Plant Pathology (2014) Doi: 10.1111/ppa.12292

Evidence for two predominant viral lineages, recombination and subpopulation structure in begomoviruses associated with yellow vein mosaic disease of okra in India

V. Venkataravanappa[ab], H. C. Prasanna[a]*, C. N. Lakshminarayana Reddy[b] and M. Krishna Reddy[c]

[a]*Indian Institute of Vegetable Research, Varanasi 221305, Uttar Pradesh;* [b]*Department of Plant Pathology, University of Agricultural Sciences, and* [c]*Indian Institute of Horticultural Research, Hessaraghatta Lake PO, Bangalore, India*

Yellow vein mosaic disease (YVMD) caused by whitefly-transmitted begomoviruses is an economically significant viral disease of okra. In this study, a survey of begomoviruses associated with YVMD was carried out in eight states and two union territories of India. A total of 92 full-length DNA-A components were sequenced and characterized. Sequence comparisons and population structure analysis revealed the existence of four begomovirus species. Two novel species were detected with several recombinationally derived genome fragments that probably originated from begomoviruses known to infect malvaceous and non-malvaceous hosts. Among the four species, *Bhendi yellow vein Maharastra virus* (BYVMaV) and *Bhendi yellow vein Madurai virus* (BYVMV) were found to be predominant in okra, with BYVMV having a pan-India distribution. There was evidence for a high degree of genetic variability and subpopulation structure within these four species. Neutrality tests suggested the occurrence of purifying selection acting upon these populations. The results of the current study have uncovered the diversity and genetic structure of okra-infecting begomoviruses in India and generated potentially useful information for developing management strategies for YVMD.

Keywords: begomovirus genetic diversity, *Bhendi yellow vein Madurai virus*, *Bhendi yellow vein Maharastra virus*, geminivirus, recombination

Introduction

Okra (*Abelmoschus esculentus*) is an annual herb of the family Malvaceae. It is an important vegetable grown extensively in tropical, subtropical and warm temperate zones of the world (Charrier, 1984). With six million tonnes of okra produced on half a million hectares (FAOSTAT, 2013), India produces more okra than any other country in the world. Okra is one of India's primary vegetable crops and is widely cultivated throughout the year in different regions. Yellow vein mosaic disease (YVMD), which is caused by whitefly-transmitted begomoviruses, is one of the major constraints of okra production in India, with the disease first reported in 1924 and the causal agent identified as a virus only in 1940 (Kulkarni, 1924; Uppal *et al.*, 1940). Later, the associated virus was morphologically identified as a geminivirus based on its serological relationship to the begomovirus species *African cassava mosaic virus* (Harrison *et al.*, 1991).

To date, many species of okra-infecting begomoviruses have been reported from both the Old World and New World. These include *Cotton leaf curl Gezira virus* (CLCuGeV) from Sudan, *Okra leaf curl Cameroon virus* (OLCuCMV) from Cameroon, *Okra yellow crinkle virus* (OYCrV) from Mali and Cameroon, *Okra yellow mosaic Mexico virus* (OYMMV) and *Okra yellow mottle Igula virus* (OKYMoIV) from Mexico, *Okra yellow vein mosaic virus* (OYVMV) from Pakistan, *Okra leaf curl Oman virus* from Oman and *Sida micrantha mosaic virus* from Brazil (Fauquet *et al.*, 2008; Kon *et al.*, 2009; Hernandez-Zepeda *et al.*, 2010; Leke *et al.*, 2013; Akhtar *et al.*, 2014). In India, distinctive monopartite begomoviruses such as *Bhendi yellow vein Madurai virus* (BYVMV), *Cotton leaf curl Allahabad virus* (CLCuAlV), *Bhendi yellow vein Bhubaneswar virus* (BYVBhV), *Cotton leaf curl Bangalore virus* (CLCuBaV), *Bhendi yellow vein Maharashtra virus* (BYVMaV) and *Okra enation leaf curl virus* (OELCuV) have been reported (Fauquet *et al.*, 2008; Brown *et al.*, 2012; Venkataravanappa *et al.*, 2012b, 2013a,b). Besides these, *Bhendi yellow vein Delhi virus* (BYVDV), a new bipartite begomovirus species, was recently found to be associated with YVMD on okra (Venkataravanappa *et al.*, 2012a).

Mechanisms that create genetic variation such as mutation, reassortment and recombination are all known to commonly occur during begomovirus evolution. Once novel genetic variants arise in nature due to these mechanisms, selection pressures and the innate fitness of the viral variants in different environments, transmission vectors and hosts drive the evolution of begomovirus populations

*E-mail: prasanahc@yahoo.com

(Seal *et al.*, 2006). Indeed, the Indo-China region exhibits a very high degree of begomovirus diversity (Nawaz-ul-Rehman & Fauquet, 2009). These complex mechanisms and factors associated with begomovirus evolution underscore the need for continuous sampling and monitoring of begomovirus diversity, particularly in major crop-growing regions with a high level of begomovirus diversity. Such activities may lead to the identification of divergent novel species or provide insights into the prevalence and origins of recombinant begomoviruses, many of which may display altered host ranges, vector specificities and/or degrees of pathogenicity.

This study surveys the major okra-growing regions of India (representing eight states and two union territories) and describes the characterization of 92 DNA-A-like sequences of okra-infecting begomoviruses and the identification of two novel species along with many variants and strains of other known begomovirus species. It also assesses the genetic diversity and uncovers the structure of okra-infecting begomovirus populations. The information generated in this study is important for understanding the evolutionary mechanisms, and for facilitating development of management strategies, of okra-infecting begomoviruses of India.

Materials and methods

Survey and collection of virus isolates

Leaf samples and twigs were collected individually from 200 plants exhibiting symptoms such as yellow mosaic, vein thickening, petiole bending, complete yellowing, upward leaf curling and stunted growth. The sampling survey covered 39 locations of eight states and two union territories of India between 2005 and 2007 (Table S1). A total of 39 leaf samples from each of these locations were also collected from symptomless plants. As the field-collected samples may be potentially infected with multiple viruses, repeated whitefly transmissions were carried out under controlled conditions on the begomovirus-susceptible okra cultivar 1685. Whiteflies maintained on healthy cotton plants were collected and released into the insect-proof cage containing the twigs of diseased plants. The viruliferous whiteflies were collected after an access period of 24 h. Inoculation of cv. 1685 at the 4-leaf stage was achieved by exposing the whole plant to the viruliferous whiteflies in a closed cage for 48 h of inoculation access. These transmissions were continued until disease symptoms remained consistent in successive transmissions. Although such successive transmissions did not necessarily yield okra plants that were infected with a single virus species/strain, it will have selected for begomovirus variants that were specifically adapted to infecting okra. The virus isolates established on cv. 1685 were maintained at the Indian Institute of Horticultural Research, Bangalore in an insect-proof glasshouse.

Total DNA extraction, PCR amplification and cloning

Total DNA was extracted from plants with symptoms using a CTAB-based method (Doyle & Doyle, 1990). Full-length viral DNA-A components were amplified using begomovirus-specific primers as described previously (Venkataravanappa *et al.*, 2012a). The full-length sequences were obtained using the amplification with three primer pairs resulting in fragments that overlapped by approximately 200 bp. PCR products of the expected size (1·5 kb) were purified from agarose gels, ligated to the plasmid vector pTZ57R (Fermentas) according to the manufacturer's instructions, and transformed into *Escherichia coli* DH5α competent cells (Novagen). Bacterial colonies containing the recombinant plasmids carrying DNA inserts corresponding to the 1·5 kb fragment were selected for sequencing. Three clones from each sample were sequenced using an automated DNA sequencer ABI PRISM 3730 (Applied Biosystems) at Anshul Biotechnologies DNA sequencing facility, Hyderabad, Andhra Pradesh, India.

Sequence comparisons and phylogenetic analysis

All the sequences obtained were initially analysed using the VECTOR NTI ADVANCE TM9 package to remove vector sequences. The full-length genome sequence of each of the begomoviruses was compared with other begomoviruses available in GenBank (http://blast.ncbi.nlm.nih.gov/Blast.cgi) by BLAST (Altschul *et al.*, 1990). The newly determined sequences (92), GenBank sequences yielding the most significant E-score matches to the newly determined sequences, and okra-infecting begomoviruses from Asia (25), Brazil (3), Mexico (3) and Africa (7) were aligned using CLUSTALW (Thompson *et al.*, 1994). A sequence identity matrix for these begomoviruses was generated using BIOEDIT sequence alignment editor (v. 5.0.9) and a maximum likelihood phylogenetic tree was generated using MEGA6 (Tamura *et al.*, 2013) with 1000 bootstrap replications. Additionally, pairwise distances were also calculated using SPECIES DEMARCATION TOOL (SDT) v. 1.0 (Muhire *et al.*, 2013).

Recombination analysis

The possible presence of reticulate evolution (i.e. such as one that occurs during recombination) was tested using the neighbour-net method implemented in SPLITSTREE4 (Huson & Bryant, 2006). The recombination detection program RDP4, which implements different recombination detection methods, namely RDP, GENECONV, MAXCHI, CHIMERA, SISCAN and 3SEQ, was used to identify recombination events (Martin *et al.*, 2010). Default RDP settings with a 0·05 *P*-value cut-off and standard Bonferroni correction were used with the option 'Reference sequence selection' set as 'internal references only'.

Population structure analysis

The population structure of okra-infecting begomoviruses was analysed as previously described (Prasanna *et al.*, 2010) using methods implemented in the computer program STRUCTURE (v. 2.0; Pritchard *et al.*, 2000). Briefly, initial burn-in of 100 000 iterations and 100 000 iterations for parameter estimation were used. To estimate the number of populations (K), the begomovirus data set comprising 92 okra-infecting begomoviruses characterized in this study and 18 previously reported begomoviruses was analysed. The value of K was allowed to vary from 1 to 7. Three independent runs were carried out for each K value. The admixture model was selected with the option of correlated allele frequencies between populations and most of the parameters were set to their default values. The optimum number of populations (K_{opt}) was determined based on maximum log probability of data ln*P*(D). The ΔK parameter was used as previously described (Evanno *et al.*, 2005) to determine the optimum number of populations when the ln*P*(D) was not indicative of K_{opt}.

Population differentiation, genetic divergence and neutrality tests

The population structure was inferred by analysis of molecular variance (AMOVA). The fixation indices (or F_{ST} statistics) estimated from the sequence analyses were tested using a non-parametric permutation approach with 10 000 permutations. AMOVA and *F* statistic tests were performed using ARLEQUIN v. 3.5 (Excoffier *et al.*, 2005). To estimate nucleotide polymorphism, several parameters such as haplotype diversity (hd), number of segregating sites (S), pairwise estimates of nucleotide divergence (Pi), and average of nucleotide differences (k), were considered. The tests for genetic differentiation, estimation of F_{ST} and sequence polymorphism were performed using DNASP v. 5.10 (Librado & Rozas, 2009). Three different types of neutrality tests were performed, namely Tajima's *D* and Fu and Li's *D* and *F*. These analyses were performed using DNASP v. 5.10 with data sets of different subpopulations uncovered in the STRUCTURE analysis.

Results

Characterization of DNA-A-like sequences

A total of 200 samples collected from 39 locations were analysed for the presence of begomoviruses using PCR with primers specific to the DNA-A/DNA-A-like components of begomovirus genomes that are known to infect malvaceous plants. Out of these, 92 samples were selected for further characterization primarily based on the geography and distribution of okra fields (Table S1). A set of full-length betasatellite sequences amplified from 36 of these samples was reported previously (Venkataravanappa *et al.*, 2011). Attempts to amplify DNA-A from DNA extracts of symptomless plants that were collected in the vicinity of samples with symptoms were uniformly unsuccessful. Similarly, attempts to amplify DNA-B components by PCR using different sets of specific primers (Venkataravanappa *et al.*, 2012a) were unsuccessful in these samples. This suggests that okra-infecting begomoviruses characterized in the present study are monopartite viruses. The complete nucleotide sequences of the 92 DNA-A-like components were determined and these ranged in size from 2672 to 2793 nt. All the sequences had predicted gene arrangements typical of Old World begomoviruses (Table S1).

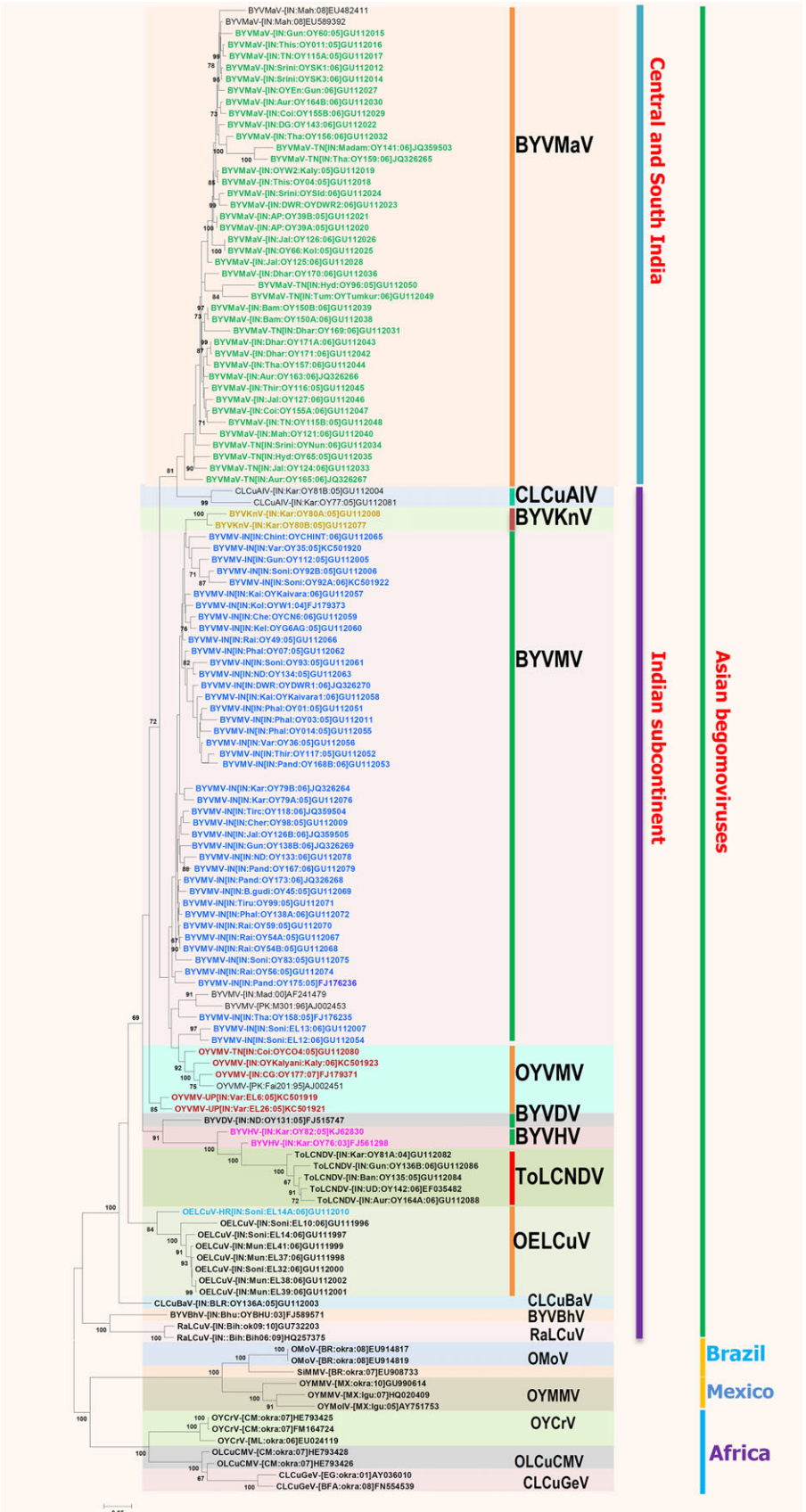
Relationships with known begomoviruses and phylogenetic analysis

A total of 89 sequences shared genome-wide similarities of 89·1 to 98·7% and three shared similarities of <89% with already known species of begomovirus (Table S1). Furthermore, the pairwise sequence analyses showed that 41 sequences shared nucleotide sequence identities of between 89·7 and 98·7% with *Bhendi yellow vein Maharashtra virus* (BYVMaV: EU589392) that has previously been reported from the Maharashtra state of India. Among these isolates, nine shared 89·7–92·5% sequence identity with a known BYVMaV isolate and were therefore classified as members of new strain BYVMaV-Tamil Nadu (BYVMaV-TN; Table S2). While 37 sequences shared between 89·1 and 93·5% nucleotide sequence identity with a *Bhendi yellow vein mosaic virus* (BYVMV: AF241479) isolate reported from Madurai, India, a further five shared 89·3–92·1% sequence identity with a BYVMV isolate reported from Pakistan (AJ002453). Therefore, the majority of the isolates sequenced in the current study were identified as members of new strain BYVMV-India (BYVMV-IN; Table S2). A further three sequences shared 89·1–92·1% sequence identity with an *Okra yellow vein mosaic virus* (OYVMV: AJ002451) isolate previously reported from Pakistan. Finally, one of the characterized sequences (EL 14A) shared 90·5% sequence identity with *Okra enation leaf curl virus* (OELCuV: GU112002) previously reported from India. Therefore EL14A is proposed as a new Haryana strain of OELCuV (OELCuV-HR; Table S2).

The phylogenetic tree generated using 92 sequences characterized in the present study and 38 monopartite and bipartite okra-infecting begomovirus sequences available in the GenBank database is presented in Figure 1. As expected from the pairwise similarity analyses, the results of phylogenetic analysis grouped all the newly characterized sequences (Table S1) into four major clusters that corresponded to previously identified species, namely BYVMV, BYVMaV, OELCuV and OYVMV (>70% bootstrap). A total of 42 sequences sampled from southern and central India that showed genetic similarity to the BYVMaV from Maharashtra clustered together (81% bootstrap). Forty-three isolates formed a cluster with BYVMV isolates from Madurai, India and Pakistan (AF241479, AJ002453). Two OYVMV sequences formed a clade with 85% bootstrap support and three other OYVMV sequences branched off from the BYVMV cluster to form a separate clade (92% bootstrap). The isolate OELCuV-[IN:SP:EL14A:06] clustered with previously identified OELCuV isolates (84% bootstrap).

Identification of novel species of okra-infecting begomoviruses

The genome of isolates OY76, OY82 and OY80A shared <89% nucleotide identity with known begomovirus sequences. OY76 and OY82 were 97% identical to each other and most closely related to isolates of *Tomato leaf curl New Delhi virus* at 82·8–87·6% nucleotide sequence identities. Isolate OY80A was most closely related to *Bhendi yellow vein mosaic virus* (AF241479, AJ002453, 88·2% identity). Therefore, in accordance with the cut-off point established by the ICTV for begomovirus species demarcation (Fauquet *et al.*, 2008), it most likely represents a new species. The analysis performed using the SPECIES DEMARCATION TOOL (SDT) also supported the classification of these isolates as belonging to novel species (Fig. S1). It is proposed here to name OY76 and OY82 as Bhendi yellow vein Haryana virus [India: Haryana:06] and OY80A as Bhendi yellow vein Karnal virus [India:Karnal:06]. The isolates of Bhendi yellow


BYVMaV-[IN:Mah:08]EU482411
BYVMaV-[IN:Mah:08]EU589392
BYVMaV-[IN:Gun:OY60:05]GU112015
BYVMaV-[IN:This:OY01:05]GU112016
BYVMaV-[IN:TN:OY115A:05]GU112017
BYVMaV-[IN:Srini:OYSK1:06]GU112012
BYVMaV-[IN:Srini:OYSK3:06]GU112014
BYVMaV-[IN:OYEn:Gun:06]GU112027
BYVMaV-[IN:Aur:OY164B:06]GU112030
BYVMaV-[IN:Coi:OY155B:06]GU112029
BYVMaV-[IN:DG:OY143:06]GU112022
BYVMaV-[IN:Tha:OY156:06]GU112032
BYVMaV-TN[IN:Madam:OY141:06]JQ359503
BYVMaV-TN[IN:Tha:OY159:06]JQ326265
BYVMaV-[IN:OYW2:Kaly:05]GU112019
BYVMaV-[IN:This:OY04:05]GU112018
BYVMaV-[IN:Srini:OYSld:06]GU112024
BYVMaV-[IN:DWR:OYDWR2:06]GU112023
BYVMaV-[IN:AP:OY39B:05]GU112021
BYVMaV-[IN:AP:OY39A:05]GU112020
BYVMaV-[IN:Jal:OY126:06]GU112026
BYVMaV-[IN:OY66:Kol:05]GU112025
BYVMaV-[IN:Jal:OY125:06]GU112028
BYVMaV-[IN:Dhar:OY170:06]GU112036
BYVMaV-TN[IN:Hyd:OY96:05]GU112050
BYVMaV-TN[IN:Tum:OYTumkur:06]GU112049
BYVMaV-[IN:Bam:OY150B:06]GU112039
BYVMaV-[IN:Bam:OY150A:06]GU112038
BYVMaV-TN[IN:Dhar:OY169:06]GU112031
BYVMaV-[IN:Dhar:OY171A:06]GU112043
BYVMaV-[IN:Dhar:OY171:06]GU112042
BYVMaV-[IN:Tha:OY157:06]GU112044
BYVMaV-[IN:Aur:OY163:06]JQ326266
BYVMaV-[IN:Thir:OY116:05]GU112045
BYVMaV-[IN:Jal:OY127:06]GU112046
BYVMaV-[IN:Coi:OY155A:06]GU112047
BYVMaV-[IN:TN:OY115B:05]GU112048
BYVMaV-[IN:Mah:OY121:06]GU112040
BYVMaV-TN[IN:Srini:OYNun:06]GU112034
BYVMaV-TN[IN:Hyd:OY65:05]GU112035
BYVMaV-TN[IN:Jal:OY124:06]GU112033
BYVMaV-TN[IN:Aur:OY165:06]JQ326267
BYVMaV
CLCuAlV-[IN:Kar:OY81B:05]GU112004
CLCuAlV-[IN:Kar:OY77:05]GU112081
CLCuAlV
BYVKnV-[IN:Kar:OY80A:05]GU112008
BYVKnV-[IN:Kar:OY80B:05]GU112077
BYVKnV
BYVMV-IN[IN:Chint:OYCHINT:06]GU112065
BYVMV-IN[IN:Var:OY35:05]KC501920
BYVMV-IN[IN:Gun:OY112:05]GU112005
BYVMV-IN[IN:Soni:OY92B:05]GU112006
BYVMV-IN[IN:Soni:OY92A:06]KC501922
BYVMV-IN[IN:Kai:OYKaivara:06]GU112057
BYVMV-IN[IN:Kol:OYW1:04]FJ179373
BYVMV-IN[IN:Che:OYCN6:06]GU112059
BYVMV-IN[IN:Kel:OYG6AG:05]GU112060
BYVMV-IN[IN:Rai:OY49:05]GU112066
BYVMV-IN[IN:Phal:OY07:05]GU112062
BYVMV-IN[IN:Soni:OY93:05]GU112061
BYVMV-IN[IN:ND:OY134:05]GU112063
BYVMV-IN[IN:DWR:OYDWR1:06]JQ326270
BYVMV-IN[IN:Kai:OYKaivara1:06]GU112058
BYVMV-IN[IN:Phal:OY01:05]GU112051
BYVMV-IN[IN:Phal:OY03:05]GU112011
BYVMV-IN[IN:Phal:OY014:05]GU112055
BYVMV-IN[IN:Var:OY36:05]GU112056
BYVMV-IN[IN:Thir:OY117:05]GU112052
BYVMV-IN[IN:Pand:OY168B:06]GU112053
BYVMV-IN[IN:Kar:OY79B:06]JQ326264
BYVMV-IN[IN:Kar:OY79A:05]GU112076
BYVMV-IN[IN:Tirc:OY118:06]JQ359504
BYVMV-IN[IN:Cher:OY98:05]GU112009
BYVMV-IN[IN:Jal:OY126B:06]JQ359505
BYVMV-IN[IN:Gun:OY138B:06]JQ326269
BYVMV-IN[IN:ND:OY133:06]GU112078
BYVMV-IN[IN:Pand:OY167:06]GU112079
BYVMV-IN[IN:Pand:OY173:06]JQ326268
BYVMV-IN[IN:B.gudi:OY45:05]GU112069
BYVMV-IN[IN:Tiru:OY99:05]GU112071
BYVMV-IN[IN:Phal:OY138A:06]GU112072
BYVMV-IN[IN:Rai:OY59:05]GU112070
BYVMV-IN[IN:Rai:OY54A:05]GU112067
BYVMV-IN[IN:Rai:OY54B:05]GU112068
BYVMV-IN[IN:Soni:OY83:05]GU112075
BYVMV-IN[IN:Rai:OY56:05]GU112074
BYVMV-IN[IN:Pand:OY175:05]FJ176236
BYVMV-[IN:Mad:00]AF241479
BYVMV-[PK:M301:96]AJ002453
BYVMV-IN[IN:Tha:OY158:05]FJ176235
BYVMV-IN[IN:Soni:EL13:06]GU112007
BYVMV-IN[IN:Soni:EL12:06]GU112054
BYVMV
OYVMV-TN[IN:Coi:OYCO4:05]GU112080
OYVMV-[IN:OYKalyani:Kaly:06]KC501923
OYVMV-[IN:CG:OY177:07]FJ179371
OYVMV-[PK:Fai201:95]AJ002451
OYVMV-UP[IN:Var:EL6:05]KC501919
OYVMV-UP[IN:Var:EL26:05]KC501921
OYVMV
BYVDV-[IN:ND:OY131:05]FJ515747
BYVDV
BYVHV-[IN:Kar:OY82:05]KJ62830
BYVHV-[IN:Kar:OY76:03]FJ581298
BYVHV
ToLCNDV-[IN:Kar:OY81A:04]GU112082
ToLCNDV-[IN:Gun:OY136B:06]GU112086
ToLCNDV-[IN:Ban:OY135:05]GU112084
ToLCNDV-[IN:UD:OY142:06]EF035482
ToLCNDV-[IN:Aur:OY164A:06]GU112088
ToLCNDV
OELCuV-HR[IN:Soni:EL14A:06]GU112010
OELCuV-[IN:Soni:EL10:06]GU111996
OELCuV-[IN:Soni:EL14:06]GU111997
OELCuV-[IN:Mun:EL41:06]GU111999
OELCuV-[IN:Mun:EL37:06]GU111998
OELCuV-[IN:Soni:EL32:06]GU112000
OELCuV-[IN:Mun:EL38:06]GU112002
OELCuV-[IN:Mun:EL39:06]GU112001
OELCuV
CLCuBaV-[IN:BLR:OY136A:05]GU112003
CLCuBaV
BYVBhV-[IN:Bhu:OYBHU:03]FJ589571
BYVBhV
RaLCuV-[IN:Bih:ok09:10]GU732203
RaLCuV-[IN::Bih:Bih06:09]HQ257375
RaLCuV
OMoV-[BR:okra:08]EU914817
OMoV-[BR:okra:08]EU914819
OMoV
SiMMV-[BR:okra:07]EU908733
OYMMV-[MX:okra:10]GU990614
OYMMV-[MX:Igu:07]HQ020409
OYMoV-[MX:Igu:05]AY751753
OYMMV
OYCrV-[CM:okra:07]HE793425
OYCrV-[CM:okra:07]FM164724
OYCrV-[ML:okra:06]EU024119
OYCrV
OLCuCMV-[CM:okra:07]HE793428
OLCuCMV-[CM:okra:07]HE793426
OLCuCMV
CLCuGeV-[EG:okra:01]AY036010
CLCuGeV-[BFA:okra:08]FN554539
CLCuGeV
0·05
Central and South India
Indian subcontinent
Asian begomoviruses
Brazil
Mexico
Africa

vein Haryana virus, namely BYVHV [IN:Kar:OY82:05] and BYVHV [IN:Kar:OY76:05] clustered with okra-infecting bipartite begomovirus *Bhendi yellow vein Delhi virus* (BYVDV) and *Tomato leaf curl New Delhi virus* (ToLCNDV). The isolates of Bhendi yellow vein Karnal virus grouped with BYVMV isolates (Fig. 1).

An analysis of these novel sequences revealed that Bhendi yellow vein Haryana virus (BYVHV) and Bhendi yellow vein Karnal virus (BYVKnV) have a genome organization similar to that of previously characterized monopartite begomoviruses of the Old World. The results of comparisons of nucleotide sequences and predicted amino acid sequences of genes of BYVHV with other begomoviruses showed the highest degrees of similarity for the C1, C2 and C4 open reading frame (ORF) with ToLCNDV (Table 1). For the CP and V2 proteins the highest degrees of similarity were with OYVMV-PK and BYVMV, respectively. ORF C5 was most similar to that of the okra-infecting CLCuAlV species. The analysis of BYVKnV indicated that the ORF C1, C4 and C5 showed high degrees of nucleotide similarity with BYVDV. Whereas V2 and C2 were most similar to OYVMV, the ORF CP and C3 were most similar to BYVMaV. The iteron sequence of BYVHV and BYVKnV shared high degrees of similarity with those of CLCuAlV and OELCuV, respectively.

Analysis of recombination

A neighbour-net analysis of the okra-infecting begomovirus sequences revealed that they contained patterns of genetic variation with high degrees of phylogenetic conflict: a clear sign that many of the analysed sequences might be recombinant. Degrees of phylogenetic conflict were notably high in the cluster of sequences containing the BYVMV and BYVMaV isolates (Fig. 2). The possibility of widespread recombination in these populations was also clearly evident in the population structure analysis (Fig. 3). The admixed strains identified in the population structure analysis performed using STRUCTURE strongly suggested that many of the subpopulations contained recombinant genomes and that members of the BYVMV and BYVMaV populations were particularly admixed, a result that would have been expected if many of these viruses were recombinants.

A search for evidence of recombination in the BYVHV and BYVKnV genomes was undertaken using a data set of begomoviruses infecting different Malvaceae, Euphorbiaceae, Solanaceae, Cucurbitaceae and Leguminaceae species. The recombination events supported by at least three methods were considered to exclude unreliable signals. It was apparent from this analysis that these two novel begomoviruses bear multiple detectable recombination events with evidence of at least six in BYVHV and seven in BYVKnV (Table 2). Most of the recombination events detected in these two sequences have probably involved begomoviruses that are most closely related to viruses known to infect okra, tomato, chilli, cotton, croton, radish, sida and malvestrum. Among the recombination events that were detected, BYVHV has apparently obtained almost the entire coat protein ORF and intergenic region (IR) from a OYVMV (minor parent) and the rest of the genome from a virus related to ToLCNDV (major parent) with a *P*-value range of $1{\cdot}11 \times 10^{-29}$ to $2{\cdot}56 \times 10^{-76}$. In the case of BYVKnV, a recombination event was detected at nucleotide coordinates 104–1288, with BYVMaV as minor parent and OELCuV as major parent. This recombination event was supported by all the six detection methods of RDP4 with *P*-values of $6{\cdot}85 \times 10^{-15}$ to $8{\cdot}1 \times 10^{-54}$.

Genetic structure of okra-infecting begomovirus populations

The parametric population structure analysis that was performed indicated the existence of at least four genetically different okra-infecting begomovirus populations. Most of the isolates that were related to four previously characterized species, namely BYVMaV, BYVMV, OELCuV and ToLCNDV, segregated together and formed four genetically cohesive populations. The populations inferred by this analysis were consistent with the overall results of the phylogenetic analysis. The two predominant populations observed included a group of viruses genetically similar to BYVMV (52 sequences) and BYVMaV (42 sequences). Most of the individual sequences of the BYVMV population (44/52) were assigned to this population with >70% support. The sequences belonging to novel species, namely OY76 and OY80A, were assigned to ToLCNDV and BYVMV populations, respectively. Almost all the sequences (37/42) of the BYVMaV group were assigned with a support of >70% to this population. A total of nine viruses formed an OELCuV population and seven segregated to form a ToLCNDV population.

These four major populations were used to study the existence of subpopulation structure. Each of the four

Figure 1 Phylogenetic tree based on the complete DNA-A and DNA-A-like nucleotide sequences of isolates of okra-infecting begomoviruses characterized in this study and the other selected begomovirus sequences. Phylogenetic analyses were performed with MEGA v. 6.01 (Tamura *et al.*, 2013) using the neighbour-joining method. Tree branch support was determined using 1000 replications (bootstrap values are displayed at the nodes). Different clusters are indicated at the right with abbreviations as follows: BYVMaV: *Bhendi yellow vein Maharashtra virus*, CLCuAlV: *Cotton leaf curl Alabad virus*, BYVKnV: Bhendi yellow vein Karnal virus, BYVMV: *Bhendi yellow vein mosaic virus*, OYVMV: *Okra yellow vein mosaic virus*, BYVDV: *Bhendi yellow vein Delhi virus*, BYVHV: Bhendi yellow vein Haryana virus, ToLCNDV: *Tomato leaf curl New Delhi virus*, OELCuV: *Okra enation leaf curl virus*, CLCuBaV: *Cotton leaf curl Bangalore virus*, BYVBhV: *Bhendi yellow vein Bhubaneswar virus*, RaLCuV: *Radish leaf curl virus*, OMoV: *Okra mottle virus*, OYMMV: *Okra yellow mosaic Mexico virus*, OYCrV: *Okra yellow crinkle virus*, OLCuCMV: *Okra leaf curl Cameroon virus*, CLCuGeV: *Cotton leaf curl Gezira virus*.

Table 1 Percentage identities for complete genome and intergenic region (IR) nucleotide and open reading frame nucleotide and amino acid sequences (in parentheses) for Bhendi yellow vein Haryana virus [BYVHV] (OY76) and Bhendi yellow vein Karnal virus [BYVKnV] (OY80A) and other okra-infecting begomoviruses

	Genome		Intergenic region (IR)		V2		CP		C1		C2		C3		C4		C5	
Begomovirus[a]	OY76	OY80A	OY76	OY80A	OY76	OY80A	OY76	OY80A	OY76	OY80A	OY76	OY80A	OY76	OY80A	OY76	OY80A	OY76	OY80A
OELCuV	73	79	61	**88**[b]	78 (71)	80 (71)	73 (75)	72 (75)	73 (78)	82 (83)	73 (57)	84 (73)	80 (75)	89 (78)	74 (51)	94 (85)	–	–
BYVBhV	72	73	50	52	77 (68)	76 (66)	78 (82)	78 (83)	73 (77)	76 (75)	77 (64)	73 (60)	76 (65)	75 (62)	72 (45)	85 (66)	79 (52)	77 (51)
BYVMV-PK	79	**88**	60	87	90 (87)	**95 (95)**	91 (97)	91 (98)	72 (75)	85 (82)	74 (64)	85 (80)	**86 (82)**	**93 (86)**	74 (45)	86 (73)	92 (82)	86 (70)
BYVMaV	80	83	69	67	90 (89)	94 (93)	92 (97)	**93 (98)**	71 (74)	75 (72)	74 (64)	86 (81)	85 (81)	**93 (85)**	75 (55)	67 (46)	93 (82)	88 (72)
BYVMV-Mad	80	**88**	64	85	**91 (51)**	94 (59)	92 (97)	92 (98)	71 (74)	85 (82)	74 (64)	88 (85)	85 (82)	**93 (90)**	73 (48)	88 (70)	93 (81)	88 (70)
CLCuAlV	78	81	**85**	60	88 (75)	92 (81)	92 (96)	92 (97)	67 (67)	70 (66)	74 (64)	85 (80)	85 (82)	92 (86)	68 (43)	68 (41)	**96 (83)**	88 (69)
OYVMV	80	87	63	71	90 (89)	**95 (96)**	**93 (98)**	92 (99)	72 (76)	82 (79)	75 (62)	**90 (83)**	85 (80)	93 (88)	73 (45)	89 (72)	97 (92)	88 (70)
ToLCNDV	**88**	71	83	59	83 (72)	75 (66)	81 (91)	78 (90)	**96 (96)**	71 (72)	**89 (82)**	66 (50)	81 (74)	67 (60)	**94 (92)**	80 (58)	81 (53)	77 (45)
BYVDV	81	81	73	57	75 (66)	68 (60)	85 (84)	86 (83)	76 (80)	**86 (82)**	75 (62)	86 (81)	84 (78)	92 (84)	75 (52)	**98 (94)**	92 (80)	**90 (75)**
CLCuBaV	75	82	57	77	78 (71)	81 (76)	82 (93)	82 (93)	74 (77)	85 (84)	73 (58)	86 (74)	79 (72)	87 (75)	75 (50)	97 (93)	85 (61)	83 (58)
RaLCuV	70	73	47	49	74 (47)	74 (47)	80 (75)	80 (75)	72 (73)	80 (78)	56 (21)	60 (28)	69 (74)	69 (43)	72 (45)	90 (78)	–	–

[a]The acronyms of begomovirus species used in the pairwise comparisons. OELCuV: *Okra enation leaf curl virus*; BYVBhV: *Bhendi yellow vein Bhubaneswar virus*; BYVMV-PK: *Bhendi yellow vein mosaic virus*; BYVMaV: *Bhendi yellow vein Maharashtra virus*; BYVMV: *Bhendi yellow vein mosaic virus*; CLCuAlV: *Cotton leaf curl Alabad virus*; OYVMV: *Okra yellow vein mosaic virus*; ToLCNDV: *Tomato leaf curl New Delhi virus*; BYVDV: *Bhendi yellow vein Delhi virus*; CLCuBaV: *Cotton leaf curl Bangalore virus*; RaLCuV: *Radish leaf curl virus*.

[b]Figures in bold are the highest value for each column.

populations was analysed separately to determine the optimum number of subpopulations that best represented within-population structure. The BYVMaV population contained five and BYVMV contained three minor subpopulations. There was evidence of potentially three and two minor subpopulations in the OELCuV and ToLCNDV populations, respectively. The majority of the isolates assigned to these subpopulations clustered within these subpopulations with more than 60% support. The evidence for population structure within these genetically distinct subpopulations was tested using AMOVA and it was found that all 13 were supported by a highly significant F_{ST} statistic (F_{ST} = 0·635; $P < 0·0001$). Overall, four major populations and 13 subpopulations accounted for 63·59% of the observed genetic diversity (data not shown). Additionally, F_{ST} statistics indicated that most of the subpopulations displayed a high degree of genetic differentiation (Table 3). A very high degree of differentiation was recorded between the BYVMaV-II and the BYVMaV-IV subpopulations (F_{ST} = 0·61). The lowest degree of differentiation (N_{ST} = 0·02) was observed between the OELCuV-I and OELCuV-II subpopulations.

Genetic variability and tests of neutrality

A summary of haplotype sequence polymorphisms and diversity is presented in Table 4. With regard to haplotype distributions, a total of 107 haplotypes were detected out of the 110 sequences analysed. The highest number of haplotypes was observed in BYVMV-I with 44 haplotypes, followed by BYVMaV-I with 27 haplotypes. All the subpopulations recorded high haplotype diversity (hd > 0·95) with the exception of BYVMaV-II. The number of segregating sites was highest in BYVMV-I, with 988 segregating sites and nearly 181 average nucleotide differences. A high degree of nucleotide diversity (0·26351) with an average of 717 nucleotide differences was noted in the ToLCNDV-II subpopulation. Tests of neutrality performed on subpopulations showed statistically significant deviations for different ORFs, mainly in the BYVMaV and BYVMV subpopulations (Table S3). Most of these statistically significant values of three different tests were negative. However, there were positive but statistically nonsignificant values for a few ORFs in different subpopulations.

Discussion

India accounts for more than 70% of worldwide okra production. Among the major okra-producing states of India, Andhra Pradesh and West Bengal produce 30% of the country's okra. The sampling survey conducted in the present study focused on the major okra-growing regions of these two states and six others, along with two union territories. During these surveys, several okra fields contained plants displaying yellow mosaic and other disease symptoms. Preliminary analyses of these samples by PCR diagnostic tests indicated that the plants with symptoms were infected with begomoviruses.

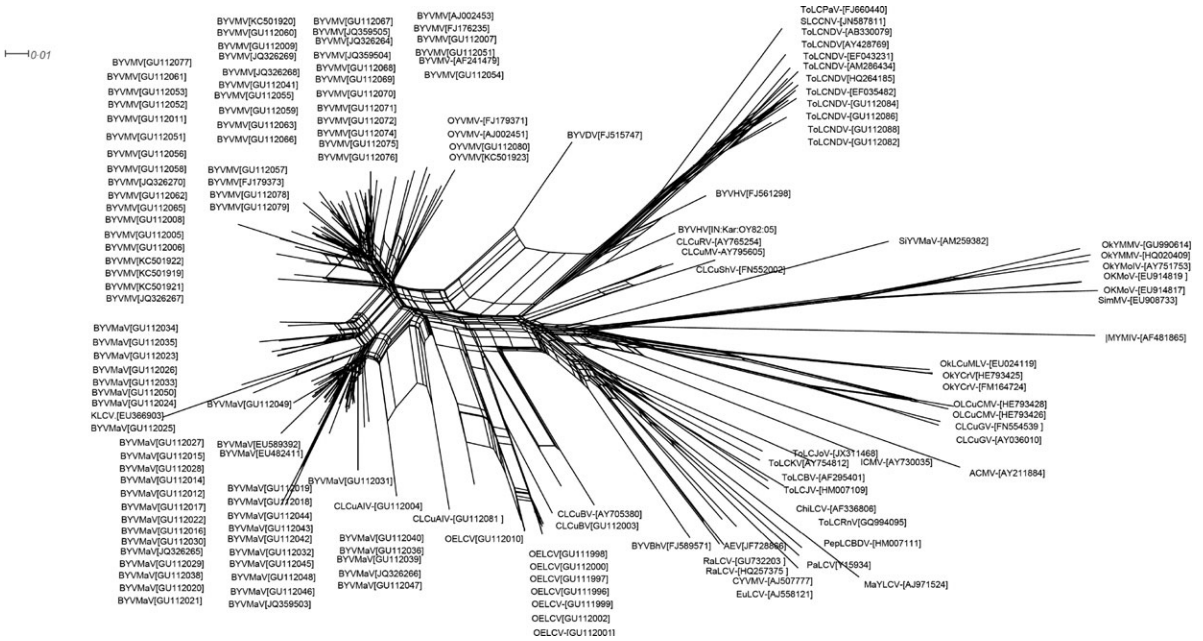


Figure 2 Neighbour net generated for okra-infecting begomoviruses characterized in this study along with other selected begomovirus sequences. Analysis was performed using SplitsTree4 (Huson & Bryant, 2006). Extensive network structures (rather than predominantly bifurcating tree-like structures) indicate that there are substantially conflicting phylogenetic signals within the analysed sequences, a situation that is consistent with the hypothesis that many of the analysed sequences are potentially recombinant. The abbreviations of isolates characterized in this study are listed in Table S1.

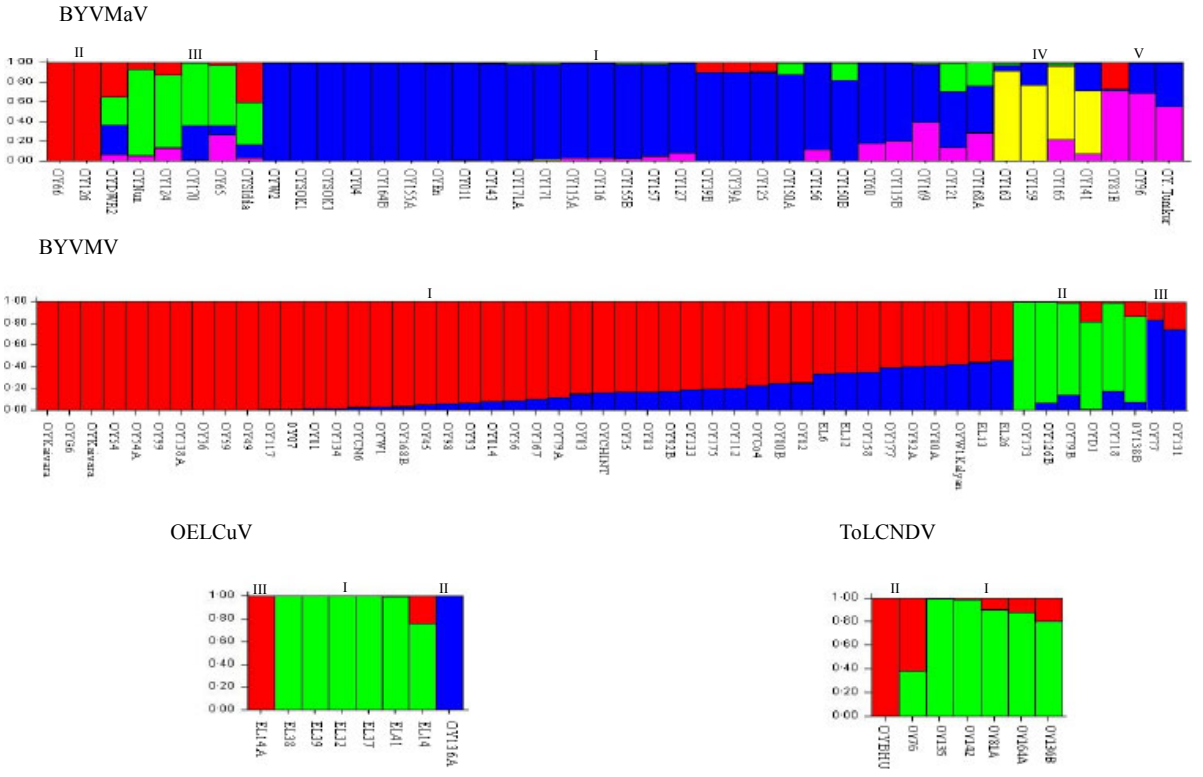


Figure 3 Subpopulation structure analysis within the okra-infecting begomovirus lineages performed using the computer program STRUCTURE (Pritchard *et al.*, 2000). Different subpopulations (k) are indicated by different colours. Each column within the population represents an isolate sorted according to membership proportion. *Bhendi yellow vein Maharashtra virus* (BYVMaV), K = 5. *Bhendi yellow vein Madurai virus* (BYVMV), K = 3. *Okra enation leaf curl virus* (OELCuV), K = 3. *Tomato leaf curl New Delhi virus* (ToLCNDV), K = 2.

Table 2 Breakpoint analysis of novel okra-infecting begomoviruses Bhendi yellow vein Haryana virus (BYVHV) and Bhendi yellow vein Karnal virus (BYVKnV) and their putative parental sequences

Virus (isolate)	Breakpoint start–end	Major parent[a] (GenBank accession)	Minor parent[a] (GenBank accession)	Method and *P*-value					
				RDP	GENECONV	MAXCHI	CHIMERA	SISCAN	3SEQ
BYVHV (OY76)	209–300	BYVMV (GU112051)	ToLCNDV (AY428769)	$5{\cdot}52 \times 10^{-5}$	$1{\cdot}11 \times 10^{-3}$	$5{\cdot}63 \times 10^{-3}$	$5{\cdot}50 \times 10^{-4}$	NS[b]	NS
	28–1235	ToLCNDV (GU112084)	OYVMV (AJ002451)	$5{\cdot}4 \times 10^{-43}$	$2{\cdot}08 \times 10^{-35}$	$6{\cdot}20 \times 10^{-32}$	$3{\cdot}20 \times 10^{-31}$	$1{\cdot}11 \times 10^{-29}$	$2{\cdot}56 \times 10^{-76}$
	789–1003	BYVMaV (EU589392)	CLCuAlV (GU112081)	$5{\cdot}13 \times 10^{-7}$	$2{\cdot}50 \times 10^{-6}$	$3{\cdot}67 \times 10^{-4}$	$1{\cdot}66 \times 10^{-4}$	NS	$1{\cdot}83 \times 10^{-6}$
	1449–1827	BYVMV (GU112051)	CLCuAlV (GU112081)	$3{\cdot}94 \times 10^{-5}$	$4{\cdot}46 \times 10^{-3}$	$4{\cdot}49 \times 10^{-6}$	$3{\cdot}16 \times 10^{-10}$	NS	$3{\cdot}26 \times 10^{-5}$
	1973–2030	BYVMV (GU112008)	CLCuMV (AY795605)	NS	NS	$2{\cdot}27 \times 10^{-2}$	$1{\cdot}81 \times 10^{-3}$	NS	$2{\cdot}60 \times 10^{-3}$
	2122–2307	CLCuMV (AY795605)	CLCuGV (AY036010)	$1{\cdot}31 \times 10^{-6}$	NS	$1{\cdot}34 \times 10^{-4}$	$6{\cdot}22 \times 10^{-4}$	$4{\cdot}26 \times 10^{-8}$	$2{\cdot}54 \times 10^{-2}$
BYVKnV (OY80A)	104–1288	OELCuV (GU112001)	BYVMaV (EU589392)	$8{\cdot}21 \times 10^{-37}$	$8{\cdot}24 \times 10^{-37}$	$9{\cdot}53 \times 10^{-24}$	$6{\cdot}86 \times 10^{-15}$	$1{\cdot}57 \times 10^{-24}$	$8{\cdot}10 \times 10^{-54}$
	127–577	BYVMaV (GU112023)	OYVMV (AJ002451)	NS	NS	$1{\cdot}78 \times 10^{-4}$	$3{\cdot}21 \times 10^{-3}$	NS	$6{\cdot}64 \times 10^{-6}$
	630–904	BYVMV (GU112051)	CLCuRV (AY765254)	$6{\cdot}84 \times 10^{-13}$	$1{\cdot}55 \times 10^{-5}$	$1{\cdot}16 \times 10^{-7}$	NS	NS	$2{\cdot}67 \times 10^{-16}$
	1564–1631	CYVMV (AJ507777)	CLCuAlV (GU112081)	$4{\cdot}71 \times 10^{-2}$	NS	$6{\cdot}34 \times 10^{-4}$	$2{\cdot}13 \times 10^{-2}$	NS	$1{\cdot}73 \times 10^{-2}$
	1742–1855	CLCuAlV (GU112081)	ToLCJoV (HM007108)	$3{\cdot}17 \times 10^{-2}$	NS	$1{\cdot}04 \times 10^{-4}$	NS	NS	$7{\cdot}25 \times 10^{-2}$
	2133–2314	CLCuRV (AY765254)	SiYVMaV (AM259382)	$3{\cdot}65 \times 10^{-1}$	NS	NS	NS	$3{\cdot}74 \times 10^{-2}$	$2{\cdot}82 \times 10^{-2}$
	2267–2885	MaYLCV (AJ971524)	RaLCV (GU732203)	$2{\cdot}94 \times 10^{-9}$	$9{\cdot}47 \times 10^{-6}$	$1{\cdot}35 \times 10^{-5}$	$8{\cdot}08 \times 10^{-6}$	$2{\cdot}50 \times 10^{-13}$	NS

[a]Full name of the begomovirus species identified as major and minor parent is as follows: BYVMaV: *Bhendi yellow vein Maharashtra virus*; ToLCNDV: *Tomato leaf curl New Delhi virus*; BYVMV: *Bhendi yellow vein mosaic virus*; CLCuMV: *Cotton leaf curl Multan virus*; OELCuV: *Okra enation leaf curl virus*; ToLCJoV: *Tomato leaf curl Joydebpur virus*; CYVMV: *Croton yellow vein mosaic virus*; CLCuAlV: *Cotton leaf curl Alabad virus*; CLCuRV: *Cotton leaf curl Rajasthan virus*; MaYLCV: *Malvastrum yellow leaf curl virus*; OYVMV: *Okra yellow vein mosaic virus*; CLCuGV: *Cotton leaf curl Gezira virus*; SiYVMaV: *Sida yellow vein Madurai virus*; RaLCuV: *Radish leaf curl virus*.
[b]NS, not significant.

Table 3 Genetic differentiation of okra-infecting begomovirus subpopulations

Subpopulation[a]	N_{ST}[b]	F_{ST}[b]
Bhendi yellow vein Maharastra virus		
BYVMaV I/BYVMaV II	0·40	0·40
BYVMaV I/BYVMaV III	0·25	0·25
BYVMaV I/BYVMaV IV	0·50	0·49
BYVMaV II/BYVMaV III	0·35	0·35
BYVMaV II/BYVMaV IV	0·62	0·61
BYVMaV III/BYVMaV IV	0·53	0·51
Bhendi yellow vein Madurai virus		
BYVMV I/BYVMV II	0·48	0·46
BYVMV I/BYVMV III	0·05	0·07
BYVMV II/BYVMV III	0·32	0·31
Okra enation leaf curl virus		
OELCuV I/OELCuV II	0·02	0·06
Tomato leaf curl New Delhi virus		
ToLCNDV I/ToLCNDV II	0·18	0·19

[a]The isolates of subpopulations are as in Figure 3.
[b]The N_{ST} and F_{ST} values < 0·05 indicate little genetic differentiation; 0·05–0·15 indicate moderate differentiation; 0·15–0·25 indicate great differentiation; > 0·25 indicate high differentiation.

Because of the wide distribution of plants with symptoms and the fact that such symptoms are commonly associated with begomovirus infections, this study set out to characterize begomoviruses infecting these plants at the molecular level, so as to uncover their diversity and population structure. Although a DNA-B molecule is absolutely required by bipartite begomoviruses, for monopartite begomoviruses, infections are often achievable with the DNA-A-like component alone or in the presence of satellite DNA molecules referred to as betasatellites and alphasatellites. The betasatellites associated with the disease samples collected in this survey have been characterized previously (Venkataravanappa *et al.*, 2011). The focus of this study was to elucidate the diversity and population structure of okra-associated begomovirus DNA-A-like component sequences.

Following the species and strain demarcation thresholds recommended for begomovirus classification by the ICTV (Fauquet *et al.*, 2008), 89 of the DNA-A-like begomovirus sequences characterized in this study could be classified as new strains or variants of previously defined begomovirus species. Two of the DNA-A-like sequences characterized from the northern state, Haryana, were less than 89% similar to any known begomovirus sequences currently deposited in GenBank. Therefore these two isolates probably represent two new species of begomoviruses, for which the names Bhendi yellow vein Haryana virus [India:Haryana:06] and Bhendi yellow vein Karnal virus [India:Karnal:06] are proposed. Furthermore, these two novel species were identified as being recombinants. Most recombination breakpoints in these new species were detected in the *Rep* gene, *cp* and *REn* genes at sites that have been previously reported to be recombination hotspots within begomovirus genomes (Lefeuvre *et al.*, 2007; Prasanna & Rai, 2007). Most of

Table 4 Descriptive genetic parameters of diversity of okra-infecting begomovirus subpopulations

Population[a]	No. of sequences	Parameter[b]							
		s	Eta	k	π	h	hd	Θ-w	Θ-Eta
Bhendi yellow vein Maharastra virus									
BYVMaV I	27	581	638	103·12	0·04	26	1·00	0·06	0·06
BYVMaV II	3	128	128	85·33	0·03	2	0·67	0·03	0·03
BYVMaV III	5	289	303	128·40	0·05	5	1·00	0·05	0·05
BYVMaV IV	4	288	308	158·83	0·06	4	1·00	0·06	0·06
BYVMaV V	3	509	537	348·67	0·13	3	1·00	0·12	0·13
Bhendi yellow vein Madurai virus									
BYVMV I	44	988	1242	180·72	0·07	44	1·00	0·08	0·11
BYVMV II	6	369	405	159·33	0·06	6	1·00	0·06	0·07
BYVMV III	2	563	563	563·00	0·21	2	1·00	0·21	0·21
Okra enation leaf curl virus									
OELCuV I	7	103	103	39·67	0·01	6	1·00	0·02	0·02
OELCuV II	2	479	479	479·00	0·18	2	1·00	0·18	0·18
Tomato leaf curl New Delhi virus									
ToLCNDV I	5	249	260	117·70	0·04	5	1·00	0·04	0·05
ToLCNDV II	2	717	717	717·00	0·26	2	1·00	0·26	0·26

[a]The member isolates of subpopulations are as presented in Figure 3.

[b]s, total number of segregating sites; Eta, total number of mutations; k, average number of nucleotide differences between sequences; π, nucleotide diversity; h, haplotype number; hd, haplotype diversity; Θ-w, population mutation rate on total number of segregating sites; Θ-Eta, population mutation rate on total number of mutations.

the recombinationally derived genome fragments present within these novel begomovirus sequences probably originated from begomovirus species known to infect both malvaceous and non-malvaceous hosts.

Recombination is one of the key mechanisms that has contributed to the diversification and evolution of begomoviruses (Lefeuvre *et al.*, 2009). A variety of different recombination analysis methods were used in this study to provide evidence for recombination in major okra-infecting begomovirus populations. The significance of begomovirus recombination has been well established. The epidemics associated with the begomovirus recombinants and the potential adaptation of recombinants to new host species and vector biotypes (Sanz *et al.*, 2000; Berrie *et al.*, 2001; Monci *et al.*, 2002) underscore the significance of begomovirus recombination. Begomoviruses infecting malvaceous and non-malvaceous crops are, in fact, known to be particularly recombinogenic and have the potential to evolve into epidemiologically significant recombinants. For example, the cotton leaf curl disease epidemic that occurred in Pakistan during the 1990s was attributed to a species complex of mostly recombinant begomovirus species (Mansoor *et al.*, 2003). This epidemic was managed using resistant varieties. However, almost a decade later a second cotton leaf curl disease epidemic has occurred in this country that is believed to be caused by a resistance-breaking recombinant begomovirus/betasatellite complex (Amrao *et al.*, 2010). Reports have been made on the displacement of cotton-infecting begomoviruses by this recombinant virus in northwestern India and Pakistan (Rajagopalan *et al.*, 2012). Therefore it is necessary to continuously survey and monitor the diversity and composition of begomovirus populations on cultivated crops in different regions. Indeed, two novel species identified in the current study were recombinants, indicating that this important evolutionary mechanism is aiding the evolution of new begomovirus species that are capable of infecting okra.

Aside from identifying two novel species and providing evidence for widespread recombination, the current study has for the first time brought into focus the diversity and distribution of the okra-infecting begomovirus populations of India. The results of the current investigation revealed that there are presently four major okra-infecting begomovirus species in the sampled regions of India: BYVMV, BYVMaV, OELCuV and OYVMV. Among these, BYVMV and BYVMaV were found to be predominant. BYVMV predominates in the majority of studied locations, from northern India (Haryana, Delhi and Chandigarh), through central (Madhya Pradesh and Maharashtra), eastern (West Bengal and Uttar Pradesh) and southern India (Tamil Nadu, Karnataka, Kerala and Andhra Pradesh). BYVMaV, on the other hand, appears to have a distribution that is limited to the southern (Andhra Pradesh and Tamil Nadu) and central (Maharashtra) Indian states. Interestingly, the okra samples analysed here did not contain representatives of previously reported monopartite and bipartite begomovirus species that have been found infecting okra in India: species such as BYVBhV, BYVDV, CLCuBaV and RaLCV. However, the evidence for an absence of any species should be interpreted with caution in the context of limitations associated with scale of overall survey. Begomoviruse species are known to exhibit both widespread geographical distributions and geographical based population structures. Therefore differences in the observed

distributions of the various begomovirus species sampled here are not particularly surprising. However, it is interesting that only two species predominate within okra, with only one of these (BYVMV) having a pan-Indian distribution. The successful survival, population establishment and diversity of begomovirus species are known to be influenced by a complex interaction of selection pressures exerted by the genetic diversity of the insect vector species, the range of host plant species infected by the virus and differential interactions between these plant species with both virus variants and vector biotypes. Therefore the differences in the observed prevalence of the various Indian okra-infecting begomoviruses could possibly reflect differences in the adaptation of the virus species to whitefly transmission and/or infection of the spectrum of host species that are available in different parts of India. Variation between begomovirus species regarding their adaptation to certain hosts, their adaptation to particular environments and their capacity to compete with other begomovirus species in these environments, have been reported elsewhere (Garcia-Andres *et al.*, 2007) and include examples of begomoviruses that infect malvaceaeous crop species (Rajagopalan *et al.*, 2012). On the vector front, some members of *Bemisia tabaci* species complex are known to transmit begomoviruses more efficiently. For example, Asia II1 was demonstrated to transmit cotton leaf curl virus more efficiently than MEAM-1 (Saleem *et al.*, 2003). Indeed, there are reports of association of Asia II1 with the severe cotton leaf curl disease epidemic that is presently occurring in Pakistan (Ahmed *et al.*, 2011).

The population structure analyses also indicated subpopulation structure within the four major okra-infecting virus populations. However, the pattern of population subdivision did not correlate with the geographical distributions of the subpopulations. The influence of factors such as variation in the distribution of host species other than okra could not be ruled out, as this study focused only on okra. Therefore, better understanding of the factors responsible for the observed patterns of genetic diversity and structure could be gained if future studies are focused on the genetic diversity of the whitefly populations feeding on okra and other host plant species as well as begomoviruses associated with non-okra plants. Neutrality tests and selection analyses indicated that all of the subpopulations identified within the four major okra-infecting begomovirus populations were evolving in a non-neutral manner with a clear evidence of purifying selection operating at the nucleotide level.

In conclusion, the survey and analyses performed here provide a picture of okra-associated begomovirus diversity in the most important okra production region of the world. The experimental approach used in this study for the characterization and elucidation of genetic structure was based on data generated using the amplification of viral genomic DNA by PCR. Although this method is effective in detecting close relatives of known viruses, it is less effective in discovering viruses with low levels of similarity to known viruses. As an alternative to PCR, the sequence independent amplification and detection of begomoviruses using rolling circle amplification is a highly sensitive and robust method that is currently being widely employed to characterize begomovirus diversity. However, both these approaches can be biased in favour of detecting the most abundant viral variants. Use of these approaches might not be ideal for the detection of rare begomovirus variants and could therefore yield underestimates of begomovirus diversity (Idris *et al.*, 2014).

Nevertheless, the information and picture of diversity uncovered in this study has the potential to aid resistance breeding, disease management and resistant variety/gene deployment programmes. Using agroinoculation and/or biolistics-based methods of inoculation, the representative species from predominant virus lineages elucidated in the current study could be used to screen okra varieties for resistance against okra-infecting begomoviruses.

Acknowledgements

The authors wish to thank Darren P. Martin for critical review and editing of the manuscript. This work was supported by the project 'Establishment of association of begomovirus species with yellow vein mosaic disease (YVMD) in wild and cultivated species of okra and identification of source of resistance to the most predominant virus' funded by National Fund for Basic and Strategic Research in Agricultural Sciences (NFBSFARA), Indian Council of Agricultural Research, Government of India, New Delhi, India. The authors have no conflict of interest to declare.

References

Ahmed MZ, De Barro PJ, Greef JM, Ren SX, Naveed M, Qiu BL, 2011. Genetic identity of the *Bemisia tabaci* species complex and association with high cotton leaf curl disease (CLCuD) incidence in Pakistan. *Pest Management Science* **67**, 307–17.

Akhtar S, Khan AJ, Singh AS, Briddon RW, 2014. Identification of a disease complex involving a novel monopartite begomovirus with beta- and alphasatellites with okra leaf curl disease in Oman. *Archives of Virology* **159**, 1199–205.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ, 1990. Basic local alignment search tool. *Journal of Molecular Biology* **215**, 403–10.

Amrao L, Amin I, Shahid MS, Briddon RW, Mansoor S, 2010. Cotton leaf curl disease in resistant cotton is associated with a single begomovirus that lacks an intact transcriptional activator protein. *Virus Research* **152**, 153–63.

Berrie LC, Rybicki EP, Rey MEC, 2001. Complete nucleotide sequence and host range of South African cassava mosaic virus: further evidence for recombination amongst begomoviruses. *Journal of General Virology* **82**, 53–8.

Brown JK, Fauquet CM, Briddon RW, Zerbini M, Moriones E, Navas-Castillo J, 2012. Geminiviridae. In: King AMQ, Adams MJ, Carstens EB, Lefkowitz EJ, eds. *Virus Taxonomy: Ninth Report of the International Committee on Taxonomy of Viruses*. London, UK: Associated Press, 351–73.

Charrier A, 1984. *Genetic Resources of the Genus* Abelmoschus. Rome, Italy: International Board for Plant Genetic Resources, 21–41.

Doyle JJ, Doyle JL, 1990. Isolation of plant DNA from fresh tissue. *Focus* **12**, 13–5.

Evanno G, Regnaut S, Goudet J, 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology* **14**, 2611–20.

Excoffier L, Laval G, Schneider S, 2005. ARLEQUIN Ver. 3.0: an integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online* **1**, 47–50.

FAOSTAT, 2013. Production Quantities by Country. Food and Agriculture Organization of the United Nations. [http://faostat3.fao.org/faostat-gateway/go/to/browse/Q/QC/E]. Accessed 9 September 2014.

Fauquet CM, Briddon RW, Brown JK *et al.*, 2008. Geminivirus strain demarcation and nomenclature. *Archives of Virology* **153**, 783–821.

Garcia-Andres S, Accotto GP, Navas-Castillo J, Moriones E, 2007. Founder effect, plant host, and recombination shape emergent population of begomoviruses that cause the tomato yellow leaf curl disease in the Mediterranean basin. *Virology* **359**, 302–12.

Harrison BD, Muniyappa V, Swanson MM, Roberts IM, Robinson DJ, 1991. Recognition and differentiation of seven whitefly transmitted geminiviruses from India and their relationships to African cassava mosaic and Thailand mungbean yellow mosaic viruses. *Annals of Applied Biology* **118**, 299–308.

Hernandez-Zepeda C, Isakeit T, Scott A Jr, Brown JK, 2010. First report of Okra yellow mosaic Mexico virus in Okra in the United States. *Plant Disease* **94**, 924.

Huson DH, Bryant D, 2006. Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution* **23**, 254–67.

Idris A, Al-Saleh M, Piatek MJ, Al-Shahwan I, Ali S, Brown JK, 2014. Viral metagenomics: analysis of begomoviruses by Illumina high-throughput sequencing. *Viruses* **6**, 1219–36.

Kon T, Rojas MR, Abdourhamane IK, Gilbertson RL, 2009. Roles and interactions of begomoviruses and satellite DNAs associated with okra leaf curl disease in Mali, West Africa. *Journal of General Virology* **90**, 1001–13.

Kulkarni GS, 1924. Mosaic and other related diseases of crops in the Bombay presidency. *Poona Agriculture College Magazine* **16**, 6–12.

Lefeuvre P, Martin DP, Hoareau M *et al.*, 2007. Begomovirus 'melting pot' in the south-west Indian Ocean islands: molecular diversity and evolution through recombination. *Journal of General Virology* **88**, 3458–68.

Lefeuvre P, Lett JM, Varsani A, Martin DP, 2009. Widely conserved recombination patterns among single-stranded DNA viruses. *Journal of Virology* **83**, 2697–707.

Leke WN, Sattar MN, Ngane EB, Ngeve JM, Kvarnheden A, Brown JK, 2013. Molecular characterisation of begomoviruses and DNA satellites associated with okra leaf curl disease in Cameroon. *Virus Research* **174**, 116–25.

Librado P, Rozas J, 2009. DNASP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25**, 1451–2.

Mansoor S, Amin I, Iram S *et al.*, 2003. Breakdown of resistance in cotton to cotton leaf curl disease in Pakistan. *Plant Pathology* **52**, 784.

Martin DP, Lemey P, Lott M, Vincent M, Posada D, Lefeuvre P, 2010. RDP3: a flexible and fast computer program for analyzing recombination. *Bioinformatics* **26**, 2462–3.

Monci F, Sanchez-Campos S, Navas-Castillo J, Moriones E, 2002. A natural recombinant between the geminiviruses *Tomato yellow leaf curl Sardinia virus* and *Tomato yellow leaf curl virus* exhibits a novel pathogenic phenotype and is becoming prevalent in Spanish populations. *Virology* **303**, 317–26.

Muhire B, Martin DP, Brown JK *et al.*, 2013. A genome-wide pairwise-identity-based proposal for the classification of viruses in the genus *Mastrevirus* (*Geminiviridae*). *Archives of Virology* **158**, 1411–24.

Nawaz-ul-Rehman MS, Fauquet CM, 2009. Evolution of geminiviruses and their satellites. *FEBS Letters* **583**, 1825–32.

Prasanna HC, Rai M, 2007. Detection and frequency of recombination in tomato-infecting begomoviruses of South and Southeast Asia. *Virology Journal* **4**, 111.

Prasanna HC, Sinha DP, Verma A *et al.*, 2010. The population genomics of begomoviruses: global population structure and gene flow. *Virology Journal* **7**, 220.

Pritchard JK, Stephens M, Donnelly P, 2000. Inference of population structure using multilocus genotypic data. *Genetics* **155**, 945–59.

Rajagopalan PA, Naik A, Katturi P, Kurulekar RS, Anandalaksmi R, 2012. Dominance of resistance breaking cotton leaf curl Burewala virus (CLCuBuV) in northwestern India. *Archives of Virology* **157**, 855–68.

Saleem HM, Bedford ID, Evans AAF, Markham PG, 2003. Geminivirus transmission by different biotypes of the whitefly *Bemisia tabaci* (Gennadius). *Pakistan Journal of Zoology* **35**, 343–51.

Sanz AI, Fraile A, García-Arenal F *et al.*, 2000. Multiple infection, recombination and genome relationships among begomovirus isolates found in cotton and other plants in Pakistan. *Journal of General Virology* **81**, 1839–49.

Seal SE, van den Bosch F, Jeger MJ, 2006. Factors influencing begomovirus evolution and their increasing global significance: implications for sustainable control. *Critical Reviews in Plant Sciences* **25**, 23–46.

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S, 2013. MEGA6: molecular evolutionary genetics analysis version 6.0. *Molecular Biology and Evolution* **28**, 2731–9.

Thompson JD, Higgins DG, Gibson TJ, 1994. CLUSTALW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research* **22**, 4673–80.

Uppal BN, Varma PM, Capoor SP, 1940. Yellow vein mosaic of bhendi. *Current Science* **9**, 227–8.

Venkataravanappa V, Reddy CNL, Swaranalatha P, Jalali S, Briddon RW, Krishna Reddy M, 2011. Diversity and phylogeography of begomovirus-associated beta satellites of okra in India. *Virology Journal* **8**, 555.

Venkataravanappa V, Reddy CNL, Jalali S, Krishna Reddy M, 2012a. Molecular characterization of distinct bipartite begomovirus infecting bhendi (*Abelmoschus esculentus* L.) in India. *Virus Genes* **44**, 522–35.

Venkataravanappa V, Lakshminarayana RC, Swarnalatha P, Devaraju A, Jalali S, Krishna Reddy M, 2012b. Molecular evidence for association of Cotton leaf curl Alabad virus with yellow vein mosaic disease of okra in North India. *Archives of Phytopathology and Plant Protection* **45**, 1–19.

Venkataravanappa V, Reddy CNL, Jalali S, Krishna Reddy M, 2013a. Molecular characterization of a new species of begomovirus associated with yellow vein mosaic of bhendi (Okra) in Bhubhaneswar, India. *European Journal of Plant Pathology* **136**, 811–22.

Venkataravanappa V, Reddy CNL, Devaraju A, Jalali S, Reddy MK, 2013b. Association of *Cotton leaf curl Bangalore virus* with yellow vein mosaic disease of okra is further evidence for natural recombination of begomoviruses in India. *Indian Journal of Virology* **24**, 188–98.

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site.

Figure S1. Two-dimensional genome-wide pairwise nucleotide sequence identity plot generated using SEQUENCE DEMARCATION TOOL (SDT v. 1.0). New virus species identified in this study are in red and green colour. Virus names used in this analysis are as in Figure 1.

Table S1. Details of geographical origin of isolates and characterization of okra-infecting begomoviruses of India.

Table S2. Details of okra-infecting begomovirus strains identified in the present study.

Table S3. Tests of neutrality for different open reading frame in sub-populations of okra-infecting begomoviruses.