Identification and characterisation of a highly divergent geminivirus : evolutionary and taxonomic implications

Pauline Bernardo, Michael Golden, M. Akram, - Naimuddin, N. Nadarajan, Emmanuel Fernandez, Martine Granier, A. G. Rebelo, Michel Peterschmitt, D. P. Martin, et al.

Virus Research 177 (2013) 35–45

Contents lists available at ScienceDirect

Virus Research

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

Identification and characterisation of a highly divergent geminivirus: Evolutionary and taxonomic implications☆

Pauline Bernardo[a,b], Michael Golden[c], Mohammad Akram[d], Naimuddin[d], Nagaswamy Nadarajan[e], Emmanuel Fernandez[a], Martine Granier[a], Anthony G. Rebelo[f], Michel Peterschmitt[a], Darren P. Martin[c], Philippe Roumagnac[a,*]

[a] CIRAD/UMR BGPI, TA A54/K, Campus International de Baillarguet, 34398 Montpellier Cedex 5, France
[b] INRA/UMR, BGPI, TA A54/K, Campus International de Baillarguet, 34398 Montpellier Cedex 5, France
[c] Computational Biology Group, Institute of Infectious Disease and Molecular Medicine, UCT Faculty of Health Sciences, Observatory 7925, South Africa
[d] Division of Crop Protection, Indian Institute of Pulses Research, Kalyanpur, Kanpur 208024, India
[e] Indian Institute of Pulses Research, Kalyanpur, Kanpur 208024, India
[f] South African National Biodiversity Institute, Kirstenbosch, Private Bag X7, Claremont, 7735 Cape Town, South Africa

ARTICLE INFO

Article history:
Received 8 April 2013
Received in revised form 8 July 2013
Accepted 9 July 2013
Available online 22 July 2013

Keywords:
DNA viruses
Geminivirus
Plant viruses
Phylogenetic analysis
Evolution
Taxonomy

ABSTRACT

During a large scale "*non a priori*" survey in 2010 of South African plant-infecting single stranded DNA viruses, a highly divergent geminivirus genome was isolated from a wild spurge, *Euphorbia caput-medusae*. In addition to being infectious in *E. caput-medusae*, the cloned viral genome was also infectious in tomato and *Nicotiana benthamiana*. The virus, named Euphorbia caput-medusae latent virus (EcmLV) due to the absence of infection symptoms displayed by its natural host, caused severe symptoms in both tomato and *N. benthamiana*. The genome organisation of EcmLV is unique amongst geminiviruses and it likely expresses at least two proteins without any detectable homologues within public sequence databases. Although clearly a geminivirus, EcmLV is so divergent that we propose its placement within a new genus that we have tentatively named Capulavirus. Using a set of highly divergent geminiviruses genomes, it is apparent that recombination has likely been a primary process in the genus-level diversification of geminiviruses. It is also demonstrated how this insight, taken together with phylogenetic analyses of predicted coat protein and replication associated protein (Rep) amino acid sequences indicate that the most recent common ancestor of the geminiviruses was likely a dicot-infecting virus that, like modern day mastreviruses and becurtoviruses, expressed its Rep from a spliced complementary strand transcript.

1. Introduction

Among plants viruses, those of the family *Geminiviridae* are responsible for a disproportionately large number of recently emergent crop diseases worldwide. They have dramatically impacted agricultural yields over the past 50 years (Moffat, 1999), and are a major threat to the food security of developing countries in the tropical and sub-tropical regions of the world (Rey et al., 2012; Rybicki and Pietersen, 1999). Most at risk are countries in sub-Saharan Africa where reports near the beginning of the 1900s of diseases in exotic introduced cultivated staple food species such as cassava and maize provided the first clear descriptions of geminivirus infections (Fuller, 1901; Warburg, 1894). Caused by at least seven different African geminivirus species, cassava mosaic disease (CMD) is today recognised as the most important biotic constraint of cassava production throughout this region (Legg and Fauquet, 2004; Patil and Fauquet, 2009). For instance, a recent CMD epidemic affected at least nine countries in East and Central Africa (spanning an area of 2.6 million square kilometres) inflicting annual economic losses of US$1.9–2.7 billion (Patil and Fauquet, 2009). Similarly, throughout sub-Saharan Africa the geminivirus species that causes maize streak disease (MSD) inflicts annual losses averaging approximately US$120–480 million (Martin and Shepherd, 2009). In addition, a range of other African geminivirus species have been described in the past three decades that, while obviously causing serious yield reductions in tomatoes, beans, and sweet-potatoes, have a currently unquantified impact on African agriculture (Rey et al., 2012).

* Corresponding author. Tel.: +33 (0)499 62 48 58.
E-mail address: philippe.roumagnac@cirad.fr (P. Roumagnac).

0168-1702/$ – see front matter © 2013 The Authors. Published by Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.virusres.2013.07.006

All characterised geminivirus genomes are composed of one or two circular single stranded DNA components each of which contains fewer than 3640 nucleotides (Loconsole et al., 2012). Each genomic component is encapsidated in a geminate (twinned) incomplete icosahedral particle. To date, seven genera have been approved within the *Geminiviridae* family: Begomovirus, Curtovirus, Topocuvirus, Mastrevirus, Becurtovirus, Turncurtovirus, and Eragrovirus (http://talk.ictvonline.org/files/ictv_official_taxonomy_updates_since_the_8th_report/m/plant-official/4454.aspx). The criteria for demarcating genera in the family *Geminiviridae* are genome organisation, insect vector, host range and sequence relatedness (Fauquet and Stanley, 2003). It is noteworthy, however, that not all of these criteria are necessary for the approval of new genera as, for example, the genus Topocuvirus has been distinguished based only on sequence relatedeness and vector species while the genus Mastrevirus includes viruses that have multiple different vector species and infect either monocotyledonous or dicotyledonous hosts. Mastrevirus genomes also have the fewest genes and these viruses express only four different proteins with two of these, the replication associated protein (Rep) and RepA, sharing identical N-termini but distinct C-termini (they are expressed from alternatively spliced versions of the same transcript). By contrast the other six genera have between 5 and 8 genes, with only the coat protein (*cp*) and *rep* genes being detectably homologous across all of the genera.

The high diversity of geminivirus genome sequences is likely facilitated by these viruses having much higher mutation and recombination rates than those seen in many other DNA viruses. Despite geminiviruses utilising host DNA polymerases during their replication, their mutation rates are as high as many RNA viruses that replicate using error prone RNA dependent RNA polymerases (Duffy and Holmes, 2008; Ge et al., 2007; Isnard et al., 1998). Whereas it is most likely that the high recombination rates of geminiviruses occur as a consequence of their replication involving a mixture of rolling circle and recombination dependent mechanisms (Jeske, 2009), the generally broad host ranges of these viruses together with the frequent occurrence in nature of mixed infections (Martin et al., 2011) has resulted in both frequent instances of inter-species recombination (Padidam et al., 1999), and occasional instances of inter-genus recombination (Briddon et al., 1996; Stanley et al., 1986). While recombination events have sometimes yielded new geminivirus species, it is also plausible that past inter-genus recombination events may have yielded new geminivirus genera (Briddon et al., 1996; Stanley et al., 1986).

The development and application of rolling circle amplification (RCA) based approaches to discover novel circular ssDNA viruses from a variety of environmental sources (Delwart, 2012; Ng et al., 2011b) has tremendously accelerated the rate at which such viruses have been discovered (Rosario et al., 2012) and, when applied to the study of plant samples, has revealed that geminivirus diversity likely far exceeds that which is currently known (Haible et al., 2006; Ng et al., 2011a; Schubert et al., 2007). Besides the characterisation of divergent curtoviruses, mastreviruses and begomoviruses, various geminivirus species have been discovered that are so divergent that they cannot be convincingly assigned to any of the four established geminivirus genera (Briddon et al., 2010; Loconsole et al., 2012; Varsani et al., 2009; Yazdi et al., 2008). Both the creation of new genera such as Becurtovirus, Eragrovirus and Turncurtovirus to accommodate some of these divergent species and perpetually growing numbers of new species within the existing "older" genera underline the steadily increasing complexity of geminivirus taxonomy and the need to recurrently re-evaluate the objectivity and meaningfulness of genus and species demarcation criteria that are applied to the members of this family (Muhire et al., 2013).

Amongst the most divergent of these newly discovered geminiviruses is Eragrostis curvula streak virus (ECSV) isolated from an uncultivated African grass species, *Eragrostis curvula* (Varsani et al., 2009). Given both that there exists a tremendous bias favouring the discovery of novel viruses in cultivated species, and that many novel geminivirus species have in the past been discovered in uncultivated hosts (Briddon et al., 2010; Tan et al., 1995; Varsani et al., 2009), it is likely that further attempts to discover divergent geminiviruses in uncultivated hosts will prove successful.

The potential risks to cultivated crops of viruses that predominantly infect only uncultivated plant species has been documented for maize streak virus (MSV) (Varsani et al., 2008), the African Mastrevirus that causes MSD. MSV is the most economically important virus of maize in Africa. Maize was introduced to West Africa by the Portuguese in the early 1500s and to southern Africa by the Dutch in the mid-1600s but probably only began manifesting evidence of severe MSD around the 1860s (Monjane et al., 2011) with the emergence of a maize-adapted recombinant of two *Digitaria*-adapted MSV strains (Varsani et al., 2008). Therefore, besides purely taxonomic reasons for characterising geminiviruses that mainly infect uncultivated species, the ever present risk that such viruses can become adapted to and cause disease in cultivated hosts is a strong incentive for cataloguing the entire range of plant viral species that are found within terrestrial ecosystems.

Here we describe a new highly divergent geminivirus species isolated from the uncultivated South African spurge, *Euphorbia caput-medusae*. This new geminivirus has a unique genome organisation and distant sequence relatedness to other known geminiviruses and likely represents a new genus-level geminivirus lineage. Using an infectious genomic clone, we show that although it causes an asymptomatic infection in its uncultivated natural host, it can cause a severe infection in an important cultivated species such as tomato. The virus was named Euphorbia caput-medusae latent virus (EcmLV) and, accordingly, we propose that the new genus within which it should be placed be named Capulavirus. Together with a selection of diverse geminivirus sequences we use this new sequence to infer, firstly, previously undetected instances of likely inter-genus recombination in the geminiviruses and, secondly, that the most recent common ancestor of the geminiviruses was possibly a dicot-infecting virus with a *rep* gene that was expressed from a spliced complementary strand transcript.

2. Materials and methods

2.1. Plant sampling

In 2010, samples were collected in the Darling region of the Western Cape from *Euphorbia caput-medusae* plants as part of a large scale survey (for which >800 plants were collected) focusing on viral diversity at the interface between a preserved Cape fynbos ecosystem (Buffelsfontein Game and Nature Reserve) and an intensively cropped agro-ecosystem. Preliminary analysis of the collection of plants that were sampled showed that an unknown geminivirus was detected in an *E. caput-medusae* sample (see below). Nine more *E. caput-medusae* plants were collected in 2011 in the Western Cape region: three from the 2010 sampling site (Buffelsfontein Reserve), and six from coastal fynbos areas, including three near Laaiplek and three in Pater Noster (Supplementary Fig. 1 and Table 1). Whereas samples from the 2010 collection were preserved on dry ice before storage at −80 °C, those from 2011 were preserved by drying them with calcium chloride. The botanical identification of the 2010 plants was carried out initially by eye and was later confirmed by sequencing the C-terminal chloroplast *ndh*F gene using the primer pair 972-F (5′-GTC TCA ATT GGG TTA TAT GAT G-3′) and 2110-R (5′-CCC CCT AYA TAT TTG ATA CCT TCT CC-3′) (Kim and Jansen, 1995).

Table 1
Description of *Euphorbia caput-medusae* samples collected in 2010 and 2011 in the Western Cap floristic region of South Africa.

Sample name	Location	Sampling date	GPS position	PCR detection	Accession number
Dar10	Darling	Sept. 2010	33°14′45.96″S18°13′48.24″E	+	HF921459
Pan1	Pater Noster	Aug. 2011	32°48′25.81″S17°53′43.22″E	–	
Pan2	Pater Noster	Aug. 2011	32°48′26.17″S17°53′43.77″E	–	
Pan3	Pater Noster	Aug. 2011	32°48′28.12″S17°53′46.09″E	–	
Lap0	Laiiplek	Aug. 2011	32°42′36.17″S18°12′35.17″E	–	
Lap11	Laiiplek	Aug. 2011	32°42′36.36″S18°12′34.62″E	+	HF921477
Lap2	Laiiplek	Aug. 2011	32°42′37.06″S18°12′33.84″E	–	
Dar11	Darling	Aug. 2011	33°15′55.91″S18°12′58.63″E	+	HF921460
Dar0	Darling	Aug. 2011	33°15′55.05″S 18°13′0.47″E	–	
Dar2	Darling	Aug. 2011	33°15′56.19″S18°12′58.32″E	–	

2.2. DNA extraction, amplification, cloning and sequencing

Plant samples were ground with ceramic beads (MP 83 biomedicals) and purified quartz (Merck) within a grinding machine (Fastprep 24, MP biomedicals). Total DNA was extracted with the DNeasy Plant Mini Kit (Qiagen) following the manufacturer's protocol. Circular DNA molecules were amplified by RCA using *Phi29* DNA polymerase (TempliPhi™, GE Healthcare, USA) as previously described (Shepherd et al., 2008). The RCA product was digested with *Eco*RI, *Xho*I or *Bam*HI for 3 h at 37 °C; *Eco*RI and *Bam*HI generated a product of about 2.7 kbp. *Eco*RI restricted products of ~2.7 kbp obtained from two samples collected in Darling – one in 2010 (Dar10) and one in 2011 (Dar11) – were gel purified with the QIAquick Gel Extraction Kit (Qiagen), cloned into the pBC plasmid and sequenced. Sequence data were obtained by standard Sanger sequencing (Beckman Coulter Genomics) using a primer walking approach. Partially overlapping PCR primers were designed according to the sequences of clones Dar10 and Dar11 (Dar-1981F forward primer 5′-CCT CAC TGA ATC CAC ATC CA-3′ and Dar-1966R reverse primer 5′-CGA GGA ATT CGG ACT TGG-3′) to generate a third clone from a sample collected near Laaiplek in 2011 (Lap11), as follows: 1 μl RCA product obtained with Lap11 was amplified in a final volume of 25 μl containing 12.5 μl of HotStarTaq Plus Mastermix, 0.5 μl of each primer (at 10 μM concentration each) and 10.5 μl of RNase free water. The following amplification conditions were used: an initial denaturation at 95 °C for 5 min, followed by 30 cycles at 94 °C for 1 min, 60 °C for 1 min, 68 °C for 3 min, and a final extension step at 72 °C for 10 min. An amplification product of ~2.7 kbp was gel purified, ligated into pGEM-T Easy (Promega Biotech) and sequenced by standard Sanger sequencing using a primer walking approach.

2.3. Sequence analysis

Sequences were assembled using BioNumerics Applied Maths V6.5 (Applied. Maths, Ghent, Belgium) and compared to database sequences using BlastN, BlastP and tBlastX (Altschul et al., 1990). Open reading frames (ORFs) were identified that could potentially express proteins larger than 50 amino acids in length. Blast results were considered as indicative of significant homology when BLAST *e*-values were smaller than 10^{-2}.

The computer programme SMART (http://smart.embl-heidelberg.de/ (Letunic et al., 2012)) was also used for the detection of known domain architectures within the ORFs that had no detectable homologues within the public sequence databases. For nucleotide sequences and ORFs that did have clear homologues within other geminivirus genomes (those corresponding to the replication origins, transcription start/termination sites, replication associated protein and coat protein genes) we inferred the locations of domains and motifs within the newly sequenced genomes and their potentially expressed proteins based on the experimentally inferred positions of these domains in these other geminiviruses. Pairwise identity scores were calculated as previously described (Muhire et al., 2013).

2.4. PCR-mediated detection of EcmLV

Two 100% Dar10-complementary primer pairs were designed with PRIMER3 (Rozen and Skaletsky, 2000) which theoretically should not hybridise to 63 representative species of the family *Geminiviridae*. The first primer pair was designed to amplify a region starting within the large intergenic region and ending within the coat protein gene: Dar-136F forward primer 5′-CGA AGA GGT CAT TGG GAC AT-3′ and Dar-730R reverse primer 5′-CGG GTC TGG CTA AGA GAG TG-3′. The second primer pair is targeted to the *rep* gene: Dar-1662F forward primer 5′-TCG-ARC-AGG-TTT-CTG-GTC-CT and Dar-2257R reverse primer 3′-ACA-CCT-TCA-CTG-CCT-TGT-CC. The 9 *E. caput-medusae* samples collected in 2011 were PCR-tested with these two pairs of primers using the HotStarTaq Plus Master Mix Kit (Qiagen) following the manufacturer's protocol. Amplification conditions consist of an initial denaturation at 95 °C for 5 min, 30 cycles at 94 °C for 1 min, 55 °C for 1 min, 72 °C for 30 sec, and a final extension step at 72 °C for 10 min. A 100% pCAMBIA2300-complementary primer pair was designed with DNAMAN (version 5.0; Lynnon BioSoft, Quebec, Canada) in its 35S promoter region: pCambia2300F forward primer 5′-TGC TTT GAA GAC GTG GTT GG-3′ and pCambia2300R reverse primer 5′-ACG ACA CTC TCG TCT ACT CC-3′. Amplification conditions were as described above but with an annealing temperature of 65 °C.

2.5. Construction of an agro-infectious clone

To test the infectivity of the cloned geminivirus genome, we used a modification of the *Agrobacterium tumefaciens* delivery system (Peterschmitt et al., 1996) as follows: the insert of clone Dar10 was released using *Eco*RI and gel purified using a Wizard SV Gel and PCR Clean-Up System (Promega) following the manufacturer's protocol. A tandem head to tail dimer of the Dar10 genome was ligated into the dephosphorilated *Eco*RI cloning site of the binary vector pCAMBIA2300. This construct was introduced into *A. tumefaciens* C58 by electroporation.

2.6. Plant inoculation

A liquid culture of the Dar10 containing *A. tumefaciens* with an OD between 2.0 and 5.0 was concentrated ten times in sterile LB medium before inoculation. Test plants were inoculated either by injection with a syringe and a needle or by infiltration with the syringe alone directly in contact with the leaf. Alternatively a 24 h-plated culture of the Dar10 containing *A. tumefaciens* was used for inoculation as follows: the tip of an 18 Gauge ×1 1/2 needle previously dipped into the solid culture of the Dar10 containing *A. tumefaciens* was pricked several times into the plant. Ten tomato plants of the cv. Monalbo were inoculated by injection, ten by

infiltration and ten by pricking. A total of 30 *N. benthamiana* plants were similarly inoculated. Two or three plants of each species were not inoculated and used as non-infected controls. *E. caput-medusae* test plants were germinated from seeds by a commercial nursery. Due to the limited number of plants which could be supplied by the nursery, we tested different inoculation techniques on the same plants. Thus, each test plant was inoculated by both injection and pricking but on separate branches. Depending on the size of the plants, inoculation was done on one branch per technique (one plant), two branches per technique (one plant) or three branches per technique (three plants). Two *E. caput-medusae* plants were used as negative controls, one was inoculated with *A. tumefaciens* containing an empty vector and one was not inoculated. Plants were maintained in insect-free containment growth chambers under 14 h light at 26 °C, and 10 h dark at 24 °C.

2.7. Detection of potential inter-genus recombination events

We attempted to detect evidence of recombination within the Dar10 sequence by analyzing it together with nineteen other geminivirus full genome sequences collectively representing the most divergent geminivirus lineages available (see the "recombination analysis.rdp" and "recombination analysis alignment.fas" files provided as supplementary material for the identities of these 19 other viruses). These sequences were aligned with MEGA (Tamura et al., 2011) and analysed for recombination with the computer programme RDP4.18 (Martin et al., 2010) using the "verify alignment consistency" procedure outlined in Varsani et al. (2006). We analysed the alignment with seven of the recombination detection methods implemented in RDP4.18 with window size settings as follows: RDP = 100; BOOTSCAN = 800; MAXCHI = 240, CHIMAERA = 160; SISCAN = 500; 3SEQ and GENECONV with unspecified window sizes. These window sizes were set to approximately four times their default values to specifically detect, with a minimum of noise, signals of large sequence transfers (>300 nucleotides) between distantly related genomes. We discounted all signals of recombination that could not be confirmed (i) with four or more of the seven recombination detection methods and (ii) following realignment in isolation from the remainder of the 20 sequences in the alignment, of the sequence triplets within which signals were discovered. Whereas simulation analyses have revealed that the consensus of four methods should yield a false recombination detection rate per dataset of far below the 5% expected for each of the methods individually (Posada and Crandall, 2001), the realignment and retesting step ensured that the detected signals were not simply caused by alignment errors (Varsani et al., 2006).

2.8. Phylogenetic analysis

For purposes of reconstructing the evolutionary relationships of the various major geminivirus lineages we focused exclusively on the *rep* and *cp* coding regions of these genomes (i.e. the homologous portions). We assembled datasets consisting of 63 Rep and 59 CP amino acid sequences representing the entire breadth of known geminivirus diversity. Besides the inferred CP and Rep amino acid sequences of Dar10, each of these datasets contained 24 Mastrevirus sequences, 18 Begomovirus sequences, 8 Curtovirus sequences, 2 Becurtovirus sequences, one Eragrovirus sequence, one Turncurtovirus sequence, one Topocuvirus sequence, and one sequence each from divergent geminiviruses recently discovered infecting citrus plants (Citrus chlorotic dwarf associated virus, CCDaV, genome accession: JQ920490) (Loconsole et al., 2012) and grapevines (Grapevine Cabernet Franc-associated virus, GCFaV, genome accession: JQ901105) (Krenz et al., 2012). In addition to these sequences, the Rep dataset contained two geminivirus-like Rep sequences recently discovered within the apple genome (genome accession: PRJNA28845; whole genome shotguns: ACYM01134023 and ACYM01134026; Martin et al., 2011). Finally, the Rep dataset also contained two divergent outlier sequences that were used to root the Rep phylogeny: one derived from the witches broom associated phytoplasmal plasmid and the other from the geminivirus-like mycovirus Sclerotinia sclerotiorum hypovirulence-associated DNA virus (SsHADV, genome accession: NC_013116).

Because of the extremely distant relationships that existed even between the inferred amino acid sequences of these proteins it was not possible to accurately align the sequences. We therefore accounted for alignment uncertainty using a Bayesian approach to simultaneously estimate phylogenetic trees and alignments with the computer programme BAli-Phy version 2.2.1 (Suchard and Redelings, 2006). Default Bali-Phy settings were used for both the Rep and CP amino acid sequences. The LG2008 + gwF (Le and Gascuel, 2008) + generalised weighted frequencies amino acid substitution model (previously determined to be the best fit amino acid substitution model for these sequences (Varsani et al., 2009)) was used in conjunction with the RS07 (Suchard and Redelings, 2006) insertion-deletion model. Convergence was checked by comparing two independent MCMC chains for each of the two sets of aligned sequences. Estimated sample sizes (ESS) of model parameters were calculated using the programme Tracer http://tree.bio.ed.ac.uk/software/tracer, Tracerv 1.4 Rambaud and Drummond (2007) and ESS scores following merging of samples from the two independent chains run for each of the alignments were all greater than 200.

3. Results and discussion

3.1. Discovery of a new highly divergent geminivirus from E. caput-medusae

DNA was extracted from one *E. caput-medusae* plant, collected in Darling (Western Cape region, South Africa) in 2010. *E. caput-medusae* is a sprawling, succulent wild spurge indigenous to the coastal regions of South Western Africa. A 2.7 kbp *Eco*RI restricted DNA fragment was obtained from the RCA product generated from this plant. The *Eco*RI restricted fragment was cloned (clone Dar10), sequenced and shown to consist of 2678 bp. BlastN and BlastX comparisons between Dar10 and all sequences in Genbank indicated that the highest identity score was detected with a recently deposited geminivirus genome isolated from French bean in India: French bean severe leaf curl virus (FbSLCV; accession number NC_018453, highest percent identity = 78%, *e*-value = 1×10^{-169}). Although FbSLCV has not been described in the peer reviewed literature, a manuscript dealing with the biological characterisation of this virus is presently being prepared by its discoverers. Nevertheless, because the FbSLCV sequence is in the public domain, we included it in our comparative analyses between the Dar10-like viruses and the rest of the known geminiviruses. It is also noteworthy that in our scan of Genbank for sequences homologous with those occurring in Dar10, the second best match to Dar10 we found was to a geminivirus-like Rep sequence that is apparently integrated into the *Malus domestica* genome (Martin et al., 2011).

Among the nine *E. caput-medusae* plants collected in 2011, two tested positive by PCR for Dar10-like viruses: one from Darling and one from Laaiplek (Table 1). One viral DNA fragment was cloned from each of these two PCR-positive plants: clone Dar11 with 2650 bp and clone Lap11 with 2677 bp. The *Eco*RI cloning sites used for clones Dar10 and Dar11 was found to be unique in the Lap11 genome generated with partially overlapping primers. The *Eco*RI site was also confirmed to be unique in Dar10 and Dar11 with the sequenced PCR products generated with one of the

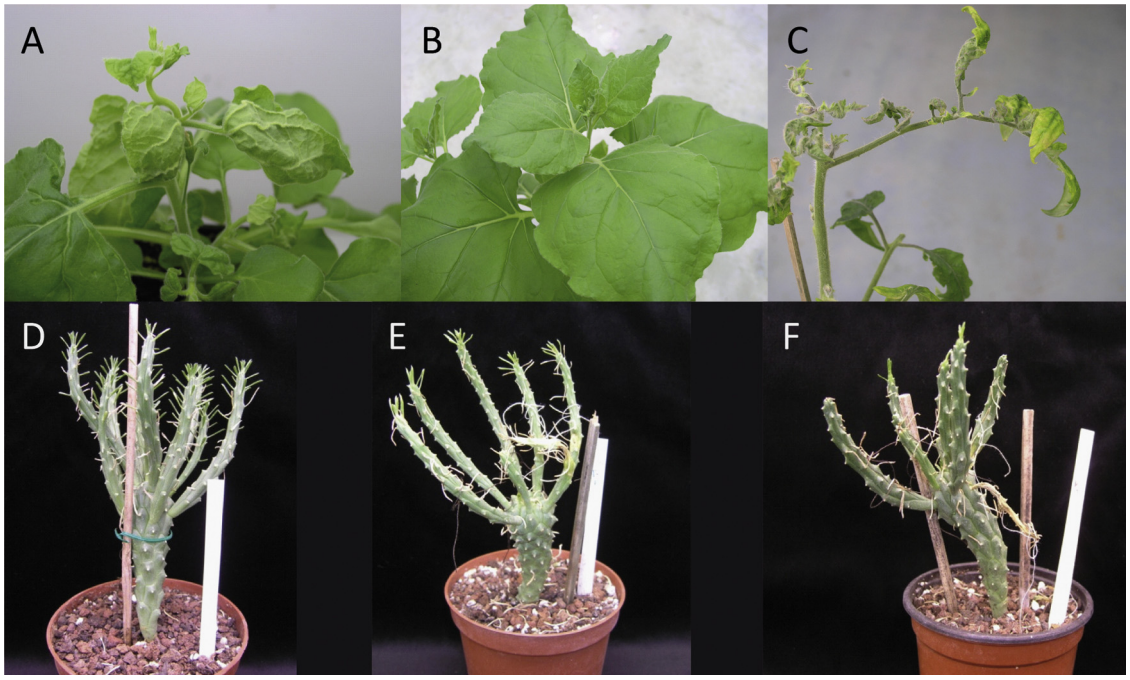


Fig. 1. Symptoms caused by EcmLV on *E. caput-medusae*, *N. benthamiana* and *S. lycopersicum* cv Monalbo plants at 21 dpi following agroinoculation with the agroinfectious clone Dar11. (A) *Nicotiana benthamiana* exhibited leaf curling, distortion and vein thickening. (B) *Nicotiana benthamiana*, which was used as non-infected control. (C) Tomato plants exhibited leaf curling, distortion, stunting and yellowing. (D) *Euphorbia caput-medusae* seedling which was used as non-infected control. (E) *E. caput-medusae* seedling agro-inoculated with the empty vector pCambia2300. It exhibited both yellowing wilt and necrosis on the branch inoculated with liquid culture. (F) *E. caput-medusae* seedling agro-inoculated with EcmLV and detected PCR-positive with EcmLV specific primers. The EcmLV inoculated plant did not exhibit any symptoms which could be related to viral infection.

specific primer pairs (Dar-1662F/Dar-2257R) spanning the *Eco*RI site. This indicated that the three cloned viral DNA fragments correspond to full length geminiviral genomes amplified by RCA. These circular DNA molecules were considered to be the complete viral genomes of geminiviruses infecting the three *E. caput-medusae* plants, because only one band was resolved by electrophoresis of the *Eco*RI and *Bam*HI digested RCA products (unique restriction sites for the viral clones) in the size range of geminivirus genome component sequences and no fragments were detected within the size ranges of known geminivirus satellite sequences. Similarly, restriction with *Xho*I, a non-cutting enzyme for any of the three clones, also did not show any DNA fragment in the satellite sequence size range.

To further confirm that the unique viral DNA sequences that we had cloned were complete genomes, we tested the infectivity of Dar10 in its natural host. One (plant 2) and two (plants 2 and 3) out of five *E. caput-medusae* plants which were agro-inoculated with clone Dar10, tested positive at 35 and 154 days post-inoculation (dpi), respectively, with the Dar10 specific primers, indicating that the clone was biologically active in its natural host. This finding suggests that the cloned 2.7 kb DNA fragment represents the complete genome of this new geminivirus (Supplementary Fig. 2A). The Dar10 positive samples were negative with a PCR detection test targeted to the pCAMBIA2300 binary vector, confirming that the positive virus detection was not simply due to the detection of the agro-inoculated construct (Supplementary Fig. 2B). The virus-infected plant did not exhibit any symptoms, which could differentiate it from the control plant which was agro-inoculated with the empty pCAMBIA2300 vector (Fig. 1). However we noticed that all inoculated plants reacted severely both to the injection of the liquid culture of *A. tumefaciens*, with yellowing wilt and necrosis of the inoculated branch, and to pricking inoculation with slight chlorosis and yellowing near the inoculation spot.

The lack of symptoms in the experimentally agro-infected plants at 35 and 154 dpi was consistent with the observation that the three wild plants of *E. caput-medusae* in which the new geminivirus was detected did not exhibit any conspicuous symptoms such as chlorotic mosaic, streaking or leaf deformation that differentiated them from plants that tested negative for Dar10-like viruses. Therefore, given that both field and experimentally infected *E. caput-medusae* plants did not exhibit any conspicuous symptoms (Fig. 1), we conclude that this geminivirus is likely latent in its natural host species, *E. caput-medusae*. This absence of visual symptoms is in line with results obtained during recent wild plant virus biodiversity surveys, which showed that most of the viral sequences discovered came from asymptomatic, healthy-looking plant samples (Melcher et al., 2008; Muthukumar et al., 2009; Roossinck et al., 2010).

Although the three full genomes isolated from the wild spurges could be aligned with the genome of FbSLCV, attempts to accurately align them with a diverse representation of other known geminiviruses proved largely unsuccessful due both to the extremely low degrees of sequence similarity in the homologous *cp* and *rep* genes, and the fact that the remainders of the genomes of these viruses were, with the exception of particular highly conserved sequence motifs (see below) not detectably homologous with those of other geminiviruses. The three sequences (Dar10; Dar11 and Lap10) showed between 93.6 and 95.3% pairwise identity scores. These identity scores are well above all of the species demarcation thresholds recommended for the various geminivirus genera by the geminivirus Study Group of the ICTV (Fauquet et al., 2008) and a recent proposal for the classification of viruses in the genus Mastrevirus (Muhire et al., 2013). The inter-clone distances are, however, unexpectedly high with respect to the relatively small geographical and temporal distances separating the three sampling sites (<60 km and approximately one year). The occurrence of long term infections in vegetatively propagated *E. caput-medusae* plants

and inefficient plant to plant viral transmission leading to deep population subdivisions, could both contribute to the presence of such divergent lineages within the sampling region. To test this hypothesis would however require additional virus samples and controlled infection and transmission experiments.

Although the geminivirus genomes isolated from *E. caput-medusae* plants are most closely related to FbSLCV, the pairwise identity score between them (71.7–72.5%) is below the lowest of the geminivirus species demarcation thresholds recommended by the ICTV (75% for mastreviruses) (Fauquet et al., 2008; Muhire et al., 2013). The name Euphorbia caput-medusae latent virus (EcmLV) is proposed for the new species.

3.2. Experimental hosts of EcmLV

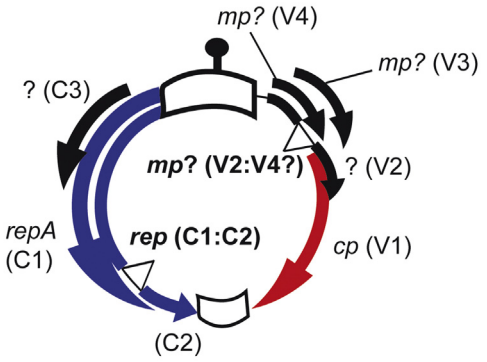
To test if EcmLV has the potential to infect species outside the family *Euphorbiaceae*, the Dar10 agroinfectious clone was inoculated into plants belonging to two solanaceous species, *Solanum lycopersicum* (the host of a large range of begomoviruses from the Old and New Worlds), and *Nicotiana benthamiana*, a very permissive host often used in experimental virology. Surprisingly, all 30 of the inoculated tomato plants exhibited severe symptoms 21 days post-inoculation, whatever the inoculation technique (injection, infiltration or pricking). The symptoms, similar to, but clearly distinct from those caused by tomato yellow leaf curl virus, consisted of curling and distortion of leaflets, leaf stunting, and prominent yellowing along leaf margins and/or interveinal regions (Fig. 1). Similarly, 27 out of 30 inoculated *N. benthamiana* plants (90%) exhibited severe symptoms 21 days post inoculation whatever the inoculation technique. Symptoms consisted of leaf curling and distortion and thickening of the veins (Fig. 1). Infected plants of both species were stunted and did not produce flowers. Whereas EcmLV was detected in symptomatic plants using the specific primers designed in this study, the asymptomatic plants and the non-inoculated plants tested negative for EcmLV.

EcmLV therefore potentially has a broad host range. This is consistent with the classical hypothesis that viruses evolving in species-rich-wild plant communities will likely adapt to infect a wide range of hosts – possibly even including those belonging to different plant families (Jones, 2009). There are however multiple factors which constrain virus infection in natural conditions, as has been revealed by a study in which five generalist virus species were each detected in a narrow range of host plants among twenty one wild species (Malpica et al., 2006).

3.3. Characteristics of the EcmLV genomes

The arrangement of open reading frames (ORF) (Fig. 2) within the 2678 bp circular DNA of the EcmLV-Dar10 clone is most similar to those reported for mastreviruses (Rosario et al., 2012), with two overlapping complementary sense ORFs (C1 and C2), and two intergenic regions: a large one (LIR) and a small one (SIR). A third ORF (C3) in the complementary sense region was also detected. Unlike mastrevirus genomes, which contain two virion sense ORFs (V1 and V2), the EcmLV circular DNA has four ORFs (V1, V2, V3 and V4) (Fig. 2) that would be predicted to encode proteins with >68 amino acids. The two other genomes obtained from Dar11 and Lap11 contained the same patterns of ORFs and intergenic regions. Except for the V4 ORF which was not detected in FbSLCV, the ORF pattern of FbSLCV is similar to that of EcmLV.

The V1 ORFs of the three EcmLV clones encode predicted proteins of 242 aa in length that share ~73% identity with the predicted coat protein of FbSLCV (*e*-value = 6×10^{-125}). The V2 ORFs of the three clones encode predicted proteins of 135 aa in length that share ~53% identity with the cognate ORF of FbSLCV. The V3 ORFs of the three clones encode proteins of 87 aa in length but the identity



ORF	coordinates	number of amino acids
V1 (*cp*)	546-1274	242
V2 (*mp*?)	129-335	68
V3	248-652	105
V4 (*mp*?)	169-432	87
C1 (*rep*A)	complement (1667-2431)	331
C2	complement (1297-1689)	130
C3	complement (1816-2181)	121

Fig. 2. Genomic organisation of EcmLV showing the arrangement of seven predicted open reading frames, putative proteins and the position of the small and large intergenic regions. Arrows indicate the positions and orientations of ORFs (V = virion sense and C = complementary sense) that are suspected to encode expressed proteins. *mp* = movement protein gene, *cp* = coat protein gene, *repA* = replication associated protein gene A, *rep*(C1:C2) = gene derived from a spliced transcript encompassing C1 and C2 ORFs, *mp?*(V2:V4) = gene derived from a spliced transcript encompassing V2 and V4 ORFs. A question mark indicates that an ORFs function is either completely unknown or only suspected. The only genes shared between all *Geminiviridae* genomes are *rep* (in blue) and *cp* (in red). Intergenic regions are represented as open blocks and the hairpin structure at the origin of virion strand replication is indicated at the 12 o'clock position. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of the article.)

with the cognate ORF of FbSLCV is very low (36%). Except for their homologues encoded by FbSLCV, the V2, V3 and V4 ORFs of Dar10 potentially encode proteins with no significant degrees of identity with any other known protein.

Despite the absence of any detectable homologues of these proteins in sequence databases, exploration of protein domains and domain architectures using SMART (http://smart.embl-heidelberg.de/ (Letunic et al., 2012)) indicated that V3 and V4 contain 22 and 23 amino acid regions, respectively, that are likely transmembrane domains. Given that similar probable transmembrane domains are present in the movement protein encoded by mastreviruses (Boulton, 2002), this suggests that EcmLV V3 and V4 may encode movement proteins that are functional analogues (if not *bona fide* homologues) of those found in mastreviruses.

Similarly to the mastrevirus MSV in which an intron was detected in V2 (Wright et al., 1997), potential splice junctions were detected in the virion sense transcript of EcmLV. However, whereas the virion sense transcript results in an in-frame deletion (76 nucleotides for MSV (Wright et al., 1997)) of the movement protein gene of MSV, splicing of the EcmLV transcript would potentially result in both the elimination of the V2 ORF start codon and a frame shift that together would cause the V2 and V4 genes to be expressed as a single protein (Supplementary Fig. 3) in much the same way as full length *rep* genes are expressed from the two complementary sense ORFs of mastreviruses and becurtoviruses. Interestingly, similar likely splice junctions are present in the virion sense transcripts of the FbSLCV (Fig. 2). While suggesting that the involvement of virion sense gene transcript splicing as a gene expression strategy possibly predated the most recent common ancestor of EcmLV, FbSLCV, the mastreviruses and the begomoviruses, the specific

characteristics of this splicing in EcmLV would differentiate its genome architectures from those of all other known geminiviruses. It is noteworthy, however, that there is no homologue of the V4 ORF in the FbSLCV genome and, although potential virion strand transcript splice junctions are present in this virus, it is unlikely that a protein homologous to a V2–V4 protein of EcmLV could be expressed by this virus from spliced virion strand transcripts.

As is the case in mastreviruses and becurtoviruses, the EcmLV and FbSLCV genomes likely express a Rep protein from a spliced complementary sense transcript (Dekker et al., 1991); Fig. 2. The spliced transcript originating from EcmLV C1–C2 encodes a putative 331 aa protein which Exhibits 79% identity to the Rep protein of FbSLCV (*e*-value = 1×10^{-117}). It is also plausible that, as is likely the case in mastreviruses, the viruses also express a RepA protein of 254 aa from an unspliced complementary sense transcript. The inferred amino acid sequences of these EcmLV and FbSLCV Rep proteins contain canonical rolling circle replication (RCR) motifs which, besides being present in all known geminivirus Reps, are also strongly conserved amongst many other rolling circle replicons (Ilyina and Koonin, 1992; Koonin and Ilyina, 1993; Rosario et al., 2012) (Supplementary Fig. 4). In addition, the inferred EcmLV and FbSLCV Reps contain the ATPase motifs Walker-A, Walker-B, and C which were shown for other geminiviruses to be included in the Rep region exhibiting helicase activity (Choudhury et al., 2006; Clerot and Bernardi, 2006) (Supplementary Figs. 3 and 4). Unlike in some mastrevirus Reps and the Rep of the divergent geminivirus ECSV (new Eragrovirus genus), the predicted EcmLV and FbSLCV Rep proteins lack the canonical LXCXE retinoblastoma binding domain (Arguello-Astorga et al., 2004). The EcmLV and FbSLCV Reps do, however, contain the so-called GRS domain identified in other geminivirus Reps (Nash et al., 2011).

As in all other known geminiviruses, the 348–375 bp LIRs of the EcmLV isolates contain a predicted hairpin-loop structure with the extremely conserved geminiviral virion strand origin of replication nonanucleotide motif (TAATATTAC) in the loop (Lazarowitz et al., 1992) (Supplementary Fig. 3). Also similar to mastreviruses, the stem sequence of this predicted hairpin structure contains a number of nucleotide mismatches (Supplementary Fig. 5). Between the first probable TATA box of the complementary sense gene promoter (located at position 2540) and the replication origin, there is a GC-rich region that might, as is the case in other geminiviruses, constitute a G-Box with a role into transcriptional regulation (Eagle and Hanley-Bowdoin, 1997).

Between the likely hairpin structure and *rep* gene start codon are two different sets of directly and inversely repeated sequences arranged similarly to analogous "iteron" sequences that have been identified as *rep* binding recognition sites in begomoviruses, curtoviruses and topocuviruses (see the sequences labelled as Type A and Type B iterons in Supplementary Fig. 3) (Londono et al., 2010). The Lap11 LIR sequence differs somewhat from that of the Dar10 and Dar11 sequences in that it has 24 fewer nucleotides between the potential iteron sequence (type A) and the potential virion-sense gene TATA box.

Finally, the EcmLV and FbSLCV sequences respectively contain 21–23 bp and 93 bp small intergenic regions (SIR) between the *rep* and *cp* stop codons – a feature shared with mastreviruses, becurtoviruses, ECSV, CCDaV and GCFaV. In common with begomoviruses, topocuviruses and curtoviruses but unlike mastreviruses and ECSV, however, in EcmLV and FbSLCV the likely polyadenylation signals of the V-sense and C sense transcripts reside within the *rep* and *cp* genes.

3.4. Analysis of recombination

Since it has long been accepted that recombination between highly divergent geminiviruses could have played a role in the genesis of some of the known geminivirus genera (Briddon et al., 1996; Rybicki, 1994; Stanley et al., 1986) we attempted to identify evidence that EcmLV and/or FbSLCV might be inter-genus recombinants using a recombination–detection and alignment consistency verification approach previously devised by Varsani et al. (2006) for the analysis of recombination between extremely divergent sequences. Although this analysis yielded no evidence that either EcmLV or FbSLCV were recombinants, it both identified previously suspected recombination events within topocuvirus and curtovirus genomes (Briddon et al., 1996; Rybicki, 1994; Stanley et al., 1986; Varsani et al., 2009) and identified previously undetected evidence of recombination within the ECSV, CCDaV and Turnip curly top virus (TCTV) genomes (Fig. 3; see also the "recombination analysis.rdp" and "recombination analysis alignment.fas" files provided as supplementary material for additional details on these detected recombination events).

3.5. Phylogenetic relations between EcmLV and others geminiviruses

Possible evolutionary relationships between EcmLV and other known geminiviruses were investigated using phylogenetic analyses of inferred amino acid sequences from a representative sampling of geminiviral coat protein and Rep sequences. In the case of the Rep analysis, geminivirus-like phytoplasmal plasmid and mycovirus derived Rep amino sequences were included as outliers so that the phylogenetic tree determined with these amino acid sequences could be properly rooted. It was not possible to root the phylogenetic tree determined from CP amino acid sequences because there exist no suitable known non-geminiviral CP homologues. Also included in the Rep dataset was the geminivirus Rep-like sequence we had earlier found integrated within the apple genome.

Since the analysed amino acid sequences were so diverse, it was not possible to reliably align them. We therefore opted to use a Bayesian phylogenetic tree construction approach that explicitly accounts for alignment uncertainty by using a Markov chain Monte Carlo sampling scheme to simultaneously infer families of almost equally plausible alignments along with their associated phylogenetic trees. The maximum clade credibility (MCC) consensus of these trees is what we present in Fig. 4.

In both these trees, EcmLV clearly clusters with FbSLCV on a branch that is neither closely associated with any sequences classified within any of the seven established geminivirus genera, nor closely associated with any of the other known divergent geminivirus sequences. It is however, notable that the EcmLV and FbSLCV CP sequences are slightly more similar to those of the begomoviruses than they are to the other geminiviruses. However, given the lack of an outgroup, we cannot conclude that these viruses share a more recent common ancestor with the begomoviruses than they do with any of the other geminivirus groups.

The rooted Rep phylogeny on the other hand clearly indicated that EcmLV and FbSLCV cluster, with a posterior probability of 0.95, with all other geminiviruses that express Rep proteins from spliced complementary sense transcripts. Among the geminivirus Reps from known free-living geminiviruses (i.e. excluding Rep sequences integrated into host genomes) the EcmLV and FbSLCV Reps are most closely related to that from a recently discovered grapevine infecting geminivirus, GCFaV (Krenz et al., 2012).

It is also noteworthy that the sequences of apparently geminivirus-like Reps that are reportedly integrated into the apple genome (Martin et al., 2011) are very clearly most closely related to the EcmLV and FbSLCV sequences. While the possibility remains that these apparent integrons might be a sequence assembly artefacts arising due to the contamination of shotgun sequenced genomic apple DNA with DNA derived from an undetected

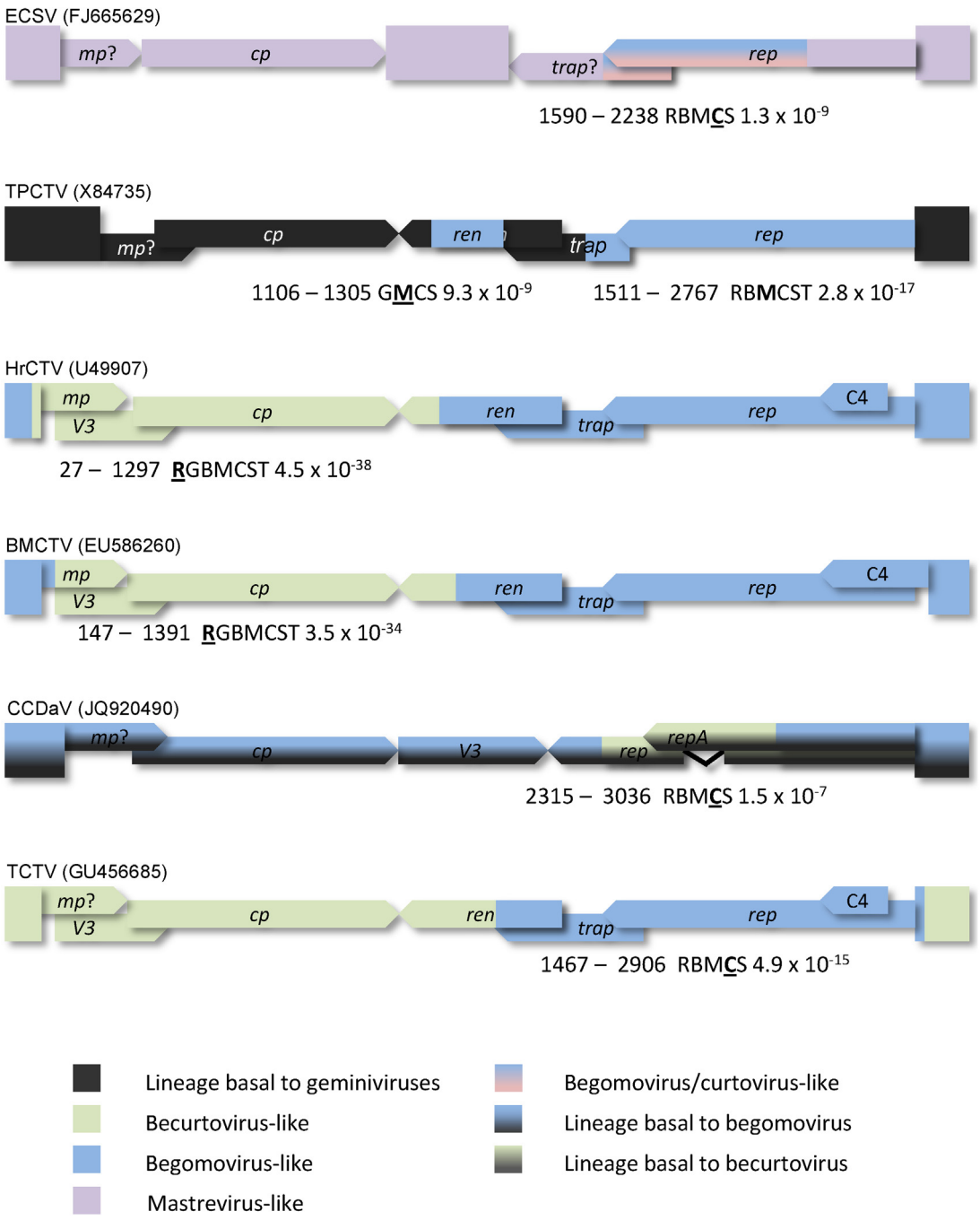


Fig. 3. Genome sequences of representative viruses with evidence of inter-genus recombination events. Bars indicate the genome sequences of representative inter-genus recombinants linearised at the virion strand origin of replication. Colours indicate the likely origins of the indicated genome regions. Also shown for each of the seven represented recombination events are the approximate beginning and ending coordinates of the recombinationally derived genome fragments (where nucleotide 1 is the first nucleotide 5′ of the virion strand origin of replication), the recombination analysis methods with which the recombination events are detectable (with an associated multiple testing corrected *p*-value <0.05; R = RDP, G = GENECONV, B = BOOTSCAN, M = MAXCHI, C = CHIMAERA, S = SISCAN, T = 3SEQ), and their associated *p*-value's (corresponding to the method indicated in bold/underlined). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of the article.)

geminivral infection, it is nevertheless interesting that EcmLV-like viruses might, in addition to infecting members of the *Euphorbiaceae* and *Fabaceae*, also be capable of infecting members of the *Rosaceae*.

Finally it is important to point out that, even considering the unknown location of the CP tree root, the Rep and CP trees have major discrepancies that are potentially indicative of the Rep and CP sequences of many of the viruses considered here having separate evolutionary histories. It is apparent from previous studies and the analysis we describe above that recombination probably underlies the discordant locations of the various curtoviruses, ECSV and CCDaV in the two trees (Varsani et al., 2009). Crucially, the additional sequences that we included in our analysis relative to the study of Varsani et al. (2009), mean that we have for the first time been able to identify ECSV as having a genome that is mostly mastrevirus-like but with a fragment of the genome corresponding to the *rep* intron region of genuine mastreviruses having been derived from what appears to be a divergent begomovirus/curtovirus-like ancestor. While explaining the discordant position of ECSV in the Rep

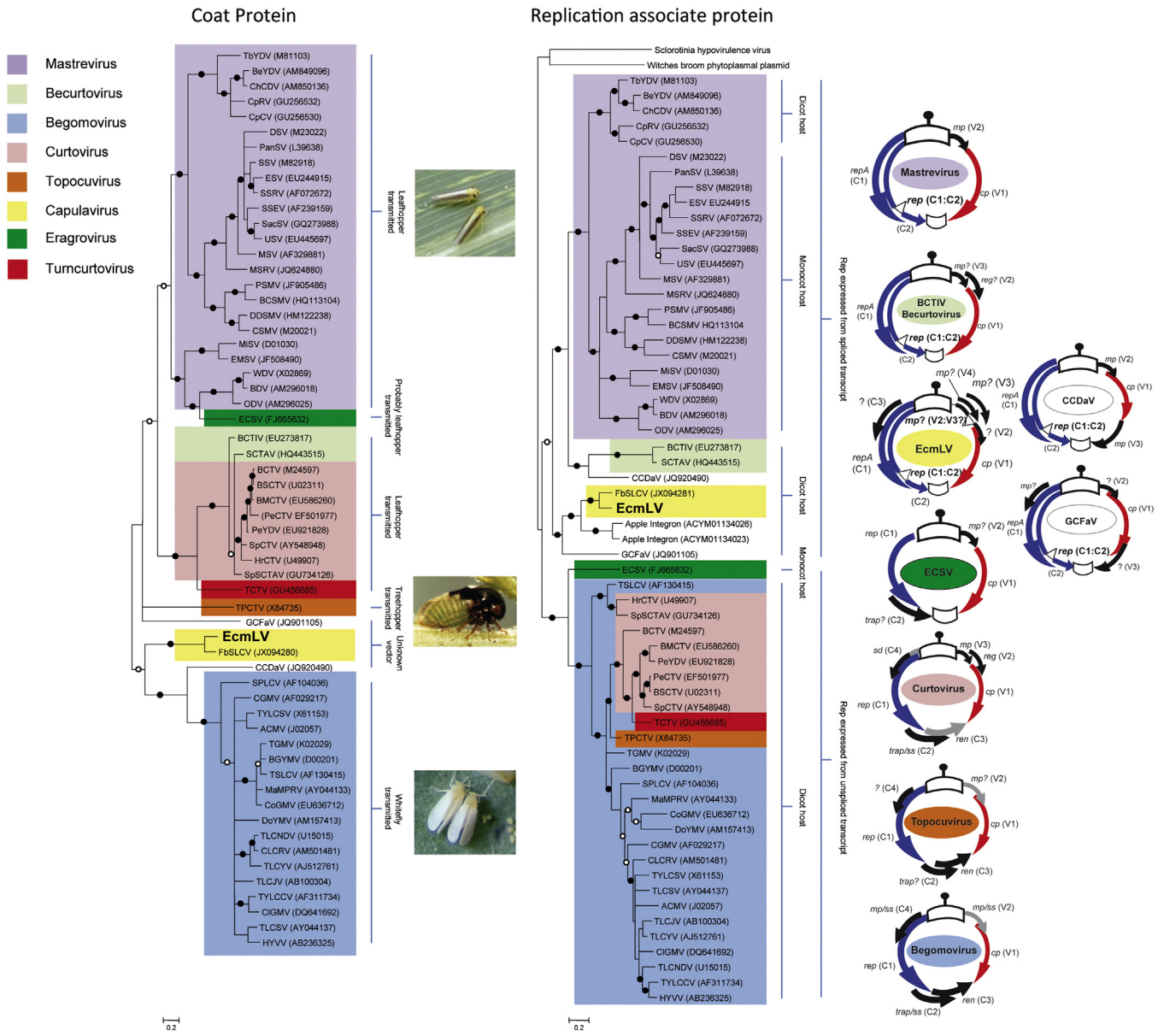


Fig. 4. Maximum clade credibility phylogenetic trees describing the evolutionary relationships between various geminivirus coat protein (CP) and replication associated protein (Rep) amino acid sequences. The trees were constructed so as to account for alignment uncertainty using simultaneous Bayesian inference of alignments and phylogenies. Species belonging to the seven established geminivirus genera (Mastrevirus, Becurtovirus, Topocuvirus, Begomovirus, Eragrovirus, Turncurtovirus and Curtovirus) and one tentative genus (Capulavirus) are indicated with coloured blocks. Branches with a filled dot have >99% posterior probability support whereas those with an empty dot have >95% posterior probability support. All branches with less than 80% posterior probability support have been collapsed. Indicated on the CP tree are the known/likely insect vectors of many of the viruses. Indicated on the Rep tree are the hosts of the viruses, the genomic organisation of the different geminivirus genera and divergent geminiviruses, and whether or not their Rep proteins are expressed from spliced complementary strand transcripts. Photo courtesy of: J.M. Lett from CIRAD (*Cicadulina mbila*), John Innes Centre (*Micrutalis malleifera* Fowler) and A. Franck from CIRAD (*Bemisia tabaci*).

and CP phylogenies this recombination event could also explain why, in common with begomoviruses and curtoviruses, it has a Rep that is expressed from an unspliced complementary strand transcript.

3.6. *What does the new data tell us about the earliest geminiviruses?*

EcmLV and the various highly divergent geminiviruses represented within our rooted Rep phylogeny permit us to infer, with the greatest accuracy yet achieved, some of the likely characteristics of the earliest geminiviruses. It is noteworthy, for example, that the monocot-infecting mastreviruses form a well-supported (with a 0.99 posterior probability) monophyletic clade within the Rep phylogeny that is nested within a much larger more divergent cluster of dicot-infecting geminiviruses with spliced Reps. It is therefore most likely that the most recent common ancestor (MRCA) of the mastreviruses infected dicots and that the host switch occurred from dicots to monocots and not the other way around as has been previously supposed (Varsani et al., 2009). The only known non-mastreviral monocot-infecting geminivirus is ECSV which, because it branches near the root of the geminivirus Rep phylogeny, might be interpreted as evidence that the MRCA of the geminiviruses could have plausibly infected either monocots or dicots (or perhaps even the common ancestor of these plant lineages). It is, however, apparent

that the basal location of ECSV in the Rep tree is potentially due to this isolate having a recombinant *rep* gene that is approximately 1/3 mastrevirus-like and 2/3 begomovirus/curtovirus-like (Fig. 3). If one accounts for this it would imply that the basal-position of ECSV in the Rep phylogeny may be largely artefactual and that the "true" position of the majority of its genome is at the base of the monocot-infecting mastrevirus lineage (as it is in the CP tree). If this is the case then one need invoke only a single dicot to monocot host switch to explain the present host distributions of all the known geminiviruses.

Given the clear separation within the Rep phylogeny of viruses with and without a *rep* intron (Fig. 4), it is similarly possible to infer that in geminiviruses there was possibly only a single instance of either mutational loss or gain of the *rep* gene intron splicing signals. It is, however, not entirely clear whether the MRCA of the geminiviruses had a *rep* gene intron or not since the viral lineages with both types of gene branch from the root of the *rep* gene phylogeny. It is perhaps significant that five of the six inter-genus recombinants identified by our recombination analyses (Fig. 3) have "intronless" *rep* genes and carry evidence of undergoing recombination events that "converted" a virus with a spliced *rep* gene into a virus with an unspliced *rep* gene. The exceptional case, CCDaV, appears to have involved the conversion of a virus with an intronless *rep* gene into one that had a *rep* gene intron. It is also noteworthy that some species of the geminivirus-like mycoviruses (represented in Fig. 4 by SsHADV) also likely express Rep proteins from spliced complementary strand transcripts (Dayaram et al., 2012) and it is conceivable therefore that the *rep* gene of the MRCA of the geminiviruses and geminivirus-like mycovirus could have also contained an intron.

3.7. *Capulavirus: a new genus of the Geminiviridae family*

EcmLV and FbSLCV clearly belong to a highly divergent geminivirus lineage. Moreover, the virion-sense genes of these new viruses exhibit a unique organisation among this family with only the *cp* gene having detectable homologues in other currently known geminivirus genomes. Given that these viruses are obviously distinguishable from the other established geminivirus genera based on sequence relatedness and genome organisation, we suggest that EcmLV and FbSLCV be placed within a new geminivirus genus. The name that we propose for this new genus is "Capulavirus". It is expected that Capulavirus may also be distinguishable based on the vector criteria, because the vector of EcmLV, if any, is expected to be different from the vectors so-far reported for other geminiviruses. Indeed, the coat proteins of EcmLV and FbSLCV, the only protein likely to be involved in their vector specificity (Briddon et al., 1990) is very distantly related to all other previously reported geminivirus coat proteins. Further studies are needed to confirm this expectation and additionally to test whether EcmLV and FbSLCV are also distinct from the other genera with respect to the breadth and/or specificity of the range of host species that they naturally infect.

Acknowledgements

This work was supported by Fondation pour la Recherche sur la Biodiversité, Direction Générale de l'Armement (Ministère de la Défense, France), Métaprogramme INRA « Meta-omics of microbial ecosystems » and CIRAD. We wish to express our sincere thanks and appreciation to Mr Paul Loubser and colleagues from Buffelsfontein Game & Nature Reserve.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.virusres.2013.07.006.

References

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. Journal of Molecular Biology 215 (3), 403–410.
Arguello-Astorga, G., Lopez-Ochoa, L., Kong, L.J., Orozco, B.M., Settlage, S.B., Hanley-Bowdoin, L., 2004. A novel motif in geminivirus replication proteins interacts with the plant retinoblastoma-related protein. Journal of Virology 78 (9), 4817–4826.
Boulton, M.I., 2002. Functions and interactions of mastrevirus gene products. Physiological and Molecular Plant Pathology 60 (5), 243–255.
Briddon, R.W., Bedford, I.D., Tsai, J.H., Markham, P.G., 1996. Analysis of the nucleotide sequence of the treehopper-transmitted geminivirus, tomato pseudo-curly top virus, suggests a recombinant origin. Virology 219 (2), 387–394.
Briddon, R.W., Heydarnejad, J., Khosrowfar, F., Massumi, H., Martin, D.P., Varsani, A., 2010. Turnip curly top virus, a highly divergent geminivirus infecting turnip in Iran. Virus Research 152 (1/2), 169–175.
Briddon, R.W., Pinner, M.S., Stanley, J., Markham, P.G., 1990. Geminivirus coat protein gene replacement alters insect specificity. Virology 177 (1), 85–94.
Choudhury, N.R., Malik, P.S., Singh, D.K., Islam, M.N., Kaliappan, K., Mukherjee, S.K., 2006. The oligomeric Rep protein of Mungbean yellow mosaic India virus (MYMIV) is a likely replicative helicase. Nucleic Acids Research 34 (21), 6362–6377.
Clerot, D., Bernardi, F., 2006. DNA helicase activity is associated with the replication initiator protein rep of tomato yellow leaf curl geminivirus. Journal of Virology 80 (22), 11322–11330.
Dayaram, A., Opong, A., Jaschke, A., Hadfield, J., Baschiera, M., Dobson, R.C.J., Offei, S.K., Shepherd, D.N., Martin, D.P., Varsani, A., 2012. Molecular characterisation of a novel cassava associated circular ssDNA virus. Virus Research 166 (1/2), 130–135.
Dekker, E.L., Woolston, C.J., Xue, Y.B., Cox, B., Mullineaux, P.M., 1991. Transcript mapping reveals different expression strategies for the bicistronic Rnas of the geminivirus wheat dwarf virus. Nucleic Acids Research 19 (15), 4075–4081.
Delwart, E., 2012. Animal virus discovery: improving animal health, understanding zoonoses, and opportunities for vaccine development. Current Opinion in Virology 2 (3), 344–352.
Duffy, S., Holmes, E.C., 2008. Phylogenetic evidence for rapid rates of molecular evolution in the single-stranded DNA begomovirus tomato yellow leaf curl virus. Journal of Virology 82 (2), 957–965.
Eagle, P.A., Hanley-Bowdoin, L., 1997. Cis elements that contribute to geminivirus transcriptional regulation and the efficiency of DNA replication. Journal of Virology 71 (9), 6947–6955.
Fauquet, C.M., Briddon, R.W., Brown, J.K., Moriones, E., Stanley, J., Zerbini, M., Zhou, X., 2008. Geminivirus strain demarcation and nomenclature. Archives of Virology 153 (4), 783–821.
Fauquet, C.M., Stanley, J., 2003. Geminivirus classification and nomenclature: progress and problems. Annals of Applied Biology 142 (2), 165–189.
Fuller, C., 1901. Mealie variegation. First report of the government entomologist 1899–1900., pp. 17–19.
Ge, L., Zhang, J., Zhou, X., Li, H., 2007. Genetic structure and population variability of Tomato yellow leaf curl China virus. Journal of Virology 81 (11), 5902–5907.
Haible, D., Kober, S., Jeske, H., 2006. Rolling circle amplification revolutionizes diagnosis and genomics of geminiviruses. Journal of Virological Methods 135 (1), 9–16.
Ilyina, T.V., Koonin, E.V., 1992. Conserved sequence motifs in the initiator proteins for rolling circle DNA replication encoded by diverse replicons from eubacteria, eucaryotes and archaebacteria. Nucleic Acids Research 20 (13), 3279–3285.
Isnard, M., Granier, M., Frutos, R., Reynaud, B., Peterschmitt, M., 1998. Quasispecies nature of three maize streak virus isolates obtained through different modes of selection from a population used to assess response to infection of maize cultivars. Journal of General Virology 79 (Pt 12), 3091–3099.
Jeske, H., 2009. Geminiviruses. In: zur Hausen, H., de Villiers, E.-M. (Eds.), Torque Teno Virus: The Still Elusive Human Pathogens, vol. 331. Springer, Berlin, pp. 185–226.
Jones, R.A.C., 2009. Plant virus emergence and evolution: origins, new encounter scenarios, factors driving emergence, effects of changing world conditions, and prospects for control. Virus Research 141 (2), 113–130.
Kim, K.J., Jansen, R.K., 1995. Ndhf sequence evolution and the major clades in the sunflower family. Proceedings of the National Academy of Sciences of the United States of America 92 (22), 10379–10383.
Koonin, E.V., Ilyina, T.V., 1993. Computer-assisted dissection of rolling circle DNA replication. Biosystems 30 (1–3), 241–268.
Krenz, B., Thompson, J.R., Fuchs, M., Perry, K.L., 2012. Complete genome sequence of a new circular DNA virus from grapevine. Journal of Virology 86 (14), 7715.
Lazarowitz, S.G., Wu, L.C., Rogers, S.G., Elmer, J.S., 1992. Sequence-specific interaction with the viral AL1 protein identifies a geminivirus DNA replication origin. Plant Cell 4 (7), 799–809.
Le, S.Q., Gascuel, O., 2008. An improved general amino acid replacement matrix. Molecular Biology and Evolution 25 (7), 1307–1320.

Legg, J.P., Fauquet, C.M., 2004. Cassava mosaic geminiviruses in Africa. Plant Molecular Biology 56 (4), 585–599.
Letunic, I., Doerks, T., Bork, P., 2012. SMART 7: recent updates to the protein domain annotation resource. Nucleic Acids Research 40 (D1), D302–D305.
Loconsole, G., Saldarelli, P., Doddapaneni, H., Savino, V., Martelli, G.P., Saponari, M., 2012. Identification of a single-stranded DNA virus associated with citrus chlorotic dwarf disease, a new member in the family *Geminiviridae*. Virology 432 (1), 162–172.
Londono, A., Riego-Ruiz, L., Arguello-Astorga, G.R., 2010. DNA-binding specificity determinants of replication proteins encoded by eukaryotic ssDNA viruses are adjacent to widely separated RCR conserved motifs. Archives of Virology 155 (7), 1033–1046.
Malpica, J.M., Sacristan, S., Fraile, A., Garcia-Arenal, F., 2006. Association and host selectivity in multi-host pathogens. PLoS ONE 1, e41.
Martin, D.P., Biagini, P., Lefeuvre, P., Golden, M., Roumagnac, P., Varsani, A., 2011. Recombination in eukaryotic single stranded DNA viruses. Viruses 3 (9), 1699–1738.
Martin, D.P., Lemey, P., Lott, M., Moulton, V., Posada, D., Lefeuvre, P., 2010. RDP3: a flexible and fast computer program for analyzing recombination. Bioinformatics 26 (19), 2462–2463.
Martin, D.P., Shepherd, D.N., 2009. The epidemiology, economic impact and control of maize streak disease. Food Security 1 (3), 305–315.
Melcher, U., Muthukumar, V., Wiley, G.B., Min, B.E., Palmer, M.W., Verchot-Lubicz, J., Ali, A., Nelson, R.S., Roe, B.A., Thapa, V., Pierce, M.L., 2008. Evidence for novel viruses by analysis of nucleic acids in virus-like particle fractions from *Ambrosia psilostachya*. Journal of Virological Methods 152 (1-2), 49–55.
Moffat, A.S., 1999. Plant pathology – Geminiviruses emerge as serious crop threat. Science 286 (5446), 1835-1835.
Monjane, A.L., Harkins, G.W., Martin, D.P., Lemey, P., Lefeuvre, P., Shepherd, D.N., Oluwafemi, S., Simuyandi, M., Zinga, I., Komba, E.K., Lakoutene, D.P., Mandakombo, N., Mboukoulida, J., Semballa, S., Tagne, A., Tiendrebeogo, F., Erdmann, J.B., van Antwerpen, T., Owor, B.E., Flett, B., Ramusi, M., Windram, O.P., Syed, R., Lett, J.M., Briddon, R.W., Markham, P.G., Rybicki, E.P., Varsani, A., 2011. Reconstructing the history of maize streak virus strain a dispersal to reveal diversification hot spots and its origin in southern Africa. Journal of Virology 85 (18), 9623–9636.
Muhire, B., Martin, D.P., Brown, J.K., Navas-Castillo, J., Moriones, E., Zerbini, F.M., Rivera-Bustamante, R., Malathi, V.G., Briddon, R.W., Varsani, A., 2013. A genome-wide pairwise-identity-based proposal for the classification of viruses in the genus Mastrevirus (family *Geminiviridae*). Archives of Virology 158 (6), 1411–1424.
Muthukumar, V., Melcher, U., Pierce, M., Wiley, G.B., Roe, B.A., Palmer, M.W., Thapa, V., Ali, A., Ding, T., 2009. Non-cultivated plants of the Tallgrass Prairie Preserve of northeastern Oklahoma frequently contain virus-like sequences in particulate fractions. Virus Research 141 (2), 169–173.
Nash, T.E., Dallas, M.B., Reyes, M.I., Buhrman, G.K., Ascencio-Ibanez, J.T., Hanley-Bowdoin, L., 2011. Functional analysis of a novel motif conserved across geminivirus Rep proteins. Journal of Virology 85 (3), 1182–1192.
Ng, T.F.F., Duffy, S., Polston, J.E., Bixby, E., Vallad, G.E., Breitbart, M., 2011a. Exploring the diversity of plant DNA viruses and their satellites using vector-enabled metagenomics on whiteflies. PLoS ONE 6 (4).
Ng, T.F.F., Willner, D.L., Lim, Y.W., Schmieder, R., Chau, B., Nilsson, C., Anthony, S., Ruan, Y.J., Rohwer, F., Breitbart, M., 2011b. Broad surveys of DNA viral diversity obtained through viral metagenomics of mosquitoes. PLoS ONE 6 (6).
Padidam, M., Sawyer, S., Fauquet, C.M., 1999. Possible emergence of new geminiviruses by frequent recombination. Virology 265 (2), 218–225.
Patil, B.L., Fauquet, C.M., 2009. Cassava mosaic geminiviruses: actual knowledge and perspectives. Molecular Plant Pathology 10 (5), 685–701.
Peterschmitt, M., Granier, M., Frutos, R., Reynaud, B., 1996. Infectivity and complete nucleotide sequence of the genome of a genetically distinct strain of maize streak virus from Reunion Island. Journal of Virology 141 (9), 1637–1650.
Posada, D., Crandall, K.A., 2001. Evaluation of methods for detecting recombination from DNA sequences: computer simulations. Proceedings of the National Academy of Sciences of the United States of America 98 (24), 13757–13762.
Rambaud, A., Drummond, A.J., 2007. Tracer, version 1.4. http://beast.bio.ed.ac.uk/Tracer
Rey, M.E.C., Ndunguru, J., Berrie, L.C., Paximadis, M., Berry, S., Cossa, N., Nuaila, V.N., Mabasa, K.G., Abraham, N., Rybicki, E.P., Martin, D., Pietersen, G., Esterhuizen, L.L., 2012. Diversity of dicotyledenous-infecting geminiviruses and their associated DNA molecules in southern Africa, including the South-west Indian ocean islands. Viruses 4 (9), 1753–1791.
Roossinck, M.J., Saha, P., Wiley, G.B., Quan, J., White, J.D., Lai, H., Chavarria, F., Shen, G.A., Roe, B.A., 2010. Ecogenomics: using massively parallel pyrosequencing to understand virus ecology. Molecular Ecology 19, 81–88.
Rosario, K., Duffy, S., Breitbart, M., 2012. A field guide to eukaryotic circular single-stranded DNA viruses: insights gained from metagenomics. Archives of Virology 157 (10), 1851–1871.
Rozen, S., Skaletsky, H., 2000. Primer3 on the WWW for general users and for biologist programmers. Methods in Molecular Biology 132, 365–386.
Rybicki, E.P., 1994. A phylogenetic and evolutionary justification for three genera of *Geminiviridae*. Archives of Virology 139 (1/2), 49–77.
Rybicki, E.P., Pietersen, G., 1999. Plant virus disease problems in the developing world. Advances in Virus Research 53 (53), 127.
Schubert, J., Habekuss, A., Kazmaier, K., Jeske, H., 2007. Surveying cereal-infecting geminiviruses in Germany – diagnostics and direct sequencing using rolling circle amplification. Virus Research 127 (1), 61–70.
Shepherd, D.N., Martin, D.P., Lefeuvre, P., Monjane, A.L., Owor, B.E., Rybicki, E.P., Varsani, A., 2008. A protocol for the rapid isolation of full geminivirus genomes from dried plant tissue. Journal of Virological Methods 149 (1), 97–102.
Stanley, J., Markham, P.G., Callis, R.J., Pinner, M.S., 1986. The nucleotide sequence of an infectious clone of the geminivirus beet curly top virus. EMBO Journal 5 (8), 1761–1767.
Suchard, M.A., Redelings, B.D., 2006. BAli-Phy: simultaneous Bayesian inference of alignment and phylogeny. Bioinformatics 22 (16), 2047–2048.
Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., Kumar, S., 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Molecular Biology and Evolution 28 (10), 2731–2739.
Tan, P.H.N., Wong, S.M., Wu, M., Bedford, I.D., Saunders, K., Stanley, J., 1995. Genome organization of ageratum yellow vein virus, a monopartite whitefly-transmitted geminivirus isolated from a common weed. Journal of General Virology 76, 2915–2922.
Varsani, A., Shepherd, D.N., Dent, K., Monjane, A.L., Rybicki, E.P., Martin, D.P., 2009. A highly divergent South African geminivirus species illuminates the ancient evolutionary history of this family. Virology Journal 6, 36.
Varsani, A., Shepherd, D.N., Monjane, A.L., Owor, B.E., Erdmann, J.B., Rybicki, E.P., Peterschmitt, M., Briddon, R.W., Markham, P.G., Oluwafemi, S., Windram, O.P., Lefeuvre, P., Lett, J.M., Martin, D.P., 2008. Recombination, decreased host specificity and increased mobility may have driven the emergence of maize streak virus as an agricultural pathogen. Journal of General Virology 89 (Pt. 9), 2063–2074.
Varsani, A., van der Walt, E., Heath, L., Rybicki, E.P., Williamson, A.L., Martin, D.P., 2006. Evidence of ancient papillomavirus recombination. Journal of General Virology 87, 2527–2531.
Warburg, O.M.D.S., 1894. Die kulturpfl anzen usambaras. Mitteilungen aus den DeutschenSchutzgebieten 7, 131.
Wright, E.A., Heckel, T., Groenendijk, J., Davies, J.W., Boulton, M.I., 1997. Splicing features in maize streak virus virion- and complementary-sense gene expression. Plant Journal 12 (6), 1285–1297.
Yazdi, H.R.B., Heydarnejad, J., Massumi, H., 2008. Genome characterization and genetic diversity of beet curly top Iran virus: a geminivirus with a novel nonanucleotide. Virus Genes 36 (3), 539–545.