Journal of General Virology (2015), **96**, 2411–2420 **DOI** 10.1099/vir.0.000173

Identification and characterization of a novel geminivirus with a monopartite genome infecting apple trees

Pengbo Liang,[1,2] Beatriz Navarro,[3] Zhixiang Zhang,[1] Hongqing Wang,[2] Meiguang Lu,[1] Hong Xiao,[1] Qingfa Wu,[4] Xueping Zhou,[1] Francesco Di Serio[3] and Shifang Li[1]

Correspondence
Shifang Li
sfli@ippcaas.cn
Francesco Di Serio
francesco.diserio@ipsp.cnr.it

[1]State Key Laboratory of 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

[2]College of Agronomy and Biotechnology, China Agricultural University, Yuanmingyuan West Road No. 2, Haidian District, Beijing 100193, PR China

[3]Istituto per la Protezione Sostenibile delle Piante, UOS Bari, Consiglio Nazionale delle Ricerche (IPSP-CNR), Via Amendola, 70126 Bari, Italy

[4]School of Life Sciences, University of Science and Technology of China, Hefei 230026, Anhui, PR China

Received 5 December 2014
Accepted 30 April 2015

A novel circular DNA virus sequence has been identified through next-generation sequencing and *in silico* assembly of small RNAs of 21–24 nt from an apple tree grown in China. The virus genome was cloned using two independent approaches and sequenced. With a size of 2932 nt, it showed the same genomic structure and conserved origin of replication reported for members of the family *Geminiviridae*. However, the low nucleotide and amino acid sequence identity with known geminiviruses indicated that it was a novel virus, for which the provisional name apple geminivirus (AGV) is proposed. Rolling circle amplification followed by RFLP analyses indicated that AGV was a virus with a monopartite DNA genome. This result was in line with bioassays showing that the cloned viral genome was infectious in several herbaceous plants (*Nicotiana bethamiana*, *Nicotiana glutinosa* and *Solanum lycopersicum*), thus confirming it was complete and biologically active, although no symptoms were observed in these experimental hosts. AGV genome structure and phylogenetic analyses did not support the inclusion of this novel species in any of the established genera in the family *Geminiviridae*. A survey of 165 apple trees grown in four Chinese provinces showed a prevalence of 7.2 % for AGV, confirming its presence in several cultivars and geographical areas in China, although no obvious relationship between virus infection and specific symptoms was found.

INTRODUCTION

Next-generation sequencing (NGS) of small RNAs (21–24 nt; sRNAs) accumulating in plant tissues is increasingly used for identifying previously reported and new viruses and viroids (Seguin *et al.*, 2014). Indeed, both viruses and viroids are targeted by the plant RNA-silencing machinery that, through the activity of Dicer-like proteins (DCLs), generates virus- and viroid-derived sRNAs (v-sRNA and vd-sRNA, respectively) of 21–24 nt, which are loaded into AGO (argonaute) proteins to guide the sequence-specific degradation or methylation of cognate RNAs and DNA, respectively (Carbonell *et al.*, 2012; Mallory & Vaucheret, 2010; Minoia *et al.*, 2014; Miozzi *et al.*, 2013; Navarro *et al.*, 2012). Both v-sRNAs and vd-sRNAs accumulating in infecting tissues are therefore molecular markers of ongoing infections by viral and subviral agents. After NGS, these sRNAs can be assembled *in silico* into larger contigs that in turn can be used for searching in databases for similar sequences, allowing identification of most viruses and viroids infecting a single plant, thus unveiling the plant virome (Coetzee

The GenBank/EMBL/DDBJ accession number for the complete genome sequence of apple geminivirus is KM386645.

Four supplementary tables and seven supplementary figures are available with the online Supplementary Material.

et al., 2010). NGS of sRNAs can also be successfully used for *de novo* assembly of genome fragments of DNA viruses, because the corresponding transcribed viral RNAs are targeted by the plant RNA-silencing machinery (Seguin *et al.*, 2014).

Geminiviruses are DNA viruses (2.5–3.6 kb in size) with a unique twinned particle morphology and with either a monopartite or a bipartite single-stranded circular genome. They are grouped in the family *Geminiviridae*, which includes seven genera (*Becurtovirus*, *Begomovirus*, *Curtovirus*, *Eragrovirus*, *Mastrevirus*, *Topocuvirus* and *Turncurtovirus*), established based on insect vector, host range and genomic characteristics of their members (Brown *et al.*, 2012; Fauquet *et al.*, 2008; Varsani *et al.*, 2014). Geminiviruses are the cause of devastating plant diseases, particularly in the tropics and subtropics (Moffat, 1999; Varma & Malathi, 2003). Although most geminiviruses infect herbaceous hosts, three geminiviruses have been reported in woody hosts, including citrus (associated with citrus chlorotic dwarf disease) (Loconsole *et al.*, 2012), grape (associated with grapevine red blotch and red leaf diseases) (Al Rwahnih *et al.*, 2013; Poojari *et al.*, 2013) and, more recently, *Jatropha multifida* (associated with foliar mosaic, distortion and necrosis in infected plants) (Polston *et al.*, 2014).

Apple (*Malus domestica* Borkh.) is a cultivated fruit crop worldwide, with China being the largest producer of apples in the world, both in terms of yield and planting area (http://faostat.fao.org/site/339/default.aspx). In an effort to determine the virome of an apple tree affected by scar skin disease, a fruit disorder known to be caused by apple scar skin viroid (ASSVd; Hashimoto & Koganezawa, 1987), other sequences ascribable to viruses and viroids in addition to ASSVd were identified. Here, we describe and discuss the molecular and biological properties of a new DNA virus from apple, provisionally named apple geminivirus (AGV).

RESULTS AND DISCUSSION

Viruses and viroids identified by NGS of the sRNA library from apple

NGS of sRNAs from the symptomatic apple sample generated 15776 701 raw reads. After trimming the adaptor and removing sequences matching with the apple genome (http://www.rosaceae.org/species/malus/malus_x_domestica/genome_v1.0), a total of 4208 333 short reads were obtained. Filtered reads were assembled by Velvet (Zerbino & Birney, 2008). BLASTN searches identified several contigs homologous to apple stem grooving virus (ASGV), apple stem pitting virus (ASPV), apple chlorotic ringspot virus (ACLSV) and ASSVd. Moreover, a novel small circular viroid-like RNA was identified using computational algorithms (Wu *et al.*, 2012; Zhang *et al.*, 2014). The actual presence of these infectious viral and subviral agents in the source plant was verified by reverse transcription (RT)-PCR and/or Northern blotting (Zhang *et al.*, 2014). In addition, four contigs with similarities to genomic regions of several members of the family *Geminiviridae* were also identified (Table S1, available in the online Supplementary Material), suggesting the presence of an additional virus with circular DNA genomic component(s) in the apple tree analysed.

Identification of a novel DNA virus in apple

Further confirmation of the presence of a circular geminivirus DNA in the apple tree tested was obtained by cloning and sequencing its full-length genome by two independent methods. The first approach relied on the use of two adjacent and complementary primers (1630-F/R; Table S2), designed according to the sequence of one contig (node 1630; Table S1), which would amplify a specific amplicon derived from a circular DNA template. When DNA preparation from the scar-skin-affected apple tree already selected for NGS of the sRNA library was tested by this primer pair, a product of ~3 kb was obtained (Fig. S1). By cloning and sequencing this amplicon, a sequence of 2932 nt in size was obtained (GenBank accession number KM386645; Fig. S2), which contained the four geminivirus-related contigs identified by NGS (Table S1).

The second approach was based on rolling circle amplification (RCA) followed by RFLP analyses (Haible *et al.*, 2006; Inoue-Nagata *et al.*, 2004), a non-sequence-specific method that allows amplification and identification of any circular DNA. RCA products were digested with several restriction enzymes that, based on the information from the first complete genome sequence, were predicted to cut at a single site. A single fragment with a size of ~3 kb was always obtained following digestion with all the selected enzymes, with no extra band ever detected (Fig. 1). These data suggested the

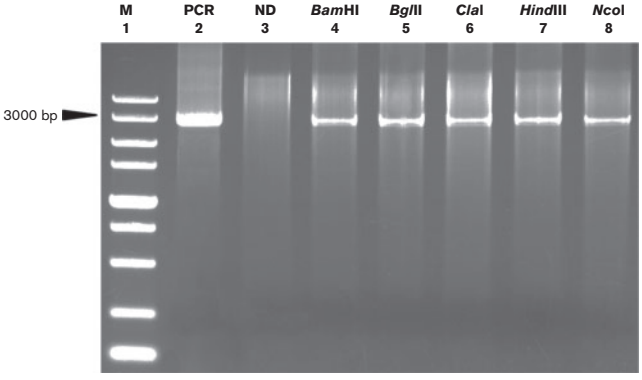

Fig. 1. Analysis of the RCA product obtained from AGV-infected apple. Lane 1 (M), DNA ladder DL5000 (TaKaRa). Lane 2, PCR product of AGV complete genome. Lane 3, the amplicon was not digested; amplicons were digested with *Bam*HI (lane 4), *Bgl*II (lane 5), *Cla*I (lane 6), *Hin*dIII (lane 7) and *Nco*I (lane 8), and loaded on 1% agarose gel stained with GelRed Nucleic Acid Gel Stain (Biotium).

absence of additional circular DNAs in the apple sample, besides the circular DNA sequenced by the PCR-based amplification method reported above. Moreover, cloning and sequencing of the amplicon obtained by RCA and digested by *Nco*I showed the same sequence of 2932 nt determined previously (Fig. S2). Attempts to obtain additional amplification products using universal primers designed to amplify β-satellites of geminiviruses (Briddon *et al.*, 2002) or degenerated primers designed to amplify fragments of DNA components from whitefly transmitted geminiviruses (Rojas *et al.*, 1993) were always unsuccessful. Altogether, the data derived from NGS, PCR and RCA experiments indicated that the novel virus identified in apple had a monopartite circular DNA genome, not accompanied by other DNA components.

Genome organization of AGV and phylogenetic analyses

Predictions by ORF Finder (http://www.ncbi.nlm.nih.gov/gorf/gorf.html) identified six putative ORFs in the AGV genome, two of which were coded on the 5′ half of the virion-sense strand and four on the 5′ half of the complementary-sense strand, resembling the genomic organization of monopartite geminiviruses (Fig. 2a, Table S3). In line with this and further results reported below, hereafter AGV ORFs are numbered and named according to Padidam *et al.* (1995). A nonanucleotide sequence (TAATATTAC, from position 2926 to position 2), highly conserved in almost all geminiviruses, was identified within the loop of the AGV intergenic region (Fig. 2b). Similarly to other members of the family *Geminiviridae*, the AGV IR contains TATA-boxes for the virion (position 178) and the complementary sense (position 2874) transcripts (Hanley-Bowdoin *et al.*, 1999), and an iterated DNA motif (iteron) (GGAG, at positions 2857 and 2864) (Argüello-Astorga & Ruiz-Medrano, 2001). However, the size of the AGV IR (398 nt) differs from that of most begomoviruses ($\sim$200–300 nt), resembling that of geminiviruses recently reported from woody plants, which is $>$300 nt (Table 1).

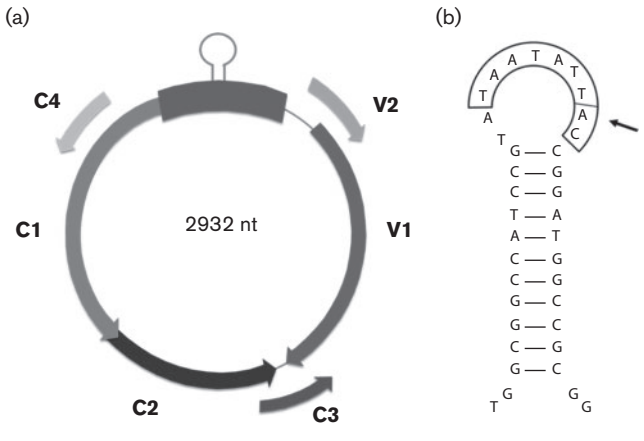

Fig. 2. (a) Schematic representation of the circular DNA genome of AGV with the relative position of the six predicted ORFs; the stem–loop in the IR is shown on the top. (b) Nonanucleotide sequence (TAATATTAC, from position 2926 to position 2), highly conserved in almost all geminiviruses, within the loop of the AGV IR. The arrow indicates position 1, which corresponds to the predicted replication origin of the viral DNA.

BLASTN searches using the full genome of AGV identified similarities with several members in the genus *Begomovirus*, the best scores always being in the C1/C4 ORFs. BLASTP analysis of the Rep protein (352 aa, encoded by C1 ORF) showed 74% amino acid sequence identity with Rep of East African cassava mosaic virus-Kenya (EACMV-KE, a begomovirus). The motifs FLTYP (motif I), HLH (motif II) and YIEKG (motif III), reported in other geminivirus Rep proteins and required for dsRNA binding, protein conformational changes and viral DNA cleavage (Nash *et al.*, 2011), were identified at aa 18, 60 and 106 of AGV Rep, respectively (Fig. 3). In addition, the GRS motif, composed of 12 consensus amino acids and possibly involved in rolling circle initiation in several geminiviruses (Nash *et al.*, 2011), was found between aa 77 and 99 of this AGV protein (Fig. 3), thus indicating that the latter had the typical features of a geminivirus Rep. With regard to the other AGV-encoded proteins, Pfam (Finn *et al.*, 2014) search detected conserved domains expected for geminiviral proteins, with the exception of AGV V2 (Table S4). In addition, the significance of Pfam analysis on V1 (putative coat protein, CP) was quite low (Table S4). In line with these data, no protein with similarity to AGV V2 was identified by BLASTP searches and in the case of AGV V1 only a very limited similarity (60/237 aa) with the CP of Euphorbia caput-medusae latent virus (EcmLV; Bernardo *et al.*, 2013) – a member of a new genus tentatively named *Capulavirus* – was obtained.

As previously reported for several begomoviruses (Zhou *et al.*, 1998), V1 and C3 ORFs of AGV overlap in 4 nt, including their respective stop codons. Pair-wise comparisons between AGV and closely related geminiviruses, including citrus chlorotic dwarf-associated virus, grapevine red-blotch-associated virus and Jatropha mosaic virus, which are geminiviruses infecting woody hosts, showed nucleotide sequence identity of $\sim$56 % when the full genomes were considered (Table 1). However, comparisons of individual encoded proteins highlighted an asymmetrical distribution of homology, with AGV C1 and C4 proteins showing the highest amino acid identity (54–75%) with several members of the genus *Begomovirus*. In contrast, the other proteins, and in particular those encoded by the virion-sense genes V1 and V2, showed extremely low identity ($\sim$14–23%) with the corresponding proteins of begomoviruses and other geminiviruses, thus indicating that AGV is only partially related to members of the genus *Begomovirus*.

This conclusion is further supported by the analyses of phylogenetic relationships between AGV and representative geminiviruses, selected to include (i) the type species of

Table 1. Nucleotide and amino acid sequence identities between AGV and other geminiviruses

Virus*	GenBank accession no.	Nucleotide sequence identity (%)		IR length (nt)	Amino acid sequence identity (%)					
		DNA	IR		V1	V2	C1	C2	C3	C4
TLCMV	AF327436	63.7	64.9	274	16.2	14.3	71.3	30.1	39.1	55.8
AYVV	KC810890	65.2	59.8	288	17.7	14.3	70.3	30.8	43.9	57.1
EACMV-KE	JF909175	64.4	58.2	330	19.3	10.2	73.9	35.3	40.9	72.7
TLCAV	AM701758	66.1	65.9	217	18.3	14.3	74.9	28.6	38.6	75.3
JaMV	KF998097	64.7	68.4	324	15.8	–	60.5	33.8	40.2	52.0
IYVV	EU839577	63.5	69.7	195	19.5	23.5	63.8	27.4	26.1	39.5
EcmLV	HF921477	59.4	62.1	388	20.4	16.3	33.5	19.2	–	19.5
GRBaV	JQ901105	56.7	56.4	453	21.2	19.4	25.0	14.0	16.0	–
CCDaV	NC_018151	58.2	57.8	413	16.7	23.5	35.7	–	–	–
TPCTV	NC_003825	61.9	62.1	294	18.3	18.4	63.4	23.9	35.3	52.0

*Geminiviruses with the best sequence matches (BLASTN and BLASTP) with AGV and those reported from other woody plants. Viruses belonging to the genus *Begomovirus* are in italics. TLCMV, tomato leaf curl Malaysia virus; AYVV, Ageratum yellow vein virus; EACMV-KE, East African cassava mosaic virus-Kenya; TLCAV, tomato leaf curl Anjouan virus; JaMV, Jatropha mosaic virus; IYVV, Ipomoea yellow vein virus; EcmLV, Euphorbia caput-medusae latent virus (a tentative capulavirus); GRBaV, grapevine red-blotch-associated virus; CCDaV, citrus chlorotic dwarf-associated virus; TPCTV, tomato pseudo-curly top virus (a topocuvirus).

each genera of the family *Geminiviridae*, (ii) other species representative of subgroups proposed in some genera of this family based on phylogenetic, biological and/or epidemiological features (Fauquet *et al.*, 2008; Varsani *et al.*, 2014), (iii) geminivirus species sharing a certain sequence identity with AGV (Table 1), and (iv) unclassified geminiviruses infecting woody plants (Al Rwahnih *et al.*, 2013; Loconsole *et al.*, 2012; Polston *et al.*, 2014). When a phylogenetic tree was reconstructed using the Rep protein encoded by these viruses, AGV was clustered together with other members of the genus *Begomovirus* (Fig. 4a). In contrast, when the phylogenetic tree was generated

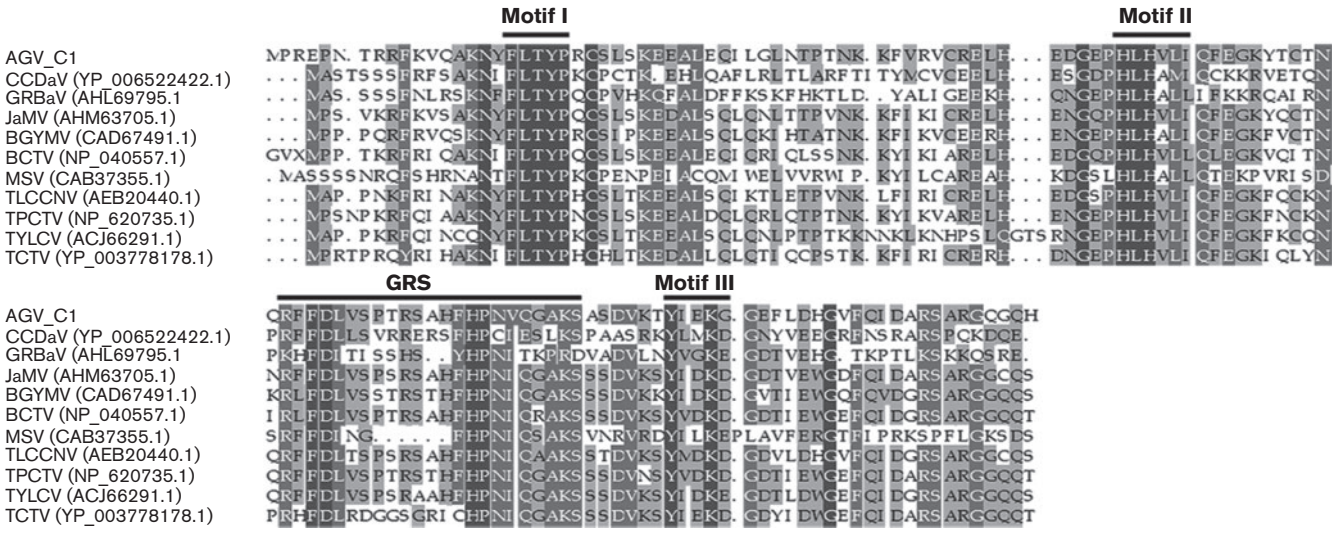

Fig. 3. Multiple alignment of the N-terminal portion of Rep protein sequences of AGV and other geminiviruses, highlighting conserved motifs I, II, III and GRS. The other geminiviruses are from woody hosts [citrus chlorotic dwarf-associated virus (CCDaV); grapevine red-blotch-associated virus (GRBaV); Jatropha mosaic virus (JaMV)], or are representative members of the genera *Begomovirus* [bean golden yellow mosaic virus (BGYMV); tomato leaf curl China virus (TLCCNV); tomato yellow leaf curl virus (TYLCV)], *Curtovirus* [beet curly top virus (BCTV)], *Topocuvirus* [tomato pseudo-curly top virus (TPCTV)], *Turncurtovirus* [turnip curly top virus (TCTV)] and *Mastrevirus* [maize streak virus (MSV)].

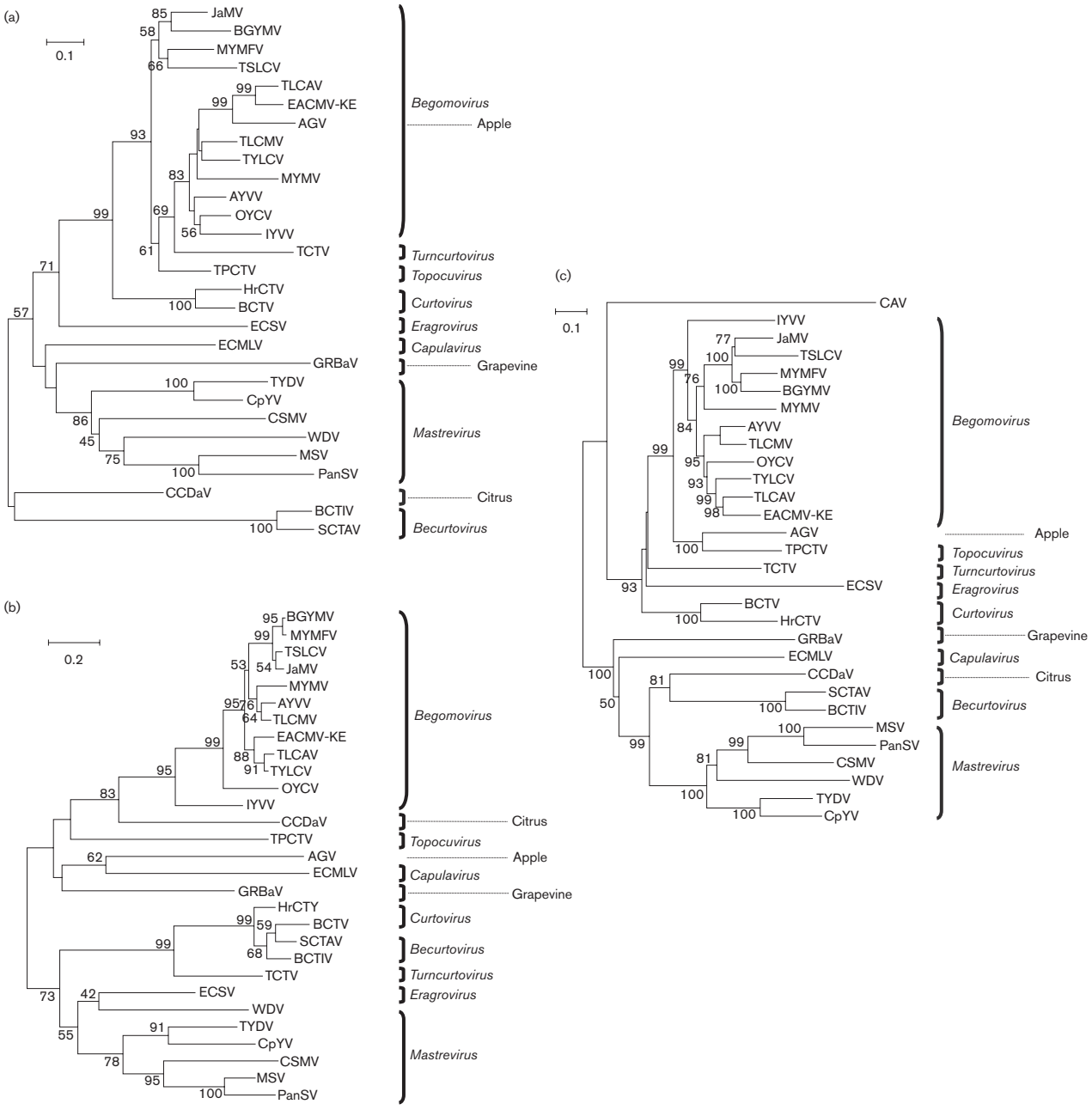

Fig. 4. Phylogenetic trees reconstructed by the neighbour-joining method from the alignment of the amino acid sequence of (a) Rep protein, (b) coat protein and (c) the full-genome nucleic acid sequence of AGV and representative geminiviruses using MEGA6. Bootstrap values (%) for 1000 replicates are indicated at the main branches, with those <40 % not reported. Branch length is proportional to number of nucleotide changes. Names of viruses and GenBank accession numbers are given either in Table 1 or below: bean golden yellow mosaic virus (BGYMV), DQ119824.1; beet curly top Iran virus (BCTIV), JQ707951.1; beet curly top virus (BCTV) M24597.2; chicken anemia virus (CAV), NC_001427.1; chickpea yellows virus (CpYV), JN989439; chlorosis striate mosaic virus (CSMV), M20021.1; Eragrostis curvula streak virus (ECSV), NC_012664.1; grapevine red-blotch-associated virus (GRBaV), JQ901105; horseradish curly top virus (HrCTV). U49907; Macroptilium yellow mosaic Florida virus (MYMFV), AY044135; maize streak virus (MSV), NC_001346.1; mungbean yellow mosaic virus (MYMV), AY271896.1; okra yellow crinkle virus (OYCV), HE793425.1; Panicum streak virus (PanSV), L39638.1; spinach curly top Arizona virus (SCTAV), HQ443515.1; tobacco yellow dwarf virus (TYDV), M81103; tomato pseudo-curly top virus (TPCTV), X84735.1; tomato severe leaf curl virus (TSLCV), AF130415.2; tomato yellow leaf curl virus (TYLCV), NC_004005.1; turnip curly top virus (TCTV), JQ742019.1; wheat dwarf virus (WDV), DQ868525.

using the CP protein, AGV was located outside the begomovirus cluster (Fig. 4b) and closer to other geminivirus species of still unassigned genera, including EcmLV, whose CP resulted in the unique match in the previous analyses based on BLASTP (see above). AGV also fell outside the begomovirus cluster when the phylogenetic tree was reconstructed using the full genomes of the same geminiviruses (Fig. 4c). In this case, AGV clustered together with tomato pseudo-curly top virus (a topocuvirus). The high bootstrap values observed at the nodes of phylogenetic trees support the significance of the observed clustering, thus supplying the first evidence for proposing a new genus in the family *Geminiviridae*, with AGV as the unique member so far.

Characterization of v-sRNAs derived from AGV

We further characterized the population of v-sRNAs derived from AGV. When the program SearchSmallRNA (de Andrade & Vaslin, 2014) was used for aligning non-redundant v-sRNAs with the AGV genomic DNA, the latter was covered on 97% of its length (Fig. S3a). v-sRNAs of 20–24 nt of both polarity strands accumulated at similar levels and, when grouped in size classes, a distribution profile with prevalent peaks of 21 and 22 nt was observed (Fig. S3b). Interestingly, v-sRNAs of 24 nt were also detected, in line with previous data from v-sRNAs accumulating in plant tissues infected by other geminiviruses, including cabbage leaf curl virus (CaLCuV) (Aregger *et al.*, 2012), tomato yellow leaf curl China virus (TYLCCNV) (Yang *et al.*, 2011) and tomato yellow leaf curl Sardinia virus (TYLCSV) (Miozzi *et al.*, 2013). Regarding AGV sRNAs biogenesis, altogether these data are consistent with the involvement of multiple DCLs, as suggested also for other geminiviruses infecting herbaceous hosts (Aregger *et al.*, 2012; Miozzi *et al.*, 2013; Yang *et al.*, 2011). Supporting this conclusion, apple genes coding for proteins homologous to *Arabidopsis* DCLs (DCL1–DCL4) have been annotated in GenBank and/or Phytozome databases (http://www.phytozome.net/apple.php).

When the 5′ terminus of each v-sRNA was mapped along the AGV genome, a distribution profile characterized by some prominent peaks was observed (Fig. S3c). v-sRNAs of both polarities mapped almost homogeneously in the coding regions of the AGV genome, with a minor density in the IR (Fig. S3c). The difference in DNA G+C content between coding regions and IRs (46.1 and 41.5%, respectively), could justify, at least partially, the limited accumulation of v-sRNAs in the AGV IR. Indeed, GC-rich regions were reported to be preferred for DCL activity (Ho *et al.*, 2007), with data supporting that a decrease of viral DNA G+C content from 50 to 42% may negatively affect DCL targeting (Ho *et al.*, 2010). A relatively similar higher accumulation of v-sRNAs in the coding regions and lower accumulation in the AU-rich IR was also reported previously in the case of CaLCuV, TYLCCNV and TYLCSV (Aregger *et al.*, 2012; Miozzi *et al.*, 2013; Yang *et al.*, 2011). Based on the accumulation and the distribution profiles of AGV sRNAs of both polarity strands along the viral genome, dsRNA substrate(s) seem(s) to play major role in their biogenesis, as previously proposed for other geminiviruses (Aregger *et al.*, 2012; Yang *et al.*, 2011).

Bioassays and field surveys

To test infectivity of AGV, *Agrobacterium tumefaciens* cells transformed with the expression vector pCAMBIA1305.1-1.9U, which contained a 1.9-mer tandem repeat of the AGV genome, were injected into the stems or petioles of herbaceous plant species, including *Nicotiana benthamiana*, *Nicotiana glutinosa*, *Nicotiana occidentalis*, *Solanum lycopersicum* and *Cucumis sativus*. At 25 days post-inoculation, although no obvious symptoms were observed, DNA preparations from the inoculated plants were first tested by PCR, using primer pair AG496F/R. Amplicons of the expected size (496 bp) were amplified from most agroinoculated *N. benthamiana*, *N.glutinosa* and *S. lycopersicum* (Fig. S4). Coincidental results were obtained by testing the same DNA preparations by Southern blot hybridization with an AGV-specific DIG-labelled probe, thus showing infection rates ranging from 73 to 80% in these three experimental hosts. In contrast, AGV was not detected, either by PCR or by Southern blot, in any of the inoculated *N. occidentalis* or *C. sativus* plants (Table 2).

Using Southern blot hybridization, the typical geminiviral DNA forms, i.e. open circular double-stranded (oc), supercoiled double-stranded (sc) and genomic single-stranded (ss), were observed in DNA preparations from *N. glutinosa*. In contrast, ocDNA and ssDNA forms were not clearly detected in apple and *S. lycopersicum* (Fig. 5). However, in some experiments in which higher quantities of total DNA from apple were loaded, a more complex pattern, with additional bands coincident with ocDNA and ssDNA forms from *N. benthamiana*, was observed (Fig. S5). We interpreted these results as indirect evidence of different accumulation of AGV DNA forms in some herbaceous and woody plants. However, considering results reported in other geminivirus–host combinations (Mason *et al.*, 2008), a bias due to the extraction method that could favour preferential recovery of some DNA forms cannot be excluded.

Table 2. Results of bioassays

Plant species	Infected/inoculated	Infection rate (%)*
N. benthamiana	24/30	80
N. glutinosa	11/15	73
S. lycopersicum	15/20	75
N. occidentalis	0/18	0
C. sativus	0/11	0

*Determined by PCR and Southern blot hybridization.

Table 3. Survey of virus and viroid infections in apple trees in four Chinese provinces

Province	AGV-infected/total	No. of infected trees (cultivar/symptoms)	Presence of other viruses and viroids
Shandong	6/74	1 (Shamaohaitang/leaf mosaic)	ASGV
		1 (Fenguozi/upward leaf curling)	ACLSV
		1 (Beni Shogun/scar skin on fruit)	ASPV, ASGV, ASSVd
		1 (Chaguozi/leaf mosaic)	ASGV
		1 (Gala/no symptoms)	ASGV
		1 (Changbasuan/no symptoms)	–
Shaanxi	3/9	1 (Gala/no symptoms)	ASGV
		1 (Gala/no symptoms)	ASGV, ASSVd
		1 (Fuji/no symptoms)	ASSVd
Jiangsu	3/30	3 (Fuji/no symptoms)	–
Liaoning	0/52	–	–

To investigate the incidence of AGV in the three Chinese provinces with major relevance in apple production (Liaoning, Shandong and Shaanxi) and in the neighbouring Jiangsu province, 165 samples were collected and first tested by PCR using the AGV-specific primer pair AG496F/R (Table S2). The expected amplicon of 496 bp was observed in 12 out of the 165 tested samples (Fig. S6a). When the PCR test was repeated using a different primer pair (AG664F/R; Table S2), the same samples tested positive (Fig. S6b). To further confirm these data, total DNA preparations from these PCR-positive samples and from 12 additional PCR-negative samples selected randomly were further assayed by Southern blot hybridization with an AGV-specific probe. AGV-specific hybridization signals were only detected in the 12 samples previously tested positive in the PCR assays (Fig. S7), thus showing conclusively that AGV infection incidence was 7.2% (Table 3). AGV was detected in several cultivars grown in different geographical areas in China, including provinces in which apple production is particularly relevant (Table 3). Notably, none of the samples originating from Liaoning province, one of the major areas for apple production in China, tested positive for AGV. We also tested, by RT-PCR, the same 12 AGV-positive samples for the presence of other viruses (ACLSV, ASPV and ASGV) and one viroid (ASSVd) known to be widespread in China (CABI, 2012; Ji *et al.*, 2013). At least four samples of two different cultivars (Fuji and Changbasuan) tested positive only to AGV, but no symptoms on leaves and fruits were observed (Table 3). These data seemed to indicate that AGV may infect apple trees without eliciting obvious symptoms, as also shown by bioassays in herbaceous hosts. However, symptoms on leaves or fruits were observed in samples simultaneously infected by AGV and at least one additional virus/viroid, thus raising the question of whether synergistic interactions may exist between the novel geminivirus and RNA viruses and/or viroids. In addition to giving a preliminary view on AGV spread in China, these experiments validated the PCR-based detection methods that will be useful for further investigating the epidemiology and pathogenicity of AGV.

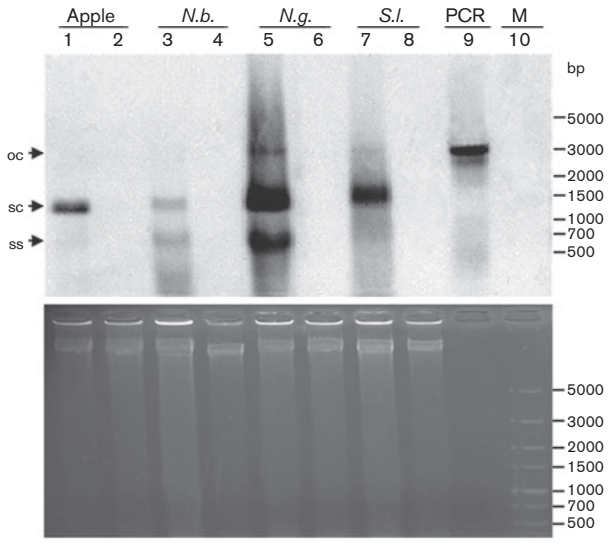

Fig. 5. Southern blot, hybridized with AGV-specific probe, of DNA preparations from AGV-infected and uninfected apple, *N. benthamiana* (*N.b.*), *N. glutinosa* (*N.g.*) and *S. lycopersicum* (*S.l.*). Lane 1, naturally infected apple; lanes 3, 5, and 7, laboratory-infected plants; lanes 2, 4, 6 and 8, healthy controls; lane 9, PCR-amplified AGV full-length genomic DNA. Lane 10 (M), DNA ladder DL5000 (TaKaRa). Arrowheads indicate the open-circular (oc), supercoiled (sc) and single-stranded (ss) DNA forms expected for AGV.

CONCLUSIONS

Only a few years ago, citrus and grapevine were reported to be the first woody plants naturally infected by different geminiviruses (Krenz *et al.*, 2012; Loconsole *et al.*, 2012), followed more recently by *J. multifida* (Polston *et al.*, 2014). In this work, using NGS and bioinformatics tools, we identified a novel geminivirus in apple trees, thus extending the natural hosts of this important virus family to another economically relevant woody crop. This finding

supports the use of NGS as a powerful technology for identifying new viruses, especially in woody plants.

Although no correlation of AGV with a specific apple disease was found in our survey, we showed that this virus already infects several apple cultivars and is present in Chinese provinces in which apple production is economically relevant. The incidence of the novel virus in apple fields appears limited and its spread could be still contained by testing the apple propagation material by the PCR detection methods developed here. Molecular and biological data, based on bioassays in herbaceous hosts, showed conclusively that AGV is a geminivirus with a monopartite genome. A conclusive statement on AGV pathogenicity in apple demands specific bioassays using virus-free apple seedlings. The availability of the infectious clone, developed here and tested on herbaceous hosts, allows us to address this issue in the near future. Several other biological properties of AGV need to be investigated, including the identification of its vector. Considering the very limited similarity of AGV CP with CPs encoded by other geminiviruses and the role of CP in the insect-mediated transmission (Briddon *et al.*, 1990), it is not possible at this stage to speculate on which insect vector could be responsible for AGV transmission in the field. Although bioassays with AGV infectious clones have shown that three different herbaceous species are experimental hosts of this virus, whether other natural host(s) of AGV may play a role in virus epidemiology is unknown. Answering this question is relevant for identifying potential risks posed by AGV to apple and other crops.

Comparisons of molecular features and potential expression strategies of the genome of AGV, which was cloned and sequenced by two independent methods, showed similarities with other geminiviruses, although sequence comparisons and phylogenetic analyses were unable to include AGV in any genus reported so far. The limited similarity with other members of the family may support the inclusion of AGV in a new genus in the family *Geminiviridae*.

METHODS

Plant materials. The sRNA library was generated using a bark sample from an apple tree (*Malus pumila* Mill. 'Fuji') with fruits showing scar skin symptoms and grown in Qixia city, Shandong province, China. For field surveys, leaf samples were collected from 165 apple trees (representative of 60 different cultivars) randomly chosen in the Shandong, Shaanxi, Liaoning and Jiangsu provinces (China) in 2014. Some of these trees produced fruits with typical symptoms of scar skin; others showed curling of leaves or were symptomless.

Construction, sequencing and analyses of the sRNA library. The RNA preparation for generating the cDNA library of sRNAs was obtained by an Illumina Small RNA Sample Preparation kit (Invitrogen) following the supplier's method. Protocols for sRNA purification, adaptor ligation, RT-PCR amplification, library purification and NGS on an Illumina HiSeq 2000 have been reported previously (Aliyari *et al.*, 2008; Wu *et al.*, 2012). The reads resulting from NGS were processed by removing the adaptor and then assembled *de novo* into larger contigs using Velvet Software 0.7.31 (Zerbino & Birney, 2008) with a *k*-mer of 17. The contigs obtained were subsequently screened for homologous sequences by BLASTN on the GenBank Virus Reference Sequence Database (ftp://ftp.ncbi.nlm.nih.gov/refseq/release/viral/). Coverage and distribution of viral-specific siRNAs were determined using the program SearchSmallRNA (de Andrade & Vaslin, 2014) and AGV (GenBank accession number KM386645) as reference virus sequence.

Total DNA extraction, cloning of full-length viral genome and sequence analyses. A cetyltrimethyl ammonium bromide (CTAB)-based extraction procedure (Murray & Thompson, 1980) was used to recover total DNA from 100 mg apple leaves. The full viral genome was determined by two independent methods. The first method was based on amplification of the full-length viral genome by a pair of adjacent primers named 1630-F/R (Table S2) and Phusion DNA polymerase (Thermo Scientific). The PCR product was cloned into pJET1.2 blunt cloning vector (Thermo Scientific); it was initially sequenced using the pJET1.2 universal sequencing primer and subsequently by genome walking using additional primers. The second cloning method was based on RCA performed using an Illustra TempliPhi 100 Amplification kit (GE Healthcare) according to the manufacturer's instructions, followed by digestion with the restriction enzyme *Nco*I, ligation within the pMDC32 vector (to generate the plasmid pMDC32-*Nco*I-AGV) and sequencing. Attempts to PCR-amplify additional genomic components possibly corresponding to DNA-B and β satellite DNA were made with the degenerate (DNA-B-F/R) and universal (β-F/R) primers reported by Rojas *et al.* (1993) and Briddon *et al.* (2002), respectively, and detailed in Table S2.

Analyses for sequences homologous to AGV were performed using BLASTN and BLASTP (Altschul *et al.*, 1990). ORFs were identified using ORF Finder (http://www.ncbi.nlm.nih.gov/gorf/gorf.html). Pfam (http://pfam.xfam.org/) database searching was used for identification of conserved protein domains (Finn *et al.*, 2014). Nucleotide and amino acid sequence alignments were generated with CLUSTAL W (Larkin *et al.*, 2007). Phylogenetic trees were reconstructed with the neighbour-joining method (Saitou & Nei, 1987) using 1000 bootstrap replicates in MEGA6 (Tamura *et al.*, 2013).

Detection of AGV by PCR and Southern blot. For detection purposes, two PCR assays were performed using separately two primer pairs (AG496F/R and AG664F/R) designed on two different AGV genomic regions (Table S2). The PCR protocol lasted for 30 cycles after an initial denaturation at 95 °C for 2 min. Each cycle consisted of denaturation at 95 °C for 30 s, primer annealing for 30 s at 53 °C and extension for 30 s at 72 °C for AG496F/R (40 s at 72 °C for AG664F/R), followed by an elongation step of 10 min at 72 °C. ACLSV, ASGV, ASPV and ASSVd were detected by RT-PCR using standard protocols and primers reported in Table S2.

Southern blot analysis was carried out by loading aliquots (10–20 μg) of DNA preparations, previously treated with RNase A, in 1% agarose gels and transferring the separated DNAs to Hybond-N + membranes (GE Healthcare) that were hybridized with the specific probe following the procedure previously (Faruk *et al.*, 2008). Prehybridization and hybridization steps were carried out at 45 °C in the buffer DIG Easy Hyb Version 11 (Roche). The AGV-specific DNA probe was labelled with DIG using a commercial PCR DIG Probe Synthesis kit (Roche), the primer pair Probe-F/R (Table S2) and the plasmid pMDC32-*Nco*I-AGV as template, according to the manufacturer's instructions.

Construction of the AGV infectious clone. Specific AGV primers *Hin*dIII-F and *Hin*dIII-R (Table S1) were designed to amplify a full-length fragment by PCR (Phusion DNA polymerase; Thermo Scientific). The PCR product was then inserted into the pJET1.2 vector (Thermo Scientific) to produce pJET1.2-1.0U. pJET1.2-1.0U was digested with *Hin*dIII and *Bam*HI (unique site in AGV genome),

and the expected 2692 bp fragment was introduced into the plant expression vector pCAMBIA1305.1, previously digested with the same enzymes, to produce pCAMBIA1305.1-0.9U. pJET1.2-1.0U was digested with *Hin*dIII to obtain a full-length unit of AGV DNA and inserted into the unique *Hin*dIII site of pCAMBIA1305.1-0.9U to produce pCAMBIA1305.1-1.9U, containing a 1.9-mer tandem repeat of the viral genome. All the amplified fragments used for the construction of infectious clones were sequenced entirely with an automated model 377 DNA sequencing system (Perkin Elmer), and sequence analyses indicated that no mutation was introduced into the clones.

Agroinoculation of plants and detection. *A. tumefaciens* cells (strain EHA105), transformed with pCAMBIA1305.1-1.9U, were grown at 28 °C for 36 h, centrifuged and resuspended in infiltration buffer (10 mM $MgCl_2$, 10 mM MES, pH 5.7, 150 μM acetosyringone). An aliquot of 0.2 ml culture (OD_{600} 1) was injected into the stems or petioles of herbaceous plants at the five-leaf stage. *N. benthamiana*, *N. glutinosa*, *N. occidentalis*, *S. lycopersicum* ('zhongzajiuhao') and *C. sativus* ('suyo') were agroinoculated and grown in an insect-free cabinet with supplementary lighting corresponding to a 16 h day length. Total DNA was extracted 25 days post-inoculation by the CTAB method (Murray & Thompson, 1980), and used for PCR analysis with primer pair AG496F/R (Table S1) and Southern blot following the protocol described above. Non-inoculated plants were used as negative controls.

ACKNOWLEDGEMENTS

We thank Xiuling Yang, Yuxin Ma and Fei Liu for help and advice during this research. We thank Huifeng Li (Shandong Institute of Pomology) for help with sample collection. 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 (201203076 and 200903004) and the National Natural Science Foundation of China (31471747 and 31371911). Work in the laboratories of B. N. and F. D. S. was partially supported by a dedicated grant (CISIA) of the Ministero dell'Economia e Finanze Italiano to the CNR (Legge 191/2009).

REFERENCES

Al Rwahnih, M., Dave, A., Anderson, M. M., Rowhani, A., Uyemoto, J. K. & Sudarshana, M. R. (2013). Association of a DNA virus with grapevines affected by red blotch disease in California. *Phytopathology* **103**, 1069–1076.

Aliyari, R., Wu, Q., Li, H. W., Wang, X. H., Li, F., Green, L. D., Han, C. S., Li, W. X. & Ding, S. W. (2008). Mechanism of induction and suppression of antiviral immunity directed by virus-derived small RNAs in *Drosophila*. *Cell Host Microbe* **4**, 387–397.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990). Basic local alignment search tool. *J Mol Biol* **215**, 403–410.

Aregger, M., Borah, B. K., Seguin, J., Rajeswaran, R., Gubaeva, E. G., Zvereva, A. S., Windels, D., Vazquez, F., Blevins, T. & other authors (2012). Primary and secondary siRNAs in geminivirus-induced gene silencing. *PLoS Pathog* **8**, e1002941.

Argüello-Astorga, G. R. & Ruiz-Medrano, R. (2001). An iteron-related domain is associated to Motif 1 in the replication proteins of geminiviruses: identification of potential interacting amino acid-base pairs by a comparative approach. *Arch Virol* **146**, 1465–1485.

Bernardo, P., Golden, M., Akram, M., Naimuddin, Nadarajan, N., Fernandez, E., Granier, M., Rebelo, A. G., Peterschmitt, M. & other authors (2013). Identification and characterisation of a highly divergent geminivirus: evolutionary and taxonomic implications. *Virus Res* **177**, 35–45.

Briddon, R. W., Pinner, M. S., Stanley, J. & Markham, P. G. (1990). Geminivirus coat protein gene replacement alters insect specificity. *Virology* **177**, 85–94.

Briddon, R. W., Bull, S. E., Mansoor, S., Amin, I. & Markham, P. G. (2002). Universal primers for the PCR-mediated amplification of DNA beta: a molecule associated with some monopartite begomoviruses. *Mol Biotechnol* **20**, 315–318.

Brown, J., Fauquet, C., Briddon, R., Zerbini, M., Moriones, E. & Navas-Castillo, J. (2012). Geminiviridae. In *Virus Taxonomy: Ninth Report of the International Committee on Taxonomy of Viruses*, pp. 351–373. Edited by A. M. Q. King, M. J. Adams, E. B. Carstens & E. J. Lefkowitz. London: Academic Press.

CABI (2012). Distribution maps of plant diseases, 2012, October, Map 1127 (edition 1). www.cabi.org/dmpd.

Carbonell, A., Fahlgren, N., Garcia-Ruiz, H., Gilbert, K. B., Montgomery, T. A., Nguyen, T., Cuperus, J. T. & Carrington, J. C. (2012). Functional analysis of three *Arabidopsis* ARGONAUTES using slicer-defective mutants. *Plant Cell* **24**, 3613–3629.

Coetzee, B., Freeborough, M. J., Maree, H. J., Celton, J. M., Rees, D. J. & Burger, J. T. (2010). Deep sequencing analysis of viruses infecting grapevines: virome of a vineyard. *Virology* **400**, 157–163.

de Andrade, R. R. & Vaslin, M. F. (2014). SearchSmallRNA: a graphical interface tool for the assemblage of viral genomes using small RNA libraries data. *Virol J* **11**, 45.

Faruk, M. I., Eusebio-Cope, A. & Suzuki, N. (2008). A host factor involved in hypovirus symptom expression in the chestnut blight fungus, *Cryphonectria parasitica*. *J Virol* **82**, 740–754.

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

Finn, R. D., Miller, B. L., Clements, J. & Bateman, A. (2014). iPfam: a database of protein family and domain interactions found in the Protein Data Bank. *Nucleic Acids Res* **42** (D1), D364–D373.

Haible, D., Kober, S. & Jeske, H. (2006). Rolling circle amplification revolutionizes diagnosis and genomics of geminiviruses. *J Virol Methods* **135**, 9–16.

Hanley-Bowdoin, L., Settlage, S., Orozco, B., Nagar, S. & Robertson, D. (1999). Geminiviruses: models for plant DNA replication, transcription, and cell cycle regulation. *Crit Rev Plant Sci* **18**, 71–106.

Hashimoto, J. & Koganezawa, H. (1987). Nucleotide sequence and secondary structure of apple scar skin viroid. *Nucleic Acids Res* **15**, 7045–7052.

Ho, T., Wang, H., Pallett, D. & Dalmay, T. (2007). Evidence for targeting common siRNA hotspots and GC preference by plant Dicer-like proteins. *FEBS Lett* **581**, 3267–3272.

Ho, T., Wang, L., Huang, L., Li, Z., Pallett, D. W., Dalmay, T., Ohshima, K., Walsh, J. A. & Wang, H. (2010). Nucleotide bias of DCL and AGO in plant anti-virus gene silencing. *Protein Cell* **1**, 847–858.

Inoue-Nagata, A. K., Albuquerque, L. C., Rocha, W. B. & Nagata, T. (2004). A simple method for cloning the complete begomovirus genome using the bacteriophage phi29 DNA polymerase. *J Virol Methods* **116**, 209–211.

Ji, Z., Zhao, X., Duan, H., Hu, T., Wang, S., Wang, Y. & Cao, K. (2013). Multiplex RT-PCR detection and distribution of four apple viruses in China. *Acta Virol* **57**, 435–441.

Krenz, B., Thompson, J. R., Fuchs, M. & Perry, K. L. (2012). Complete genome sequence of a new circular DNA virus from grapevine. *J Virol* **86**, 7715.

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A. & other authors (2007). CLUSTAL W and CLUSTAL X version 2.0. *Bioinformatics* **23**, 2947–2948.

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**, 162–172.

Mallory, A. & Vaucheret, H. (2010). Form, function, and regulation of ARGONAUTE proteins. *Plant Cell* **22**, 3879–3889.

Mason, G., Caciagli, P., Accotto, G. P. & Noris, E. (2008). Real-time PCR for the quantitation of *Tomato yellow leaf curl Sardinia virus* in tomato plants and in *Bemisia tabaci*. *J Virol Methods* **147**, 282–289.

Minoia, S., Carbonell, A., Di Serio, F., Gisel, A., Carrington, J. C., Navarro, B. & Flores, R. (2014). Specific argonautes selectively bind small RNAs derived from potato spindle tuber viroid and attenuate viroid accumulation *in vivo*. *J Virol* **88**, 11933–11945.

Miozzi, L., Pantaleo, V., Burgyán, J., Accotto, G. P. & Noris, E. (2013). Analysis of small RNAs derived from tomato yellow leaf curl Sardinia virus reveals a cross reaction between the major viral hotspot and the plant host genome. *Virus Res* **178**, 287–296.

Moffat, A. (1999). Geminiviruses emerge as serious crop threat. *Science* **286**, 1835.

Murray, M. G. & Thompson, W. F. (1980). Rapid isolation of high molecular weight plant DNA. *Nucleic Acids Res* **8**, 4321–4326.

Nash, T. E., Dallas, M. B., Reyes, M. I., Buhrman, G. K., Ascencio-Ibañez, J. T. & Hanley-Bowdoin, L. (2011). Functional analysis of a novel motif conserved across geminivirus Rep proteins. *J Virol* **85**, 1182–1192.

Navarro, B., Gisel, A., Rodio, M. E., Delgado, S., Flores, R. & Di Serio, F. (2012). Small RNAs containing the pathogenic determinant of a chloroplast-replicating viroid guide the degradation of a host mRNA as predicted by RNA silencing. *Plant J* **70**, 991–1003.

Padidam, M., Beachy, R. N. & Fauquet, C. M. (1995). Classification and identification of geminiviruses using sequence comparisons. *J Gen Virol* **76**, 249–263.

Polston, J. E., Londoño, M. A. & Capobianco, H. (2014). The complete genome sequence of New World jatropha mosaic virus. *Arch Virol* **159**, 3131–3136.

Poojari, S., Alabi, O. J., Fofanov, V. Y. & Naidu, R. A. (2013). A leafhopper-transmissible DNA virus with novel evolutionary lineage in the family *Geminiviridae* implicated in grapevine redleaf disease by next-generation sequencing. *PLoS One* **8**, e64194.

Rojas, M., Gilbertson, R., Russel, D. & Maxwell, D. (1993). Use of degenarate primers in polymerase chain reaction to detect whitefly-transmitted geminiviruses. *Plant Dis* **77**, 340–347.

Saitou, N. & Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* **4**, 406–425.

Seguin, J., Rajeswaran, R., Malpica-López, N., Martin, R. R., Kasschau, K., Dolja, V. V., Otten, P., Farinelli, L. & Pooggin, M. M. (2014). *De novo* reconstruction of consensus master genomes of plant RNA and DNA viruses from siRNAs. *PLoS One* **9**, e88513.

Tamura, K., Stecher, G., Peterson, D., Filipski, A. & Kumar, S. (2013). mega6: molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol* **30**, 2725–2729.

Varma, A. & Malathi, V. G. (2003). Emerging geminivirus problems: a serious threat to crop production. *Ann Appl Biol* **142**, 145–164.

Varsani, A., Navas-Castillo, J., Moriones, E., Hernández-Zepeda, C., Idris, A., Brown, J. K., Murilo Zerbini, F. & Martin, D. P. (2014). Establishment of three new genera in the family *Geminiviridae*: *Becurtovirus*, *Eragrovirus* and *Turncurtovirus*. *Arch Virol* **159**, 2193–2203.

Wu, Q., Wang, Y., Cao, M., Pantaleo, V., Burgyan, J., Li, W. X. & Ding, S. W. (2012). Homology-independent discovery of replicating pathogenic circular RNAs by deep sequencing and a new computational algorithm. *Proc Natl Acad Sci U S A* **109**, 3938–3943.

Yang, X., Wang, Y., Guo, W., Xie, Y., Xie, Q., Fan, L. & Zhou, X. (2011). Characterization of small interfering RNAs derived from the geminivirus/betasatellite complex using deep sequencing. *PLoS One* **6**, e16928.

Zerbino, D. R. & Birney, E. (2008). Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs. *Genome Res* **18**, 821–829.

Zhang, Z., Qi, S., Tang, N., Zhang, X., Chen, S., Zhu, P., Ma, L., Cheng, J., Xu, Y. & other authors (2014). Discovery of replicating circular RNAs by RNA-seq and computational algorithms. *PLoS Pathog* **10**, e1004553.

Zhou, X., Liu, Y., Robinson, D. J. & Harrison, B. D. (1998). Four DNA-A variants among Pakistani isolates of cotton leaf curl virus and their affinities to DNA-A of geminivirus isolates from okra. *J Gen Virol* **79**, 915–923.