07), KuYMV (*S* =222 for a sample size of N=03), and RhYMV (*S* =192; N=03) showed more segregating sites in virus genome (Table 1). Thus, LYMV population as whole was found to show very high (*S* =898) number of segregating sites due to its high nucleotide diversity (π=0.15089). Haplotype analysis was performed to identity single nucleotide polymorphism variants within species, genotypes and in whole population. The uniqueness of haplotypes present in the virus population was inferred from the parameter haplotype diversity (Hd). High level of haplotype diversity in the population and among the genotypes was observed along with a notable exception of RhYMIV (Hd= 0.667) which showed relatively low level of haplotype diversity (Table 1). Analysis of DNA B genomic components of LYMVs, deduced that nucleotide diversity (π) was found to be high in MYMV (π=0.16591) than MYMIV (π=0.06009) despite comparable number of isolates of MYMIV and MYMV under study. The observed nucleotide diversity of MYMV DNA B has also been corroborated with relatively high no. of polymorphic sites (S= 627). Number of polymorphic sites was found to be more in DNA B genomic components of MYMV, MYMIV, HgYMV, and KuYMV even though DNA B of entire LYMV population displayed relatively low polymorphic sites (Table 1).

Molecular phylogeny of DNA A and DNA B genomic components of LYMVs was inferred employing MEGA 6 (Tamura et al 2013). Phylogeny of DNA A revealed that the cluster-I comprise major LYMVs (MYMIV, MYMV) along with HgYMV isolates forming a basal, distinct sub-clades. Within the cluster-I, MYMIV and MYMV are represented as sub-clades arising from one large branch. Other LYMVs, such as RhYMIV, RhYMV and KuMV formed separate clades basal to HgYMV. Interestingly, DoYMV formed a separate, basal cluster to all the above mentioned LYMVs (Fig. 1). Phylogeny reconstruction of DNA B genomic components revealed similar pattern with notable exception that DoYMV formed part of a cluster formed by RhYMIV, RhYMV and KuMV (Fig. 2). LYMVs such as SoCSV, RhGMV, formed a separate, basal cluster to all other LYMVs indicating their genetic distinctness (Fig. 1 and Fig. 2)

Viral population and selection pressure

Analysis of population statistic parameters to test the theory of neutral evolution revealed that all species of LYMVs except RhYMIV and RhYMV showed negative Tajima's D indicating the operation of purifying selection and population expansion (Table 2). Among the virus species, MYMIV (-1.31011), MYMV (-1.18985), DoYMV (-0.36806) and HgYMV (-1.04892) showed negative Tajima's D hence population of these species undergo purifying selection. However, population selection analysis revealed MYMIV,

Table 1. Genetic diversity of genomic components (DNA A and DNA B) of legume infecting begomoviruses [Mungbean Yellow Mosaic India Virus (MYMIV); Mungbean Yellow Mosaic Virus (MYMV); Dolichos yellow mosaic virus (DoYMV); Horsegram yellow mosaic virus (HgYMV); Kudzu yellow mosaic virus (KuYMV); Rhynchosia yellow mosaic India virus (RhYMIV); Rhynchosia yellow mosaic virus (RhYMV)] . Genetic diversity is ascertained from the parameters *viz*., nucleotide diversity (π), haplotype diversity (Hd) and number of polymorphic sites (S) wherein higher values imply greater diversity of the virus species at the level of nucleotide sequences.

	DNA A				DNA B			
Virus species	No. of isolates (N)	No. of polymorphic sites(S)	Nucleotide diversity (π)	Haplotype diversity (Hd)	No. of isolates (N)	No. of polymorphic sites(S)	Nucleotide diversity (π)	Haplotype diversity (Hd)
MYMIV	57	624	0.03875	0.999	38	642	0.06009	1.000
MYMV	22	222	0.03103	0.996	30	627	0.16591	0.998
DoYMV	07	309	0.04729	0.952	02	02	0.00049	0.667
HgYMV	06	145	0.02074	1.000	06	181	0.02620	1.000
KuYMV	03	222	0.04232	0.833	03	499	0.10036	0.833
RhYMIV	02	01	0.00024	0.667	04	10	0.00187	1.000
RhYMV	03	192	0.04684	0.833	03	424	0.08422	0.833
All (DNAA)	100	898	0.15089	0.999	86	479	0.15356	0.999

Table 2. Test of neutral evolution of genomes of begomoviruses infecting legumes (N-refers to number of isolates, Tajimas's D, Fu &Li's D, and Fu &Li's F refer to the test statistic parameters evaluating theory of neutral evolution. Tajimas's D and Fu & Li's F >0 balancing selection and low frequency of rare alleles; Tajimas's D and Fu & Li's F <0 purifying selection and high frequency of rare alleles; Fu & Li's D>0 lack of singletons; Fu & Li's D<0 excess of singletons)

Virus	DNA A				DNA B			
	No. of isolates (N)	Tajimas's D	Fu &Li's D	Fu &Li's F	No. of isolates (N)	Tajimas's D	Fu &Li's D	Fu &Li's F
MYMIV	57	-1.31011	-2.77314	-2.63462	38	-0.83485	-1.47908	-1.48765
MYMV	22	-1.18985	-0.95237	-1.20288	30	-0.10319	-0.97578	-0.80657
DoYMV	07	-0.36806	-0.15367	-0.22401	02	1.89306	1.89306	1.61138
HgYMV	06	-1.04892	-0.95908	-1.07576	06	-0.99336	-0.98513	-1.08554
KuYMV	03	-0.66158	-0.60484	-0.66239	03	-0.60903	-0.49128	-0.55333
RhYMIV	02	1.63299	1.63299	1.27657	04	-0.83379	-0.83379	-0.83370
RhYMV	03	2.28092	2.29753	2.46898	03	-0.7800	-0.67750	-0.75109
All	100	-0.76435	0.78597	0.11606	86	-0.81407	0.94211	0.22452

Table 3. Genetic differentiation and gene flow estimates among the species of LYMVs by analysing DNA A (P-value: 0.3167; $\chi 2$: 100)and DNA B (P-value: 0.3310; $\chi 2$: 86) genomic components [Hs, Hst: Haplotype based statistic to estimate genetic differentiation; Kst, Snn, Z: Nucleotide based test statistic to estimate the genetic differentiation (Kst value close to zero indicates no differentiation; Snn value close to one indicates differentiation); Fst: Statistic estimates the extent of gene flow between various genotypes (Value close to zero indicates free gene flow or panmixis value close to one indicates genotypic groups are closed to gene flow)].

	DNA A					DNA B				
Genotypes	H_s	H_{st}	K_{st}*	S_{nn}	F_{st}	H_s	H_{st}	K_{st}*	S_{nn}	F_{st}
MYMIV vs others	0.99786	0.00113	0.10804	1.00000	0.42769	0.99751	0.00112	0.05999	1.00000	0.24363
MYMV vs others	0.99804	0.00095	0.08770	1.00000	0.49862	0.99750	0.00113	0.04427	0.98837	0.15361
DoYMV vs others	0.99663	0.00236	0.04691	1.00000	0.77806	0.99885	-0.00022	0.01093	1.00000	0.7701
HgYMV vs others	0.99890	0.00009	0.02521	1.00000	0.52457	0.99849	0.00014	0.03653	0.98837	0.58749
KuYMV vs others	0.99894	0.00005	0.01424	1.00000	0.64280	0.99855	0.00008	0.01566	1.00000	0.56378
RhYMIV vs others	0.99895	0.00004	0.00808	1.00000	0.73492	0.98722	0.01143	0.04426	1.00000	0.79332
RhYMV vs others	0.99568	0.00332	0.01497	1.00000	0.61858	0.99855	0.00008	0.01627	1.00000	0.59437

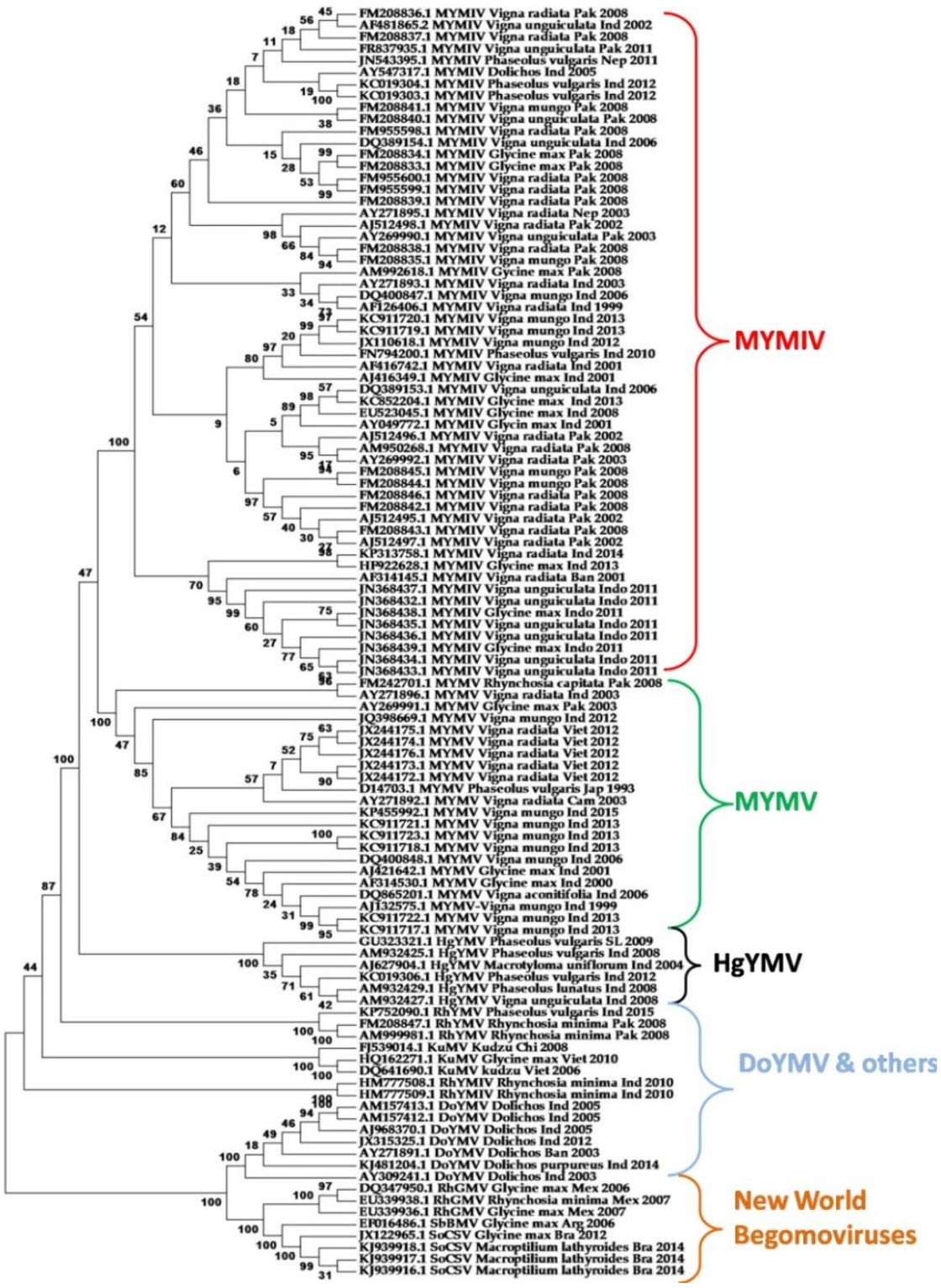

Fig 1. Molecular phylogeny reconstruction of DNA A genomic components of legume infecting begomoviruses was inferred by Maximum Likelihood method. The values on the node represent percentages of bootstrap. Evolutionary analyses were conducted in MEGA6 with default parameters and 1000 replications in bootstrap analysis.

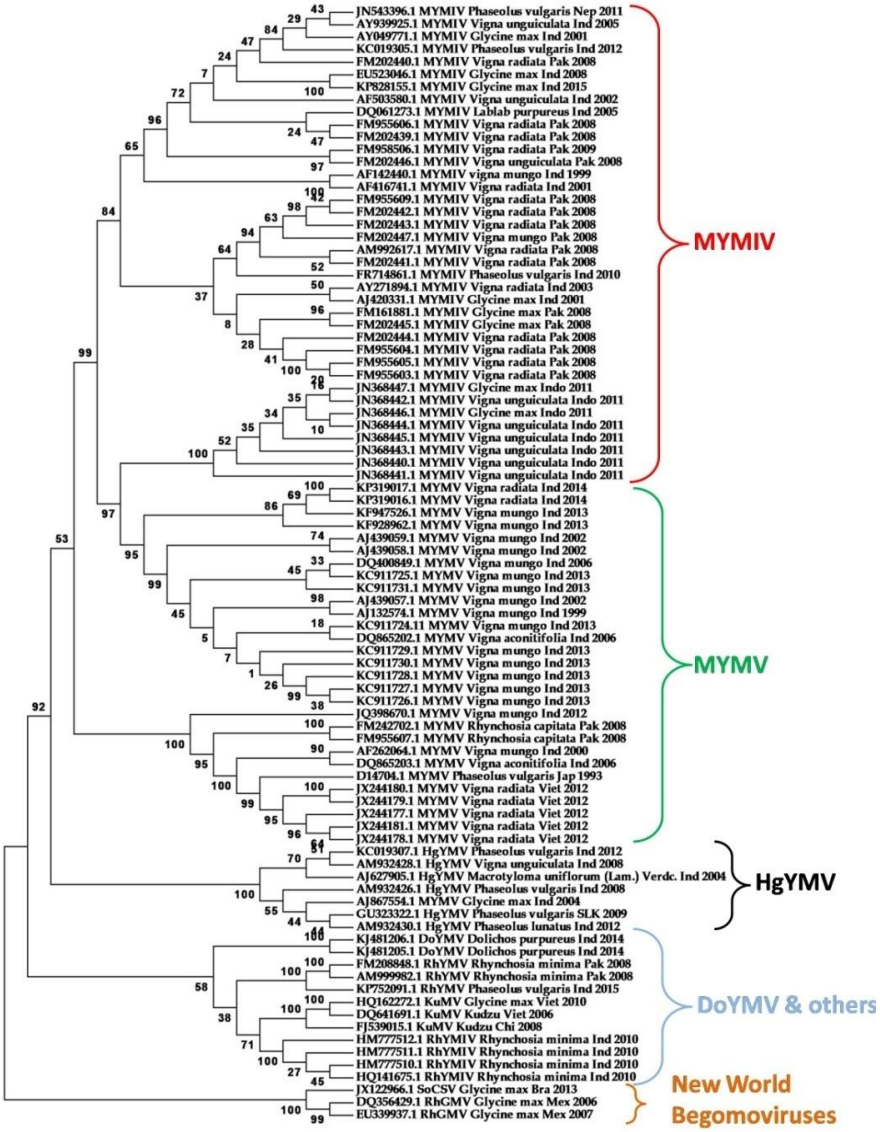

Fig 2. Molecular phylogeny reconstruction of DNA B genomic component of legume infecting begomoviruses was inferred by Maximum Likelihood method. The values on the node represent percentages of bootstrap. Evolutionary analyses were conducted in MEGA6 with default parameters and 1000 replications in bootstrap analysis.

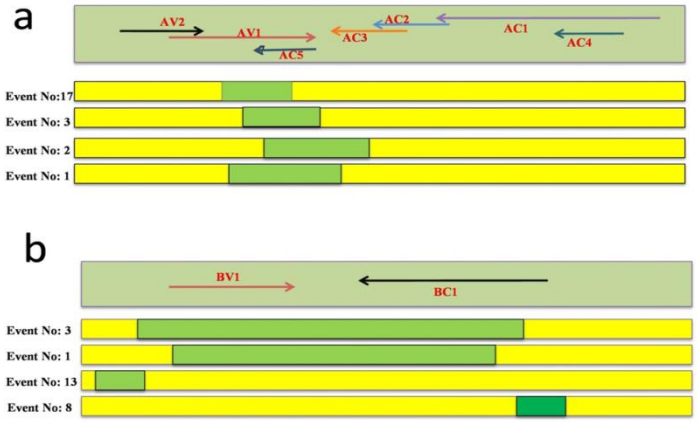

Fig.3. Recombination detection analysis in complete DNA A (a) and DNA B (b) genomic segments of legume yellow mosaic viruses (LYMVs) using RDP 4 Beta 4.27. The consensus orientations and genomic positions of DNA A and DNA B encoded viral genes are presented for reference. Genomic region involved in recombination is highlighted in green. (Virus isolates and corresponding accession numbers are presented in Supplementary table 3 and Supplementary table 4 respectively).

MYMV and HgYMV undergo more purifying selection than entire LYMV population (-0.76435) (Table 2). Similarly, analysis of other population statistic parameters like Fu &Li's D and Fu &Li's F, in MYMIV, MYMV, HgYMV, DoYMV and KuYMV showed negative values reiterating the operation of purifying selection and population expansion. Nevertheless, RhYMIV and RhYMV showed positive values for Tajima's D, Fu &Li's D and Fu &Li's F indicating the operation of neutral selection in these virus populations. Interestingly, LYMV population as whole show positive Fu &Li's D and Fu &Li's F values (0.78597 and0.11606) (Table 2).

DNA B genomic component of LYMVs (MYMIV, MYMV, HgYMV, KuYMV) revealed similar characteristic features as DNA A when analyzed for neutral evolution. However, DoYMV population showed positive values of Tajima's D, Fu & Li's D and Fu &Li's F suggesting the operation of neutral selection in population. Furthermore, RhYMIV and RhYMV genotypes showed negative Tajima's D, Fu & Li's D and Fu &Li's F indicating the operation of purifying selection and population expansion. On the whole, genomic components of LYMV derived DNA A and DNA B show similar trends in the evolution when analyzed for neutral theory of evolution. Combination of low negative Tajima's D and positive Fu &Li's D and Fu &Li's F values, indicate that the population size of LYMVs is decreasing and the population is undergoing an act of balancing or neutral selection (Table 2).

Genetic recombination events in LYMV population

Recombination Detection Program-4 (RDP 4 Beta 4.16) (Martin et al., 2015) detected 4 different recombinant events among the DNA A genomic components of LYMVs. Event no. 3 was predominant as the event was fixed in the MYMIV population which is followed by recombination events in MYMV and DoYMV species [Event: 17 (DoYMV), Events: 1 and 2 (MYMV)]. Soybean isolate of MYMIV described in this study (KC852204) appears to be a recombinant arising from a MYMV isolate infecting *Vigna radiata* as major parent (JX244176) and RhYMIV isolate infecting *Rhynchosia* (HM777508.1) as a minor parent. This event (Event no: 3) was detected by RDP (4.359E-7), BootScan (2.915E-7), MaxChi (1.718 E-7), Chimaera (1.428E-9) and SisScan (1.120E-2) methodologies (Fig. 3a and Supplementary Table 3). Analysis of DNA A viral genomic region and recombination events showed that genes AV1, and AC5 are prone to genetic exchanges. Recombination detection among the DNA B genomic components of LYMV's identified four recombination events (Event no: 3, 1, 13, and 8). Three events among them were found to be fixed within MYMV population (Fig. 3b and Supplementary Table 4). The fourth recombinant (KP828155.1_MYMIV_ *Glycine_max*_Ind_2015) has been found to be soybean isolate of MYMIV described in this study. This recombination event was detected by RDP (1.107E-10), GENECONV (2.512E-7), BootScan (2.939E-13), and MaxChi (1.975E-2) methodologies. Furthermore, analysis of recombination events showed that DNA B encoded genes, MP and NSP are prone to genetic recombination.

Extent of gene flow in LYMV population

Genetic differentiation within the population of LYMVs was deduced from the haploid-based statistics (Hs and Hst) along with nucleotide test statistics such as Ks, Kst (Kst value close to zero indicates no differentiation) and Snn (Snn value close to one indicate differentiation)(Hudon et al., 1992a; Hudson et al., 1992b). In addition, the direction and extent of gene flow among the population of various LYMV genotypes were estimated by Fst statistic (value close to zero indicates panmixis, value close to one indicates infrequent gene flow)(Hudon et al., 1992a).

Genetic differentiation between MYMIV and other LYMVs based on nucleotide test statistic (Kst) was found to be high (Kst: 0.10804), when compared to genetic differentiation observed between MYMV and other LYMVs (0.08770), DoYMV vs other LYMVs (0.04691), HgYMV vs other LYMVs (0.02521), KuYMV vs other LYMVs (0.01424), RhYMIV vs other LYMVs (0.00808), RhYMV vs other legumes (0.01497) (Table 3). Further the test statistic Snn also supports the observed genetic differentiation between major LYMVs (MYMIV and MYMV) and other LYMVs/. Gene flow estimates obtained from the test statistic (F_{st}) reveal that MYMIV (0.42769) and MYMV (0.49862) infecting legumes, show moderate gene flow with other LYMVs. Greater gene flow was observed between DoYMV and LYMVs (0.77806) followed by RhYMIV vs other LYMVs (0.73492). Similarly, genetic differentiation among the DNA B genomes of LYMVs demonstrates high values for MYMIV (0.05999) compared to other LYMVs. However, the genetic differentiation of MYMIV derived DNA B genomes was low than found in DNA A components. Thus it is inferred that MYMIV is relatively diverse than MYMV and other begomoviruses. Similarly, gene flow estimates involving DNA B genomes reveal frequent gene flow between DoYMV vs other LYMVs(0.7701) and RhYMIV vs other LYMVs (0.79332). Despite MYMIV's higher genetic differentiation, the least gene flow was observed between MYMV vs LYMVs (0.15361) and MYMIV vs LYMVs (0.24363) (Table 3).

Discussion

Legume infecting begomoviruses cause serious damage to the cultivation of grain legumes. The economic loss due to yellow mosaic virus infection in soybean and other legumes accounts for 300m US $ (Varma and Malathi 2003)**.** Central India is a major hub of soybean production and processing as the region harbours more than 80% of soybean cultivated area of the country. Therefore, yellow mosaic disease is a potential threat to the cultivation of soybean in the region. The problem of YMD was accentuated further due to sudden outbreak of the disease in the central Indian region (Soybean News, 2015). Genome characterization of YMV causing YMD in central India and devising suitable disease management measures are immediate necessity. Hence, molecular characterization of genome of yellow mosaic virus infecting soybean and global analysis of LYMV population to delineate molecular phylogeny, diversity, and population selection analysis were performed. In this study complete genome of *Mungbean yellow mosaic India virus* (MYMIV) isolate infecting soybean in the central Indian region is described. Earlier reports indicate two species of LYMVs are involved in the etiology of soybean yellow mosaic disease (Usharani et al., 2004; Girish and Usha 2005; Ramesh et al., 2013). Soybean isolates of Mungbean yellow mosaic virus (MYMV) and Mungbean yellow mosaic India virus (MYMIV) were known to cause the disease respectively in Southern and Northern region of the country(Usharani et al., 2004; Ramesh et al., 2016). Partial genome characterization of LYMV infecting soybean in the central Indian region identified it as MYMIV (Ramesh et al., 2013). Genome sequence information revealed that isolate of MYMIV

infecting soybean differs little in the genome features of isolates infecting soybean, and other legumes.

Global genomic analysis of legume infecting begomovirus is essential to delineate the dynamics of viral evolution, factors driving the genetic diversity and to categorize the viral population structure. Here we performed the genetic variability studies and evolutionary genomic analysis of the begomoviruses infecting legumes. Availability of 108 DNA A and 89 DNA B genomic sequences of LYMVs in GenBank database (Supplementary Table 1 and Supplementary Table 2) and the menace of legume begomoviruses in tropical conditions warrant a global analysis to delineate evolutionary and population dynamics.

Among the LYMVs, MYMIV, MYMV and DoYMV showed high nucleotide diversity, eventhough nucleotide diversity of MYMIV is little higher than MYMV. Nucleotide diversity of DNA B genomic components showed similar trend however DoYMV displayed low diversity. Molecular phylogeny revealed DoYMV formed a basal cluster to the main cluster comprising all other LYMVs (Fig. 1 and Fig. 2). However, all other legume yellow mosaic viruses are basal to all Old World begomoviruses. Previous studies on phylogeny of Mungbean infecting YMV in Pakistan revealed presence of Old World begomoviruses MYMIV and MYMV and the former was shown to be a major infectious agent (Hameed and Robinson 2004). Coat protein gene based diversity of LYMVs in southern India revealed MYMV and Horse gram yellow mosaic virus (HgYMV) are the two different species of begomoviruses causing disease (Maheswari et al., 2014).

Furthermore the distinctness of MYMIV and MYMV from other legume infecting begomoviruses has been proven from the population selection and test of neutral evolution studies. Test of neutral evolution discovered that population of both MYMIV and MYMV is under purifying selection hence population expansion is observed. However, LYMV population as a whole showed decrease in size owing to operation of balancing or neutral selection this could be due to the effect of RhYMIV and RhYMV (DNA A) and DoYMV (DNA B) population which are undergoing balancing selection.

Genetic differentiation studies indicate that MYMIV and MYMV are genetically more diverse from each other. Gene flow estimates identified infrequent flow of genetic material between MYMIV vs LYMVs, and MYMV vs LYMVs. However, frequent gene flow was observed between DoYMV vs LYMVs, and RhYMIV vs LYMVs. Despite infrequent gene flow with other LYMVs, MYMIV isolates were frequently identified as recombinants arising from other LYMVs due to genetic exchanges. Thus, recombination events have been fixed in the population of MYMIV. Furthermore, recombination detection analysis identified that MYMIV genomic components (both DNA A and DNA B) reported in this study are recombinants. The situation warrants serious management techniques to tackle YMD disease of soybean in central India. In this context, it is pertinent to mention begomoviruses and their beta satellites which were not considered as pathogens of grain legumes were found to involve in genetic recombination while analyzing genetic diversity and phylogeography in Pakistan (Ilyas et al., 2010). Likewise, a novel legumovirus *Soybean chlorotic blotch virus* (SbCBV) found to be infecting soybean in Nigeria was characterized with genomic components identical to the virus infecting wild species *Centrosema pubescens* (Alabi et al., 2010). These findings thus, reiterate the importance of genetic interaction between cultivated grain legumes and wild counterparts. In addition, expanded host range (*V. hainiana and V. trilobata*.) for MYMIV infection has also been reported in Indian context (Naimuddin and Pratap 2011). Thus this infringement of genetic isolation of legume yellow mosaic viruses (LYMVs) (Qazi et al., 2007) is a serious threat for the cultivation of grain legumes including soybean.

The hypothesis of genetic isolation of legume yellow mosaic viruses (LYMVs) has been put forward hence genetic recombination or genomic components exchanges are rare (Qazi et al., 2007; Ilyas et al., 2010). Supporting the hypothesis, all the recombination events detected in this study are between Old World begomoviruses (Fig 3a and Fig 3b and Supplementary table 3 and Supplementary table 4). The reason for genetic isolation is attributed to its limited host range. However, results of this study indicate that legume infecting begomoviruses are prone to genetic exchanges and hence genetic variation which might lead to the development of virus genotypes with devastating potential. Analysis of recombination hot spot in LYMV DNA A affirmed AC5 and AV1 genes are more prone to genetic recombination. Similarly, genetic recombinants in MYMIV population (Girish and Usha 2005) and within the rep (replication associated protein) gene of Geminiviruses have also been demonstrated (Vadivukarasi 2007). This genetic variation and genetic recombination based on DNA-A genomic components emphasizes its role in generating variability contrary to the high genetic diversity observed among the DNA B component of the begomoviruses (Briddon et al., 2010). Despite the infrequent gene flow estimates between MYMIV and other LYMVs, it is observed that genomic components of MYMIV were more prone to genetic recombination (Fig. 3). Furthermore, MYMIV was found to be major disease causing agent in the central Indian region where soybean is grown extensively (Ramesh et al., 2013). Hence, in light of these findings breeding efforts have to be harnessed to identify or to breed soybean genotypes that show stable resistance against MYMIV.

Materials and Methods

Sample collection, and characterization of virus genome

Symptomatic soybean leaves from the fields of ICAR-Indian Institute of Soybean Research, Indore, Madhya Pradesh, India were collected. Total DNA from infected soybean leaves and healthy leaf samples (negative control) were extracted as described previously (Doyle and Doyle 1987). Extracted DNA was used as template for PCR amplification of MYMIV and MYMV specific coat protein (AV-1) region (Ramesh et al., 2016) using forward primers (MYMIV F 5' GCATCAAGTCCGTGTACATTAC 3'and MYMV F 5'GTGTTAAGTCTATCTGGG3') respectively. However, a common reverse primer (YMV R 5'CACAGGATTTGATG-CATGAG 3' was used for detection of YMV species (Ramesh et al., 2016).

Rolling circle amplification was performed in a 20 µl reaction volume using 25ng of extracted total DNA and comprising 2µl of Phi 29 DNA polymerase buffer (10X), 2µl of exo-resistant random primers (500µM) and 2µl of dNTPs (10mM). The template DNA was denatured for a brief period of 3 min at 94^{o}C and cooled down to room temperature. It was followed by addition of 4 µl of pyrophosphatase (0.1U/µl) and 0.7µl of Phi 29, DNA polymerase (10 U/µl) and incubated at 30^{o} C for 18 hrs (Fermentas, Massachusetts, USA). The amplification reaction was stopped by enzyme inactivation at 65^{o}C for 10 min. The RCA derived DNA was digested with selected restriction endonucleases (*Bam*HI,

*Eco*RI, *Hin*dIII, *Pst*I) individually to release unit-length viral genomes. The resultant fragments of size ~2.7Kb were gel eluted using QIAquick Gel Extraction Kit™ (Venlo, Limburg, Netherlands) according to manufacturer's instructions. The eluted DNA fragments were ligated with restricted pUC118-35SP-T vector and recombinant clones were generated. Virus genome sequence information was obtained through primer walking strategy (Merck Biosciences, Bengaluru, India). Complete genome sequences were submitted to GenBank, NCBI, USA (KC852204 and KP828155).

LYMV genomic data and phylogeny reconstruction

Complete genome sequence of yellow mosaic virus isolate infecting soybean grown in Central Indian region was annotated using BioEdit sequence alignment editor software (Hall 1999). Since, the genome sequences of other known LYMVs were ~2750 nts in length BioEdit sequence alignment editor was used for annotation. Genome sequence of MYMIV obtained from infected soybean plant was analyzed along with complete genome sequences of other legume infecting begomoviruses obtained from GenBank database. A total of 108 complete DNA A genome segments (Supplementary Table 1) and 89 complete DNA B genome segments (Supplementary Table 2) of LYMVs (including genome sequences reported in this study-) were obtained from GenBank (KC852204 and KP828155). Sequence alignment and phylogenetic trees were generated in MEGA 6 using ClustalW algorithm (Tamura et al., 2013). The phylogenetic tree was constructed in MEGA 6 with default parameters and 1000 replicates in the bootstrap analysis using the Maximum likelihood method.

Genetic diversity and neutrality tests

In order to study the nucleotide diversity and DNA polymorphism among the genomic sequences of LYMVs, the computational tool-DnaSP (Librado and Rozas 2009) was used. Similarly, to test the theory of neutral evolution of begomovirus population, test statistics like Tajimas's D (Tajima 1989), Fu & Li's D and Fu & Li's F (Fu 1997; Fu and Li 1993) were inferred by employing DnaSP software (Librado and Rozas 2009).

Recombination detection

Recombination among the genome sequences of legume yellow mosaic viruses was detected using Recombination Detection Program-4 (RDP 4 Beta 4.16) (Martin et al., 2015). Multiple sequence alignment of genome sequences created in MEGA 6 was used as a query in the recombination detection program with in-built algorithms (RDP, BootScan, GENECONV, MAXCHI, CHIMAERA, SISCAN, LARD, PhylPro and 3SEQ). The highest acceptable p-value was set at 0.05. Further, to increase the stringency, recombination events detected atleast by 3 and more of these methods alone were considered for further analysis.

Genetic differentiation and gene flow estimates

Genetic differentiation among the species of begomoviruses infecting legumes was estimated from the Hudson's test statistics such as Ks, Kst and Snn (Hudson 2000). Haplotype statistics such as Hs and Hst (Hudson et al., 1992a; Hudson et al., 1992b) were computed using DnaSP (Librado and Rosas 2009). DnaSP was used to study the extent of gene flow among the begomovirus populations by estimating statistic Fst (Hudson et al., 1992b).

Conclusion

Legume yellow mosaic viruses (LYMVs) impose serious damage to the cultivation of grain legumes causing yellow mosaic disease (YMD). Hence in this study complete genome sequence of YMV infecting soybean in the central Indian region was reported. Evolutionary lineage analysis of DNA A and DNA B genomic components of LYMVs revealed that population of MYMIV and MYMV is undergoing purifying selection and population expansion. However, LYMVs as a whole, show decrease in population size and act of balancing or neutral selection. Frequent gene flow was observed-by the way of genetic recombination- among the LYMVs infecting legumes. Genetic isolation of LYMVs has been reaffirmed; nevertheless phylogeny reconstruction showed genetic distinctness of Old World begomoviruses from New World begomoviruses. Thus the molecular evolutionary genomics analysis identifies factors driving the process of natural selection in LYMV population and their genetic variability. The findings of this study have potential implications for devising suitable disease management strategies.

Competing Interests

The authors declare that they have no competing interests

Acknowledgements

Funding provided by Indian Council of Agricultural Research (ICAR) (Project code: DSR 1.24/12) is gratefully acknowledged.

References

Alabi OJ, Kumar PL, Mgbechi Ezeri JU, Naidu RA (2010) Two new 'legumoviruses' (genus *Begomovirus*) naturally infecting soybean in Nigeria. Arch Virol.155:643-656.

Briddon RW, Patil BL, Bagewadi B, Nawaz-ul-Rehman MS, Fauquet CM (2010) Distinct evolutionary histories of the DNA-A and DNA-B components of bipartite begomoviruses. BMC Evol Biol. 10:97.

Doyle JJ, Doyle JL (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem Bull. 19:11–15.

Fauquet C, Briddon R, Brown J, Moriones E, Stanley J, Zerbini M, Zhou X (2008) Geminivirus strain demarcation and nomenclature. Arch Virol.153:783–821.

Fauquet CM, Stanley J (2003) Geminivirus classification and nomenclature; Progress and problems. Ann Appl Biol.142:165–189.

Fiallo Olive E, Marquez Martin B, Hassan I, Chirinos DT, Geraud Pouey F, Navas Castillo, J, Moriones E, (2013) Complete genome sequences of two novel begomoviruses infecting common bean in Venezuela. Arch Virol.158:723-727.

Fu Y (1997) Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. Genetics. 147:915–925.

Fu Y, Li WH (1993) Statistical tests of neutrality of mutations. Genetics. 133: 693–709

Girish KR, Usha R (2005) Molecular characterization of two soybean-infecting *Begomoviruses* from India and evidence for recombination among legume-infecting *Begomoviruses* from South-East Asia. Virus Res. 108:167–176.

Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symposium Series. 41:95–98.

Hameed S, Robinson DJ (2004) Begomoviruses from mungbeans in Pakistan: epitope profiles, DNA A sequences and phylogenetic relationships. Arch Virol.149:809-819.

Hudson RR (2000) A new statistic for detecting genetic differentiation. Genetics. 155: 2011–2014.

Hudson RR, Boos DD, Kaplan NL (1992a) A statistical test for detecting geographic subdivision. Mol Biol. 9:138–151.

Hudson RR, Slatkin M, Maddison WP (1992b) Estimations of levels of gene flow from DNA sequence data. Genetics. 132:583–589.

Ilyas M, Qazi J, Mansoor S, Briddon RW (2010) Genetic diversity and phylogeography of begomoviruses infecting legumes in Pakistan. J Gen Viro. 91:2091-2101.

Librado P, Rozas J (2009) DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. Bioinformatics. 25:1451–1452.

Maheshwari R, Panigrahi G, Angappan K (2014) Molecular characterization of distinct YMV (Yellow mosaic virus) isolates affecting pulses in India with the aid of coat protein gene as a marker for identification. Mol Biol Rep. 41: 2635-2644.

Martin DP, Murrell B, Golden M, Khoosal A, Muhire B (2015) RDP4: Detection and analysis of recombination patterns in virus genomes. Virus Evol. 1: vev003

Naimuddin MA, Pratap A (2011) First report of natural infection of Mungbean yellow mosaic India virus in two wild species of *Vigna*. New Disease Reports 23, 21 [doi:10.5197/j.2044-0588.2011.023.021]

Qazi J, Ilyas M, Mansoor S, Briddon RW (2007) Legume yellow mosaic viruses: genetically isolated begomoviruses. Mol Plant Pathol. 8:343-348.

Ramesh SV, Bhaskale R, Admane N, Gupta, GK, Husain SM (2013) Multiply Primed Rolling Circle Amplification (MPRCA) of *Yellow mosaic virus* genome from infected soybean in central Indian region divulges it as *Mungbean yellow mosaic Indian virus*-[sb] and its implications for RNAi mediated virus resistance. Paper presented at the IXth World Soybean Research Conference, Durban, South Africa.17-22 February, 2013.

Ramesh SV, Chouhan BS, Gupta GK, Ramteke RK, Chand S, Husain SM (2016) Molecular diversity analysis of coat protein gene encoded by legume begomoviruses and PCR assay to detect yellow mosaic viruses infecting soybean in India. Br Biotech J. 3: 1-10.

Stanley J, Bisaro DM, Briddon RW, Brown JK, Fauquet CM, Harrison BD, Rybicki EP, Stenger DC (2005) Geminiviridae. Virus Taxonomy. VIIIth Report of the International Committee on Taxonomy of Viruses, pp. 301–326. In: Fauquet CM, Mayo MA, Maniloff J, Desselberger U and Ball LA (Eds). Elsevier/ Academic Press, London

Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics. 123: 585–595.

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. Mol Biol Evol. 30:2725-2729.

Usharani KS, Surendranath B, Haq QMR, Malathi VG (2004) Yellow mosaic virus infecting soybean in northern India is distinct from the species infecting soybean in southern and western India. Curr Sci. 86:845–850.

Vadivukarasi T, Girish KR, Usha R (2007) Sequence and recombination analyses of the geminivirus replication initiator protein. J Biosci. 32:17-29.

Varma A, Malathi VG (2003) Emerging Geminivirus problems A serious threat to crop production. Ann Appl Biol. 142:145-164