

The complete genome sequence of Piper yellow mottle virus (PYMoV)

U. Hany · I. P. Adams · R. Glover ·
A. I. Bhat · N. Boonham

Received: 27 March 2013 / Accepted: 5 July 2013 / Published online: 5 September 2013
© Crown Copyright 2013

Abstract This study reports the first complete genome sequence of Piper yellow mottle virus (PYMoV, KC808712) identified in black pepper. The genome is 7,622 nucleotides long, possessing four open reading frames (ORFs). ORF1, ORF2 and ORF4 of PYMoV are reported as hypothetical proteins of unknown function with a predicted molecular mass of 15.7, 17.1 and 17.9 kDa, respectively. ORF3 of PYMoV encodes a polyprotein of 218.6 kDa and consists of a viral movement protein (MP), trimeric dUTPase, zinc finger, retropepsin, RT-LTR, and RNase H. Detailed PYMoV genome analysis confirmed that it is a member of the family *Caulimoviridae*, genus *Badnavirus*. Fragments of two additional novel sequences resembling those found in members of the family *Caulimoviridae* were also identified in the black pepper sample, and the viruses from which they were derived were tentatively named *Piper* DNA virus 1 and 2.

Introduction

Black pepper (*Piper nigrum* L, Piperaceae), originating in the tropical evergreen forests of Western Ghats of India, is one of the most ancient spice crops cultivated for its berries [12]. Lockhart et al. [9] reported a new mealybug-transmitted badnavirus, Piper yellow mottle virus (PYMoV), infecting black pepper in Malaysia, Thailand, the Philippines and Sri Lanka. Subsequently, its occurrence was also

reported in India [2, 7]. The disease caused by this virus is characterized by leaf distortion, mosaic, mottling, shortened internodes, and poor filling of spikes, leading to reduced yield. Since black pepper is propagated vegetatively, the presence of viruses is a matter of concern, as viruses could be transmitted to subsequent generations. The perennial nature of the crop further aggravates the situation, as the virus inocula remain in the field for long periods, rendering the whole plantation vulnerable to disease spread by mealybug vectors. To date, only partial sequences of ORF I (700 bp) and ORF III (600 bp) of PYMoV are available [4, 7]. In this study, the full genome sequence and genome structure of PYMoV were determined using next-generation sequencing. Two other distinct sequences resembling those found in members of the family *Caulimoviridae* were also identified in the infected plant.

Viral particles and sequencing

A PYMoV-infected black pepper plant (cv. Karimunda) from Calicut, Kerala, India, was used as the source of virus material. The plant was showing symptoms such as yellow mottling, deformation, reduction in size, and curling of leaves with shortened internode length, leading to stunting of the plant. The symptoms observed were not different from those caused by previous PYMoV isolates. Fresh leaves from the infected plant were collected, freeze-dried immediately and stored at 4 °C until further use. Enrichment of virus particles was done using the method described by de Silva et al. [4]. Viral and plant DNA were isolated from the resulting virus-enriched fraction using a DNeasy Plant Mini Kit (QIAGEN, Chatsworth, USA) following the manufacturer's instructions. The presence of

U. Hany (✉) · I. P. Adams · R. Glover · N. Boonham
Food and Environment Research Agency, Sand Hutton, York,
UK
e-mail: Ummey.Hany@Fera.gsi.gov.uk

A. I. Bhat
Indian Institute of Spices Research, Calicut, India

the virions of PYMoV was confirmed by amplifying the genomic DNA with primers specific for ORF1 and ORF3 of PYMoV [7]. The DNA was sequenced using a Roche 454 GS-FLX Titanium genome sequencer according to the manufacturer's protocol. Sequencing was done on a quarter of a picotitre plate, which produced 316,640 reads averaging 360 bp in length. These reads were assembled using gsAssembler software (Newbler) version 2.5 (Roche), and the contigs and remaining singletons were subjected to BLAST [3] analysis against the GenBank nucleotide sequence database [1]. The comparisons showed that one of the larger contigs was made up of 324 individual reads with homology to PYMoV. To confirm that the 7,622-bp linear contig represented the whole circular genome of PYMoV, PCR primers (GCAGTACCAAGATGCCCTTGA and AGATTGGCGCAAGTTGCCT) were designed targeting the ends of the contig, and the original infected plant DNA was subjected to amplification using the primer pair. The amplified PCR product was sequenced by the Sanger method (Eurofins, Germany), and the sequence was shown to overlap and join the two ends of the linear genome sequence. An extra 59-bp sequence was found at the junction of the two genome ends and was added to the 5' and 3' ends of the sequence.

PYMoV genome

The complete viral genome of PYMoV consists of 7622 bp of double-stranded circular DNA (GenBank accession no. KC808712). Open reading frame analysis using Vector NTi 11 (Invitrogen) revealed four ORFs (Fig. 1). To determine the relationship of PYMoV to the badnaviruses sequenced to date, Mega 5 [15] was used to construct a phylogenetic tree for the polyprotein of PYMoV and different members of the family *Caulimoviridae*, including different badnavirus sequences available in the GenBank database (Fig. 2). The robustness of the trees was determined by performing bootstrap sampling of the multiple alignments (1,000 sets). The phylogenetic tree showed that PYMoV is closely related to cacao swollen shoot virus (CSSV), citrus yellow mosaic virus (CYMoV) and *Dioscorea* bacilliform virus (DBV). DNA-level comparisons revealed a very high level of homology between the PYMoV genome and all but two of the 14 PYMoV DNA sequences found in the GenBank database. Comparisons revealed that the sequence with accession number DQ836230 is similar to the *Piper nigrum* plant genome, and the other one with accession number DQ836233 showed no similarity to any published sequences. From the sequence similarity search, it is likely that these two sequences were misidentified as originating from a virus and are more likely to be of plant origin.

ORF1 of PYMoV (position 720-1,124) encodes a 135-aa putative protein of unknown function, with a predicted molecular mass of 15.7 kDa [6, 13]. ORF1 includes a conserved protein domain identified as pfam 07028 [5], which is a c106184 superfamily protein that is restricted to badnaviruses [10, 14]. ORF1 of PYMoV has 45 % identity to the ORF1 protein of fig badnavirus (YP_006273073). It also shares 42-44 % identity with different isolates of citrus yellow mosaic virus (CYMV) (AC055656) and cacao swollen shoot virus (CSSV) (CAE76626).

The ORF2 of PYMoV (position 1,124-1,585) encodes a putative protein of 154 aa with a predicted molecular mass of 17.1 kDa, which is slightly larger in size than those found in CYMV (NP_569152) and *Dioscorea* bacilliform virus (DBV) (AB147985.1) [13]. No putative conserved domains were identified in ORF2; however, comparisons revealed that the PYMoV ORF-2 is 44 % identical to that of CYMV (137 aa) (NP_569152) and 39 % identical to that of DBV (AB147985.1).

ORF3 of PYMoV (position 1,512-7,286) encodes a putative polyprotein of 1,925 aa with a predicted molecular weight of 218.6 kDa that is 50 % identical to the ORF3 of DBV (YP_001036293) and 55 % identical to that of citrus yellow mosaic virus (CYMV; NP_569153). A similarity search in the NCBI conserved domain database (CDD) [10] revealed that the amino acid sequences of ORF3 consists of regions specific for a viral movement protein (MP), trimeric dUTPase, zinc finger, retropepsin, RT-LTR, and RNase H. The PYMoV MP protein was identified as a movement protein [10], but it did not contain the expected conserved DxR motif [8]. The zinc-binding motif of the zinc finger protein is composed of CGGKGFHFGDDC in PYMoV. Next to the zinc finger protein is a retropepsin, which possesses a conserved domain consisting of a catalytic motif, an inhibitor binding site, a catalytic residue, and an active site flap. The conserved RT LTR domain found in PYMoV contains a putative active site, a putative nucleic-acid-binding site, and a putative NTP-binding site. At the end of OFR3 is an RNase-H-like protein consisting of an active site and a DNA/RNA-binding site.

ORF4 of PYMoV encodes a putative 158-amino-acid protein with 78 % identity to a hypothetical protein whose sequence was previously submitted as a partial PYMoV genome (ABI30246).

Considering sequence similarities of all the ORFs and predicted genome annotation data, PYMoV should be considered a member of a unique species of the genus *Badnavirus* [8]. No recombinant PYMoV sequences were detected, and although the presence of such recombinants cannot be excluded, the described DNA sequence is likely to be the predominant form present within the plant studied.

Fig. 1 Proposed genome arrangement of PYMoV, showing domains identified within ORF3

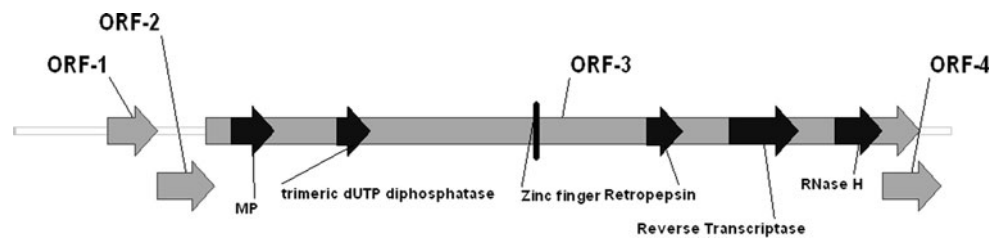
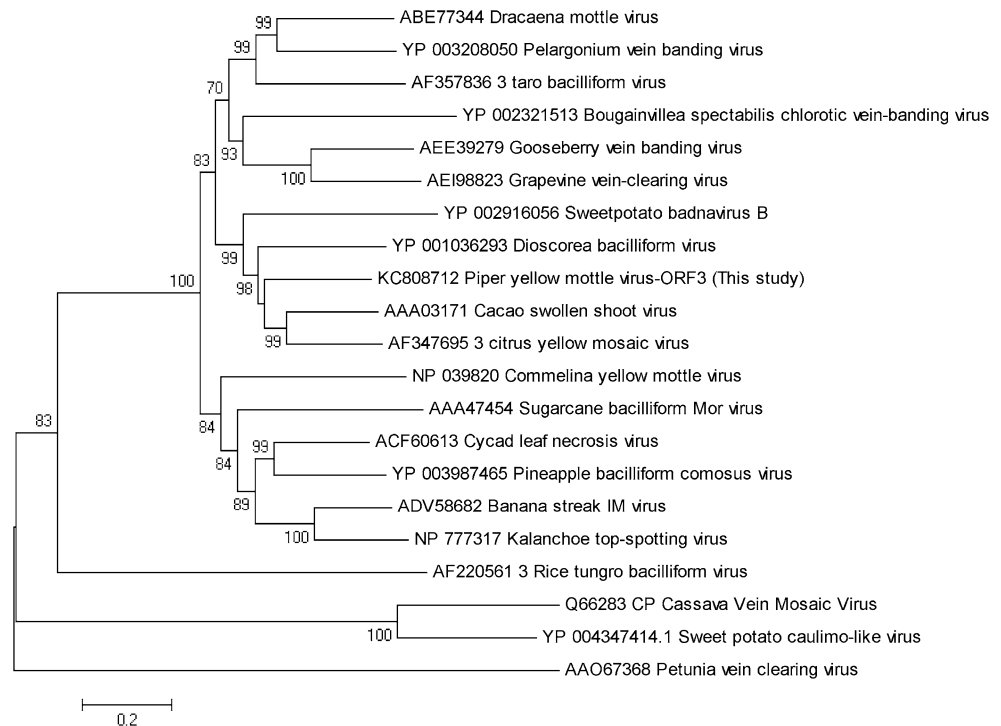


Fig. 2 A phylogenetic tree showing the relationship between PYMoV and 20 of the most closely related badnaviruses. The alignment was produced using full-length amino acid sequences of the polyprotein. Phylogenetic and molecular evolutionary analyses were conducted using MEGA 5. The tree was constructed using the neighbour-joining algorithm. The horizontal branch lengths are proportional to the genetic distance, and numbers shown at branch points indicate bootstrap values from 1000 replicates. Sequences of different members of the family *Caulimoviridae* (belonging to the genera *Badnavirus*, *Caulimovirus*, *Tungrovirus*, *Soymovirus*, *Cavemovirus*, *Petuvirus*) were obtained from the GenBank database



Two further contigs were identified as having virus-like properties. It is not known if these constitute infectious viruses or endogenous pararetroviral sequences.

The first sequence was 7,178 bp long and was derived from 213 reads, and a sequence similarity search against the GenBank Nr database revealed weak homology to rice tungro bacilliform virus, a tungrovirus. Further examination of the nucleotide sequence (JX40674) revealed a single ORF (position 2-6085). Alignment of the hypothetical polypeptide produced from this sequence with the P194 polyprotein of RTBV (AFQ62092) revealed 32 % homology, and alignment with the 216-kDa polyprotein of the type member of the genus *Badnavirus*, commelina yellow mottle virus (NC001343) (ComYMV) revealed 25 % homology. This comparison revealed that the new sequence contains RNA-binding, reverse transcriptase and RNase H domains, as described by Melberry et al. [11], and the alignment suggested that the putative polypeptide lacks an N-terminus. We have tentatively named this virus-like sequence *Piper DNA virus 1* (PDV-1).

A second 800-bp contig, derived from 20 reads (JX406742) was identified as having 44 % homology at the nucleotide level to ComYMV, 41 % homology to RTBV, and 70 % to PDV-1. This sequence was too short to analyse in any detail but may be a second distinct virus of the family *Caulimoviridae*, with the tentative name *Piper DNA virus 2* (PDV-2).

The levels of homology between PYMoV, PDV-1 and PDV-2 were quite low, and there was no evidence of any recombination between them.

The complete genome of the virus PYMoV has now been described, confirming that PYMoV is a distinct member of the genus *Badnavirus*. The presence of two further potential viruses of the family *Caulimoviridae* in the same sample opens the question as to whether the symptoms described to be caused by PYMoV are caused by a single virus or a combination of multiple viruses. The sequence data presented in the current study will allow this question to be answered by studying other plants identified as infected with PYMoV and looking for the presence of PDV-1 and PDV-2.

Acknowledgments This project was funded by EU-QBOL and Defra.

References

1. Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW (2010) GenBank. *Nucleic Acids Res* 38:D46–D51
2. Bhat AI, Siljo A, Devasahayam S (2012) Occurrence of symptomless source of *Piper yellow mottle virus* in black pepper (*Piper nigrum* L.) varieties and a wild *Piper* species. *Arch Phytopathol Plant Protect* 45:1000–1009
3. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL (2009) BLAST+: architecture and applications. *BMC Bioinform* 10:421
4. De Silva DPP, Jones P, Shaw MW (2002) Identification and transmission of *Piper yellow mottle virus* and *Cucumber mosaic virus* infecting black pepper (*Piper nigrum*) in Sri Lanka. *Plant Pathol* 51:537–545
5. Finn RD, Mistry J, Tate J, Coghill P, Heger A, Pollington JE, Gavin OL, Gunasekaran P, Ceric G, Forslund K, Holm L, Sonnhammer ELL, Eddy SR, Bateman A (2010) The Pfam protein families database. *Nucleic Acids Res* 38:D211–D222
6. Hagen LS, Jacquemond M, Lepingle A, Lot H, Tepfer M (1993) Nucleotide sequence and genomic organization of cacao swollen shoot virus. *Virology* 196:619–628
7. Hareesh PS, Bhat AI (2008) Detection and partial nucleotide sequence analysis of *Piper yellow mottle virus* infecting black pepper in India. *Indian J Virol* 19:160–167
8. King E, Lefkowitz E, Adams M, Carstens E (2011) *Virus taxonomy: ninth report of the international committee on taxonomy of viruses*. Elsevier, London
9. Lockhart BEL, Kiratiya-Angul K, Jones P, Eng L, Silva PD, Olszewski NE, Lockhart N, Deema N, Sangalang J (1997) Identification of *Piper yellow mottle virus*, a mealybug-transmitted badnavirus infecting *Piper* spp. in Southeast Asia. *Eur J Plant Pathol* 103:303–311
10. Marchler-Bauer A, Lu S, Anderson JB, Chitsaz F, Derbyshire MK, DeWeese-Scott C, Fong JH, Geer LY, Geer RC, Gonzales NR, Gwadz M, Hurwitz DI, Jackson JD, Ke Z, Lanczycki CJ, Lu F, Marchler GH, Mullokandov M, Omelchenko MV, Robertson CL, Song JS, Thanki N, Yamashita RA, Zhang D, Zhang N, Zheng C, Bryant SH (2011) CDD: a conserved domain database for the functional annotation of proteins. *Nucleic Acids Res* 39:D225–D229
11. Medberry SL, Lockhart BE, Olszewski NE (1990) Properties of *Commelina yellow mottle virus*'s complete DNA sequence, genomic discontinuities and transcript suggest that it is a pararetrovirus. *Nucleic Acids Res* 18:5505–5513
12. Ravindran PN (2000) *Black pepper (Piper nigrum L)* Hardwood Academic Publishers, The Netherlands
13. Seal S, Muller E (2007) Molecular analysis of a full-length sequence of a new yam badnavirus from *Dioscorea sansibarensis*. *Arch Virol* 152:819–825
14. Sether DM, Melzer MJ, Borth WB, Hu JS (2012) Pineapple bacilliform CO virus: Diversity, Detection, Distribution, and Transmission. *Plant Dis* 96:1798–1804
15. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* 28:2731–2739