

The complete genome sequence of a *Passion fruit woodiness virus* isolate from Australia determined using deep sequencing, and its relationship to other potyviruses

Stephen J. Wylie · Michael G. K. Jones

Received: 12 August 2010 / Accepted: 19 October 2010 / Published online: 13 November 2010
© Springer-Verlag 2010

Abstract The complete genome sequence (9,858 nucleotides) of the *Passion fruit woodiness virus* isolate MU-2 was determined using Illumina sequencing. The large open reading frame (ORF) encodes a polyprotein containing 3,086 amino acids, with an AUG start codon and UAA stop codon. The polyprotein yielded 11 proteins (P1, HC-Pro, P3, PIPO, 6K1, CI, 6K2, NIa-VPg, NIa-Pro, NIb and CP). Putative cleavage sites between them were identified by sequence comparison to those of other known potyviruses. Accuracy of the genome sequence information was provided by 42–1691-fold sequence coverage, and viral RNA accounted for 7.38% of total polyadenylated RNA from the host plant.

In Australia, viruses belonging to the species *Passion fruit woodiness virus* (family *Potyviridae*, genus *Potyvirus*) cause mild to severe disease in *Passiflora edulis* (edible passion fruit), *P. caerulea* (blue passion flower), *P. suberosa* (corky passionfruit), *P. foetida* (stinking passion flower), and the indigenous *P. aurantia* (golden passion flower), as well some legumes including *Glycine max* (soybean), *Arachis hypogaea* (peanut) and *Phaseolus vulgaris* (common bean) [2, 8, 10]. Although *P. edulis* and other passiflora species are grown internationally, most *Passion fruit woodiness virus* isolates (PWV) for which sequences are available in the GenBank database originate from Australia (P32574, P32575, P32576, AY461662, AJ430527, DQ898215, DQ898216, DQ898217,

DQ898218, U67149, U67150, U67151). A sequence from Taiwan labeled as passionfruit woodiness virus (AF208662) is actually East Asian passiflora virus (EAPV) [1], and cowpea aphid-borne mosaic virus (CABMV) infection of passionfruit outside Australia is commonly called passionfruit woodiness disease [7]. Further confusion arises because phylogenetic analyses of PWV isolates from within Australia indicate two distinct potyviruses [10, 12].

Leaf tissue from a wild plant of *P. caerulea* showing mild leaf mosaic, distortion and purple discoloration symptoms was collected at the campus of Murdoch University, Perth, Western Australia. Total RNA was extracted from 100 mg of ground leaf tissue using a Qiagen RNeasy Plant RNA kit. RNA was quantified and 20 µg suspended in ethanol and submitted to MacroGen Inc. (Seoul, South Korea) for synthesis of cDNA from polyadenylated RNA followed by deep sequencing using Illumina Solexa GA IIx technology, where a data set of 31,494,884 sequence reads, each of 78 nucleotides (nt), was generated using one lane of a cell.

Five one-million-read files were chosen at random from the complete data set and assembled de novo in both directions using the assembly tool in Geneious Pro 5.0.4 [4], and mean numbers per million reads were calculated. A mean of 10,675 contigs per million reads was generated from 10 or more reads. The longest contig of approximately 10,000 nt had 99.4% pairwise identity between reads, and the consensus sequence was calculated by Blastn and Blastx analysis to represent the complete genome of PWV isolate MU-2. The contig was edited manually, and the genome sequence was revealed to be 9,858 nt in length, excluding the poly(A) tail. The virus consensus sequence was generated from a mean of 73,831 reads per million reads analysed (7.38%), with 584-fold overall sequence coverage. Coverage was not evenly distributed over the genome, being lowest at nt 292 (42-fold coverage) and

S. J. Wylie (✉) · M. G. K. Jones
Plant Virus Section, Western Australian State Agricultural
Biotechnology Centre, Murdoch University,
Perth, WA 6150, Australia
e-mail: s.wylie@murdoch.edu.au

highest at nt 9249 (1,691-fold coverage) (Fig. 1). The viral RNA molecule was the longest and most highly expressed polyadenylated RNA detected. The two longest non-viral contigs were a putative magnesium chelatase subunit H (4,606 nt) and a phosphoenolpyruvate carboxylate (3,401 nt) mRNA. The two most highly expressed non-viral contigs were a putative chlorophyll A/B binding protein (1.39%) and a ribulose-1,5-bisphosphate small subunit (0.94%) mRNA.

Open reading frame 1 (ORF1) of isolate MU-2 encodes a large polyprotein consisting of 3,086 amino acids (aa) with an AUG start codon and UAA stop codon. A small ORF2 encoding the PIPO protein [3] is embedded within ORF1 in the P3 cistron. The large polyprotein gives rise to 10 proteins (P1, HC-Pro, P3, 6K1, CI, 6K2, VPg, NIa-pro, NIb and CP), and putative cleavage sites of each protein were identified by sequence comparison to those of three other potyviruses. The 5' untranslated region (UTR) was of 346 nt and the 3' UTR 250 nt (Fig. 2). Expected potyvirus motifs of FRNK in the HC-Pro, GDD in the NIb, and DAG in the CP are conserved. The aa sequence of ORF1 was globally aligned with those of 19 other potyviruses using the alignment tool in Geneious Pro 5.0.4 set with a Blossum62 cost matrix, a gap open penalty of 12 and a gap extension penalty of 3. Sequence alignment data was bootstrapped with 1,000 resamplings of alignments to assess the robustness of the lineages in the trees. The evolutionary distances were computed using the Maximum Composite Likelihood method. Trees were drawn to scale, with branch lengths in the same units as those of evolutionary distances used to infer the phylogenetic tree branch length in nucleotide substitutions per site.

Sequence identity at the aa level between those viruses and PWV was found to be 23–35% in P1, 48–70% in HC-Pro, 27–54% in P3, 30–76% in 6K2, 55% in PIPO (soybean mosaic virus), 55–81% in CI, 26–75% in 6K2, 49–71% in VPg, 50–76% in NIa-Pro, 60–76% in NIb, and 56–69% in the CP region. The complete genome sequence of PWV most closely resembled those of isolates of wisteria vein mosaic virus, dasheen mosaic virus, East Asian passiflora virus (EAPV), and other potyviruses that belong to the bean common mosaic virus subgroup (Fig. 3a). The CP

sequence of isolate PWV MU-2 fell into a group of PWV isolates (S, M, TB, 299, CoP-1, Gld-1, Ku-1 and Car-1) that shared 93–99% nt identity and 87–99% aa identity with one another (Fig. 3b). Isolate 388 (AY461662) also shared high identity but was not used here because the sequence was too short. On the other hand, CP sequences of PWV isolates U. Sydney, K, Culnes shared much lower identity (mean of 76% nt and 81% aa identities) with isolate MU-2. Isolate Newrybar (U67151) also clustered with these isolates, but its sequence was too short to be included. PWV isolate Taiwan was even less similar; it shared only 69% nt and 70% aa identity with PWV isolate MU-2. Other viruses to which PWV MU-2 was closely related were Clitoria virus Y (79% nt, 82% aa identities), Siratro 1 virus Y (78% nt, 81% aa identities), Siratro 2 virus Y (77% nt, 79% aa identities), Hibbertia virus Y (77% nt, 81% aa identities), and Hardenbergia mosaic virus (77% nt, 81% aa identities) (Fig 3b), all potyviruses described only from Australia and mainly from native plants [5, 9], which lends support to the proposition that passionfruit woodiness virus, and the other viruses listed, evolved in Australia [12].

It was proposed by Webster et al. [12] that the virus described from PWV isolates U. Sydney, Culnes and Newrybar be renamed *Passiflora mosaic virus* because CP sequence identity between them and other viruses of the same name was slightly below the species demarcation point [1]. We suggest retaining the name *Passion fruit woodiness virus* for isolates of the group into which isolate MU-2 fits. Isolates M, S, TB, and 299 from this group were the first isolate sequences named from *Passion fruit woodiness virus* [8]. Members of the passion fruit woodiness virus group share high CP nt sequence identity with one another, but identities with isolates of Australian viruses Clitoria virus Y, Hardenbergia mosaic virus, Hibbertia virus Y, *Passiflora mosaic virus*, passionfruit woodiness virus, Siratro 1 virus Y and Siratro 2 virus Y hover close to the species demarcation point of 76–77% nt (82% aa) sequence identity [1]. It is therefore legitimate to question their status as distinct taxa, and possibly to consider them all as one diverse species [6].

The high-throughput sequencing method used to obtain the full genome sequence of the PWV isolate MU-2 was

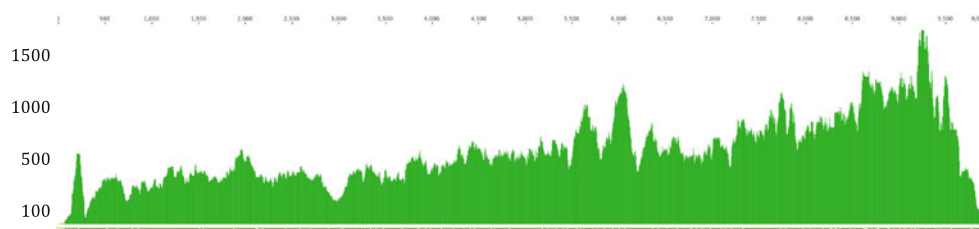


Fig. 1 Sequence coverage and nucleotide position along the genome of the *Passion fruit woodiness virus* isolate MU-2. The image was generated in Geneious Pro 5.0.4 using the coverage feature

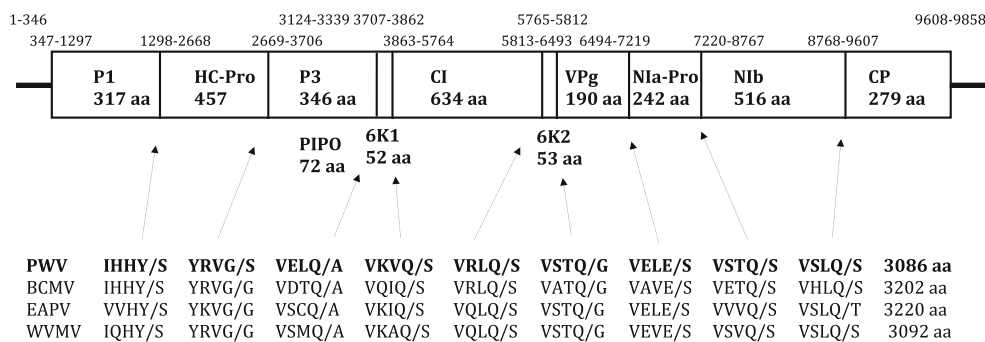


Fig. 2 Representation of the organisation of the genome of *Passion fruit woodiness virus* isolate MU-2 and the putative protein cleavage sites compared to three other potyviruses. Proteolytic cleavage sites

within the polyprotein are indicated by *arrows*. Each cleavage site is indicated by a *slash symbol*. The lengths of the amino acid and nucleotide sequences of the mature proteins are indicated

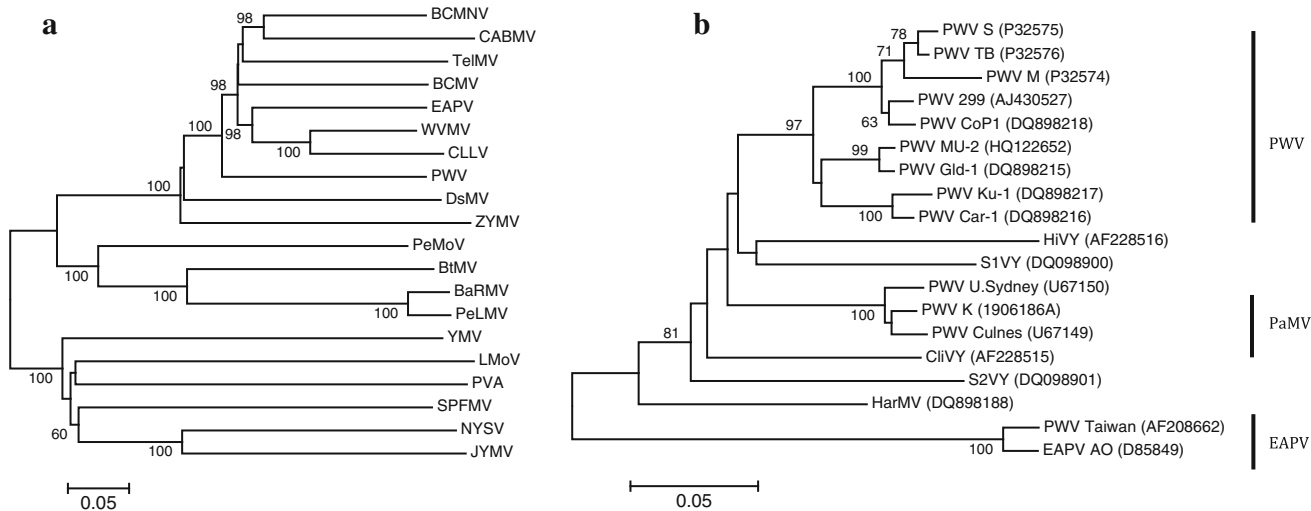


Fig. 3 Phylogenetic trees of amino acid sequences of (a) complete polyproteins of passionfruit woodiness virus (PWV) and 19 other potyviruses, and (b) complete and nearly complete coat proteins of isolates of PWV, passiflora mosaic virus (PaMV), East Asian Passiflora virus (EAPV), clitoria virus Y (CliVY), hardenbergia mosaic virus (HarMV), hibbertia virus Y (HiVY), siratro 1 virus Y (S1VY), and siratro 2 virus Y (S2VY) (GenBank accession codes given in parentheses). Complete polyprotein sequences used: basella rugose mosaic virus (BaRMV, NC_009741); bean common mosaic virus (BCMV, NC_003397); bean common mosaic necrosis virus (BCMNV, NC_004047); beet mosaic virus (BtMV, NC_005304); calla lily latent virus (CLLV, EF105299); cowpea aphid-borne mosaic

virus (CABMV, NC_004013); dasheen mosaic virus (DsMV, NC_003537); East Asian Passiflora virus (EAPV, NC_007728); Japanese yam mosaic virus (JYMV, NC_000947); lily mottle virus (LMoV, NC_005288); narcissus yellow stripe virus (NYSV, NC_011541); passionfruit woodiness virus (HQ122652); peanut mottle virus (PeMoV, NC_002600); peace lily mosaic virus (PeLMV, NC_009743); potato virus A (PVA, NC_004039); sweet potato feathery mottle virus (SPFMV, NC_001841); telosma mosaic virus (TelMV, NC_009742); wisteria vein mosaic virus (WVMV, NC_007216); yam mosaic virus (YMV, NC_004752); zucchini yellow mosaic virus (ZYMV, NC_003224)

simple, rapid, highly accurate, and relatively inexpensive compared to traditional sequencing strategies. Here, total RNA was not enriched for viral RNA as others have reported doing [8], thereby simplifying sample preparation. A further advantage of sequencing total polyadenylated RNA rather than viral-enriched RNA is that both the viral RNA and plant transcripts can be quantified [11]. However, sequencing polyadenylated RNA species precludes detection of non-polyadenylated RNA viruses, viroids, and DNA viruses. The complete genome was obtained with high sequence coverage from only 1 million reads (and

confirmed using another 4 million) of the 31 million reads obtained. This leaves ample scope for multiplexing tens to hundreds of samples, thereby reducing cost per sample. Before mixing and sequencing, multiple samples can be uniquely labeled to allow subsequent matching of viruses with host plants [8], or they could be mixed and sequenced without labeling. In the latter case, PCR-based assays can be designed later from resulting virus sequences and used to screen source plants.

The complete genome sequence of passionfruit woodiness virus was deposited with GenBank (accession code

HQ122652). This is the first information on the complete genome sequence and genomic structure of a *Passion fruit woodiness virus* isolate.

Acknowledgments SW was supported by an Australian Research Council Linkage Grant (LP0882671), funded partially by industry partners Alcoa Australia Pty Ltd, BHP Billiton Worsley Alumina Pty Ltd, Kings Park and Botanic Garden, Department of Agriculture and Food Western Australia, and the Murdoch University Institutes for Crop and Plant Science Research and Sustainable Ecosystems Research.

References

1. Adams MJ, Antoniw JF, Fauquet CM (2005) Molecular criteria for genus and species discrimination within the family Potyviridae. *Arch Virol* 150:459–479
2. Buchen-Osmond C, Crabtree K, Gibbs A, McLean G (1988) Viruses of plants in Australia. Australian National University Printing Service, Canberra
3. Chung BYW, Miller WA, Atkins JF, Firth AE (2008) An overlapping essential gene in the Potyviridae. *Proc Natl Acad Sci USA* 105:5897–5902
4. Drummond AJ, Ashton B, Cheung M, Heled J, Kearse M, Moir R, Stones-Havas S, Thierer T, Wilson A (2010) Geneious v5.0, <http://www.geneious.com>
5. Gibbs A, Mackenzie A (1997) A primer pair for amplifying part of the genome of all potyvirids by RT-PCR. *J Virol Meth* 63:9–16
6. Gibbs A, Ohshima K (2010) Potyviruses and the digital revolution. *Annu Rev Phytopathol* 48:205–223
7. Nascimento AVS, Santana EN, Braz ASK, Alfenas PF, Pio-Ribeiro G, Andrade GP, de Carvalho MG, Murilo Zerbini F (2006) *Cowpea aphid-borne mosaic virus* (CABMV) is widespread in passionfruit in Brazil and causes passionfruit woodiness disease. *Arch Virol* 151:1797–1809
8. Roossinck MJ, Saha P, Wiley GB, Quan J, White JD, Lai H, Chavarría F, Shen G, Roe BA (2010) Ecogenomics: using massively parallel pyrosequencing to understand virus ecology. *Mol Ecol* 19(Suppl 1):81–88
9. Shukla DD, McKern NM, Ward CW (1988) Coat protein of potyviruses. 5. Symptomatology, serology, and coat protein sequences of three strains of passionfruit woodiness virus. *Arch Virol* 102:221–232
10. Sokhandan N, Gillings MR, Bowyer JW (1997) Polymerase chain reaction detection and assessment of genetic variation in New South Wales isolates of passionfruit woodiness potyvirus. *Aust Plant Pathol* 26:155–164
11. Tsibris AM, Korber B, Arnaout R, Russ C, Lo CC, Leitner T, Gaschen B, Theiler J, Paredes R, Su Z, Hughes MD, Gulick RM, Greaves W, Coakley E, Flexner C, Nusbbaum C, Kuritzkes DR (2009) Quantitative deep sequencing reveals dynamic HIV-1 escape and large population shifts during CCR5 antagonist therapy in vivo. *PLoS One* 4:e5683
12. Webster C, Jones RAC, Coutts BA, Jones MGK, Wylie SJ (2007) Virus impact at the interface of an ancient ecosystem and a recent agroecosystem: studies on three legume-infecting potyviruses in the southwest Australian floristic region. *Plant Pathol* 56:729–742