Identification and molecular characterization of a novel monopartite geminivirus associated with mulberry mosaic dwarf disease

Yuxin Ma,1 Beatriz Navarro,2 Zhixiang Zhang,1 Meiguang Lu,1 Xueping Zhou,1 Shengqi Chi,3 Francesco Di Serio2 and Shifang Li1

1State Key Laboratory for Biology of Plant Diseases and Insect Pests, Institute of Plant Protection, Chinese Academy of Agricultural Sciences, Yuanmingyuan West Road No. 2, Haidian District, Beijing 100193, PR China
2Istituto per la Protezione Sostenibile delle Piante CNR, UO Bari, Via Amendola, 70126 Bari, Italy
3College of Agronomy and Plant Protection, Qingdao Agricultural University, Changcheng Road No. 700, Chengyang District, Qingdao 266000, PR China

High-throughput sequencing of small RNAs allowed the identification of a novel DNA virus in a Chinese mulberry tree affected by a disease showing mosaic and dwarfing symptoms. Rolling-circle amplification and PCR with specific primers, followed by sequencing of eleven independent full-length clones, showed that this virus has a monopartite circular DNA genome (~2.95 kb) containing ORFs in both polarity strands, as reported previously for geminiviruses. A field survey showed the close association of the virus with diseased mulberries, so we tentatively named the virus mulberry mosaic dwarf-associated virus (MMDaV). The MMDaV genome codes for five and two putative proteins in the virion-sense and in the complementary-sense strands, respectively. Although three MMDaV virion-sense putative proteins did not share sequence homology with any protein in the databases, functional domains [coiled-coil and transmembrane (TM) domains] were identified in two of them. In addition, the protein containing a TM domain was encoded by an ORF located in a similar genomic position in MMDaV and in several other geminiviruses. As reported for members of the genera Mastrevirus and Becurtovirus, MMDaV replication-associated proteins are expressed through the alternative splicing of an intron, which was shown to be functional in vivo. A similar intron was found in the genome of citrus chlorotic dwarf-associated virus (CCDaV), a divergent geminivirus found recently in citrus. On the basis of pairwise comparisons and phylogenetic analyses, CCDaV and MMDaV appear to be closely related to each other, thus supporting their inclusion in a putative novel genus in the family Geminiviridae.

INTRODUCTION

The family Geminiviridae comprises insect-transmitted viruses that may cause significant economic losses to many crops (Moffat, 1999; Varma & Malathi, 2003). These viruses have circular, single-stranded, monopartite or bipartite DNA genomes replicating in the nucleus through a combination of rolling-circle and recombination-mediated mechanisms (Gutierrez, 1999; Preiss and Jeske, 2003). Based on genome organization and number of components, the type of vector and host plants, seven geminivirus genera (Mastrevirus, Curtovirus, Topocuvirus, Becurtovirus, Eragrovirus, Turncurtovirus and Begomovirus) have been identified that comprise most of the known species (Brown et al., 2012; Fauquet et al., 2008;Varsani et al., 2014a). Most geminiviruses belong to the genus Begomovirus, which includes species with one or two genomic components, whereas all the other geminiviruses have a single genomic component. Members of the genera Mastrevirus and Becurtovirus differ from the others because of the expression strategy of two replication-associated proteins (Rep), which relies on the alternative splicing of the complementary-sense transcripts (Varsani et al., 2014a; Wright et al., 1997). A few highly divergent and still unclassified geminiviruses have recently been described from Euphorbia (Euphorbia...
caput-medusae-associated virus, EcmLV) (Bernardo et al., 2013), citrus (citrus chlorotic dwarf-associated virus, CCDaV) (Locomonte et al., 2012), grapevine (grapevine red blotch-associated virus, GRbV) (Al Rwahnih et al., 2013; Krenz et al., 2012; Poojari et al., 2013).

Viral RNAs, including those of geminiviruses (Seguin et al., 2014), are targeted by host defence mechanisms based on RNA silencing and, through the activity of Dicer-like enzymes (DCLs), become the source of virus-derived small RNAs of 21–24 nt (v-sRNAs). Recent developments in next-generation sequencing (NGS) of sRNAs and bioinformatics have supplied powerful tools for identifying hitherto unknown viral and subviral agents in plants (Al Rwahnih et al., 2009, 2013; Giampetruzzi et al., 2012; Ito et al., 2013; Massart et al., 2014). Thus, NGS technology was used to investigate the possible involvement of viral agent(s) in a disease denoted mulberry mosaic dwarf (MMD) that affects Morus alba in China. Of the five viruses reported from mulberry in this country by Kuai (2010), only two species, the carlavirus Mulberry latent virus and the nepovirus Mulberry ringspot virus, are listed in the Ninth Report of the International Committee on Taxonomy of Viruses (King et al., 2012), whereas another putative nepovirus was recently found in plants showing symptoms resembling those of MMD, but the cause/effect relationship with this disease was not established (Lu et al., 2015). Thus, the aetiology of MMD is still undetermined.

Based on NGS, a novel, highly divergent geminivirus was discovered in plants affected by MMD, the genome of which and expression strategy is described and discussed in the present paper.

RESULTS

Discovery of a new geminivirus in mulberry affected by mosaic dwarf disease

MMD is a graft-transmissible syndrome long known in China (Xia & Lu, 2004), where it can seriously affect silk farming (Huang et al., 1992). Mulberry plants were observed in Shaanxi province that showed the severe dwarfing (Fig. 1a) accompanied by mottling, deformation, curling and puckering of the leaves (Fig. 1c–e) typical of MMD. To identify viral agents potentially associated with this disease, sRNA libraries were generated from a symptomatic and a symptomless tree (AK2 and AK0, respectively), and sequenced using an Illumina platform. Approximately 37 million and 32 million raw reads were obtained, respectively, from symptomatic and symptomless samples which, after removal of the adapters, size selection (18–30 nt) and filtering, generated a dataset of 5.3 million and 5.8 million unique reads, respectively. Filtered reads were assembled by Velvet (Zerbino & Birney, 2008). Exclusively in the library from the symptomatic sample, BLASTX identified five contigs (out of a total of 503 contigs) ranging in size from about 90 to 400 nt that shared significant nucleotide and amino acid similarity with genes encoded by viruses of the family Geminiviridae. In particular, CCDaV and chickpea chlorotic dwarf virus were the viruses showing the highest identity with three and one contigs, respectively (Table S1, available in the online Supplementary Material). In contrast, contigs from the symptomless library did not show any significant similarity with previously reported viruses.

Complete genome sequence and sequence variability of a novel geminivirus from mulberry

Based on the sequence of one of the contigs with similarity to geminiviruses (node 19, Table S1), and considering that geminiviruses have a circular DNA genome, two overlapping and complementary primers (1228fw/1233rv, Table S2) were designed for PCR amplification of the circular component of the potential DNA virus from mulberry. A PCR product of about 3000 bp was amplified using a total DNA preparation from the AK2 symptomatic mulberry and another representative mulberry tree (isolate AK1) showing the same symptoms (Fig. 2a). In contrast, no amplification product was obtained from the healthy control (Fig. 2a). Cloning of the amplicons obtained from both symptomatic isolates and sequencing of four and six independent clones, respectively, yielded inserts 2952 nt in size with sequences almost identical to each other (99.8–100% similarity), differing in only few positions (Table S3). The sequence variant (GenBank accession no. KP303687) found in four of six clones from isolate AK1 was considered as the prevalent (master) variant. This variant differed from those cloned from isolate AK2 for either six or seven nucleotide changes that did not affect the amino acid composition of the potential proteins encoded by the circular DNA (Table S3).

A second set of overlapping primers (621fw/626rv) designed to target a different DNA region of the potential new virus was tested using DNA preparations from another mulberry tree (isolate AK3) affected by MMD. Full-length sequencing of a single clone of the approximately 3 kb amplification product obtained by using this set of primers confirmed the presence in the symptomatic plant of a circular DNA with a nucleotide sequence almost identical (99.7–99.8% similarity) to those of isolates AK2 and AK1. In a parallel experiment, no amplicon was obtained from a DNA preparation of a healthy control, thus providing additional support to the possible association of the circular DNA with MMD. Also, the use of this second amplification product allowed the precise sequence of the priming sites of primers 1228fw/1233rv to be resolved. The consensus sequence generated by aligning the sequence variants from the three isolates (Table S3) corresponded to the master sequence from isolate AK1, chosen as reference variant (GenBank accession no. KP303687) for further analyses in this study. Interestingly, the circular DNA variants from diseased mulberry trees contained the sequences of the four contigs from isolate AK2 that shared sequence
similarity with geminiviruses, confirming that these contigs derived from a single circular DNA molecule.

BLASTN analysis using this reference sequence as a query revealed that the most closely related sequences in the databases were two small fragments of 115 and 202 nt of CCDaV (with 77% and 68% sequence identity, respectively), thus supporting the likelihood that the circular DNA from mulberry is the genome of a previously unreported geminivirus. Since this circular DNA was found exclusively in mulberry trees affected by MMD, it was thought to be the genome of a virus for which the name mulberry mosaic dwarf-associated virus (MMDaV) is proposed.

At this stage, the presence of additional genomic components of MMDaV, as well as of DNA satellites similar to those reported in some begomoviruses with monopartite genomes (Zhou, 2013), could not be excluded. To gain insights into this, rolling-circle amplification (RCA) followed by RFLP analyses (Haible et al., 2006; Inoue-Nagata et al., 2004) was performed using DNA preparations from AK1 and a healthy mulberry control. Based on sequencing data reported above, the restriction enzymes

![Fig. 1. Symptoms of MMD disease: (a) dwarfing; (b–e) leaf curling, mosaic and yellowing; (f) leaves of a symptomless mulberry tree.](http://vir.sgmjournals.org)
SacI, EcoRI, XhoI, SpeI and XmaI were predicted to cut at a single site the RCA products of MMDaV circular DNA. Digestion of RCA products with these restriction enzymes always generated a single fragment of about 3 kb, showing the same electrophoretic mobility as the PCR product of the full-length MMDaV DNA molecule used as a loading.

**Fig. 2.** (a) Agarose gel electrophoresis of PCR products obtained using primers 1228fw/1233rv from symptomatic AK1 and AK2 mulberry trees (lanes 1 and 2, respectively) and a symptomless mulberry plant (lane 3); lane M, AL 5000 DNA Marker (Aidlab Biotechnologies), with sizes of DNA fragments indicated on the left. (b) Agarose gel electrophoresis of RCA products from symptomless (lane 8) and symptomatic mulberry trees digested with restriction enzymes SacI, EcoRI, XhoI, SpeI, XmaI (lanes 2–6, respectively), or not digested (lane 7). Lane M, DL 10 000 DNA Marker (TAKARA), with sizes of DNA fragments indicated on the left. (c) Southern blot hybridization of the gel shown in (b) with MMDaV-specific DIG-labelled probe; positions of non-digested (ND) and digested MMDaV amplified products are indicated on the left.
control (Fig. 2b). Additional bands were never detected in the agarose gel stained with ethidium bromide and no fragments of 3 kb or smaller sizes were observed in the negative controls (Fig. 2b). Southern blots using MMDaV-specific digoxigenin (DIG)-labelled probe showed hybridization signals corresponding to the bands of the digested and undigested RCA products in the infected sample, but no hybridization signals in the healthy sample (Fig. 2c). Altogether, these data suggest the absence of additional circular DNAs in the symptomatic mulberry besides the genomic component of MMDaV previously sequenced, thus supporting the conclusion that MMDaV is a newly identified geminivirus with a monopartite genome.

Genome organization of MMDaV

Seven ORFs were identified in both strands of the MMDaV circular DNA using the program ORF Finder (http://www.ncbi.nlm.nih.gov/gorf/gorf.html). In particular, the 5' half of the complementary-sense strand of the MMDaV genome codes for two putative proteins and is separated from the 5' half of the virion-sense strand coding for the remaining five putative proteins by an intergenic region (IR) (Fig. 3a) containing transcription regulatory sequences, including TATA boxes (at positions 242–245 and 2844–2847, for transcription of virion-sense and complementary-sense virus transcripts, respectively) and the nonanucleotide sequence (TAATATTAC, at positions 2946–2952) conserved in almost all geminiviruses and positioned within a short palindromic sequence forming a stem–loop structure (Fig. 3b). In line with this and further results shown below, which unveil a genome organization resembling that of most geminiviruses with monopartite genomes, MMDaV ORFs will be numbered and named according to Padidam et al. (1995).

Based on BLASTP analysis (Table 1) and pairwise comparisons with related geminiviruses (Table S4), the predicted V1 protein of MMDaV had the highest amino acid sequence identity with the coat protein (CP) of CCDaV. Accordingly, SMART analysis recognized the V1 protein of MMDaV as a geminivirus CP (Pfam E-value: $1.4 \times 10^{-28}$), in which a nuclear localization signal was identified at aa 50–58 (ARRKRRPIN) by the NucPred program (E-value: 0.56). Also, MMDaV V2 shared the highest sequence identity with the cognate V2 movement protein of CCDaV (Tables 1 and S4), although, in this case, the identity level was lower and no conserved domains were identified in MMDaV V2 by SMART analyses.

Despite the absence in databases of proteins with sequence homology to MMDaV V3, examination by using the SMART program of potential functional motifs contained in this protein revealed the presence of a transmembrane (TM) domain of 23 aa (Table 1). Similar analyses of the genomes of several representative geminiviruses, integrated by search with the TMPRED program (Hofmann & Stoffel, 1993),
<table>
<thead>
<tr>
<th>ORF</th>
<th>Nucleotide coordinates</th>
<th>Orientation</th>
<th>No. amino acids</th>
<th>(M_r) (kDa)</th>
<th>BLASTP (E-value)</th>
<th>Predicted domains (SMART)</th>
<th>Amino acid coordinates of predicted domain</th>
</tr>
</thead>
<tbody>
<tr>
<td>V1</td>
<td>607–1344</td>
<td>Sense</td>
<td>245</td>
<td>28.0</td>
<td>Putative coat protein-like protein of CCDaV, ID: YP_006522419.1 (2e(-43))</td>
<td>Gemini_coat: Geminivirus coat-protein/nuclear export factor BR1 family (1.4e(-28))*</td>
<td>10 to 245</td>
</tr>
<tr>
<td>V2</td>
<td>369–818</td>
<td>Sense</td>
<td>149</td>
<td>17.0</td>
<td>V2 of CCDaV, ID: AHL20275.1 (0.009)</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>V3</td>
<td>236–553</td>
<td>Sense</td>
<td>105</td>
<td>12.0</td>
<td>–</td>
<td>Transmembrane domain†</td>
<td>26 to 48</td>
</tr>
<tr>
<td>V4</td>
<td>1383–1775</td>
<td>Sense</td>
<td>130</td>
<td>15.0</td>
<td>–</td>
<td>Coiled-coil domain‡</td>
<td>85 to 113</td>
</tr>
<tr>
<td>V5</td>
<td>1414–1698</td>
<td>Sense</td>
<td>94</td>
<td>10.8</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>C1</td>
<td>2842–2048</td>
<td>Complementary</td>
<td>264</td>
<td>30.4</td>
<td>Putative RepA-like protein of CCDaV, ID: YP_0065224221 (1e(-77))</td>
<td>Gemini_AL1: Geminivirus Rep catalytic domain (1.1e(-34))*</td>
<td>7 to 120</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Gemini_AL1M: Geminivirus Rep protein central domain (2e(-08))*</td>
<td>Gemini_AL1M: Geminivirus Rep protein central domain (2.3e(-12))*</td>
<td>123 to 227</td>
</tr>
<tr>
<td>C1:C2</td>
<td>2842–2237:2135–1737§</td>
<td>Complementary</td>
<td>34</td>
<td>38.7</td>
<td>Putative RepA-like protein of C1:C2 protein of CCDaV (8e(-140))‖</td>
<td>Gemini_AL1: Geminivirus Rep catalytic domain (1.9e(-34))*</td>
<td>7 to 120</td>
</tr>
</tbody>
</table>

*Domains and E-value obtained from Pfam analysis.
†Domain identified by the TMHMM v2.0 program.
‡Domain identified by the program.
§Start and stop of C1:C2 exons.
‖Putative C1:C2 protein generated by alternative splicing of complementary-sense transcript of CCDaV (GenBank accession no. NC_018151).
allowed the identification of TM domains within movement and putative proteins encoded by viral genes in a similar genomic position in members of the genera *Mastrevirus*, *Curtovirus* and *Becurtovirus*, and by EcmLV of the proposed genus *Capulavirus* (Table 2 and Fig. S1), suggesting a possible role in virus trafficking also for the proposed genus *Capulavirus* (Bernardo et al., 2013), the FGENESH program predicted the presence of 103, was identified in the V4 putative protein by SMART

The last two ORFs in the virion-sense strand are V4 and V5, which code for putative proteins composed of 130 and 94 aa (molecular mass 15 and 10.8 kDa), respectively, for which no homologous protein in the databases was found. Although the function of these proteins remains unknown, a coiled-coil domain, spanning from aa 85 to 94 aa (molecular mass 15 and 10.8 kDa), respectively, was identified in the V4 putative protein by SMART analyses (Table 1).

The complementary-sense strand of MMDaV codes for RepA and Rep proteins translated from unspliced (C1 transcript) and spliced (C1:C2 transcript) viral RNAs, respectively. In fact, similarly to viruses within the genera *Mastrevirus* and *Becurtovirus* (Dekker et al., 1991; Morris et al., 1992; Mullineaux et al., 1990; Schalk et al., 1989; Varsani et al., 2014b; Wright et al., 1997), and to EcmLV (Bernardo et al., 2013) and GRBaV (Al Rwahnih et al., 2013), the FGENESH program predicted the presence of an intron in the complementary-sense transcript (from positions 2236 to 2136 in the MMDaV genome) (Fig. 4a). This intron contains the consensus 5’ and 3’ splice sites and branchpoints characteristic of the U2 plant introns (Lewandowska et al., 2004) and has an RNA U + A content (59.4%) close to the minimum needed for its efficient splicing (Goodall & Filipowicz, 1989; Simpson & Brown, 1993). Interestingly, an intron with similar features can also be predicted in a similar position of the complementary-sense strand of the CCDaV genome. MMDaV and CCDaV intron splicing signals differ slightly from those reported for the other geminivirus introns (Fig. S2).

According to SMART analyses, ORFs C1 and C1 : C2 of MMDaV code for proteins that contain the typical domains of geminivirus Rep proteins and share the highest sequence identity with the putative RepA-like and Rep-like proteins of CCDaV (Tables 1 and S4). A more detailed analysis showed that Rep proteins encoded by MMDaV contain the motifs reported in other geminivirus Reps (Fig. S3), which include: (i) the rolling-cycle replication (RCR) motif I (FLTFP), required for specific dsDNA binding; (ii) the RCR motif II (HFH), a metal-binding site that may be involved in protein conformation and DNA cleavage; (iii) the RCR motif III (YIQKE), a catalytic site for DNA cleavage (Nash et al., 2011); (iv) the Walker A (GPT-RSGKT) and the Walker B (LYNVIDDI) domains that are crucial components of the nucleotide-binding site (Walker et al., 1982) (Fig. S3). However, the canonical LxCxE ribonuclease binding domain (Arguello-Astorga et al., 2004) previously identified in the Reps of some mastreviruses (Xie et al., 1995) and ECSV (genus *Eragrovirus*) (Varsani et al., 2009) was not identified in MMDaV Rep.

### Table 2. Transmembrane domains predicted in movement and putative proteins of several geminiviruses

<table>
<thead>
<tr>
<th>Virus</th>
<th>Genus</th>
<th>Protein</th>
<th>Transmembrane domain*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Start (aa position)</td>
<td>End (aa position)</td>
<td>Length (aa)</td>
</tr>
<tr>
<td>MMDaV</td>
<td>Unassigned</td>
<td>x†</td>
<td>29</td>
</tr>
<tr>
<td>MSV</td>
<td><em>Mastrevirus</em></td>
<td>MP</td>
<td>31</td>
</tr>
<tr>
<td>CSMV</td>
<td><em>Mastrevirus</em></td>
<td>MP‡</td>
<td>57</td>
</tr>
<tr>
<td>PanSV</td>
<td><em>Mastrevirus</em></td>
<td>MP‡</td>
<td>35</td>
</tr>
<tr>
<td>WDV</td>
<td><em>Mastrevirus</em></td>
<td>MP‡</td>
<td>31</td>
</tr>
<tr>
<td>BCTV</td>
<td><em>Curtovirus</em></td>
<td>MP‡</td>
<td>3</td>
</tr>
<tr>
<td>HrCIV</td>
<td><em>Curtovirus</em></td>
<td>MP‡</td>
<td>3</td>
</tr>
<tr>
<td>SCTV</td>
<td><em>Curtovirus</em></td>
<td>MP‡</td>
<td>3</td>
</tr>
<tr>
<td>BCTIV</td>
<td><em>Becurtovirus</em></td>
<td>MP‡</td>
<td>1</td>
</tr>
<tr>
<td>SCTAV</td>
<td><em>Becurtovirus</em></td>
<td>MP‡</td>
<td>3</td>
</tr>
<tr>
<td>EcmLV</td>
<td><em>Capulavirus</em></td>
<td>MP‡</td>
<td>10</td>
</tr>
<tr>
<td>CCDaV</td>
<td>Unassigned</td>
<td>x†</td>
<td>34</td>
</tr>
<tr>
<td>ECSV</td>
<td><em>Eragrovirus</em></td>
<td>x†</td>
<td>7</td>
</tr>
<tr>
<td>GRBaV</td>
<td><em>Graingemovirus</em></td>
<td>x†</td>
<td>11</td>
</tr>
</tbody>
</table>

*Transmembrane domains were predicted by TMPRED software.
†Putative protein proposed in this article.
‡Encoded protein previously identified as movement protein (MP).
§Tentative genus.
MMDaV complementary-sense transcript is efficiently spliced in vivo

To investigate whether the intron identified in the complementary-sense strand of the MMDaV genome actually plays a role in the expression strategy of MMDaV ORFs, we performed reverse transcription (RT)-PCR experiments using a primer pair flanking the identified intron (1858F/2335R, Table S2) and mulberry RNA preparations in which DNA was degraded by exhaustive DNase treatment. While no amplicon was obtained from the healthy control, two amplification products were detected in RNA preparations from MMDaV-infected mulberry (isolate AK1), the larger one (about 480 bp) co-migrated with the loading control for the cDNAs of the unspliced transcripts, the shorter one (about 380 bp) had the sizes expected for the cDNAs of the spliced MMDaV transcripts (Fig. 4b). Moreover, sequencing of the shorter fragment showed that it actually corresponded to the 377 bp cDNA of the viral transcript spliced according to prediction, thus confirming the presence in plants of both spliced and unspliced complementary-sense MMDaV transcripts (Fig. 4a).

Analyses of v-sRNAs derived from MMDaV (MMDaV-sRNAs) infecting the symptomatic mulberry further supported this conclusion. In line with other geminiviruses (Aregger et al., 2012; Miozzi et al., 2013; Yang et al., 2011), MMDaV-sRNAs of both polarities and with a size distribution profile with prevalent peaks of 21, 22 and 24 nt were identified in the sRNA library from the symptomatic mulberry (Fig. S4a), suggesting that MMDaV-sRNAs are generated by multiple DCLs homologous to DCL1 to -4 of Arabidopsis thaliana. Similarly to other geminiviruses, when the 5' terminus of each MMDaV-sRNA was mapped along the virus genome, a distribution profile characterized by prominent peaks located in coding regions was obtained (Fig. S4b). However, this analysis also showed that a total of 1880 and 139 MMDaV-sRNA reads mapped at the 5' and 3' intron-exon borders of the viral complementary-sense transcript, respectively. Interestingly, a total of 136 MMDaV-sRNAs spanning the exon-exon junctions.

**Fig. 4.** (a) Sequences of the intron (boxed) and flanking exons identified in the complementary-sense transcript of MMDaV. The 5' and 3' splicing signals characteristic of the U2 plant introns are indicated in bold. MMDaV-sRNAs spanning the exon-exon junctions are indicated by lines covering the corresponding sequences, with the number of reads indicated on the right. (b) Agarose gel electrophoresis of RT-PCR products obtained from MMDaV-infected (lane 1) and symptomless (lane 3) mulberry plants using the primer pair 1858F/2335R flanking the intron; PCR product using viral genomic DNA as template is a size marker for the cDNAs of unspliced transcripts (lane 2); M, DNA Marker I (Aidlab Biotechnologies).
predicted for the spliced Rep transcripts were also identified (Fig. 4a). These findings provide additional indirect evidence that unspliced and spliced MMDaV transcripts are actually generated in vivo and they also indicate that MMDaV spliced RNA transcripts are additional sources of v-sRNAs, possibly after serving as template for the synthesis of dsRNAs mediated by host RNA-dependent RNA polymerase(s).

**Phylogenetic relationships between MMDaV and other geminiviruses**

Based on BLASTP searches and pairwise comparisons of amino acid sequences of several proteins and nucleotide sequences of full-length genomes, CCDaV was identified as the gemivirus most closely related to MMDaV (Tables 1 and S4). In fact, when phylogenetic trees were generated by the maximum-likelihood method using the CP and Rep proteins encoded by MMDaV and representative members of all genera in the family Geminiviridae, MMDaV and CCDaV clustered together in the same branch, clearly separated from all the other geminiviruses (Fig. 5). These two viruses were also grouped in the same cluster in a tree generated using the full genome of representative geminiviruses (Fig. S5). A similar clustering was observed in phylogenetic trees generated with the neighbour-joining method (data not shown).

**Field survey of MMDaV in mulberry trees**

To investigate the distribution of MMDaV in Shaanxi province and its association with mosaic dwarf disease, a total of 101 mulberry trees, 92 of which displayed typical disease symptoms, were tested by Southern blotting using a MMDaV-specific DIG-labelled DNA probe. No hybridization signal was detected in samples from symptomless mulberry plants, while clear signals, corresponding to different forms of MMDaV DNA (Fig. 6), were given by 85 of 92 symptomatic samples (92.4%), confirming the close association of MMDaV with mulberry mosaic dwarf disease.

**DISCUSSION**

In the last few years, identification of divergent geminiviruses infecting citrus, grapevine and Jatropha multifida (Al Rwahnih et al., 2013; Krenz et al., 2012; Loconsole et al., 2012; Polston et al., 2014; Poojari et al., 2013) has shown that the host range of geminiviruses is not limited to herbaceous plants. In this study we report the identification and characterization of a novel gemivirus infecting mulberry trees, as an addition to the list of woody plants infected by geminiviruses in nature. A circular DNA was amplified, cloned and completely sequenced from three different symptomatic mulberry trees. Sequencing of eleven independent clones identified very low nucleotide variability in the circular DNA, substantially consisting of silent mutations in the prevailing variants. Searches in databases for similar sequences and comparisons with other viruses (Tables 1 and S4) were consistent with the possibility that such a circular DNA molecule was the genomic component of a gemivirus related to CCDaV, a divergent gemivirus recently identified in citrus (Loconsole et al., 2012).

Similarly to CCDaV, RCA followed by RFLP and Southern blot analyses (Fig. 2) confirmed that MMDaV has a single genomic component and excluded the presence of associated satellites in the mulberry plants tested. The genomic organization of MMDaV is shared by all the other monopartite geminiviruses. Although the identity levels between the full-length genomic sequences and the amino acid composition of the putative CP and Rep proteins support a closer relationship between MMDaV and CCDaV than with other geminiviruses (Tables 1 and S4), ORF V2 of MMDaV shares only a limited sequence identity with the putative movement protein (V2) encoded by the CCDaV genome. No match in the databases was found for the proteins coding for the remaining MMDaV ORFs (V3–V5) (Table 1), a situation also reported for other recently characterized geminiviruses (Bernardo et al., 2013; Briddon et al., 2010; Krenz et al., 2012; Varsani et al., 2009). However, some functional domains already known to play major roles in the infectivity of plant viruses were identified in some of these MMDaV putative proteins. This is the case of the coiled-coil and TM domains found in V4 and V3, respectively. Coiled-coil domains were previously reported in viral proteins involved in movement, insect transmission or RNA silencing suppressor activity of DNA or RNA viruses (Bragg & Jackson 2004; Hohn, 2013; Singh et al., 2014; Sun et al., 2013). TM domains were previously reported in viral movement proteins (MP) of other viruses, including mastreviruses (Boulton, 2002) and EcmLV (Bernardo et al., 2013), and we showed that a TM domain is also present in the MP of members of the genera Curtovirus (Stanley, 2008) and Becurtovirus (Yazdi et al., 2008). In the respective genomes of these geminiviruses, MP is encoded in the virus-sense ORF located close to the IR, in a position similar to that occupied by MMDaV V3 ORF (Fig. S1). Moreover, an extensive search for TM domains in most geminiviruses showed that an ORF coding for a putative protein containing a TM domain is located at the same position within the genome of other divergent geminiviruses, including ECSV (genus Eragrovirus; Varsani et al., 2014a), GRBaV (tentative genus Graingenvirus; Poojari et al., 2013) and CCDaV (unclassified genus; Loconsole et al., 2012). Possibly this ORF remained unnoticed so far because it is not present in the genome of other well-characterized geminiviruses used for comparisons. Identification, in a similar position within the genome of so many different geminiviruses, of an ORF coding for a putative protein that contains the same functional domain (TM) (Table 2 and Fig. S1) supports a genetic constraint preserving such an ORF in the viral genomes, thus suggesting that it is actually translated in vivo into a small membrane-associated protein playing a functional role, possibly related to virus trafficking.
Fig. 5. Phylogenetic trees generated by the maximum-likelihood method from the alignment of CP (a) and Rep (b) amino acid sequences of MMDaV and representative members of the genera Begomovirus, Bemovirus, Curtovirus, Eragrovirus, Mastrevirus, Topocurivirus, Turncurtovirus and unclassified viruses within the family Geminiviridae. Numbers indicate bootstrap percentages (higher than 50%) for each node. Database accession numbers are given in brackets. Bars, 0.2 substitutions per amino acid position. ACMV, African cassava mosaic virus; BCTIV, beet curly top Iran virus; BCTV, beet curly top virus; BGYMV, bean golden yellow mosaic virus CpCDV; chickpea chlorotic dwarf virus; CPRV, chickpea redleaf virus; CPVV, chickpea yellow mosaic virus; CSVMV, citrus striate mosaic virus; CCaV, citrus chlorotic dwarf-associated virus; CIGMV, Clerodendron golden mosaic virus; CoGMV, Corchorus golden mosaic virus; CLCRV, cotton leaf curl Rajasthan virus; CGMV, cowpea golden mosaic virus; DDSMV, Digitaria didactylia striate mosaic virus; DSV, Digitaria streak virus; DoYMV, Dolichos yellow mosaic virus; ECSV, Eragrostis curvula streak virus; EMSV, Eragrostis minor streak virus; ESV, Eragrostis streak virus; EcmLV, Euphorbia caput-medusae latent virus; FBSCaV, French bean severe leaf curl virus; GRBaV, grapevine red blotch-associated virus; HYVV, honeymoon strawberry vein vein virus; HCTV, horseradish curly top virus; MaMPRV, Macroptilium mosaic Puerto Rico virus; MSRV, maize streak Reunion virus; MSV, maize streak virus; MiSV, Miscanthus streak virus; MMDaV, mulberry mosaic dwarf-associated virus; ODV, oat MlaV, dwarf virus PanSV, Panicum streak virus; PSMV, Paspalum striate mosaic virus; SacSV, Saccharum streak virus; SCTAV, spinach curly top Arizona virus; SSCTV, spinach severe curly top SacSV, top virus; SDEV, sugarcane streak Egypt virus; SSRV, sugarcane streak Reunion virus; SSV, sugarcane streak virus; SPLCV,
MMDaV complementary-sense transcripts are spliced in vivo, thus showing that MMDaV adopts the same expression strategy for the Rep protein previously reported for mastreviruses (Wright et al., 1997) and predicted for becurtoviruses (Heydarnejad et al., 2013) and for other divergent geminiviruses, including GRBaV (Al Rwahnih et al., 2013; Krenz et al., 2012) and EcmLV (Bernardo et al., 2013). In our study, both the sequencing of RT-PCR products generated by primers flanking the identified intron and the identification of MMDaV-sRNAs spanning the unspliced intron–exon borders and the predicted intron-spliced junctions are consistent with the coexistence in vivo of unspliced and spliced viral transcripts. Incidentally, our study showed that NGS data can be exploited for further exploration of the expression strategy of geminiviruses. The splicing signals found at the border of the MMDaV intron are identical to those conserved in the characteristic U2 plant introns (Lewandowska et al., 2004), but they differ slightly from those reported in all the other geminiviruses known to bear an intron in the complementary-sense transcript (Fig. S3). In this context, of particular interest appears the identification in the complementary-sense strand of CCDaV of a putative intron that could be involved in the expression of Rep protein of this virus because splicing signals of this intron are identical to those found in the MMDaV intron (Fig. S2), thus strongly suggesting that the similarity between CCDaV and MMDaV can be extended to the expression strategy of their respective Rep proteins.

Overall, the most solid evidence of a close relationship between CCDaV and MMDaV relies on phylogenetic analyses. In all phylogenetic trees, regardless of whether they were generated by maximum-likelihood or neighbour-joining methods, and whether the full-length genome or the CP and Rep proteins of representative geminiviruses were considered, MMDaV and CCDaV always clustered in the same clade, clearly separated from all the other geminiviruses with high bootstrap values (Figs. 5 and S5). Therefore, the close similarities in genome composition, organization and expression of MMDaV and CCDaV, and their clear phylogenetic relationships supports that MMDaV is the second member of a new genus in the family Geminiviridae proposed by Loconsole et al. (2012).

Biological data on MMDaV are scanty at this stage. In the case of CCDaV, whitefly (Parabemisia myricae)-mediated transmission has been proposed (Kersting et al., 1996). Whether the same applies to MMDaV remains to be ascertained. Bioassays are also needed for conclusively determining host range and pathogenicity of MMDaV. However, the data supporting the association of MMDaV with mulberry mosaic dwarf disease were confirmed by a field survey in which 92% of symptomatic trees proved to contain the virus when analysed by Southern blot hybridization, a detection method more reliable than PCR. Data reported in this study will be seminal for further studying the biological and epidemiological features of this novel virus. Incidence of MMDaV in mulberry in China and other countries can be now monitored and virus spread through infected propagation material can be limited or avoided with appropriate certification programs.

METHODS

Sample collection, DNA extraction and Southern blotting. Mulberry leaf samples used for generating the cDNA libraries of sRNAs sequenced by NGS were from one mulberry tree (isolate AK2) showing the typical symptoms of MMD disease and one symptomless mulberry tree growing in Ankang City (32.7° N 108.8° E), Shaanxi Province (China). A total of 101 samples of leaves from symptomatic (92) and symptomless (9) mulberry trees growing in several fields of Shaanxi province were collected for survey.

Total DNA was extracted using the CTAB (cetyltrimethylammonium bromide) method (Murray & Thompson, 1980), followed by phenol extraction and ethanol precipitation, and analysed in 1% agarose gel. For Southern blot analyses, after washing in denaturation (0.15 M NaOH, 1.5 M NaCl) and neutralization (1 M NaCl, 0.5 M Tris base, 0.45 M HCl, pH 7.2) solutions, the gel was blotted on Amersham Hybond-N+ membrane (GE Healthcare) that was pre-hybridized at 65°C for 4 h in DIG Easy Hyb solution (Roche) and hybridized overnight at 45°C in the same solution to which the DIG-labelled MMDaV-specific DNA probe was added. The probe was synthesized using a PCR DIG Probe Synthesis kit (Roche), a full-length clone of MMDaV DNA and primers 978F and 1879R to amplify a genomic DNA probe.

sRNA library construction, NGS and analyses of sRNAs. sRNA libraries were constructed from 1.0 µg of total RNA using the ‘Small RNA v1.5 Sample Prep’ kit from Illumina and sequenced in a single lane on a Genome Analyser IIx. The output files were processed with Illumina’s CASAVA pipeline (version 1.8). Raw Illumina sRNA reads were first processed to trim adaptor and barcode sequences. Trimmed sRNA sequences between 18 and 30 nt were aligned to the mulberry genome (http://morus.snu.edu.cn/morusdb/) using the BLAST program (Li & Durbin, 2009) to further exclude non-host sRNAs that were then assembled de novo into larger contigs using Velvet software (Zerbino & Birney, 2008) with a k-mer of 17. The resulting final contigs were compared against the GenBank Virus Reference Database (http://www.ncbi.nlm.nih.gov) using the BLAST programs (Altschul et al., 1990).

Sequencing and analyses of full-length viral genomic DNA. Total DNA preparations were diluted to five- and 50-fold and used for PCR amplification with the Phusion High-Fidelity DNA Polymerase (Thermo Fisher Scientific) (isolates AK1 and AK2) or Long Taq DNA Polymerase (TIANGEN, China) (isolate AK3) and the overlapping primers 1228fw/1233rv or the primers 621fw/626rv (Table S2). The thermocycling conditions included 1 cycle at 98°C
for 30 s; 35 cycles at 98 °C for 8 s, 67.6 °C for 25 s and 72 °C for 75 s; and 1 cycle at 72 °C for 7 min. The amplified product was purified using a Wizard SV Gel and PCR Clean-Up System (Promega) and was cloned into a blunt cloning vector using a pEASY-Blunt Simple Cloning kit (TransGen Biotech). The nucleotide sequence was determined by primer walking on both strands.

The reconstructed genome was subjected to standard sequence analysis: (i) prediction of the ORFs using ORF Finder (http://www.ncbi.nlm.nih.gov/projects/orf/); (ii) assessment of the presence of transcriptional control signals using prediction program TSSP within the Softberry website (http://linux1.softberry.com/berry.phtml?topic=tssp&group=programs&subgroup=promoter/); (iii) prediction of putative introns using fGENESH 2.6 software (Solovyev et al., 2006); (iv) identification of conserved and functional domains in the predicted proteins using SMART tool (http://smart.embl.de) (Letunic et al. 2015); (v) transmembrane domains were identified using the TMHMM v2.0 program (within SMART) and by TMPRED software (Hofmann & Stoffel, 1993); (vi) multiple sequence alignments using CLUSTAL W and reconstruction of phylogenetic trees using the MEGAS5.2 package (Tamura et al., 2011).

**ACKNOWLEDGEMENTS**

We thank Professor G. P. Martelli for critical reading of the manuscript and Xiuling Yang, Pengbo Liang and L Qin for the help and advice during this research. We are also grateful for the help of Professor Guitian Gao and to the experts at the Ankang Silkworm industry development centre for providing samples. This work was supported by grants from the Earmarked Fund for China Agriculture Research System (CARS-31), the Special Fund for Agro-scientific Research in the Public Interest (nos 201203076 and 200903004), and the National Natural Science Foundation of China (nos 31471747 and 31371911). Work in the laboratories of B.N. and F.D.S. has been partially supported by a dedicated grant (CISIA) of the Ministero dell’Economia e Finanze Italiano to the CNR (Legge n. 191/2009) and collaboration between Italian and Chinese laboratories has been favoured by a CNR/CAAS bilateral project (to B.N.).

**REFERENCES**


**Fig. 6.** Southern blot hybridization with a specific DIG-labelled DNA probe using total DNA preparations from representative symptomless [(a) lanes 1–9 and (b) lane 10] and symptomatic [(a) lane 10 and (b) lanes 1–9] mulberry trees. Arrows indicate hybridization signals of MMDaV DNA forms observed only in symptomatic plants, with the uppermost one likely designating the open circular and single stranded MMDaV DNA forms, which almost co-migrate under our loading conditions, and the lowest one the supercoiled MMDaV DNA form. Lane M, DL 10 000 DNA Marker (TAKARA), with sizes of DNA fragments indicated on one side. Ethidium bromide (EtBr) staining of the gel is shown below as loading control.


