

Archives of Virology, Annotated Sequence Record

Complete genome sequence of Habenaria mosaic virus, a new potyvirus infecting a terrestrial orchid (*Habenaria radiata*) in Japan

Hideki Kondo^{1,*}, Takanori Maeda², I Wayan Gara¹, Sotaro Chiba¹, Kazuyuki Maruyama¹, Tetsuo Tamada¹ and Nobuhiro Suzuki¹

¹Institute of Plant Science and Resources (IPSR), Okayama University, Kurashiki 710-0046, Japan.

² College of Bioresource Sciences, Nihon University, Fujisawa 252-8510, Japan

*Corresponding author:

Hideki Kondo

e-mail: hkondo@rib.okayama.u-ac.jp

Tel. +81(86) 434-1232

Fax. +81(86) 434-1232

Running Title: Complete genome sequence of HaMV

Key words: habenaria mosaic virus, potyvirus, complete genome sequence, *Habenaria radiata*, terrestrial orchid

Word counts: 126 words (Abstract); 1760 words (Main Text)

Figures: 2; Supplementary Table: 1; Supplementary Figure: 2

Abstract

The complete genomic sequence of Habenaria mosaic virus (HaMV), that infects a terrestrial orchid (*Habenaria radiata*), has been determined. The genome is composed of 9,499 nucleotides excluding the 3'-terminal poly (A) tail, encoding a large polyprotein of 3,054 amino acids with the genomic features typical of a potyvirus. Putative proteolytic cleavage sites were identified by sequence comparison to those of known potyviruses. The HaMV polyprotein showed 58% amino acid sequence identity to that encoded by the most closely related potyvirus, tobacco vein banding mosaic virus. Phylogenetic analysis of polyprotein amino acid and its coding sequences confirmed that HaMV formed a cluster with the chilli veinal mottle virus group, most of which infect solanaceous plants. These results suggest that HaMV is a distinct member of the genus *Potyvirus*.

Introduction

Habenaria radiata (Thurnberg) K. Spreng. is a wild terrestrial orchid growing in wet grasslands in both warm and cold temperate regions in Japan. This orchid is cultivated as an ornamental plant because of its beautiful bird-shaped flower petals [18]. At present, the occurrence of two viruses, Habenaria mosaic virus (HaMV, unassigned virus) and watermelon mosaic virus (WMV, potyvirus, formerly WMV-2), has been reported from *H. radiata* plants in Japan [8, 12]. HaMV causes mosaic symptoms in *H. radiata* plants [12], whereas the symptoms of WMV are stunting in addition to leaf mosaics and distortions [8]. HaMV is sap-transmissible to some experimental plant species, but *H. radiata* is currently the only known systemic host [12]. Despite the lack of sequence information on the HaMV genome, the virus has long been believed to be a potyvirus because of its filamentous particle length of ca. 750 nm, non-persistent mode of transmission by aphids and formation of typical cylindrical inclusions in infected cells [12]. Furthermore, HaMV shows distant serological relationships to several potyviruses including WMV, and is thus suspected to be a new potyvirus species [12]. Here, we report the complete genomic sequence of HaMV. The phylogenetic relationship of HaMV with other potyviruses revealed that HaMV is sufficiently distinct

from known potyviruses and is a possible new species of the genus *Potyvirus*.

Materials and methods

An original isolate of HaMV (Ha-1) was obtained from a commercially available *H. radiata* plant showing mosaic symptoms in Okayama Prefecture, Japan in 1974, by N. Inouye [12]. HaMV (Ha-1) had been isolated by single legion transfer using an experimental host *Chenopodium quinoa*. The virus was purified from mechanically inoculated leaves of *C. quinoa*, essentially as described by Gara et al. [8]. Viral RNA was extracted from purified virus particles using proteinase K and SDS followed by phenol extraction and ethanol precipitation. First and second cDNA strands of HaMV genome were synthesized using the TimeSaver cDNA Synthesis Kit (Amersham Pharmacia Biotech, Uppsala, Sweden) with an oligo(dT) primer or virus-specific reverse primers designed according to the newly determined sequences. Two regions (nucleotide positions at 530–3710 and 3606–5766) with some sequence ambiguity were amplified by RT-PCR using virus-specific primer sets and sequenced. PCR was performed using KOD or KOD FX DNA polymerase (Toyobo, Osaka, Japan). The 5' end of the RNA genome was determined by RACE using the 5'RACE system (Invitrogen, Carlsbad, CA, USA). All cDNA and RACE fragments were cloned into the pGEM-T or pZErO-2 vector (Invitrogen, San Diego, CA, USA) according to the manufacturer's instructions. These plasmid clones were used to transform competent *Escherichia coli* DH5 alpha or TOP10F' cells. The DNA sequences were determined by the dideoxynucleotide chain-termination method with a DNA sequencer, model ABI377 or ABI3100 (Applied Biosystems, Foster City, CA, USA). Sequences of the primers used in this study are available upon request.

Sequence assembly was performed using the AutoAssembler program (PE Applied Biosystems). Sequence data were analyzed using GENETYX-MAC (Software Development Co., Tokyo, Japan) or Enzyme X (version 3, Mek & Tosj). The putative translation products were compared with the corresponding regions of other potyvirus sequences available in GenBank using BLAST (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>). Phylogenetic tree construction was based on a maximum-likelihood (ML) method as described previously with minor modification [5, 14]. The potyviral polyproteins or their

ORF sequences were aligned with MAFFT version 7 [13] and gap-cleaned using Gblocks version 0.91b [20]. Selection of the best-fit model for the data sets was performed using jmodeltest 2 [7, 10] and ProtTest 2.4 [1] for nucleotide and amino acid sequences, respectively. ML phylogenetic trees were generated in PhyML 3.0 using the appropriate substitution mode [11]. The resulting ML-trees were visualized using Figtree version 1.3.1 (<http://tree.bio.ed.ac.uk/software/figtree/>).

Sequence properties

The full-length genome sequence of HaMV Ha-1 is 9,499 nucleotides (nt), excluding the poly (A) tail at the 3' end (Fig. 1). The sequence has been deposited in the GenBank database as accession number AB818538. The genome organization of HaMV is typical of a potyvirus, containing a large open reading frame (ORF, 135–9,299 nt) that would encode a polyprotein of 3,054 amino acid (aa) residues with an estimated molecular weight of 345.4 kDa (Fig. 1). Nine putative protease cleavage sites were predicted, based on a multiple alignment of polyproteins encoded by closely related potyviruses (supplementary Fig. S1). The HaMV polyprotein is potentially cleaved into ten functional mature products (P1, HC-Pro, P3, 6K1, CI, 6K2, VPg, NIa-Pro, NIb and CP) by the viral-encoded proteases [2, 22] (Fig. 1). In these proteins, conserved motifs of potyviruses were detected, such as the $^{170}\text{Hx}_8\text{Dx}_{32}\text{S}^{213}$ (x, any amino acid residue) motif in the P1 protein, the $^{313}\text{KITC}^{316}$ and $^{442}\text{FRNK}^{445}$ motifs in HC-Pro, the $^{1207}\text{GSGKSx}_3\text{P}^{1251}$ and $^{1293}\text{DExH}^{1296}$ motifs in CI and the $^{2599}\text{GDD}^{2601}$ motif in the NIb [22] (supplementary Fig. S1). However, the Ha-1 isolate has $^{2774}\text{DAT}^{2776}$, instead of the DAG motif important for aphid transmission, at the N-terminus of the CP [16] (supplementary Fig. S1). A point mutation (G to T in the triplet) presumably occurred during several rounds of transfer through mechanical inoculation under green house conditions (data not shown). In addition, the presence of a small overlapping ORF (PIPO, expressed after ribosomal frameshift or transcriptional slippage at motif $\text{G}_{1-2}\text{A}_{6-7}$ [$^{2757}\text{GAAAAAA}^{2763}$], Fig. 1, arrow) was also predicted within the P3 cistron [6].

A BLASTp search with the sequence of the polyprotein encoded by HaMV Ha-1 revealed that it shared significant aa sequence identity (53–58% identities; 95–96% query coverage: QC) to the polyproteins encoded by members of chilli veinal

mottle virus (ChiVMV) group (subgroup 5) and their two relatives [9, 19], i.e., ChiVMV, chilli ringspot virus (ChiRSV), pepper veinal mottle virus (PVMV), tobacco vein banding mosaic virus (TVBMV), and wild tomato mosaic virus (WTMV) as the subgroup 5 members, and yam mild mosaic virus (YMMV) and christmas bell potyvirus (YMMV, an unclassified potyvirus) as the relatives. For each gene product of HaMV, aa sequence identities to the ChiVMV subgroup (subgroup 5) and two related viruses are 28–42% (37–96% QC) in P1, 61–64% (100% QC) in HC-Pro, 28–35% (95–100% QC) in P3, 32–57% (31–98% QC) in PIPO, 67–78% (96–98% QC) in 6K1, 58–62% (100% QC) in CI, 47–58% (98–100% QC) in 6K2, 59–69% (94–100% QC) in VPg, 52–58% (100% QC) in NIa, 61–70% (92–100% QC) in NIb and 61–70% (98–99% QC) in the CP region. Based on pairwise comparisons of the complete genome sequences, HaMV Ha-1 shares identities of 55–59% at the nt sequence level with ChiVMV and its related viruses. These values are far below the current species demarcation criteria (CP aa sequence identity less than about 80%; and nt sequence identity less than 76% either in the CP or over the whole genome), set by the International Committee on Taxonomy of Viruses [3, 4], which supports the designation of *Habenaria* mosaic virus as a novel species in the genus *Potyvirus*.

An ML phylogenetic tree was constructed by PhyML with GTR+I+G as the best-fit model selected by jmodeltest 2, based on an alignment of the nucleotide sequences of polyprotein ORF from the 78 selected potyviruses (see supplementary Table S1). As shown in Fig. 2, HaMV is clustered within the ChiVMV subgroup (subgroup 5), which includes ChiRSV, ChiVMV, PVMV, TVBMV and WTMV [9]. A similar ML-tree was obtained using an alignment of the polyprotein aa sequences (supplementary Fig. S2 and Table S1). These phylogenetic analyses suggest that HaMV might have a common ancestor with the ChiVMV subgroup members, which are mainly distributed in Asian countries and whose primary hosts are solanaceous plants [19].

A search for possible recombination events using the Recombination Detection Program (RDP3) [17] revealed that no detectable recombination between HaMV and WMV, both of which could infect *H. radiata* (data not shown).

In conclusion, this is the first report of the complete nt sequence and genome organization of a potyvirus isolated from a terrestrial orchid, *H. radiata*. Sequence comparisons and phylogenetic analyses indicated that HaMV should be classified as a new member of the genus *Potyvirus*, within the previously established ChiVMV

subgroup (subgroup 5).

Acknowledgments

We are grateful to Kouji Mitsuata for technical assistance. We thank Dr. Narinobu Inouye for helpful advice. This work was partially supported by a Grant-in-Aid for Scientific Research [KAKENHI] from the Japanese Ministry of Education, Culture, Sport, Science and Technology, and by Yomogi Inc.

References

1. Abascal F, Zardoya R, Posada D (2005) ProtTest: selection of best-fit models of protein evolution. *Bioinformatics* 21:2104–2105
2. Adams MJ, Antoniw JF, Beaudoin F (2005) Overview and analysis of the polyprotein cleavage sites in the family *Potyviridae*. *Mol Plant Pathol* 6:471–487
3. Adams MJ, Antoniw JF, Fauquet CM (2005) Molecular criteria for genus and species discrimination within the family *Potyviridae*. *Arch Virol* 150:459–479
4. Adams MJ, Zerbini FM, French R, Rabenstein F, Stenger DC, Valkonen JPT (2011) Family *Potyviridae*. In: King AMQ, Adams MJ, Carstens EB, Lefkowitz EJ (eds) *Virus taxonomy: ninth report of the international committee on taxonomy of viruses*. Elsevier, New York, pp 1069–1089
5. Chiba S, Kondo H, Tani A, Saisho D, Sakamoto W, Kanematsu S, Suzuki N (2011) Widespread endogenization of genome sequences of non-retroviral RNA viruses into plant genomes. *PLoS Pathog* 77:e1002146. doi:10.1371/journal.ppat.1002146
6. Chung BYW, Miller WA, Atkins JF, Firth AE (2008) An overlapping essential gene in the *Potyviridae*. *Proc Natl Acad Sci USA* 105:5897–5902
7. Darriba D, Taboada GL, Doallo R, Posada D (2012) jModelTest 2: more models, new heuristics and parallel computing. *Nat Methods* 9:772–772
8. Gara IW, Kondo H, Maeda T, Tamada T, Inouye N (1997) Stunt disease of *Habenaria radiata* caused by a strain of watermelon mosaic virus 2. *Ann Phytopathol Soc Jpn* 63:113–117
9. Gibbs A, Ohshima K (2010) Potyviruses and the digital revolution. *Ann Rev Phytopathol* 48:205–223
10. Guindon S, Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* 52:696–704
11. Guindon S, Dufayard JF, 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. *Syst Biol* 59:307–321
12. Inouye N, Maeda T, Mitsuhashi K, Gara IW (1998) *Habenaria* mosaic potyvirus, a new member of potyvirus from *Habenaria radiata* orchid. *Bull Res Inst*

- Bioresour Okayama Univ 5:155–168
13. Katoh K, Toh H (2008) Recent developments in the MAFFT multiple sequence alignment program. *Brief Bioinform* 9:286–298
 14. Kondo H, Chiba S, Toyoda K, Suzuki N (2013) Evidence for negative-strand RNA virus infection in fungi. *Virology* 435:201–209
 15. Kozak M (1989) The scanning model for translation: an update. *J Cell Biol* 108:229–241
 16. Lopez-Moya JJ, Wang RY, Pirone TP (1999) Context of the coat protein DAG motif affects potyvirus transmissibility by aphids. *J Gen Virol* 80:3281–3288
 17. Martin DP, Lemey P, Lott M, Moulton V, Posada D, Lefevre P (2010) RDP3: a flexible and fast computer program for analyzing recombination. *Bioinformatics* 26:2462–2463
 18. Mitsukuri K, Arita T, Johkan M, Yamasaki S, Mishiba K, Oda M (2009) Effects of type of explant and dark preconditioning on bud formation in *Habenaria radiata* (Thunb.) *in vitro*. *Hortscience* 44:523–525
 19. Rabelo FDC, Nicolini C, Resende RD, de Andrade GP, Pio-Ribeiro G, Nagata T (2013) The complete genome sequence of a Brazilian isolate of yam mild mosaic virus. *Arch Virol* 158:515–518
 20. Talavera G, Castresana J (2007) Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Sys Biol* 56:564–577
 21. Turpen T (1989) Molecular cloning of a potato virus Y genome : nucleotide sequence homology in non-coding regions of potyviruses. *J Gen Virol* 70:1951–1960
 22. Urcuqui-Inchima S, Haenni AL, Bernardi F (2001) Potyvirus proteins: a wealth of functions. *Virus Res* 74:157–1755

Figures

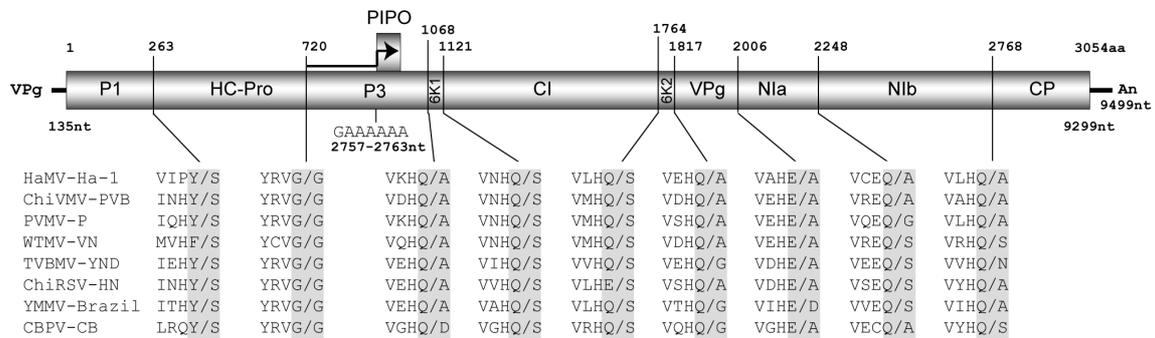


Fig. 1. Schematic representation of the genome organization of HaMV Ha-1. The 5'- and 3'-untranslated regions (UTRs) are shown as bold lines. Two highly conserved blocks, the potyboxes 'a' (14 ACACAAaCAU 25) and 'b' (42 UCAAGCA 49) [21], are present in the 5'-UTR. The box represents a large ORF, encoding a viral polyprotein. Six polymorphic sites at positions 297 (C/T), 736 (G/A), 1,436 (C/T), 4,995 (T/C), 6,314 (G/A), and 8,466 (T/C) were identified (data not shown). Of those, three sites 297 (P55S), 736 (G201E) and 8,466 (S2,778P) were non-synonymous substitutions, whereas the rest were synonymous. Numbers below the genome represent the first and the last nucleotide (nt) position of the ORF and the full-length genomic sequence excluding the poly (A) tail. The first initiation codon within the genome, at position 135, is within a favorable context for initiation (131 AGAGAUGGC 139 , initiator codon underlined) as described by Kozak [15]. The polyprotein is predicted to encode ten mature proteins: P1 (first protein/protease), HC-Pro (helper component protease), P3 (third protein), 6K1 (first 6-kDa protein), CI (cytoplasmic inclusion protein), 6K2 (second 6-kDa protein), VPg (viral protein genome-linked), NIa (nuclear inclusion a protein/protease), NIb (nuclear inclusion b protein) and CP (coat protein). The numbers above the genome indicate the amino acid (aa) position of the predicted proteolytic cleavage sites by the viral proteinases. The aa residues around the cleavage site are indicated below the diagram in comparison with those of closely related potyviruses (see Supplementary Fig. S1). A small ORF created by ribosomal frameshift or transcriptional slippage is shown in the figure above the P3 region as a small box (PIPO) [6]. The position and nt sequences of the heptanucleotide motif are indicated.

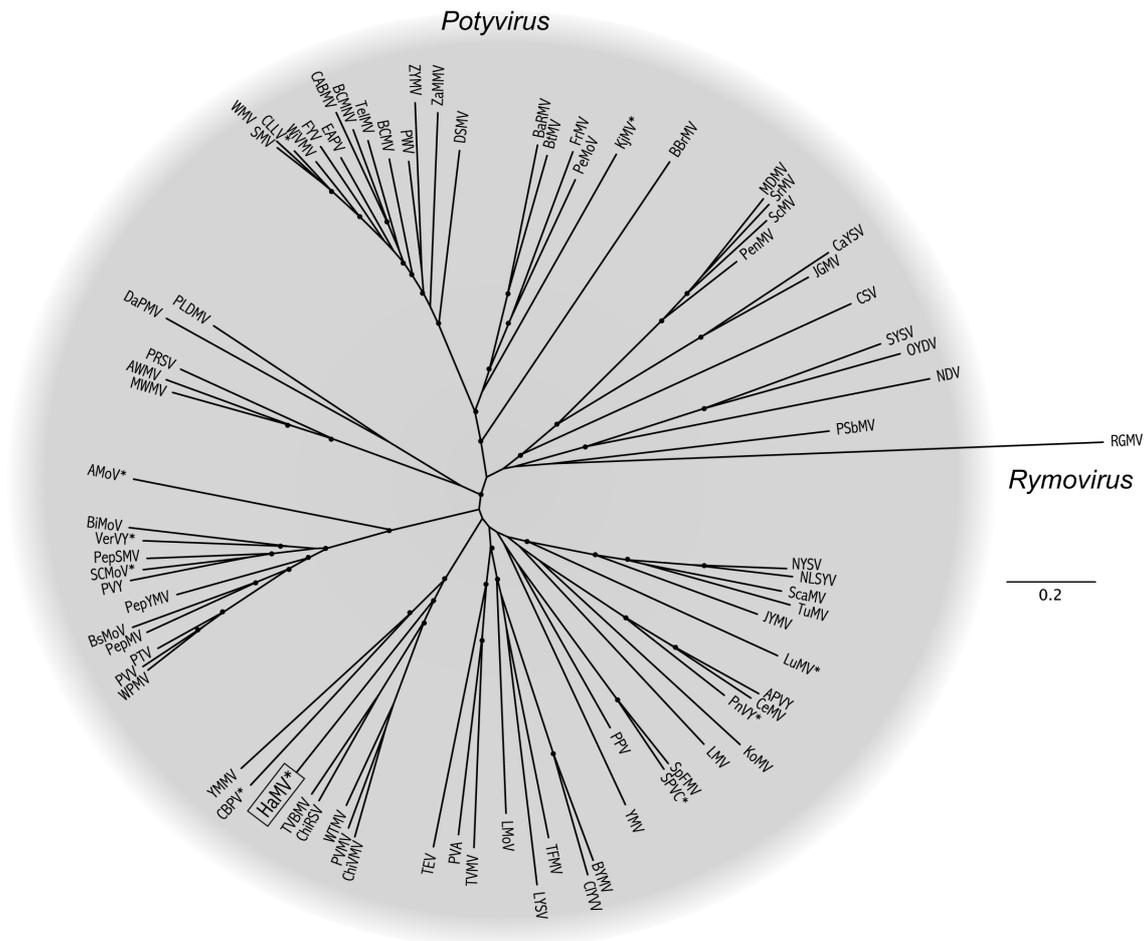


Fig. 2. Phylogenetic relation of HaMV and other potyviruses. A maximum-likelihood (ML) tree was constructed using PhyML 3.0, based on the multiple nucleotide sequence alignment of the viral polyprotein ORF. Virus names and GenBank accession numbers of their genome sequences (presented as acronyms) are listed in Supplementary Table S1. Viruses with asterisks indicate unassigned species. RGMV, genus *Rymovirus*, is used as an outgroup. 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).

Supplementary Table S1. GenBank/Refseq accession numbers of sequences compared in Figs. 1 and S1.

Virus names-isolate names, abbreviation (GenBank/Refseq accession no.)
Genus <i>Potyvirus</i>
Algerian watermelon mosaic virus-Algeria: H4, AWMV-H4 (NC_010736)
Apium virus Y-USA: Ce, ApVY-Ce (NC_014905)
Banana bract mosaic virus-Philippines, BBrMV-PH (NC_009745)
Basella rugose mosaic virus-Taiwan:AC, BaRMV-AC (NC_009741)
Bean common mosaic necrosis virus-USA:NL-3, BCMNV-NL3 (AY282577)
Bean common mosaic virus (=blackeye cowpea mosaic virus), BICMV (NC_003397)
Bean yellow mosaic virus-MB4, BYMV-MB4 (NC_003492)
Beet mosaic virus-Wa, BtMV-Wa (NC_005304)
Bidens mottle virus-Taiwan:B12, BiMoV-B12 (EU250210)
Brugmansia suaveolens mottle virus-Brazil, BsMoV-BR (NC_014536)
Canna yellow streak virus-UK, CaYSV-UK (NC_013261)
Celery mosaic virus-USA, CeMV-USA (NC_015393)
Chilli ringspot virus-China:HN/14, ChiRSV-HN/14 (NC_016044)
Chilli veinal mottle virus-pepper vein banding virus, ChiVMV-PVB (NC_005778)
Clover yellow vein virus-30, CIYVV-30 (NC_003536)
Cocksfoot streak virus-Germany, CSV-DE (NC_003742)
Cowpea aphid-borne mosaic virus-Zimbabwe, CABMV-ZM (NC_004013)
Daphne mosaic virus-Czech Republic, DapMV-CZ (NC_008028)
Dasheen mosaic virus-China: M13, DsMV-M13 (NC_003537)
East Asian passiflora virus-Japan:AO, EAPV-AO (NC_007728)
Freesia mosaic virus-South Korea, FreMV-KO FreMV-Kr_GU214748
Fritillary virus Y-China:Pan'an, FVY-PA (NC_010954)
Japanese yam mosaic virus-mild, JYMV-mild (NC_000947)
Johnsongrass mosaic virus-Australia, JGMV-AU (NC_003606)
Konjac mosaic virus-Japan: F, KoMV-F (NC_007913)
Leek yellow stripe virus-China:Yuhang, LYSV-YH (NC_004011)
Lettuce mosaic virus-E, LMV-E (NC_003605)
Lily mottle virus-China:Sb, LMoV-Sb (NC_005288)
Maize dwarf mosaic virus-Bulgaria, MDMV-BU (NC_003377)
Moroccan watermelon mosaic virus-Tunisia:TN05-76, MWMV-TN05-76 (EF579955)
Narcissus degeneration virus-China:Zhangzhou, NDV-ZZ (NC_008824)
Narcissus late season yellows virus- China: Zhangzhou2, NLSYV-HZ2 (JQ326210),
Narcissus yellow stripe virus-China:Zhangzhou, NYSV-ZZ (NC_011541)
Onion yellow dwarf virus-China:Yuhang, OYDV-YH (NC_005029)
Papaya ringspot virus-Hawaii, PRSV-HAT (X67673)
Passion fruit woodiness virus-Australia:MU2, PWV-MU2 (NC_014790)
Pea seed-borne mosaic virus-DPD1, PSbMV-DPD1 (NC_001671)
Peanut mottle virus-M, PeMoV-M (NC_002600)
Pennisetum mosaic virus-China:B, PenMV-B (NC_007147)
Pepper mottle virus-California, PepMoV-Cal (NC_001517)
Pepper severe mosaic virus-South Korea, PepSMV-KO (NC_008393)
Pepper veinal mottle virus-P, PVMV-P (NC_011918)
Pepper yellow mosaic virus-Brazil:Pi-15, PepYMV-Pi15 (NC_014327)
Peru tomato mosaic virus-Peru:PPK13, PTV-PPK13 (NC_004573)
Plum pox virus-NAT, PPV-NAT (NC_001445)
Potato virus A-Hungary: B11, PVA-B11 (NC_004039)

Pokeweed mosaic virus-MS-FR03, PkMV-MS-FR03 (NC_008393)
 Potato virus V-UK:DV 42, PVV-DV42 (NC_004010)
 Potato virus Y-France:O, PVY-O (NC_001616)
 Scallion mosaic virus-China:Hangzhou, ScaMV-HZ (NC_003399)
 Shallot yellow stripe virus-China:ZQ2, SYSV-ZQ2 (NC_007433)
 Sorghum mosaic virus-China:Xiaoshan, SrMV-Xiaoshan (NC_004035)
 Soybean mosaic virus-N, SMV-N (NC_002634)
 Sugarcane mosaic virus-China:Hangzhou, SCMV-HZ (NC_003398)
 Sweet potato feathery mottle virus-S, SPFMV-S (NC_001841)
 Telosma mosaic virus-Vietnam:Hanoi, TelMV-VN (NC_009742)
 Thunberg fritillary mosaic virus-China: Ningbo, TFMV-NB (NC_007180)
 Tobacco etch virus-HAT, EV-HAT (NC_001555)
 Tobacco vein banding mosaic virus-China:YND, TVBMV-YND (NC_009994)
 Tobacco vein mottling virus-S, TVMV-S (U38621)
 Turnip mosaic virus-UK1, TuMV-UK1 (NC_002509)
 Watermelon mosaic virus-Fr, WMV-Fr (NC_006262)
 Wild potato mosaic virus-Peru, WPMV-Peru (NC_004426)
 Wild tomato mosaic virus-Vietnam: Laichau, WTMV-VN (NC_009744)
 Wisteria vein mosaic virus-China:Beijing, WVMV-BJ (NC_007216)
 Yam mild mosaic virus-Brazil, YMMV-Brazil (NC_019412)
 Yam mosaic virus-Ivory Coast, YMV-IC (NC_004752)
 Zantedeschia mild mosaic virus-Taiwan, ZaMMV-TW (NC_011560)
 Zucchini yellow mosaic virus-Taiwan:TN3, ZYMV-TN3 (NC_00322)

Unassigned related viruses

Arracacha mottle virus- Brazil:C-17, AMoV-C-17 (NC_018176)
 Calla lily latent virus-Taiwan:m19, CLLV-m19 (EF105297)
 Christmas bell potyvirus-Taiwan:CB, CBPV-CB (EF427894)
 Keunjong mosaic virus-South Korea:Cheongwon, KjMV-CW (NC_016159)
 Lupin mosaic virus-Czech Republic:Lu2, LuMV-Lu2 (NC_014898)
 Panax virus Y-China:2, PhVY-2 (NC_014252)
 Sunflower chlorotic mottle virus-Argentina:Common SCMoV-C (NC_014038)
 Sweet potato virus C-Peru:C1, SPVC-C1 (NC_014742)
 Verbena virus Y-USA:Michigan, VerVY-Michigan (NC_010735)

Rymoviruses (outgroup):

Ryegrass mosaic virus, RGMV (NC_001814).

Supplementary Fig. S1. Multiple amino acid sequence alignment of the polyprotein of HaMV and closely related potyviruses (ChiVMV subgroup and their two relatives). The putative catalytic residues, cleavage sites and other conserved motifs of potyviral polyproteins are highlighted in blue, green and red, respectively.

HaMV* VRVLEKIYTEDLEASWRELQFVEKCYIMLQRLRWRRIIVELSESASISFKKVFHEHCSTG
ChiVMV-PVB LHIMEKIYKEDLRALWCELSWSEKLRFKLHVLRSRKQYIRDSREIISGTLKEALHCSTRF
PVMV-P LSVLEKICRDDLDQWRDLSLSAKSRIWQSLRSRKSCTGGLRDTACSIKRVFESSTAY
WTMV-VN LHLMEKIYKADLEAQWCELSWLEERSYVIYYTFKSRIQCMRDLSQDKSQQILKQTFKCSSTAF
TVBMV-YND LALIEKIYQDLEELWQGLPPEKCSILWQLRHRRKLQVELGENTQDLSKALVTFKH
ChiRSV-HN/14 FHLVEKIYQEDLEAQWRELPLFQKLYFTAWQLRHRRALVEMLGEKDTKEHCKLLHSLKRN
YMMV-Brazil* NSLTEKIYVEELKASWRELPLSEKLLSTCRSLQVCERFGRKCREERQETLRNVSKSVTQY
CBPV-CB* NQFVEKIYQDELAMCWDDELPLSSKLSATIVRYKWRAYIENFSKTIKDGTLSTPCKLYMKP
.*** : * * * : .

HaMV* LHLAARPIVKIAKCTDKFSAIVRSTHTRLLSGFIYGFRCVFRDLFTFVQVLAINIFLT
ChiVMV-PVB VQEKIVLIPTPLRYVSGKIGCGIAKVRKNVWSIYGRAVHNFRDALRFIHTLAILLSILL
PVMV-P VQGAQKIPAPFMYMLSKASKSASIRSSVLRNMYSIYAFHDAFQFIHTLAILLSVLLT
WTMV-VN VQDRMKVIPNSVQSVCSKSVCIKSIHRVRYKRYCAVNTFSDAFQFLQTMALISILLS
TVBMV-YND AQDNFDRILSPIHRSYVVTGAPMRLRGVRCVVGTFSYLSDLLRFVQLAILSVLLA
ChiRSV-HN/14 ASHCLRKTLPFKAGYSKCTQVPGFVHKMRCKLLHLSYFFSDALRFIQLVAISILMS
YMMV-Brazil* LCGGLTKMKTGVKSTKCTYKVVMSLRTVGGAFNINLVYTPFLRTARVVAISLFLS
CBPV-CB* LYMAKEKFKQAAQRGRNRMSTYREMFQRYFMQSFMSRLRYIAPELSTLVRTITIFSTLIG
: : : : : .

HaMV* ILDSLRLRSAYIANARQVYMRERQNRDKLEKLYNLIKCKLGVEPTFEEYKEFVAGVNP
ChiVMV-PVB VYTNIVDIKNKHKMNLHIEEQEMKIQSRINKHYNDLTLHNSPPSEVEFIQYVREHDP
PVMV-P IYTNILNIRNLNKRALLLVKEADRLKSERIERCFEISKLLNGAPTEQQVFQPIQEKDP
WTMV-VN VFNANLIDIKNYRNSVRIKDEKMDLEDKSIFKHYTDLKIKNGVKPSEDEFSEYKLERDP
TVBMV-YND IFNNIYAFTCQYREFKMEASREEEKLESKLDKLYAHLKYKLGESPTFEFFRQFVSACNP
ChiRSV-HN/14 ILQQLYDSMVYRSYKILAIRQEDSKDEALERLHSHLYKLGVLPTIDEFCCKVQEQSP
YMMV-Brazil* VFSKLRQIVNEQLHQRALLRDVQLATNWKIETHYETLTKKLQNTPTIDEFAQYLKESNA
CBPV-CB* MLNAAKMVVVIQOHKGRMAQQKEEQFAALEKLYHYICGKIGDQPTYEEFYEFVKGTNS
: : : : : .

P3/6K1

HaMV* ELAKQLESSEE-LEVKHQAANKRESETRLEQIVAFIALVLMVFDNERSDCVYRVMNKLKNV
ChiVMV-PVB SALAYWSTQE--KSVHQASKRPEAKLEQIVAFVALLMMVFDTRSDCVYKVLNKLKNV
PVMV-P QALQYSEQN--DGVKHQASKRPAEAKLEQIVAFVALLMMVFDTRSDCVYKVLNKLKNV
WTMV-VN DAFVHWFGKD--LKVHQASKRPEAKLEQIVAFIALVLMVFDGDRGDCVYKVLNKLKNV
TVBMV-YND QLVEFLEVSYA-PVVEHQAVKRASEVRLEQIVAFIALVMMIFDNDRSDCVYRVLNKFKN
ChiRSV-HN/14 ELLDHMGYHGENVEVHQ-AKRESEVRLEQIVAFVALLMMIFDNDRSDCVYRVLNKFKN
YMMV-Brazil* ELYKEFVEVYQLPVEHQ-AKRESEVRLEQIVAFIALVMMVFDNERSDCVYKVLNKLKNL
CBPV-CB* SLLCILEFSFS-EFVGHQDFKSENK-RIEQILAFVVLIMMLIDAERSDCVYKILNKFQV
.*** : : : : : .

6K1/CI

HaMV* MSVAEQD-VNHQSMDDDEMAFDD-NATISFELECEDPVRAYPSSSTLEQWWDNQLALNRT
ChiVMV-PVB MGVVDNDAVNHQSLDTILENFDERDEMIIEFEITAPDAKSIAYKSSTFQWWDNQLALNRT
PVMV-P MGVVDNDAVNHQSLDTILENFDERDEMIIEFEITAPDAKSIAYKSSTFQWWDNQLALNRT
WTMV-VN MGSVDNEAVNHQSLDTIVENFEETNEMIEFEITAPDAKSIAYKSSTFQWWDNQLALNRT
TVBMV-YND VGVADQEV-IHQSLDDIKTNFDESNETIDFELVTGETAPTYPKSTVFSDDWNSVNLNMGRT
ChiRSV-HN/14 VSTAEQDVV-HQSLDDIQTNFEERNETIDFELTDETPATPFKSTATFESWWDKQLEMGHT
YMMV-Brazil* MNTAEP--VAHQSLDDIIPIFEK-EQLIDFELDTHDSIYAYKSSTFQWWDNQLALNRT
CBPV-CB* IRSIEP--VGHQSLDDIAPDFEF-NETIEFSLDTEGLPELHLKLTTFSGWWDNQLALNRT
: : : : : .

GSGKS P

HaMV* IPHYRTEGYFMEFRANCAQVINEIVHNEHKDILLRGAVGSGKSTGLPAGLSTRGKVLVLL
ChiVMV-PVB ISHYRTEGHFVEFTRDTSQVADSIANSMDKDFLVRGAVGSGKSTGLPSALCKRGRVLLV
PVMV-P ISHYRTEGHFIQFTRETSQVADSISSSEIRDFLIRGVPVSGKSTGLPSALCKRGRVLLV
WTMV-VN ISHYRTEGHFIEFSRDRAAQVANEISTSDIRDYLRGAVGSGKSTGLPHLLCKKGVLLV
TVBMV-YND IPHYRTEGHFMEFRDVAATVVSNIIQSEKRDFLIRGAVGSGKSTGLPAQLAKKGVLLV
ChiRSV-HN/14 VPHYRTEGHFMEFRDVAATVVSNIIQSEKRDFLIRGAVGSGKSTGLPAQLAKKGVLLV
YMMV-Brazil* IPHYRNEGHFMGFRSAGTASEIACSDHKDILLRGAVGSGKSTGLPFLSKHGHVLLV
CBPV-CB* TSHYRTEGHFLEFRATCASVANQIITSEHTNFLIRGAVGSGKSTGLPNLARQGRVLLV
.*** : : : : : .

HaMV* ESTKPLSRNVNQLRQDPFHLSPSLMMDRSTFGSTPITIMTSGYAFHYFANNARKLHDY
ChiVMV-PVB EPTRPLAENVHAQLSASPFFHLNPTLMRNKSVFGSTPITVMTSGYALHYLANDAQRKLF
PVMV-P EPTRPLAENVHAQLSASPFFHLNPTLMRNKSVFGSTPISVMTSGYALHYLANNAHRLKLF
WTMV-VN EPTRPLAENVHSQLSQPFHHLNPTLMRNKSVFGSTPITIMTSGYALHYLANNSHRLQEF
TVBMV-YND EPTRPLSENVTQLRSQPFHLSPTLLMRNSHFGSTPISIMTSGYALHYLANSGTALSTF
ChiRSV-HN/14 EPTRPLSQNVNQLRSQPFHLSPTLLMRNSHFGSTPISIMTSGFALHYLANSGTALSTF
YMMV-Brazil* EPTRPLVENVYQLRGAPPHASPTMMRHATSFGSPVPTIMTSGFVIHYLANNSKISSY
CBPV-CB* ESTRPLAENVYQLQCSPPFHLNPTLMMDRVSFGASPIVMTSGFALHYFANNYNKLQDF
* : : : : : .

DE H

HaMV* QFIMIDECHVLDANAMAFRSLLEHEYQGKIIVKSATPPGREVEFTTQHKVEIRIEDSLS
ChiVMV-PVB SFILFDECHVLDASAMAFKSLLDVREFEGKILKVSATPPGRETEFTTQFPVKLRTEHLS
PVMV-P SFILFDECHVLDASAMAFRSLLDADRAFEKILKVSATPPGRETEFTTQFPVKLRTEHLS
WTMV-VN AFIMFDECHVLDASAMAFRSLLDADRAFEKILKVSATPPGRETEFTTQFPVKLRTEHLS
TVBMV-YND DFIMDECHVLDANAMALYSLLHDREYTGKILKVSATPPGREVEFTTQFPVKLRTEHLS
ChiRSV-HN/14 DYVILDECHVLDANAMALYSLLERDFNGKILKVSATPPGREVEFTTQFPVELRVEESLS
YMMV-Brazil* GYVILFDECHVEDANAMALRCLLDSVAFDGVKILKVSATPPGREVEFTTQFPVELRTEHLS
CBPV-CB* KYIIDECHVQDATAIAFNLLVARSFEGKILKVSATPPGRETEFTTQFPVELRTRDSLT
: : : : : .


```

vHaMV*          VDATNSFGKQAQAKGSETDGSSSRDGQQKNSVRRREGSTPIAIP---DRDINAGTTGTF
ChiVMV-PVB     VDAGSVKGEDSSSKSAD--KQATEKKNKVSQQAQQSRQSEMEVPPQVRDRDQVNVGTSGTF
PVMV-P         IDAG-----KSSGKEVTASKQPTTSQPTTEAKTVL-DRDQVNVGTSGTF
WTMV-VN        VDAG-----KNTGLVKDPTPNKDKQVMQSQPPTK-----DKDQVNVGTSGTF
TVBMV-YND      VDAG-----KNAQSNQKQRNAQAGNTSSGVVKDK----DKDQVNVGTSGTF
ChiRSV-HN/14   VDAG-----KSTQSSEKPPANFASVSKVVAETPR---DRDQVNVGTAGTF
YMMV-Brazil*   LDAG-----QFSGKQPQSQGTNSSEGS-----GRDQVNVGTSGTF
CBPV-CB*       VDAG-----VLGQSKGTAGQSGSGSQAQMR-----DKDQVNVGTSGTF
: **          :                               : : * . * * * *

HaMV*          TVPKLKGMSKLTIPKVKGVVNLQHLQYTPDQEKLSNTFATDEQFAIYWNVKSDYE
ChiVMV-PVB     TIPRLKGI FSKLTI PKVKT KAVVNL EHL LDYAPDQI HLSNTRALQSQFASWYEGVKN DYD
PVMV-P         TIPRLKGMSSKLTLPKVSSQSVVNL EHL LNYKPDQV HLSNTRALQSQFASWYNGVKN DYD
WTMV-VN        SIPRLKGISSKLTLPKTSAGMVVNL EHL LEYKPDQI HLSNTRALNSQFQSWYDGVKN DYD
TVBMV-YND      SIPRLGLSTKLNLPRIKGEVNLQHLLEYTPDQVLSNTRALNSQFASWYTVGVKN DYD
ChiRSV-HN/14   SVPRKGISSKLNLPINKAVINLDHLLKYTPDQVLSNTRALNSQFASWYEGVKT DYD
YMMV-Brazil*   SIPRIKTPMSKLTLPKLGKVLVNL EHL VEYEPDQTDI SNKRASQEQLGQWVEAVKTSYD
CBPV-CB*       AVPRLRHLSKLSVPKLKGESVNL EHL LHYQPNQDRI SNTRATDSQFQLWYDGVKSDYD
: * : : :      : * : * : . : * : * : * * : * : * * : * : * : * : * : * :

HaMV*          VSDDEMQUIILNGLMVWCIENGTSPNLGVMWMDGDEQIYPIKPLLDHAKPTFRQIMHH
ChiVMV-PVB     VTDEQMQIILNGLMVWCIENGTSPNINGVWVMDGDEQVEYPIKPLLDHAKPSFRQIMAH
PVMV-P         VDDEQMKIILNGLMVWCIENGTSPNINGMVMVDGEEQIEYPIKPLLDHAKPSFRQIMAH
WTMV-VN        VDDEQMKIIMNGLMVWCIENGTSPNINGMVMVDGEEQVEYPIKPLLDHAKPSFRQIMAH
TVBMV-YND      LDDAQMEIVL NGLMVWCIENGTSPNLGMWVMDGDEQVEYPIKPLLEHAKPTFRQIMAH
ChiRSV-HN/14   LNSQMEIVL NGLMVWCIYVGTSPNLGMWVMDGDEQVEYPIKPLLDHARPTFRQIMAH
YMMV-Brazil*   VDDEQLKIILNGLMVWCIENGTSPNINGFWMIEDGEQIKFPLKPIVENAKPTLRQIMAH
CBPV-CB*       VNDEEMKIILNGLMVWCIENGTSPNINGFWMLENDQIEFFPIKPLLDHARPTFRQIMSR
: * : : * : * : * * * * * * * * * * * * * * * * * * * * * * * * * * * *

HaMV*          FSNLAEAYIEKRNYSYMPRYGRNRLTDMSLARYAFDFYAITSRTPERAKEAHMQMKA
ChiVMV-PVB     FSNLAEAYIEKRNSEKPYMPRYGLQRNLTDMSLARYAFDLYEMTSKTPVRAREAHIQMKA
PVMV-P         FSNLAEAYIEKRNSEKSYMPRYGLQRNLTDMSLARYAFDFYEMTSKTPVRAREAHIQMKA
WTMV-VN        FSNLAEAYIERNSEKPYMPRYGLQRNLTDMSLARYAFDFYEMTSKTPSRAREAHIQMKA
TVBMV-YND      FSNLAEAYIEKRNAEKPYMPRYGLQRNLTDMTLARYAFDFYEMTSKTPVRAREAHIQMKA
ChiRSV-HN/14   FSNLAEAYIEKRNAEKPYMPRYGLQRNLTDMTLARYAFDFYEMTSKTPARAREAHIQMKA
YMMV-Brazil*   FSDLAEAYIEKRNKAYVPGYGLQRNLDYSLARYAFDFYEMTSKTPVRAREAHMQMKA
CBPV-CB*       FSDLAEAYIEKRNFERAYMPRYGLQRNLTDMSLARYAFDFYEMTSKAPARAREAHIQMKA
* : * * * * * * * * * * * * * * * * * * * * * * * * * * * *

HaMV*          AALRNTSSRMFGLDGKVGTVQEDTERHTAEDVNRNMHLLGVRGV
ChiVMV-PVB     AALRGVSNRMFGLDGRVGTQEEDTERHTAEDVNRNMHLLGVRGL
PVMV-P         AALRNANRMFGLDGKVGTVQEDTERHTAEDVNRNMHLLGVRGV
WTMV-VN        AALRNANRMFGLDGKVGTVQEDTERHTAEDVNRNMHLLGVRGV
TVBMV-YND      AAVANSKNNLFGLDGNVSTKEENTERHTATDVRNMHLLGVRGV
ChiRSV-HN/14   AAVMHAKNNMFGLDGNVSTKEENTERHTATDVRNRMHLLGVRGV
YMMV-Brazil*   AALRNTRRLFGLDGSVGNNDENTERHTSDVNRDMHSLGVRNI
CBPV-CB*       AALRNTRKRMFGLDGKVGTVQEDTERHTSDVVTGGIHSLLGVRGL
* * : : * * * * * * * * * * * * * * * * * * * * * * * *

```

Supplementary Fig. S2. Phylogenetic relation of HaMV and other potyviruses. A maximum-likelihood (ML) tree was constructed using PhyML 3.0, based on the multiple amino acid sequence alignment of the full-length viral polyproteins, as described previously with minor modification [5, 14]. Subgrouping of the clades (eleven groups) is based on the report of Gibbs and Ohshima [9]. Distantly related species that formed well-supported monophyletic clades were collapsed into a triangle (subgroup 8–11). Virus names and GenBank accession numbers of their genome sequences (presented as acronyms) are listed in Supplementary Table S1. Viruses with asterisks indicate unassigned species. RyMV, genus *Rymovirus*, is used as an outgroup. 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).

