Virus Research

Characterization of burdock mottle virus, a novel member of the genus *Benyvirus*, and the identification of benyvirus-related sequences in the plant and insect genomes

Hideki Kondo[1], Shuichi Hirano[1], Sotaro Chiba[1], Ida Bagus Andika[2], Makoto Hirai[3], Takanori Maeda[4] and Tetsuo Tamada[1,†]

(1) Institute of Plant Science and Resources (IPSR), Okayama University, 2-20-1, Chuo, Kurashiki 710-0046, Japan
(2) Institute of Virology and Biotechnology, Zhejiang Academy of Agricultural Sciences, Hangzhou 310021, P. R. China
(3) Department of Parasitology, Graduate School of Medicine, Gunma University, Maebashi 371-8511, Japan
(4) Formerly College of Bioresource Sciences, Nihon University, 1866 Kameino, Fujisawa, Kanagawa 252-8510, Japan

†Current address for T. Tamada: Kita 10, Nishi 1, 13-2-606, Sapporo, 001-0010, Japan.

Corresponding authors**;**
H. Kondo & T. Tamada
E-mail: hkondo@rib.okayama-u.ac.jp; tyrotamada@oboe.ocn.ne.jp
Tel: +81-86-434-1232 / Fax: +81-86-434-1249

Word counts: 237 words (Abstract); 7113 words (Main Text)
Table:1; Figures: 7; Supplementary Table: 2; Supplementary Figure: 4
Figures that will appear in color reproduction on the Web: 4 (Figures 1, 3, 5 and 7)

Running title: A new member of the genus *Benyvirus* and its relative sequences in eukaryotic genomes

ABSTRACT

The complete nucleotide sequence of the burdock mottle virus (BdMoV) isolated from an edible burdock plant (*Arctium lappa*) in Japan has been determined. BdMoV has a bipartite genome, whose organization is similar to RNA1 and RNA2 of benyviruses, beet necrotic yellow vein virus (BNYVV), beet soil-borne mosaic virus (BSBMV), and rice stripe necrosis virus (RSNV). BdMoV RNA1 (7,038 nt) contains a single open reading frame (ORF) encoding a 249-kDa polypeptide that consists of methyl-transferase, helicase, papain-like protease, AlkB-like, and RNA-dependent RNA polymerase domains. The AlkB-like domain sequence is not present in the proteins encoded by other known benyviruses, but is found in replication-associated proteins of viruses mainly belonging to the families *Alfaflexiviridae* and *Betaflexiviridae*. BdMoV RNA2 (4,315 nt) contains six ORFs that are similar to those of benyviruses: these are coat protein (CP), CP readthrough, triple gene block movement and cysteine-rich proteins. Phylogenetic analyses showed that BdMoV is more closely related to BNYVV and BSBMV than to RSNV. Database searches showed that benyvirus replicase-related sequences are present in the chromosomes of a chickpea plant (*Cicer arietinum*) and a blood-sucking insect (*Rhodnius prolixus*). Some other benyvirus-related sequences are found in the transcriptome shotgun libraries of a few species of plants and a bark beetle. Our results show that BdMoV is a distinct species of the genus *Benyvirus* and that ancestral and extant benyviruses may have infected or currently infect a wide range of hosts, including plants and insects.

Keywords: *Benyvirus*; Burdock mottle virus; AlkB, Paleovirology; Endogenous viral element; Transcriptome shotgun assembly

1. Introduction

Burdock mottle virus (BdMoV) was first isolated from edible burdock plants (*Arctium lappa* L.) at Soja City, Okayama Prefecture, Japan in 1970 (Inouye, 1973). Similar viruses have been found in several prefectures (i.e., Saitama, Kanagawa, Akita and Tokyo) (Yamashita et al., 2008), but it have not been reported outside of Japan. Burdock is one of the domestic root vegetable crops of Japan and has been grown for a long time.

A. lappa is a biennial plant belonging to the *Asteraceae* (*Compositae*) family and native to the Eurosiberian region and subcosmopolitan in distribution (Lopez-Vinyallonga et al., 2011). This and allied plants also have been cultivated as a traditional medicinal herb in China and other countries (Chan et al., 2011).

Burdock is the only known natural host of BdMoV. The virus usually causes mild, chlorosis or mottling symptoms on the leaves of burdock plants. It is transmissible by mechanical inoculation of sap to several plant species, but its natural vector is unknown (Inouye, 1973). BdMoV has straight rod-shaped particles, about 250 nm (a major peak) and 380 nm (a minor peak) in length and about 17 nm in width, which are similar to the particles of viruses in the genus *Benyvirus* as well as the family *Virgaviridae*. BdMoV induces the formation of spherical electron-dense, viroplasm-like inclusion bodies in the cytoplasm of cells (Inouye, 1973; Yamashita et al., 2008). Our molecular study has shown that the genome of BdMoV is similar to that of beet necrotic yellow vein virus (BNYVV) (Hirano et al., 1999), and therefore BdMoV is a tentative species of the genus *Benyvirus* (Gilmer and Ratti, 2012; Koenig and Lesemann, 2005; Rush, 2003; Tamada, 1999; Tamada and Kondo, 2013).

Benyviruses have polyadenylated, segmented plus-sense RNA genomes (2~5). Most have two segments, RNA1 and RNA2, which carry important house-keeping genes required for replication, encapsidation/assembly, viral movement/transmission and RNA silencing suppression, and satellite-like RNAs depending on virus strains, which play important roles in disease development and spread in nature (Gilmer and Ratti, 2012; Heidel et al., 1997; Lee et al., 2001). At present, the genus *Benyvirus* includes three species, *Beet necrotic yellow vein virus* (type member), *Beet soil-borne mosaic virus* and *Rice stripe necrosis virus* (Gilmer and Ratti, 2012). Taxonomically, the benyvirus differs from rod-shaped viruses belonging to the family *Virgaviridae* in replicase phylogeny, genome organization and expression strategies (Adams et al., 2009; Tamada, 1999). For example, the replicase-associated protein of benyviruses shows greater similarity to that of the viruses in the families *Togaviridae* and *Hepeviridae* than to those of other rod-shaped plant viruses (Koonin et al., 1992). Most recently, an official taxonomic proposal (TaxoProp) was submitted by Gilmer et al. (2013) to the International Committee on the Taxonomy of Viruses (ICTV) requesting the creation of

the new family *Benyviridae* that contains a new species *Burdock mottle virus* in the genus *Benyvirus.*

BNYVV has spread to most of sugar beet growing areas worldwide since first found in Italy in the 1950s, but its origin is suggested to be in East Asia (Chiba et al., 2011a). Beet soil-borne mosaic virus (BSBMV) occurs in almost all the major sugar beet growing regions of the United States (Rush, 2003). Both viruses have similarly limited host ranges and are vectored by *Polymyxa betae* (Plasmodiophoraceae, a fungoid protist) (Rush, 2003). Rice stripe necrosis virus (RSNV) occurs in the Ivory Coast in Africa (Louvel and Bidaux, 1977) and in South and Central America (Morales et al., 1999). RSNV is transmitted by *Polymyxa graminis* (Morales et al., 1999). Furthermore, two benyvirus-related rod-shaped viruses are known; Nicotiana velutina mosaic virus (NVMV) that was isolated from *Nicotiana velutina* plants in Australia having a bipartite genome (Randles and Rohde, 1990) and Chara australis virus (CAV) that was isolated from fresh water algae (*Chara australis*) in Australia having a monopartite genome, whose replicase is most closely related to that of benyviruses (Gibbs et al., 2011). This implies wide distribution of benyvirus-like species in various eukaryotic organisms.

Recently, the presence of several non-retroviral RNA virus sequences (NRVSs, syn. endogenous virus elements) related to plus-sense RNA viruses in host genomes. This includes some members of three plant virus genera *Citrivirus* (family *Betaflexiviridae*), *Tobamovirus* (family *Virigaviridae*), *Potyvirus* (family *Potyviridae*) and *Cilevirus* (unassigned family), one insect virus genus *Flavivirus* (family *Flaviviridae*) (Chiba et al., 2011b; Crochu et al., 2004; Cui and Holmes, 2012; Katzourakis and Gifford, 2010; Tanne and Sela, 2005). Examples of NRSVs endogenization extend to minus-sense and double-stranded RNA viruses (Chiba et al., 2011b; Horie et al., 2010; Liu et al., 2010; Katzourakis and Gifford, 2010; Kondo et al., 2013; Taylor and Bruenn, 2009). It is anticipated that a similar approach allows detection of benyvirus-like sequences in eukaryotes not reported to be their hosts.

In this study, we compared the complete nucleotide sequence of BdMoV to genomes of other viruses. Our analyses suggest that BdMoV is a new member of the *Benyvirus* genus with a bipartite genome. Bioinformatic analyses led to the identification of

benyvirus replicase-related sequences in plant and insect chromosomes, which suggested that the ancient benyvirus sequences have been integrated as endogenous virus-like elements into their genomes. We also found some other benyvirus-related sequences in the transcriptome shotgun libraries from plants and insects. These data suggest that benyviruses and their related viruses may have been distributed in a wide range of hosts since ancient times.

2. Materials and methods

2.1. Virus isolate, propagation and inoculations

The S isolate of BdMoV originated from burdock leaves that were collected from a field in Soja, near Kurashiki, in 1970 (Inouye, 1973). This virus was propagated in *Chenopodium quinoa* leaves in glasshouses. Locally infected leaves of *C. quinoa* were harvested approximately 1−2 weeks after inoculation and stored at −80 °C until used.

Foliar rub-inoculation and root vector inoculation using *P. betae* were conducted as described previously (Rahim et al., 2007; Tamada, 2007).

2.2. Virus purification and electron microscopy

Virus particles were purified from mechanically infected leaves of *C. quinoa*. The virus-infected leaves were homogenized in 0.5 M borate–HCl (pH 8.0), containing 0.1% thioglycolic acid, 1 mM phenylmethylsulfonylfluoride, 20 mM sodium iodoacetate, and 20 mM Na_2-EDTA. The extract was filtrated through cheesecloth and clarified by 8.5% (w/v) *n*-butanol. After low speed centrifugation, 1% Triton X-100 (w/v) was added to the supernatant, and the virus was precipitated by adding 4% PEG6000. The preparation was resuspended in 0.5 M borate buffer (pH 8.0) and concentrated by ultracentrifugation. The resulting pellets were resuspended in 0.01 M phosphate buffer (pH 7.0) and further purified through a sucrose-cesium chloride gradient. The light-scattering bands were collected from the gradient and pelleted by ultracentrifugation. To obtain electron micrograph of BdMoV virions, purified preparations were negatively stained with 2% uranyl acetate and observed in a Hitachi

model H-7100 transmission electron microscope (Hitachi, Tokyo, Japan).

2.3. Electrophoretic analysis of viral RNA

Viral RNA extraction from the purified virion and poly(A)-positive RNA fractionation were conducted as described previously (Andika et al., 2006; Kondo et al., 2009). Viral RNAs were separated on 1.4% agarose horizontal submarine gels in MOPS/EDTA buffer (pH 7.0) containing 0.22 M formaldehyde. Virion RNAs from Dendrobium mosaic virus-Japanese strain (DeMV-J, an unassigned potyvirus) were used for a poly(A)-positive RNA control and from cucumber mosaic virus Y strain (CMV-Y, a cucumovirus) and tobacco mosaic virus OM strain (TMV-OM, a tobamovirus) for poly(A)-negative controls.

2.4. SDS-PAGE and Western blot analyses

SDS-PAGE and Western blot analyses were conducted as described previously (Kondo et al., 2009). A rabbit antisera to BdMoV virions was used for immunodetection (Maeda et al., 1996).

2.5. cDNA synthesis and cloning

BdMoV cDNAs were synthesized using cDNA Synthesis Module (Amersham Pharmacia Biotech, Buckinghamshire, UK) primed with random hexanucleotides or a synthetic oligonucleotide $(dT)_{18}$ with a *Not*I site . The double-stranded cDNAs were inserted into the *Eco*RV site of pZErO^{TM}-2 (Invitrogen, San Diego, CA, U.S.A.). The plasmids were introduced into competent *Escherichia coli* strain TOP 10F' (Invitrogen). For cloning of the 5' terminal regions of both RNA1 and RNA2, 5' RACE (rapid amplification of cDNA ends) was performed according to the methods of Kondo et al. (2006). The sequences of the primers used in the BdMoV genome analysis are available upon request.

2.6. Plant and insect materials and PCR amplification

The blood-sucking bug (*Rhodnius prolixus*) individuals, which were kindly provided by H. Kanuka of Jikei University School of Medicine (Tokyo, Japan), and the commercially available seeds of chickpeas (*Cicer arietinum* cv. unknown) were used. Total genomic DNA was purified using the DNeasy ® Blood and Tissue Kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. Purified DNA solutions (TE buffer) were stored at 4 °C until use. To amplify the virus-like sequence fragments from these DNA samples by PCR, primer pairs were designed based on the virus-related sequences and their flanking sequences (Supplementary Table S1 and data not shown).

2.7. Nucleotide sequencing and sequence analysis

Nucleotide sequences of selected clones were determined by dideoxy chain termination using Applied Biosystems 377 or 3100 DNA sequencer (Applied Biosystems, Foster City, CA). Sequence data were assembled and analyzed by Auto Assembler^{TM} DNA Sequence Assembly Software (Applied Biosystems), and GENETYX-MAC/ATSQ (GENETYX co., Tokyo, Japan) and Enzyme X version 3 (http://nucleobytes.com/index.php/enzymex). Sequence similarities among viruses were searched using the BLAST program available from the National Center for Biotechnology Information (NCBI).

BLAST (tblastn) searches were conducted against genome sequence databases available from the NCBI (nucleotide collection, nr/nt; genome survey sequences, GSS; high-throughput genomic sequences, HTGS; whole-genome shotgun contigs, WGS; non-human, non-mouse expressed sequence tags, EST; transcriptome shotgun assembly, TSA, others). For the searches, we used BdMoV and other benyviruses as queries. Plant and insect genome sequences that matched viral peptides with e-values of < 0.01 were extracted. Transposable element sequences were detected using the Censor provided by Genetic Information Research Institute, USA (http://www.girinst.org/censor/index.php) (Kohany et al., 2006).

Phylogenetic tree construction was based on a maximum-likelihood (ML) method as described previously with minor modification (Kondo et al., 2013). Multiple alignments

of amino acid (aa) sequences were conducted using MAFFT ver 7 under the default parameters (http://mafft.cbrc.jp/alignment/server/phylogeny.html) (Katoh and Toh, 2008). The program GapStreeze was used to remove columns that contained gaps from the alignment (Los Alamos HIV Sequence Database; http://www.hiv.lanl.gov/content/hiv-db/GAPSTREEZE/gap.html). Selection of the best-fit model for a data set was performed using ProtTest 2.4 (Abascal et al., 2005) based on the Akaike information criterion (AIC). ML trees were constructed by PhyML 3.0 using the appropriate substitution mode (http://www.phylogeny.fr/) (Guindon et al., 2010). The branch support values were estimated by the approximate likelihood ratio test (aLRT) with a Shimodaira–Hasegawa–like (SH-like) algorithm (Anisimova and Gascuel, 2006). The obtained trees were visualized using FigTree v1.3.1 software (http://tree.bio.ed.ac.uk/software/).

3. Results

3.1. Biological and chemical properties of BdMoV

Inouye (1973) reported that BdMoV could infect 10 of 40 species from 10 families; systemic infection in burdock, *C. quinoa*, *C. murale, N. clevelandii* and *N. rustica*, and local infection in *C. amaranticolor, Beta vulgaris, Spinacia oleracea, Cucumis sativus* and *Tetragonia expansa.* Our additional inoculation tests showed that *N. benthamiana* is a systemic host, while *Gomphrena globosa* is a local lesion host. BdMoV also produced only local chlorotic lesions (no systemic infection) on *B. macrocarpa* and *B. vulgaris* subsp. *maritima* M8 plants (data not shown) that are the systemic hosts of BNYVV (Tamada, 2007). BdMoV-infected *N. benthamiana* plants showed mild mosaic or mottle symptoms with slight leaf distortion (data not shown), which is similar to those caused by BNYVV (Rahim et al., 2007).

We attempted to allow *P. betae* to acquire BdMoV from the roots of *N. benthamiana* plants systemically infected by foliar rub-inoculation and to transmit to healthy *N. benthamiana* and *C. quinoa* plants. As a result, *P. betae* failed to transmit the virus to these plants (data not shown). In contrast, *P. betae* could transmit BNYVV, a positive control, from virus-infected *N. benthamiana* roots to healthy counterparts.

BdMoV particles are about 17 nm in diameter, with two predominant lengths of 250 nm and 380 nm (Inouye, 1973). Purified preparations contained straight rod-shaped about 250-nm long particles (Fig. 1A, asterisks) and mostly aggregated end-to-end (Fig. 1A). The purified BdMoV particles were transmitted by leaf-rub inoculation, and induced a mild systemic mottle symptom in burdock and typical chlorotic spots on *C. quinoa*-inoculated leaves, respectively (Figs. 1B and C). Electrophoresis of RNAs prepared from the purified particles showed two species of single-stranded RNA, designated RNA1 (7 kb) and RNA2 (4 kb) (Fig. 1D, left panel, BdMoV and data not shown). Both RNA1 and 2 were found in the poly(A)-positive fraction, which suggests that BdMoV RNAs have a poly(A) tail, as the positive control, DeMV-J (Fig. 1D, right panel, BdMoV and DeMV-J). In SDS-PAGE from purified virus particles, a single major polypeptide (CP) of approximately 23 kDa was detected (Fig. 1E, left panel, Virions). Western blot analysis showed that the CP protein (from the purified preparation) strongly reacted with an antiserum to the BdMoV virion (Fig. 1E, right panel, Virions). CP was detected in protein extracts from BdMoV-infected *C. quinoa* leaves, but not in those from healthy leaves (Fig. 1E, right panel, Infected and Mock). It is noted that a protein with higher molecular weight (HMW, ~60 kDa) reacted with the BdMoV antiserum (Fig. 1E, right panel, BdMoV-Infected).

3.2. Nucleotide sequence of the BdMoV genome

Two independent sequence contigs (RNA1 and RNA2) were obtained from the assembly of a random cDNA library of BdMoV RNAs (data not shown). Their 5′ termini were determined by 5′ RACE. The total number of nucleotide of BdMoV RNA1 and RNA2 is 7,038 nt (Genebank accession AB818898) and 4,315 nt (Genebank accession AB818899), respectively excluding the 3′-terminal poly(A) tail. Computer-assisted sequence analyses identified that there were only large open reading frames (ORFs) in 5′-to-3′ orientation (positive strand) of the RNA1 and RNA2 cDNAs (Fig. 2 and see below for details). The BdMoV RNA1 sequence contains one large ORF encoding a polypeptide of 249 kDa (replicase; p249) (Fig. 2). The 5' and 3' non-coding regions of RNA1 are 151 and 233 nt, respectively excluding the 3'-poly(A). BdMoV RNA2 contains six ORFs and the 5' and 3' non-coding regions of RNA2 are 147 and

227 nt, respectively (Fig. 2). The first 8 nt (AAAUUCAU) at the 5'-terminal sequence of both segments are identical (Supplementary Fig. S1A). The 3'-terminal of two segments share high similarity (85% identical, Supplementary Fig. S1B). In particular, both the 3'-proximal 70 nt of BdMoV and BNYVV RNAs can be folded into a possibly similar stem-loop structure (data not shown) (Richards and Tamada, 1992).

3.3. RNA1 encoded protein

BdMoV RNA1 encodes a single protein with an estimated molecular weight of 249-kDa and a putative replicase function (Fig. 2). This replicase-associated protein contains five domains ; methyltransferase (MTR), RNA helicase (HEL), papain-like proteinase (Pro), AlkB-like protein (AlkB) and RNA-dependent RNA polymerase (RdRp) (Fig. 2, Supplementary Fig. S2A). Except for the AlkB-like, those domains are similarly found in benyvirus replicases (see below): BNYVV p237, BSBMV 239 kDa and RSNV 236 kDa proteins (Bouzoubaa et al., 1987; Lee et al., 2001; Lozano and Morales, 2009; Saito et al., 1996). In fact, a BLAST search using the deduced aa sequence of BdMoV 249 kDa protein for the NCBI Conserved Domain Database revealed HEL (Viral_helicase [pfam01443]; *E*-value = $1e^{-28}$), Pro (Peptidase_C36 [pfam05415]; *E*-value = $1e^{-15}$) and RdRp (RdRP_2 [pfam00978]; *E*-value = $5e^{-11}$) domains and no MTR or AlkB-like protein domains (Fig. 2).

A phylogenetic (ML) tree based on the three concatenated domain sequences (MTR, HEL and RdRp) for benyviruses and representative viruses of different families is shown in Fig. 3A. It is clear that BdMoV is included within the cluster of three benyviruses, BNYVV, BSBMV and RSNV. This benyvirus cluster is more related to clades of hepeviruses (*Hepeviridae*), omegatetraviruses and betatetravirus (*Tetraviridae*), and a novel mycovirus (Sclerotinia sclerotiorum RNA virus L, SsRV-L) than to rubivirus and alphavirus (*Togaviridae*), and other rod-shaped viruses belonging to the *Virgaviridae* and filamentous viruses (*Closteroviridae*) associated with plant hosts. A similar topology was found in ML trees constructed with individuals above three motifs of BdMoV and selected viruses (data not shown).

As shown in Table 1, three distinct domains (MTR, HEL and RdRp) of the BdMoV

replicase are 44 to 81% identical to the corresponding domains of three other benyviruses, and 30 to 35% identical to those of CAV excluding its MTR domain which is undetermined (Gibbs et al., 2011). The BdMoV Pro domain shows 33 to 44% identity with that of three benyviruses.

When compared with other viruses, BdMoV HEL domain shares modest level of identity with the two insect omegatetraviruses, Dendrolimus punctatus virus (GeneBank accession AAT27317; E-value = $3e^{-25}$, identity = 25%) and Helicoverpa armigera stunt virus (GeneBank accession AAC98529; *E*-value = $3e^{-4}$, identity = 24%). This domain also shows aa sequence similarities (GeneBank accession AER13447; *E*-value = $2e^{-9}$, identity = 27%) with the TGB1 helicase of hibiscus green spot virus, which is a novel bacilliform virus distantly related to the member of the genus *Cilevirus* (unassigned family). Furthermore, the BdMoV RdRp domain shares 30% identity (*E*-value = $6e^{-12}$) with hepatitis E virus (HEV, family *Hepeviridae*), while it shows lower identities (*E*-value = 0.008−$5e^{-10}$, identities = 23–30%) with some members of the *Virgaviridae* and *Closteroviridae* (data not shown).

Interestingly, the BdMoV 249-kDa replicase contains an additional sequence, that is the so-called AlkB-like domain, which is a member of the 2-oxoglutarate- and Fe(II)-dependent oxygenase superfamily (Bratlie and Drablos, 2005; van den Born et al., 2008). This domain is located between the Pro and RdRp domains (aa positions 1439 to 1592) in the BdMoV replicase, where no corresponding motif is present in the replicases of three other benyviruses (Fig. 3B, Supplementary Fig. S2A). Due to the presence of this domain and its surrounding sequences, the size of the BdMoV replicase is slightly larger than that of other benyviruses.

AlkB-like protein genes are widespread in eukaryotes and bacteria, and have also been identified in replication-associated proteins of numerous viruses largely belonging to the families *Betaflexiviridae* and *Alfaflexiviridae* (Bratlie and Drablos, 2005; Martelli et al., 2007). BLAST analysis shows that BdMoV AlkB-like regions represent a weak similarity to other AlkB genes; e.g., viral AlkB-like protein sequences from the American hop latent virus (genus *Carlavirus*, family *Betaflexiviridae*) (GeneBank accession AFI61519; *E*-value = 0.005, identity = 29%, query cover = 80%) and

actinidia virus A (genus *Vitivirus*, family *Betaflexiviridae*) (GeneBank accession AET36885; *E*-value = 0.008, identity = 28%, query cover = 99%) and AlkB protein of proteobacteria *Shewanella woodyi* ATCC 51908 (GeneBank accession ACA87888; *E*-value =0.015, identity = 27%, query cover = 80%).

The ML phylogenetic tree based on the BdMoV AlkB and selected AlkB-like sequences from plant viruses showed that the BdMoV sequence forms a clade with several other viral AlkBs (Fig. 3B). However, the tree topology did not reflect the viral phylogeny based on their replication-associated protein as described previously (van den Born et al., 2008). In the aa alignment of the selected AlkBs, the BdMoV AlkB-like sequence appeared to contain a less-conserved region in the middle (Supplementary Fig. S4). This feature was similarly observed in AlkB sequences of carlaviruses such as potato virus M, potato virus S and hippeastrum latent virus, and some other betaflexiviruses, although evolutional implications are unclear (Bratlie and Drablos, 2005; van den Born et al., 2008).

3.4. RNA2 encoded proteins

BdMoV RNA2 encodes a 20-kDa coat protein (CP) at its 5' extremity, containing the conserved CP motif (TMV_coat [pfam00721]; *E*-value = $7e^{-10}$), followed by an in-phase ~46-kDa ORF (readthrough domain, RTD) which is presumably expressed by translational readthrough of the CP cistron amber termination codon (66-kDa RT protein) (Fig. 2). Three following partially overlapping ORFs, encoding the proteins of 38 kDa (p38), 12kDa (p12) and 13 kDa (p13), which form a cassette of cell-to-cell movement protein-specifying genes known as triple gene block (TGB) (Fig. 2). BdMoV TGB1 and TGB2 proteins possess the RNA helicase (pfam01443; *E*-value = $1e^{-22}$) and the plant viral movement protein (Plant_vir_prot [pfam01307]; *E*-value = $1e^{-16}$) motifs, respectively. The 3'-proximal ORF encodes a cysteine-rich 13-kDa protein (p13 CRP) (Fig. 2). Accordingly, the genome organization of BdMoV RNA2 is quite similar to that of benyvirus RNA2 (Bouzoubaa et al., 1986; Lee et al., 2001; Lozano and Morales, 2009; Saito et al., 1996).

The phylogenetic (ML) analyses of aa sequences of CPs from BdMoV and other related

rod-shaped plant viruses clearly demonstrate that BdMoV is clustered with benyviruses which are more closely related to the furoviruses and pomoviruses than to pecluviruses and hordeivirus (Fig. 4A). BdMoV CP has 31–39% identity to benyvirus CPs (Table 1). The CP of CAV is closely related to those of tobamoviruses, although the CAV replicase is more similar to benyvirus replicase (Gibbs et al., 2011) (Fig. 4A).

As shown in Table 1, the aa sequence similarities of benyviruses in the RT domain are not very large. The aa sequence alignment showed that the BdMoV RT domain was somewhat different from those of BNYVV and BSBMV (Supplementary Fig. S2B, CP-RT), but their hydropathy profiles are similar to each other (data not shown), suggesting a structural homology. There are two hydrophobic regions at both the N- and C-termini of the RT domain. These are most likely the transmembrane regions (TM1 and TM2) commonly found in the RT domains or P2 regions of viruses with plasmodiophorid vectors. A KTER motif in the TM2 of BNYVV and BSBMV, which is important for efficient transmission (Tamada et al., 1996), was not detected in the TM2 of BdMoV and RSNV. The C-terminal regions of the RT proteins consist of relatively large numbers of serine residues, 16 out of the terminal 66 aa (BdMoV), 19 out of 83 aa (BNYVV), 21 out of 105 aa (BSBMV) and 22 out of 100 aa (RSNV).

The central region of BdMoV RNA2 contains the TGB-like structure that is possibly required for benyvirus cell-to-cell movement (Lauber et al., 1998) (Fig. 2). A similar TGB-like structure is found in NVMV (Randles and Rohde, 1990). The first TGB protein (TGB1) contains the conserved dNTP binding-site motif GxxGxGKS (x=any aa), in positions 92 to 99 (Richards and Tamada, 1992; Solovyev et al., 2012) (Supplementary Fig. S2B, TGB1). BNYVV and to a lesser extent BSBMV TGB1 have a K/R-rich cluster on the N-terminal 30 residues, whereas this region is absent in the BdMoV and RSNV TGB1. Hydrophobicity plot analysis for BdMoV TGB2 and TGB3 proteins shows that both proteins have two hydrophobic regions (data not shown). These regions are predicted as potential transmembrane domains using the TMHMM Server v.2.0 (www.cbs.dtu.dk/services/TMHMM), supporting the hypothesis that the TGB2 and TGB3 proteins bind to membranes (Supplementary Fig. S2B, TGB2 and TGB3). BdMoV TGB2 contains a semi-conserved sequence, $GDx_6GGxYxDG$, in the central region, whereas TGB3 does not have the proposed motifs, $Ax_7Px_{12}KxxDA$ in the

center and WFWxH in the C-terminal regions (Morozov and Solovyev, 2012; Solovyev et al., 2012).

The ML tree of the TGB1 aa sequences showed that BdMoV is clustered with benyviruses, which is distantly related to the pecul-, hordei- and pomoviruses (Fig. 4B). It is noted that BdMoV TGB1 is more closely related to the NVMV-coding protein. The BdMoV TGB1 protein is 42 to 48% identical to those of other benyviruses and 32% identical to NVMV TGB1 (Table 1). Similar trends were observed in the ML trees of TGB2 and TGB3 (data not shown).

Database searches for BdMoV CRP did not reveal substantial identity with any known protein (Table 1). When aligned with other benyvirus CRPs, four conserved cysteine residues (aa positions 55, 58, 92 and 94) were identified in their central regions, as described previously (Chiba et al., 2013) (Supplementary Fig. S2B, CRP). The CRPs of BdMoV, BNYVV and BSBMV have been shown to function as the weak transgene silencing suppressors in leaf tissue (Andika et al., 2012; Chiba et al., 2013; Guilley et al., 2009). BdMoV p13 CRP accumulated in the nucleus (data not shown), whereas BNYVV p14 and BSBMV 14-kDa CRPs were distributed in the nucleolus and cytoplasm in which nucleolar localization is not required for its silencing-suppression activity (Chiba et al., 2013).

3.5. The discovery of benyvirus replicase-like sequences (BRLS) in plant and insect genomes.

We searched for the presence of benyvirus-related sequences in the eukaryotic genomes that are available in NCBI databases. Using benyvirus sequences as queries, we found six significant matches to BdMoV replicase aa sequences in the whole-genome shotgun (WGS) assemblies of the chromosomes of the chickpea plant (*Cicer arietinum* L., cultivar = CDC Frontier, family *Fabaceae*) and the blood-sucking bug (*Rhodnius prolixus*, family *Triatomidae*) (Fig. 5A). Based on the nomenclature system for integrated non-retrovirus RNA viral sequences (Chiba et al., 2011b), these novel sequences were termed *Cicer arietinum* benyvirus replicase protein-like sequences 1–3 (CaBRLS1–3) and *Rhodnius prolixus* BRLSs 2–3 (RpBRLS2–3) (Figs. 5B and 5C).

Reverse BLAST analysis using their potentially encoded proteins revealed 26–50% aa identities (*E*-value = $2e^{-5}$–$2e^{-33}$) for CaBRLSs and 33 to 39% aa identities (*E*-valu e= $5e^{-5}$–$8e^{-43}$) for RpBRLSs in the corresponding region in the BNYVV replicase protein (237 kDa) (Fig. 5A).

In addition, we also found that another WGS library onto the *C. arietinum* chromosome (cultivar ICC4958), which was very recently released from the NCBI site, contains mostly identical sequences to each CaBRLS from the chickpea cultivar CDC Frontier. All BRLSs have distinguishable flanking sequences of the host origin in which CaBRLS2 and RpBRLS2 contain trace putative plant (DNA/Mariner class) and insect (LTR/Copia class) transposable elements, respectively (Figs. 5B and C).

To confirm the presence of BRLSs in the chickpea plant and the blood-sucking insect chromosomes, genomic PCR detection was performed. To this end, different sets of primers for each of the CaBRLSs and RpBRLSs were designed (Fig. 5B and C, indicated by arrows). As expected, PCR products of the expected sizes were obtained for all primer sets on genomic DNA from the chickpea plant (Fig. 6A). For RpBRLSs, a DNA fragment was only amplified by PCR using the RpBRLS3-1 primer set specific for the blood-sucking bug genomic DNA, but not using RpBRLS2-2 and 3-2 primer sets (Fig. 6B). The same results were obtained by using additional sets of RpBRLS2-2 or 3-2 primers (data not shown). The sequences of PCR fragments derived from the chickpea and blood-sucking bug samples were identical to the corresponding regions of the CaBRLSs and RpBRLS3-1 sequences from databases (data not shown). These results suggested that the CaBRLSs are probably widely present in chickpea cultivars, whereas RpBRLSs and/or their harboring *R. prolixus* chromosomes are more divergent in the strain or isolate level.

In order to evaluate the relationships of BRLSs and related viruses, we conducted ML analysis based on aa alignment of each of three replicase domains (MTR, HEL and RdRp) of the viruses. This analysis included replicase sequences of benyvirus-related viruses, such as animal hepeviruses, insect tetraviruses, animal rubiviruses, a mycovirus (SsRV-L) and an alga virus (CAV). As shown in Fig. 7, it is clear that CaBRLSs and RpBRLS are closely related to the clade of benyviruses. Notably, the BRLSs show a

closer relationship to the replicase of CAV than to those of other animal, insect and fungal viruses, and this provides details of the genealogy of this unique alga virus and benyviruses (Fig. 7).

Several studies have demonstrated that the transcriptome analysis based on next-generation sequencing technologies is a useful new research tool for the discovery of RNA viruses (Kristensen et al., 2010; Radford et al., 2012). During BLAST searching for BRLSs using the benyvirus replicase aa sequences as queries (e.g., BNYVV), interestingly, we found several significant matches in the transcriptome shotgun assembly (TSA) library databases (Fig. 5A): i.e., four matches (GeneBank accessions GACH01061598, GACH01134787, GACH01053674, GACH01115352; *E*-value = $2e^{-5}$–$4e^{-14}$, identities = 40–47%) in the TSA library of the Siam weed (*Chromolaena odorata*, family *Asteraceae*), one match (GeneBank accession JV146193; *E*-value = $4e^{-10}$, identity = 30%) in the TSA library of the European silver fir tree (*Abies alba*, family *Pinaceae*), and one match (GeneBank accession GACR01000094; *E*-value = $1e^{-30}$, identity = 40%) in the TSA library of a tree killing bark beetle (*Ips typographus*, family *Scolytidae*). In addition, one EST sequence (GeneBank accession FG537634) from the shoot apical meristem of the pea plant (*Pisum sativum*) also showed a moderate aa identity (*E*-value = $7e^{-17}$, identities = 32%) but with possible internal deletion. However, no significant hits for CAV or benyvirus protein-like sequences in the available databases of the algae were revealed. Instead, a sequence (GeneBank accession DV668917) in the EST library of a brown alga (*Sargassum binderi*, Phaeophyta) was detected to share a moderate level of identity with closteroviruses (family *Closteroviridae*), i.e. citrus tristeza virus (*E*-value = $4e^{-19}$, identity = 31%).

4. Discussion

In this work, we have determined the complete nucleotide sequences of BdMoV that was isolated from burdock plants in Japan. On the basis of sequence similarities and genome organization, we propose that BdMoV is a new species in the genus *Benyvirus* and could represent the fourth member of this genus. Phylogenetic analyses show that BdMoV is more closely related to BNYVV and BSBMV than RSNV (Table 1, Fig. 3A).

BNYVV and BSBMV usually require smaller RNA species RNA3 and RNA4 for disease development and viral survival in nature, whereas BdMoV and probably RSNV do not contain such smaller RNAs (Lozano and Morales, 2009), although this requires confirmation with a number of natural virus isolates. BSBMV supported BNYVV-RNA3 replication (Ratti et al., 2009), however attempts to superinfect BdMoV-infected plants at 20–30 days after inoculation with RNA3 or RNA4 failed (T. Tamada, unpublished results). Considering that BNYVV RNA1 and RNA2 are sufficient for viral replication, assembly, movement and transmission (Richards and Tamada, 1992; Tamada, 1999), it is clear that the benyviruses intrinsically have a bipartite genome.

Three benyviruses BNYVV, BSBMV and RSNV are transmitted by *Polymyxa* species belonging to the family Plasmodiophoraceae. Our preliminary experiments showed that BdMoV was not transmitted by *Polymyxa betae*, which is a vector of BNYVV and BSBMV. We also found that the burdock plant is not a suitable host for *P. betae* (T. Tamada, unpublished results). Thus, it seems unlikely that *P. betae* is the natural vector of BdMoV. The vector of BdMoV still needs to be determined; however, owing to a similarity in the genome structure and especially the presence of the two transmembrane regions in the RT domain (Adams et al., 2001; Tamada and Kondo, 2013), it is possible that BdMoV is also transmitted by the plasmodiophorid species. In this respect, we could not rule out the possibility of a deletion mutant in the RT domain, because BdMoV inoculum source used in this study has been propagated on inoculated leaves of *C. quinoa*. Further studies using other natural isolates or infectious clones are required to verify the role of this domain in vector transmission.

In regard to cytopathic effects, BdMoV induces the formation of viroplasm-like inclusion bodies in the cytoplasm of infected cells. These inclusion bodies are spherical-shaped, electron-dense, and granular structures which contain electron-lucent voids with various sizes (Inouye, 1973; Yamashita et al., 2008). Rod-shaped particles of BdMoV are usually arranged radially from the surface of these inclusions (like sea urchin spine) but are not observed within the inclusion body. Apart from the inclusion body, small bundles of particles which are arranged in parallel are occasionally

observed in the cytoplasm of infected cells (Inouye, 1973). So far it has been unknown whether these kinds of viroplasm-like structures are induced by infection of BNYVV or other rod-shaped viruses (Yamashita et al., 2008). In the case of BNYVV, viral particles are usually scattered or clustered in aggregates in the cytoplasm of infected cells, but sometimes dense masses of particles arranged in parallel or angle-layer arrays are observed (Putz and Vuittenez, 1980; Tamada, 2002).

Compared to other benyviruses, a most notable feature of BdMoV is the presence of the AlkB-like domain in the replicase protein (Fig. 2). The AlkB-like domain is not present in rod-shaped viruses belonging to the family *Virgaviridiae*. Viral AlkB proteins and their homologues have been identified in a number of plant RNA viruses, which include some viruses in the *Alfaflexiviridae* and most viruses in the *Betaflexiviridae* (Bratlie and Drablos, 2005; Martelli *et al.*, 2007). It is noted that AlkB is present in some, but not all, viruses that belong to the same genus. Outside of these two families, only a few numbers of known AlkB-containing viruses are recognized; these include grapevine leafroll-associated virus-3 and little cherry virus-1 (family *Closteroviridae*), blackberry virus Y (family *Potyviridae*), and black raspberry necrosis virus (BRNV, genus *Sadwavirus*) (Halgren et al., 2007; Susaimuthu et al., 2008). The viral AlkB proteins are reported to have a possible role in maintaining the integrity of the viral RNA genome through removal of deleterious RNA damage (van den Born et al., 2008).

The common features of AlkB-containing viruses are: (1) they have filamentous virions with a single exception of BRNV with isometric particles, (2) the AlkB domain is present in the replicase polyprotein of the viruses, (3) most, if not all, AlkB-containing viruses infect woody or perennial plants, and (4) the viral AlkB phylogeny does not reflect the viral species phylogeny, suggesting that AlkB may have been randomly distributed during virus evolution (van den Born et al., 2008). Indeed, the BdMoV AlkB-like domain is clearly present in the replicase polyprotein, and its sequence forms a clade with the viral AlkBs (Fig. 3B). Thus, BdMoV differs from the above AlkB-containing viruses in having rod-shaped particles and infecting a biennial burdock plant (not woody or perennial plant). It has been suggested that viral AlkBs might be acquired into the viral genome via a recombination event between two co-infecting viruses in the same host plant (Martelli et al., 2007).

The finding of the presence of BRLSs in plant and insect genomes provides additional evidence for benyvirus evolution. NRVSs that share sequence similarities to the genome of viruses in the families *Hepeviridae* and *Alphatetraviridae* have not been identified so far in the genomes of their possible hosts and, therefore, this is the first evidence of the endogenization of their associated viral sequences into a host genome. Although benyviruses have a limited host range of plants and are transmitted by specific plasmodiphorid vectors, footprints of their ancestral viral infections in both plants and insects are recorded in the form of endogenous molecular fossils (BRLSs). This implies the presence of unforeseen benyviruses in a wide range of host organisms. Interestingly, it has been reported that NRVSs having a sequence similarity to plant viruses with plus-sense RNA genomes are integrated into the genomes of diverse insects, including mosquitos, fruit flies, bees, ants, silkworms, pea aphids, Monarch butterflies, and wasps (Cui and Holmes, 2012). These NRVSs are closely or distantly related to the sequences of cileviruses or tobamoviruses, respectively (Cui and Holmes, 2012). Although it is unknown how such NRVSs were endogenized, it is speculated that progenitor viral sequences might have been integrated after reverse-transcription in aid of retrotransposons, because some NRVSs are embedded in or adjacent to sequence of transposable elements as is the case for CaBRLS2 (Fig. 5B) (Ballinger et al., 2012; Geuking et al., 2009). Indeed, evidence for recombination of retrotransposons and exogenous RNA viruses during infection/reverse transcription is reported and, moreover, such a recombinant has a potential to be integrated into the genome of host cells (Geuking et al., 2009; Goic et al., 2013; Horie et al., 2010).

NRVSs of benyvirus (BRLSs) were discovered in chickpea and blood-sucking bug chromosomes, and benyvirus-like transcript sequences were also identified in Siam weed, pea, European silver fir and bark beetle from NCBI TSA or EST databases, which might be derived from the genomic RNA of extant novel viruses or from endogenous transcripts corresponding to integrated host chromosomal virus-like sequences. Taken together with the close relationship of benyviruses to CAV isolated from algae (Fig. 7), these findings suggest that ancestral and extant benyviruses may have infected a broad range of hosts (algae, plants and insects). Thus, our findings provide an interesting insight into the origin and evolution of benyviruses and their related viruses.

Acknowledgements

We thank Kazuyuki Maruyama and Kouji Mitsuhata for technical assistance, Hirotaka Kanuka for providing insect materials, Tadao Inouye for helpful comments, and Nobuhiro Suzuki for fruitful discussion and comments on the manuscript. We also thank two anonymous reviewers for their valuable suggestions. This work was supported in part by a Grant-in-Aid for Scientific Research from the Japanese Ministry of Education, Culture, Sports, Science and Technology (KAKENHI 24580064 to H.K., KAKENHI 18580042 to T.T.), the Ohara Foundation for Agricultural Research (H.K), and by Yomogi Inc. (T.T). This research was also supported by the Japan Advanced Plant Science Network.

References

Abascal, F., Zardoya, R., Posada, D., 2005. ProtTest: selection of best-fit models of protein evolution. Bioinformatics 21(9), 2104–2105.

Adams, M.J., Antoniw, J.F., Kreuze, J., 2009. *Virgaviridae*: a new family of rod-shaped plant viruses. Archives of Virology 154(12), 1967–1972.

Adams, M.J., Antoniw, J.F., Mullins, J.G.L., 2001. Plant virus transmission by plasmodiophorid fungi is associated with distinctive transmembrane regions of virus-encoded proteins. Archives of Virology 146(6), 1139–1153.

Andika, I.B., Kondo, H., Rahim, M.D., Tamada, T., 2006. Lower levels of transgene silencing in roots is associated with reduced DNA methylation levels at non-symmetrical sites but not at symmetrical sites. Plant Molecular Biology 60(3), 423-435.

Andika, I.B., Kondo, H., Nishiguchi, M., Tamada, T., 2012. The cysteine-rich proteins of beet necrotic yellow vein virus and tobacco rattle virus contribute to efficient suppression of silencing in roots. Journal of General Virology 93(8), 1841–1850.

Anisimova, M., Gascuel, O., 2006. Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. Systematic biology 55(4), 539–552.

Ballinger, M.J., Bruenn, J.A., Taylor, D.J., 2012. Phylogeny, integration and expression

of sigma virus-like genes in *Drosophila*. Molecular Phylogenetics and Evolution 65(1), 251–258.

Bouzoubaa, S., Quillet, L., Guilley, H., Jonard, G., Richards, K., 1987. Nucleotide-sequence of beet necrotic yellow vein virus RNA-1. Journal of General Virology 68(3), 615–626.

Bouzoubaa, S., Ziegler, V., Beck, D., Guilley, H., Richards, K., Jonard, G., 1986. Nucleotide-sequence of beet necrotic yellow vein virus RNA-2. Journal of General Virology 67(8), 1689–1700.

Bratlie, M.S., Drablos, F., 2005. Bioinformatic mapping of AlkB homology domains in viruses. Bmc Genomics 6, 1. doi:10.1186/1471-2164-6-1.

Chan, Y.S., Cheng, L.N., Wu, J.H., Chan, E., Kwan, Y.W., Lee, S.M., Leung, G.P., Yu, P.H., Chan, S.W., 2011. A review of the pharmacological effects of *Arctium* lappa (burdock). Inflammopharmacology 19(5), 245–254.

Chiba, S., Hleibieh, K., Delbianco, A., Klein, E., Ratti, C., Ziegler-Graff, V., Bouzoubaa, S., Gilmer, D., 2013. The benyvirus RNA silencing suppressor is essential for long-distance movement, requires both zinc-finger and nols basic residues but not a nucleolar localization for its silencing-suppression activity. Molecular Plant-Microbe Interactions 26(2), 168–181.

Chiba, S., Kondo, H., Miyanishi, M., Andika, I.B., Han, C.G., Tamada, T., 2011a. The evolutionary history of *Beet necrotic yellow vein virus* deduced from genetic variation, geographical origin and spread, and the breaking of host resistance. Molecular Plant-Microbe Interactions 24(2), 207–218.

Chiba, S., Kondo, H., Tani, A., Saisho, D., Sakamoto, W., Kanematsu, S., Suzuki, N., 2011b. Widespread endogenization of genome sequences of non-retroviral RNA viruses into plant genomes. Plos Pathogens 7(7), e1002146 doi: 10.1371/journal.ppat.1002146.

Crochu, S., Cook, S., Attoui, H., Charrel, R.N., De Chesse, R., Belhouchet, M., Lemasson, J.J., de Micco, P., de Lamballerie, X., 2004. Sequences of flavivirus-related RNA viruses persist in DNA form integrated in the genome of *Aedes* spp. mosquitoes. Journal of General Virology 85(7), 1971–1980.

Cui, J., Holmes, E.C., 2012. Endogenous RNA viruses of plants in insect genomes. Virology 427(2), 77–79.

Firth, A.E., Wills, N.M., Gesteland, R.F., Atkins, J.F., 2011. Stimulation of stop codon

readthrough: frequent presence of an extended 3' RNA structural element. Nucleic Acids Research 39(15), 6679–6691.

Geuking, M.B., Weber, J., Dewannieux, M., Gorelik, E., Heidmann, T., Hengartner, H., Zinkernagel, R.M., Hangartner, L., 2009. Recombination of retrotransposon and exogenous RNA virus results in nonretroviral cDNA integration. Science 323(5912), 393–396.

Gibbs, A.J., Torronen, M., Mackenzie, A.M., Wood, J.T., Armstrong, J.S., Kondo, H., Tamada, T., Keese, P.L., 2011. The enigmatic genome of Chara australis virus. Journal of General Virology 92(11), 2679–2690.

Gilmer, D., Ratti, C., 2012. *Benyvirus*. In: King, A.M.Q., Adams, M.J., Carstens, E.B., Lefkowitz, E.J. (Eds.), Virus Taxonomy: Classification and Nomenclature of Viruses: Ninth Report of the International Committee on Taxonomy of Viruses. Elsevier, San Diego, CA, U.S.A., Page 1134–1138.

Gilmer, D., Ratti, C., Tamada, T., Andika, I.B., Kondo, H., 2013. Create 1 new species in the genus *Benyvirus* and assign the genus to the new family *Benyviridae*. Taxonomy Proposals – Plant, International Committee on Taxonomy of Viruses. On line:http://talk.ictvonline.org/files/proposals/taxonomy_proposals_plant1/m/plant01/4561.aspx.

Goic, B., Vodovar, N., Mondotte, J.A., Monot, C., Frangeul, L., Blanc, H., Gausson, V., Vera-Otarola, J., Cristofari, G., Saleh, M.C., 2013. RNA-mediated interference and reverse transcription control the persistence of RNA viruses in the insect model *Drosophil*a. Nature Immunology 14(4), 396–403.

Guilley, H., Bortolamiol, D., Jonard, G., Bouzoubaa, S., Ziegler-Graff, V., 2009. Rapid screening of RNA silencing suppressors by using a recombinant virus derived from beet necrotic yellow vein virus. Journal of General Virology 90(10), 2536–2541.

Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W., Gascuel, O., 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Systematic Biology, 59(3), 307–321.

Halgren, A., Tzanetakis, I.E., Martin, R.R., 2007. Identification, characterization, and detection of Black raspberry necrosis virus. Phytopathology 97(1), 44–50.

Heidel, G.B., Rush, C.M., Kendall, T.L., Lommel, S.A., French, R.C., 1997.

Characteristics of beet soilborne mosaic virus, a furo-like virus infecting sugar beet. Plant Disease 81(9), 1070–1076.

Hirano, S., Kondo, H., Maeda, T., Tamada, T., 1999. Burdock mottle virus has a high genome similarity to beet necrotic yellow vein virus. Proceedings of the fourth Symposium of the International Working Group on Plant Viruses with Fungal Vectors, 33–36.

Horie, M., Honda, T., Suzuki, Y., Kobayashi, Y., Daito, T., Oshida, T., Ikuta, K., Jern, P., Gojobori, T., Coffin, J.M., Tomonaga, K., 2010. Endogenous non-retroviral RNA virus elements in mammalian genomes. Nature 463(7277), 84–90.

Inouye, T., 1973. Host range and electron microscopy of burdock mottle virus, a rod-shaped virus from *Arctium lappa* L.: Studies on the viruses of plants in C*ompositae* in Japan. Ber Ohara Inst Landw Biol, Okayama Univ. 15, 207–218.

Katoh, K., Toh, H., 2008. Recent developments in the MAFFT multiple sequence alignment program. Briefings in Bioinformatics 9(4), 286–298.

Katzourakis, A., Gifford, R.J., 2010. Endogenous viral elements in animal genomes. Plos Genetics 6(11), e1001191. doi:10.1371/journal.pgen.1001191.

Koenig, R., Lesemann, D.E., 2005. Genus *Benyvirus*. In: Fauquet, C.M., Mayo, M.A., Maniloff, J., Desselberger, U., Ball, L.A. (Eds.), Virus Taxonomy, Eighth Report of the International Committee on Taxonomy of Viruses. Elsevier Academic Press, London, pp. 1043–1048.

Kohany, O., Gentles, A.J., Hankus, L., Jurka, J., 2006. Annotation, submission and screening of repetitive elements in Repbase: RepbaseSubmitter and Censor. BMC Bioinformatics 7(1), 474. doi:10.1186/1471-2105-7-474.

Kondo, H., Chiba, S., Toyoda, K., Suzuki, N., 2013. Evidence for negative-strand RNA virus infection in fungi. Virology 435(2), 201–209.

Kondo, H., Maeda, T., Shirako, Y., Tamada, T., 2006. Orchid fleck virus is a rhabdovirus with an unusual bipartite genome. Journal of General Virology 87(8), 2413–2421.

Kondo, H., Maeda, T., Tamada, T., 2009. Identification and characterization of structural proteins of orchid fleck virus. Archives of Virology 154(1), 37–45.

Koonin, E.V., Gorbalenya, A.E., Purdy, M.A., Rozanov, M.N., Reyes, G.R., Bradley, D.W., 1992. Computer-assisted assignment of functional domains in the nonstructural polyprotein of hepatitis-E virus - Delineation of an additional

group of positive-strand RNA plant and animal viruses. Proceedings of the National Academy of Sciences of the United States of America 89(17), 8259–8263.

Kristensen, D.M., Mushegian, A.R., Dolja, V.V., Koonin, E.V., 2010. New dimensions of the virus world discovered through metagenomics. Trends in Microbiology 18(1), 11–19.

Lauber, E., Bleykasten-Grosshans, C., Erhardt, M., Bouzoubaa, S., Jonard, G., Richards, K.E., Guilley, H., 1998. Cell-to-cell movement of beet necrotic yellow vein virus: I. Heterologous complementation experiments provide evidence for specific interactions among the triple gene block proteins. Molecular Plant-Microbe Interactions 11(7), 618–625.

Lee, L., Telford, E.B., Batten, J.S., Scholthof, K.B.G., Rush, C.M., 2001. Complete nucleotide sequence and genome organization of *Beet soilborne mosaic virus*, a proposed member of the genus *Benyvirus*. Archives of Virology 146(12), 2443–2453.

Liu, H.Q., Fu, Y.P., Jiang, D.H., Li, G.Q., Xie, J.T., Cheng, J.S., Peng, Y.L., Ghabrial, S.A., Yi, X.H., 2010. Widespread horizontal gene transfer from double-stranded RNA viruses to eukaryotic nuclear genomes. Journal of Virology 84(22), 11876–11887.

Lopez-Vinyallonga, S., Romaschenko, K., Susanna, A., Garcia-Jacas, N., 2011. Systematics of the Arctioid group: Disentangling *Arctium* and *Cousinia* (Cardueae, Carduinae). Taxon 60(2), 539–554.

Louvel, D., Bidaux, J.M., 1977. Observation de nouveaux symptômes pathologiques sur des variétés précoces de riz en Cote-d'Ivoire. Agronomie Tropicale 32, 257–261.

Lozano, I., Morales, F., 2009. Molecular characterisation of *Rice stripe necrosis virus* as a new species of the genus *Benyvirus*. European Journal of Plant Pathology 124(4), 673–680.

Maeda, T., Hirano, S., Kondo, H., Tamada, T., 1996. Characteristics of burdock mottle virus. Annals of the Phytopathological Society of Japan 62, 628 (abstract in Japanese).

Marcilla, A., Bargues, M.D., Ramsey, J.M., Magallon-Gastelum, E., Salazar-Schettino, P.M., Abad-Franch, F., Dujardin, J.P., Schofield, C.J., Mas-Coma, S., 2001. The ITS-2 of the nuclear rDNA as a molecular marker for populations, species, and

phylogenetic relationships in Triatominae (Hemiptera : Reduviidae), vectors of Chagas disease. Molecular Phylogenetics and Evolution 18(1), 136–142.

Martelli, G.P., Adams, M.J., Kreuze, J.F., Dolja, V.V., 2007. Family *Flexiviridae*: a case study in virion and genome plasticity. Annual Review of Phytopathology 45, 73–100.

Morales, F.J., Ward, E., Castano, M., Arroyave, J.A., Lozano, I., Adams, M.J., 1999. Emergence and partial characterization of rice stripe necrosis virus and its fungus vector in South America. European Journal of Plant Pathology 105(7), 643–650.

Morozov, S.Y., Solovyev, A.G., 2012. Did silencing suppression counter-defensive strategy contribute to origin and evolution of the triple gene block coding for plant virus movement proteins? Frontiers in Plant Sciencent 3, 00136. doi: 00110.03389/fpls.02012.00136.

Putz, C., Vuittenez, A., 1980. The intracellular location of beet necrotic yellow vein virus. Journal of General Virology 50(1), 201–204.

Radford, A.D., Chapman, D., Dixon, L., Chantrey, J., Darby, A.C., Hall, N., 2012. Application of next-generation sequencing technologies in virology. Journal of General Virology 93(9), 1853–1868.

Rahim, M.D., Andika, I.B., Han, C., Kondo, H., Tamada, T., 2007. RNA4-encoded p31 of beet necrotic yellow vein virus is involved in efficient vector transmission, symptom severity and silencing suppression in roots. Journal of General Virology 88(5), 1611–1619.

Randles, J.W., Rohde, W., 1990. *Nicotiana velutina* mosaic virus: Evidence for a bipartite genome comprising 3 kb and 8 kb RNAs. Journal of General Virology 71(5), 1019–1027.

Ratti, C., Hleibieh, K., Bianchi, L., Schirmer, A., Autonell, C.R., Gilmer, D., 2009. *Beet soil-borne mosaic virus* RNA-3 is replicated and encapsidated in the presence of BNYVV RNA-1 and-2 and allows long distance movement in *Beta macrocarpa*. Virology 385(2), 392–399.

Richards, K.E., Tamada, T., 1992. Mapping functions on the multipartite genome of beet necrotic yellow vein virus. Annual Review of Phytopathology 30, 291–313.

Rush, C.M., 2003. Ecology and epidemiology of *Benyviruses* and plasmodiophorid vectors. Annual Review of Phytopathology 41(1), 567–592.

Saito, M., Kiguchi, T., Kusume, T., Tamada, T., 1996. Complete nucleotide sequence of the Japanese isolate S of beet necrotic yellow vein virus RNA and comparison with European isolates. Archives of Virology 141(11), 2163–2175.

Solovyev, A.G., N.O., K., Morozov, S.Y., 2012. Recent advances in research of plant virus movement mediated by triple gene block. Frontiers in Plant Science 3, 00276. doi: 00210.03389/fpls.02012.00276.

Susaimuthu, J., Tzanetakis, I.E., Gergerich, R.C., Martin, R.R., 2008. A member of a new genus in the *Potyviridae* infects *Rubus*. Virus Research 131(2), 145–151.

Tamada, T., 1999. Benyvirus, Encyclopedia of Virology, 2nd ed. Academic Press London, pp. 154–160.

Tamada, T., 2002. *Beet necrotic yellow vein virus*, CMI/AAB Description of Plant Viruses. Vol. No. 391. Wellesbourne: Association of Applied Biologists.

Tamada, T., 2007. Susceptibility and resistance of *Beta vulgaris* subsp. *maritima* to foliar rub-inoculation with *Beet necrotic yellow vein virus*. Journal of General Plant Pathology 73(1), 76–80.

Tamada, T., Kondo, H., 2013. Biological and genetic diversity of plasmodiophorid-transmitted viruses and their vectors. Journal of General Plant Pathology in press. doi: 10.1007/s10327-013-0457-3

Tamada, T., Schmitt, C., Saito, M., Guilley, H., Richards, K., Jonard, G., 1996. High resolution analysis of the readthrough domain of beet necrotic yellow vein virus readthrough protein: A KTER motif is important for efficient transmission of the virus by *Polymyxa betae*. Journal of General Virology 77(7), 1359–1367.

Tanne, E., Sela, I., 2005. Occurrence of a DNA sequence of a non-retro RNA virus in a host plant genome and its expression: evidence for recombination between viral and host RNAs. Virology 332(2), 614–622.

Taylor, D.J., Bruenn, J., 2009. The evolution of novel fungal genes from non-retroviral RNA viruses. BMC Biology 7, 88. doi:10.1186/1741-7007-7-88

van den Born, E., Omelchenko, M.V., Bekkelund, A., Leihne, V., Koonin, E.V., Dolja, V.V., Falnes, P.O., 2008. Viral AlkB proteins repair RNA damage by oxidative demethylation. Nucleic Acids Research 36(17), 5451–5461.

Yamashita, S., Doi, Y., Kitajima, E.W. 2008. Comparative electron microscopic observation of viroplasms induced by plant viruses. Japanese Journal of Phytopathology 74(2): 97–109 (Japanese with English abstract).

Figure legends

Fig. 1. (A) Electron micrographs of the purified virus particles. Samples were stained in uranyl acetate. The bar represents 100 nm. Asterisks: particle about 250 nm in length. (B and C) Symptoms in burdock (B) and *Chenopodium quinoa* (C) infected with BdMoV. (D) Denaturing gel electrophoretic analysis of the BdMoV RNAs before and after poly(A)$^+$ fractionation. Each sample was resolved on 1.2% agarose gels followed by ethidium bromide staining. Lane 1: Dendrobium mosaic virus Japanese isolate (DeMV-J); lane 2: cucumber mosaic virus (CMV-Y); 3: BdMoV; 4 tobacco mosaic virus (TMV-OM). The positions of the BdMoV RNAs 1 and 2 are indicated on the right. (E) SDS-PAGE and Western blot analysis of partially purified preparations. Each sample was resolved on 12% SDS-PAGE gels followed by Coomassie brilliant blue-R250 staining (CBB-staining, left panel). The resolved proteins were transferred to nitrocellulose and reacted with rabbit polyclonal antiserum to BdMoV (@BdMoV). Lane 1:total proteins of the BdMoV-infected *C. quinoa* leaf (Infected); 2: BdMoV virion; 3: total proteins of the mock plant leaf (Mock). The positions of the CP and high-molecular-weight protein (HMW, ~60 kDa) are indicated on the right. The positions of the marker proteins are shown on the left.

Fig. 2. Genomic organization of BdMoV. RNA1 contains a single open reading frame (ORF, rectangle). The methyltransferase (MTF), RNA helicase (HEL), papain-like protease (Pro) and RNA-dependent RNA polymerase (RdRp) domains are indicated by bold bars on the 249-kDa replicase. The AlkB-like domain is shown by a dashed bar. The arrowhead shows the location of the putative cleavage site for polyprotein. RNA2 contains six ORFs. The coat protein (CP; 20 kDa) cistron is separated from a long in-phase ORF (46-kDa readthrough domain, RTD) by a single amber termination codon (star, 693 nt), which permits expression of a 66-kDa fusion protein. The third to fifth ORFs (38, 12 and 13 kDa, respectively) are presumably associated with viral spreading since they are similar to triple gene block (TGB). The HEL motif on the 38-kDa protein is indicated by the bold bar. The sixth ORF (13 kDa) on RNA2 is similar to the cysteine-rich protein (CRP) of benyviruses. The Cap? represents the putative 5'-cap structure and the A(n) indicates the 3' poly(A) tail. Comparisons of the flanking leaky UAG (amber, 693) codon in BdMoV with the corresponding sequences in benyviruses

revealed that these are mostly similar and a possible stem–loop-like structure 3'-adjacent to the UAG was conserved (Firth et al., 2011) (see Supplementary Fig. S1C). The position (in nucleotides) of the start and stop codons of each predicted ORF are shown.

Fig. 3. (A) Phylogenetic (ML) tree calculated from the concatenated amino acid sequences of three domains (Met–Hel–RdRp) in the putative replicase proteins of BdMoV and benyviruses together with some other related plant and animal RNA viruses. Asterisks show the unassigned viruses. (B) ML tree of BdMoV and selected viral AlkB-like domains rooted by the proteobacteria outgroup. Putatively dysfunctional AlkBs are indicated with an asterisk (van den Born et al., 2008). For both tree constructions, the alignments of the amino acid sequences were generated with MAFFT (Supplementary Figs. S3A and B). The trees were constructed by the ML method using PhyML 3.0. The branch support values were estimated using the approximate likelihood ratio test (aLRT) with a Shimodaira–Hasegawa-like (SH-like) algorithm (only values greater than 0.9 are shown as filled circles). The scale bars represent the amino acid distances. The sequences of the three replicase-associated and AlkB-like domains of RNA viruses were obtained from the EMBL/DDBJ/GenBank database and are listed in Supplementary Table S2.

Fig. 4. Phylogenetic (ML) trees calculated from the coat proteins (A) and TGB1 proteins (B) of BdMoV and benyviruses together with selected rod-shaped viruses included in the family *Virgaviradae*. The alignments of the amino acid sequences were generated with MAFFT (Supplementary Fig. S3C and D). The trees were constructed by the ML method using PhyML 3.0. The nodes with filled circles are supported by aLRT values of greater than 0.9. The scale bars represent the amino acid distances. Carlaviruses (potato virus M [PVM]: AAP76201, red clover vein mosaic virus [RCVMV]: ACN58188) were used as the outgroup for TGB1 analysis. The sequence data of other viruses used for analyses were obtained from the EMBL/DDBJ/Genbank and are listed in Supplementary Table S2. Asterisks show the unassigned viruses.

Fig. 5. Schematic representation of *Cicer arietinum* (CaBRLS) and *Rhodnius prolixus* (RpBRLS) benyvirus replicase protein-like sequences and their flanking regions. (A)

The corresponding positions of BRLSs on the benyvirus replicase. The box shows a diagrammatic representation of the conserved domains of BNYVV replicase protein (B and C) BRLSs found in the whole-genome shotgun (WGS) database of chickpea plant (*C. arietinum* L) (B) and blood-sucking insect (*R. prolixus*) (C) are shown to match the replicase protein from BNYVV in (A). The potential coding regions of BRLSs and flanking ORFs are shown as boxes. Retrotransposon-like sequences are shown by thick gray lines. The positions of the primers used for genomic PCR (see Fig. 6) are shown by arrows. The WGS-assembled sequences and undetermined sequences are shown by solid and dashed thin-lines, respectively. Symbols referring to mutations are also shown; kinked line, major deletion; asterisk, internal stop codon; F, frame-shift.

Fig. 6. Genomic PCR analysis of BRLSs. CaBRLSs of the chickpea plant (A) and blood-sucking insect samples (B) were amplified using a primer set specific for each BRLS (Figs. 5B and C). Two primer sets, At-IRS-FW and At-IRS-RV (Chiba et al., 2011b), and 5.8T Rp-ITS2F and 28T Rp-ITS2R (Marcilla et al., 2001), were used for amplification of the plant and insect (data not shown) ITS regions, respectively.

Fig. 7. Maximum likelihood (ML) phylogeny of the replicase amino acid sequences of benyvirus and their related CaBRLSs and RpBRLSs. Phylogenetic trees were constructed using PhyML 3.0 based on the multiple amino acid sequence alignment of the methyltransferase (MTF) (A), RNA helicase (HEL) (B), and RNA-dependent RNA polymerase (RdRp) (C) domains shown in Supplementary Figures S3E, F and G, respectively. A model LG with + I + G + F was selected as best-fit model for each alignment. The nodes with filled circles are supported by aLRT values of greater than 0.9. The scale bars represent the amino acid distances. Formal virus names and their GenBank accession numbers of the sequences (presented as acronyms) are listed in Supplementary Table S2. Asterisks show the viruses with unassigned species.

Legend of Supplementary Figures

Supplementary Fig. S1. (A and B) Alignment of the 5'- and 3'-untranslated terminal regions (UTRs) of RNA1 and RNA2 of BdMoV and benyviruses. The 5'-UTR sequences were compared separately for RNA1 and RNA2 (A), while 3'-UTR sequences preceding the poly(A) tail from both segments were compared together (B). Asterisks represent identical nucleotides among four viruses and are highlighted by gray boxes. Strictly conserved first 8 nt at the 5'-terminal sequence of RNAs 1 and 2 are boxed. Conserved sequence domains at the 3' extremities of both segments are indicated by thick bars above the alignment (Bouzoubaa et al., 1986). Nucleotide sequences are presented as DNA (C) Comparison of the CP read-through regions of BdMoV and other benyviruses. The amber stop codons for CP-ORFs are gray-shaded. The potential stems can be formed immediately downstream of CP-stop codons by two underlined regions.

Supplementary Fig. S2. Amino acid sequence alignment of the RNA1-encoded replicase (A) and RNA2-encoded proteins (B–F) of BdMoV and benyviruses. (A) The conserved motifs of Methyltransferase, RNA-helicase and RdRp are shown with green, blue and pink highlights, respectively. The regions spanning those domains and the papain-like protease motif, Y domains and AlkB-like sequence (underlined in red) are indicated by starting and terminating arrow sets. Conserved C and H residues in the protease domains and predicted cleavage sites are highlighted by orange. The NTP-binding motif (GxxGxGKS, where x represents any aa) and some other highly conserved residues in the helicase domains are boxed. Asterisks indicate identical amino-acid sequences among four viruses. (B) CP and CP read-through domain (RTD), (C) TGB1, (D) TGB2, (E) TGB3, (F) Cysteine-rich protein (CRP). The conserved sequence of CP-RTD, RNA-helicase domain, and CRP are shown with gray, blue and yellow highlights, respectively. The tentative transmembrane domains (TM1 and TM2) in RTD, TGB2 and TGB3 are denoted by red lines above the alignment. The KTER motif in TM2 of RTD, which is required for efficient viral transmission is underlined (black). N terminal, positively charged regions in TGB1 are specifically indicated.

Supplementary Fig. S3 Multiple amino-acid sequence alignment of the several proteins or their domains encoded by BdMoV, benyviruses and other related viruses.

(A) MTR+HEL+RdRp concatenate, (B) AlkB-like domains, (C) Coat protein, (D) TGB1 protein, (E) MTR protein + BRLS1, (F) HEL protein + BRLS2, (G) MTR protein + BRLS3.

Supplementary Fig. S4 Multiple alignments of sequence regions corresponding to the AlkB-like domain of the BdMoV and selected plant RNA viruses. The alignment was created using MUSCLE, version 3.7 (http://www.phylogeny.fr/). Similar residues are colored according to BLOSUM62. Members of different viral families are colour-coded: blue, alphaflexiviruses (*Tymovirales*); green, betaflexiviruses (*Tymovirales*); orange, closterovirus; purple, potyvirus; brown, secovirus (*Picornavirales*). The conserved residues among the AlkB proteins are indicated (a single F and Y, the HxD motif, a single H, and the Rx_5R motif, respectively). Putative inactivated AlkBs are indicated with an asterisk (van den Born et al., 2008). AlkB-like domain sequences of plant RNA viruses obtained from the EMBL/DDBJ/GenBank database are listed in Supplementary Table S2.