Koh, S.H., Li, H., Admiraal, R., Jones, M.G.K. and Wylie, S.J. (2015) Catharanthus mosaic virus: A potyvirus from a gymnosperm, Welwitschia mirabilis. Virus Research, 203 . pp. 41-46.

Title: Catharanthus mosaic virus: a potyvirus from a gymnosperm, *Welwitschia mirabilis*

Author: Shu Hui Koh Hua Li Ryan Admiraal Michael G.K. Jones Stephen J. Wylie

PII: S0168-1702(15)00125-2
DOI: http://dx.doi.org/doi:10.1016/j.virusres.2015.03.007
Reference: VIRUS 96567

To appear in: *Virus Research*

Received date: 23-2-2015
Revised date: 12-3-2015
Accepted date: 14-3-2015

Please cite this article as: Koh, S.H., Li, H., Admiraal, R., Jones, M.G.K., Wylie, S.J.,Catharanthus mosaic virus: a potyvirus from a gymnosperm, *Welwitschia mirabilis*, *Virus Research* (2015), http://dx.doi.org/10.1016/j.virusres.2015.03.007

Catharanthus mosaic virus: a potyvirus from a gymnosperm, *Welwitschia mirabilis*

Shu Hui Koh[1,2], Hua Li[1], Ryan Admiraal[2], Michael G.K. Jones[1] and Stephen J. Wylie[1]

[1] Plant Biotechnology Group –Plant Virology, Western Australian State Agricultural Biotechnology Centre, School of Veterinary and Life Sciences, Murdoch University, Perth, Western Australia 6150, Australia.

[2] School of Engineering and Information Technology, Mathematics & Statistics, Murdoch University, Perth, Western Australia 6150, Australia.

Abstract

A virus from a symptomatic plant of the gymnosperm *Welwitschia mirabilis* Hook. growing as an ornamental plant in a domestic garden in Western Australia was inoculated to a plant of *Nicotiana benthamiana* where it established a systemic infection. The complete genome sequence of 9636 nucleotides was determined using high-throughput and Sanger sequencing technologies. The genome sequence shared greatest identity (83 % nucleotides and 91 % amino acids) with available partial sequences of catharanthus mosaic virus, indicating that the new isolate belonged to that taxon. Analysis of the phylogeny of the complete virus sequence placed it in a monotypic group in the genus *Potyvirus*. This is the first record of a virus from *W. mirabilis*, the first complete genome sequence of catharanthus mosaic virus determined, and the first record from Australia. This finding illustrates the risk to natural and managed systems posed by the international trade in live plants and propagules, which enables viruses to establish in new regions and infect new hosts.

Key words

Catharanthus mosaic virus, *Welwitschia mirabilis*, Gymnosperm, Potyvirus

Catharanthus mosaic virus (CatMV) was first described from the Madagascar Periwinkle (*Catharanthus roseus*) in Brazil (Maciel et al., 2011). Like many potyviruses, CatMV appears to have a limited host range. CatMV has been identified naturally only from species belonging to the family Apocynaceae: *C. roseus* plants in Brazil (Maciel et al., 2011) and a cultivar of *Mandevilla* in North America (Mollov et al., 2014). Experimentally it also infects *Nicotiana benthamiana* systemically (family Solanaceae), and *Chenopodium amaranticolor* and *C. quinoa* locally (family Amaranthaceae) (Maciel et al., 2011). CatMV infection in *C. roseus* typically induces moderately severe symptoms of leaf mosaic patterns and deformation, leaf blade reduction and reduced seed fertility (Maciel et al., 2011), and in *Mandevilla* mosaic symptoms and deformation in leaves, premature leaf senescence and vine dieback (Mollov et al., 2014).

Welwitschia mirabilis is a monotypic species in the monotypic order Welwitschiales (Division Gnetophyta) endemic to the Namib Desert of Namibia and Angola in south-west Africa. Welwitschia is known to be one of the longest-lived plants on Earth, living up to 3000 years old (Jacobson and Lester, 2003). There is fossil evidence that members of the family Welwitschiaceae existed in South America in the Mesozoic era, and its current distribution probably reflects its Gondwanan origins and climatic changes during the Tertiary and Quaternary (Jacobson and Lester, 2003). There are only two true leaves on a *Welwitschia* plant, and these split to form several leaf strips, which grow longitudinally along the ground. Welwitschia's peculiar morphology and natural history makes it an unusual and interesting ornamental plant. To date, there is no record of viruses infecting *Welwitschia*.

There are two CatMV sequences available in GenBank, which comprise the partial replicase (NIb), the complete coat protein (CP) and the 3' untranslated region (UTR) of the genome (Maciel et al., 2011; Mollov et al., 2014). Here, the first complete genome sequence of an isolate of CatMV from *Welwitschia* in Australia was generated, demonstrating that CatMV has a broader host range and wider geographical distribution than previously recognized.

Leaf tissue was collected from a single *Welwitschia* plant growing in a domestic garden in the village of Bremer Bay, in southern Western Australia. The plant exhibited mild streaking on the leaves, resembling those induced by some viral infections. Macerated leaf tissue was mechanically inoculated onto leaves of healthy *N. benthamiana* seedlings (accession RA-4) with 0.1 M phosphate buffer (pH 7.0) and diatomaceous earth (Sigma Corp). Symptoms of chlorosis, leaf deformation and stunting were observed on inoculated plants 12-20 days post inoculation.

Total nucleic acids were extracted from infected *N. benthamiana* leaves and enriched for dsRNA using a cellulose based method (Morris and Dodds, 1979). cDNA was synthesized using GoScript™ reverse transcriptase (Promega) with a random primer. An index sequence was added by randomly-primed PCR (Table S1) using the following cycling conditions: 95 ºC for 3 min, 30 cycles of 95 ºC for 30 s, 60 ºC for 30 s, 72 ºC for 30 s and the final extension of 72 ºC for 7 min. PCR products were cleaned using QIAquick PCR purification columns (Qiagen).

Library preparation and high-throughput sequencing was done on an Illumina HiSeq2000 machine by Macrogen, South Korea. Three sequencing runs were done using the same sample, and sequence data from each run were used to confirm the viral genome sequence.

Analysis of sequence data was done after trimming off the index and primer sequences at the 5' and 3' ends. Trimmed reads were then assembled *de novo* using default parameters in CLC Genomics Workbench (Qiagen) to form contigs, followed by interrogation of GenBank (NCBI) nucleotide and protein databases using Blastn and Blastx (Altschul et al., 1990) to identify virus-like contigs. Contigs resembling virus sequences were imported into Geneious v7.0.6 (Kearse et al., 2012) for further analysis, including sorting of contigs into a group of those most closely resembling known potyvirus genome sequences and contigs with long open reading frame (ORF) were analysed through blastp (Altschul et al., 1990).

Twenty potyvirus-like contigs ranging from 263 nt to 1233 nt were identified with >60 % sequence identity to complete genome sequences of isolates of plum pox virus (PPV) or >98 % sequence identity to CatMV available in the database. Contigs were then assembled into longer contigs (parameters for the assembly were 25 % minimum overlap of read length and 10 % maximum gaps per read) and mapped to PPV by pairwise alignment. This enabled the contigs to be placed in approximate order with respect to one another, and for putative gaps to be identified in the genome sequence. Primers were designed on either side of putative gaps to amplify the missing sequences (Table S2). All primers were designed from sequences obtained from deep sequencing except the CP reverse primer, which was designed from the CatMV-Mandevilla sequence (GenBank accession KM243928.1). After PCR amplification of gap sequences, the amplicons were sequenced using the Sanger method. These sequences enabled the entire genome sequence to be assembled. A subsequent (third) Illumina sequencing run confirmed the sequence generated by previous Illumina and Sanger sequencing was correct.

The 5' UTR of 145 nt was obtained by *de novo* assembly of Illumina sequencing reads. Conserved 'Poty box' motifs within the 5' untranslated regions (UTR) of potyviruses (Shukla et al., 1994) were identified, confirming that the complete or near-complete 5' UTR was obtained. Poty box A (ACACAACA) was predicted at nt 7-15, and Poty box B at either nt 37-45 (TCAAAGCA) or nt 77-84 (TCAAGCA). The 3' UTR region was 326 nt (excluding the polyprotein stop codon), and the extent of its length was confirmed when the 3' poly-(A) tract was obtained.

Constructed from 5,784,246 sequence reads, the final consensus sequence of the virus sequence obtained from the infected *N. benthamiana* was 9636 nucleotides in length. When mapped to the consensus sequence, the mean coverage of raw sequence reads obtained from Illumina sequencing was 23,463.9 (S.D. 112,475.8). As observed previously, the depth of coverage across the whole genome was not constant, so that some regions had much higher or lower coverage then the mean (Harismendy et al., 2009). Thus, the minimum coverage was 0-fold for the regions

P3 (2669-2700 nt and 2758-2759 nt) and NIb (7152-7154 nt) while the highest coverage was 1,295,475-fold at the CI region (5411-5423 nt). The sequences of the regions of minimum coverage were verified by Sanger sequencing using primers designed from flanking regions (Table 1).

The genome encoded a large open reading frame of 9165 nt, calculated to encode a polyprotein of 3054 aa with a calculated molecular weight (MW) of 348 kDa. Conserved protease cleavage sites typical of other potyviruses were present, and are predicted to cleave the polyprotein into the 10 mature proteins (P1, HC-Pro, P3, 6K1, CI, 6K2, VPg, NIa, NIb and CP) post-translationally (Fig 1). The calculated MW of each polyprotein is P1: 34.242 kDa, HC-Pro: 51.898 kDa, P3: 40.039 kDa, 6K1: 5.701 kDa, CI: 71.742 kDa, 6K2: 6.084 kDa, VPg: 21.482 kDa, NIa-Pro: 27.606 kDa, NIb: 59.819 kDa and CP: 29.549 kDa. The small ORF, PIPO (Chung et al., 2008) of 204 nt, encoding a putative peptide of 68 aa (MW 8.142 kDa) occurred in the +2 ORF within the putative P3 cistron. Conserved potyvirus motifs were identified in CatMV-Welwitschia: FRNK (at 1571 – 1582 nt), involved in symptom development (Gal-On, 2000; Shiboleth et al., 2007) in the HC-Pro; G--SG---T---NS (from 7892 – 7933 nt) and GDD (at 8021 – 8029 nt), essential in RNA polymerase activity in the NIb (Li and Carrington, 1995) ; DAG (at 8555 - 8563 nt), involved in aphid transmission (Atreya et al., 1991) in the CP; and three conserved motifs, MVWCIENGTSP (at 8852 – 8884 nt), AFDF (at 9101 - 9112 nt) and QMKAAAL (at 9161 – 9181 nt), in the CP (Bejerman et al., 2008; Maciel et al., 2011; Marchler-Bauer et al., 2015; Miglino et al., 2010). The locations of the post-transcriptional cleavage sites were estimated by comparison with cleavage recognition sequences from other potyviruses . Predicted protease cleavage sites of CatMV-Welwitschia are P1: MTHY/S, HC-Pro: YNVG/G, P3: VEHQ/S, 6K1: VYHQ/S, CI: VQHQ/S, 6K2: VQHE/G, VPg: VLHE/G, NIa: VIEQ/G, NIb: VYHQ/S.

A comparison of percent identity of nucleotide (nt) and amino acid (aa) sequences of catharanthus mosaic virus isolate Welwitschia was done with genome sequences of other known potyviruses from GenBank. Complete genome comparison was performed through EMBOSS Water (local alignment) while the

individual nucleotide and protein regions was analysed using EMBOSS Needle alignment (McWilliam et al., 2013) (Table 1)**.** Of the analysed potyvirus in table 1, the most similar potyvirus genomes to the CatMV-Welwitschia genome sequence were turnip mosaic virus and plum pox virus/ turnip mosaic virus with 54.7 % and 47.8 % respectively in nt and aa. The predicted CP sequence shared 81.7 % and 97.2 % nucleotide and 89.5 % and 97.7 % amino acid sequence identity with those of the CatMV isolates reported previously from Brazil and the USA, respectively (Table 1). It is interesting to note that the North American and Brazilian isolates are from closely related plants located geographically close to one another, yet the Australian isolate infecting a gymnosperm is genetically closer to the North American isolate (Table 1, Fig 2), suggesting they share a more recent common ancestor than the Brazilian isolate. CP sequence identities were above the theoretical potyvirus species demarcation limits of >76 % nt and >80 % aa identities assigned for the CP region (King et al., 2012). For these reasons the new sequence was named catharanthus mosaic virus isolate Welwitschia. The complete genome sequence was granted GenBank accession code KP742991.

To confirm that the catharanthus mosaic virus isolate was derived from Welwitschia, RT-PCR was done on total RNA extracted from the infected Welwitchia plant using CatMV-specific primers (Table S2), followed by Sanger sequencing. The sequence was identical to that gained from infected *N. benthamiana* plants.

Phylogenetic analysis was carried out on the 'coherently evolving coat protein' (cCP) (Fig 2) and on the polyprotein sequence (Fig 3) of CatMV isolate Welwitschia sequence and other potyviruses. The cCP region is the CP coding region minus the N terminal region, which is repetitive and variable, thus often requiring gaps and evoking large penalty scores to align (Gibbs et al., 2008) (Table 1 on CP and cCP region). The alignment of sequences was done using ClustalW (Thompson et al., 1994) with default parameters. Maximum likelihood analysis was used with the LG (+F) model of evolution (for polyprotein) and LG model of evolution (for cCP region) using 1000 bootstrap replication within

MEGA6.06 (Tamura et al., 2013). Agropyron mosaic virus and hordeum mosaic virus (genus *Rymovirus*) were used as outgroups (French and Stenger, 2005).

Analysis of the cCP phylogeny clearly placed CatMV-Welwitschia in the same monotypic clade as those of the two available CatMV sequences with high bootstrap support (>92%) (Fig 2). Comparison of the complete polyprotein sequence of CatMV with those of other potyviruses showed with high support that CatMV-Welwitschia is a distinct virus (Fig 3).

The genome sequence of CatMV-Welwitschia was checked for evidence of recombination events against complete genomes of the 34 potyviruses that were used for phylogenetic analysis using the RDP4 package (Martin et al., 2010). Seven programs were used in the package with default parameters; RDP (Martin and Rybicki, 2000), GENECONV (Padidam et al., 1999), MaxChi (Smith, 1992), Chimaera (Posada and Crandall, 2001) and 3Seq (Boni et al., 2007), BootScan (Martin et al., 2005) and SiScan (Gibbs et al., 2000). A region was considered to be positive for recombination if four programs detected the same recombination event with high probability. No evidence of recombination was discovered within the genome sequence of CatMV-Welwitschia.

Until now, CatMV was known to naturally infect only members of the angiosperm family Apocynaceae. The virus' presence in the gymnosperm Welwitschia is surprising, since it is genetically distant from its previously recognized host (Chaw et al., 2000). However, CatMV is not the first virus reported to infect members of both the angiosperms and gymnosperms. The nepovirus cycas necrotic stunt virus (CNSV) was described from the gymnosperm *Cycas revoluta* in Japan (Han et al., 2002). Later, Wylie and associates identified an isolate of CNSV from the monocot angiosperm *Lilium longiflorum* in Australia, a species also indigenous to Japan (Wylie et al., 2012). Similarly, the tobamovirus tomato mosaic virus, originally isolated from the angiosperm *Solanum lycopersicum*, was inoculated to three gymnosperm species of spruce and fir, where the virus apparently spread naturally to other spruce and fir seedlings via root contact (Jacobi and Castello, 1992). These cases illustrate that some viruses are able to span the apparently large biological gap between

members of the angiosperms and gymnosperms. The division between the two groups is thought to have occurred in the carboniferous period between about 360-300 million years ago (Doyle and Donoghue, 1986). The only other gymnosperm-infecting virus described so far is Pinus sylvestris cryptovirus (genus *Partitvirus*), identified only from Scots Pine from Hungary (Veliceasa et al., 2006). To our knowledge, CatMV is the first potyvirus to be described infecting a gymnosperm.

W. mirabilis is not indigenous to Australia, nor is CatMV. Thus, the virus arrived in Australia in either *W. mirabilis* seed or plants, or it arrived in another species and subsequently infected the *W. mirabilis* host plant, perhaps via aphids vectors.

The presence of CatMV in Australia illustrates how the international trade in live plants and propagules serves as a vehicle for viruses to invade new lands and to encounter new hosts (Wylie et al., 2014). Further, this study highlights the difficult task of effectively screening live plants and propagules for viruses at international borders, especially those that are unexpected or new to science.

Acknowledgements This study was funded in part by Australia Research Council Linkage grant LP110200180 and by a studentship granted to SHK by Murdoch University.

References

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. J. Mol. Biol. 215(3), 403-410.

Atreya, P.L., Atreya, C.D., Pirone, T.P., 1991. Amino acid substitutions in the coat protein result in loss of insect transmissibility of a plant virus. Proc. Natl. Acad. Sci. U. S. A. 88(17), 7887-7891.

Bejerman, N., Giolitti, F., Lenardon, S., 2008. Molecular characterization of a novel Sunflower chlorotic mottle virus (SuCMoV) strain. Proceedings of the XVII International Sunflower Conference, 8-12.

Boni, M.F., Posada, D., Feldman, M.W., 2007. An exact nonparametric method for inferring mosaic structure in sequence triplets. Genetics 176(2), 1035-1047.

Chaw, S.-M., Parkinson, C.L., Cheng, Y., Vincent, T.M., Palmer, J.D., 2000. Seed plant phylogeny inferred from all three plant genomes: monophyly of extant gymnosperms and origin of Gnetales from conifers. Proc. Natl. Acad. Sci. U. S. A. 97(8), 4086-4091.

Chung, B.Y.-W., Miller, W.A., Atkins, J.F., Firth, A.E., 2008. An overlapping essential gene in the Potyviridae. Proc. Natl. Acad. Sci. U. S. A. 105(15), 5897-5902.

Doyle, J.A., Donoghue, M.J., 1986. Seed plant phylogeny and the origin of angiosperms: an experimental cladistic approach. Bot. Rev. 52(4), 321-431.

French, R., Stenger, D., 2005. Genome sequences of Agropyron mosaic virus and Hordeum mosaic virus support reciprocal monophyly of the genera Potyvirus and Rymovirus in the family Potyviridae. Arch. Virol. 150(2), 299-312.

Gal-On, A., 2000. A point mutation in the FRNK motif of the potyvirus helper component-protease gene alters symptom expression in cucurbits and elicits protection against the severe homologous virus. Phytopathology 90(5), 467-473.

Gibbs, A., Mackenzie, A., Wei, K.-J., Gibbs, M., 2008. The potyviruses of Australia. Arch. Virol. 153(8), 1411-1420.

Gibbs, M.J., Armstrong, J.S., Gibbs, A.J., 2000. Sister-scanning: a Monte Carlo procedure for assessing signals in recombinant sequences. Bioinformatics 16(7), 573-582.

Han, S., Karasev, A., Ieki, H., Iwanami, T., 2002. Nucleotide sequence and taxonomy of Cycas necrotic stunt virus. Arch. Virol. 147(11), 2207-2214.

Harismendy, O., Ng, P.C., Strausberg, R.L., Wang, X., Stockwell, T.B., Beeson, K.Y., Schork, N.J., Murray, S.S., Topol, E.J., Levy, S., 2009. Evaluation of next generation sequencing platforms for population targeted sequencing studies. Genome Biol 10(3), R32.

Jacobi, V., Castello, J.D., 1992. Infection of red spruce, black spruce, and balsam fir seedlings with tomato mosaic virus. Can. J. For. Res. 22(7), 919-924.

Jacobson, K., Lester, E., 2003. A first assessment of genetic variation in Welwitschia mirabilis Hook. J. Hered. 94(3), 212-217.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28(12), 1647-1649.

King, A., Adams, J., Carstens, E., Lefkowitz, E., 2012. Virus Taxonomy: classification and nomenclature of viruses: Ninth Report of the ICTV. Elsevier/Academic Press, Amsterdam.

283 Li, X.H., Carrington, J.C., 1995. Complementation of tobacco etch potyvirus
284 mutants by active RNA polymerase expressed in transgenic cells. Proc. Natl.
285 Acad. Sci. U. S. A. 92(2), 457-461.

286 Maciel, S.C., Silva, R.F.d., Reis, M.S., Jadão, A.S., Rosa, D.D., Giampan, J.S.,
287 Kitajima, E.W., Rezende, J.A.M., Camargo, L.E., 2011. Characterization of a new
288 potyvirus causing mosaic and flower variegation in Catharanthus roseus in Brazil.
289 Sci. Agric. 68(6), 687-690.

290 Marchler-Bauer, A., Derbyshire, M.K., Gonzales, N.R., Lu, S., Chitsaz, F., Geer,
291 L.Y., Geer, R.C., He, J., Gwadz, M., Hurwitz, D.I., 2015. CDD: NCBI's
292 conserved domain database. Nucleic Acids Res. 43(D), 222-226.

293 Martin, D., Posada, D., Crandall, K., Williamson, C., 2005. A modified bootscan
294 algorithm for automated identification of recombinant sequences and
295 recombination breakpoints. AIDS Res. Hum. Retrov. 21(1), 98-102.

296 Martin, D., Rybicki, E., 2000. RDP: detection of recombination amongst aligned
297 sequences. Bioinformatics 16(6), 562-563.

298 Martin, D.P., Lemey, P., Lott, M., Moulton, V., Posada, D., Lefeuvre, P., 2010.
299 RDP3: a flexible and fast computer program for analyzing recombination.
300 Bioinformatics 26(19), 2462-2463.

301 McWilliam, H., Li, W., Uludag, M., Squizzato, S., Park, Y.M., Buso, N., Cowley,
302 A.P., Lopez, R., 2013. Analysis tool web services from the EMBL-EBI. Nucleic
303 Acids Res. 41(1), 597-600.

304 Miglino, R., Druffel, K., Pappu, H., 2010. Identification and molecular
305 characterization of a new potyvirus infecting Triteleia species. Archives of
306 virology 155(3), 441-443.

307 Mollov, D., Guaragna, M.A., Lockhart, B., Rezende, J.A.M., Jordan, R., 2014.
308 First report of Catharanthus mosaic virus in Mandevilla in the United States. Plant
309 Dis. 99(1), 165.

310 Morris, T., Dodds, J., 1979. Isolation and analysis of double-stranded RNA from
311 virus-infected plant and fungal tissue. Phytopathology 69(8), 854-858.

312 Padidam, M., Sawyer, S., Fauquet, C.M., 1999. Possible emergence of new
313 geminiviruses by frequent recombination. Virology 265(2), 218-225.

314 Posada, D., Crandall, K.A., 2001. Evaluation of methods for detecting
315 recombination from DNA sequences: computer simulations. Proc. Natl. Acad. Sci.
316 U. S. A. 98(24), 13757-13762.

317 Shiboleth, Y.M., Haronsky, E., Leibman, D., Arazi, T., Wassenegger, M.,
318 Whitham, S.A., Gaba, V., Gal-On, A., 2007. The conserved FRNK box in HC-
319 Pro, a plant viral suppressor of gene silencing, is required for small RNA binding
320 and mediates symptom development. J. Virol. 81(23), 13135-13148.

321 Shukla, D.D., Ward, C.W., Brunt, A.A., 1994. The Potyviridae. Cab International,
322 Wallingford, UK.

323 Smith, J.M., 1992. Analyzing the mosaic structure of genes. J. Mol. Evol. 34(2),
324 126-129.

325 Tamura, K., Stecher, G., Peterson, D., Filipski, A., Kumar, S., 2013. MEGA6:
326 molecular evolutionary genetics analysis version 6.0. Mol. Biol. Evol. 30(12),
327 2725-2729.

328 Thompson, J.D., Higgins, D.G., Gibson, T.J., 1994. CLUSTAL W: improving the
329 sensitivity of progressive multiple sequence alignment through sequence
330 weighting, position-specific gap penalties and weight matrix choice. Nucleic
331 Acids Res. 22(22), 4673-4680.

Veliceasa, D., Enünlü, N., Kós, P.B., Köster, S., Beuther, E., Morgun, B., Deshmukh, S.D., Lukács, N., 2006. Searching for a new putative cryptic virus in Pinus sylvestris L. Virus Genes 32(2), 177-186.

Wylie, S.J., Li, H., Saqib, M., Jones, M.G., 2014. The global trade in fresh produce and the vagility of plant viruses: A case study in garlic. PloS ONE 9(8), e105044.

Wylie, S.J., Luo, H., Li, H., Jones, M.G., 2012. Multiple polyadenylated RNA viruses detected in pooled cultivated and wild plant samples. Arch. Virol. 157(2), 271-284.

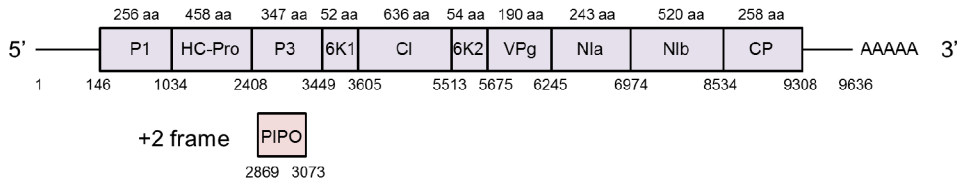

Figure 1 Genome organisation of catharanthus mosaic virus isolate Welwitschia. The calculated length for each protein is indicated (not to scale). The PIPO protein, encoded in the +2 open reading frame is located within the P3.

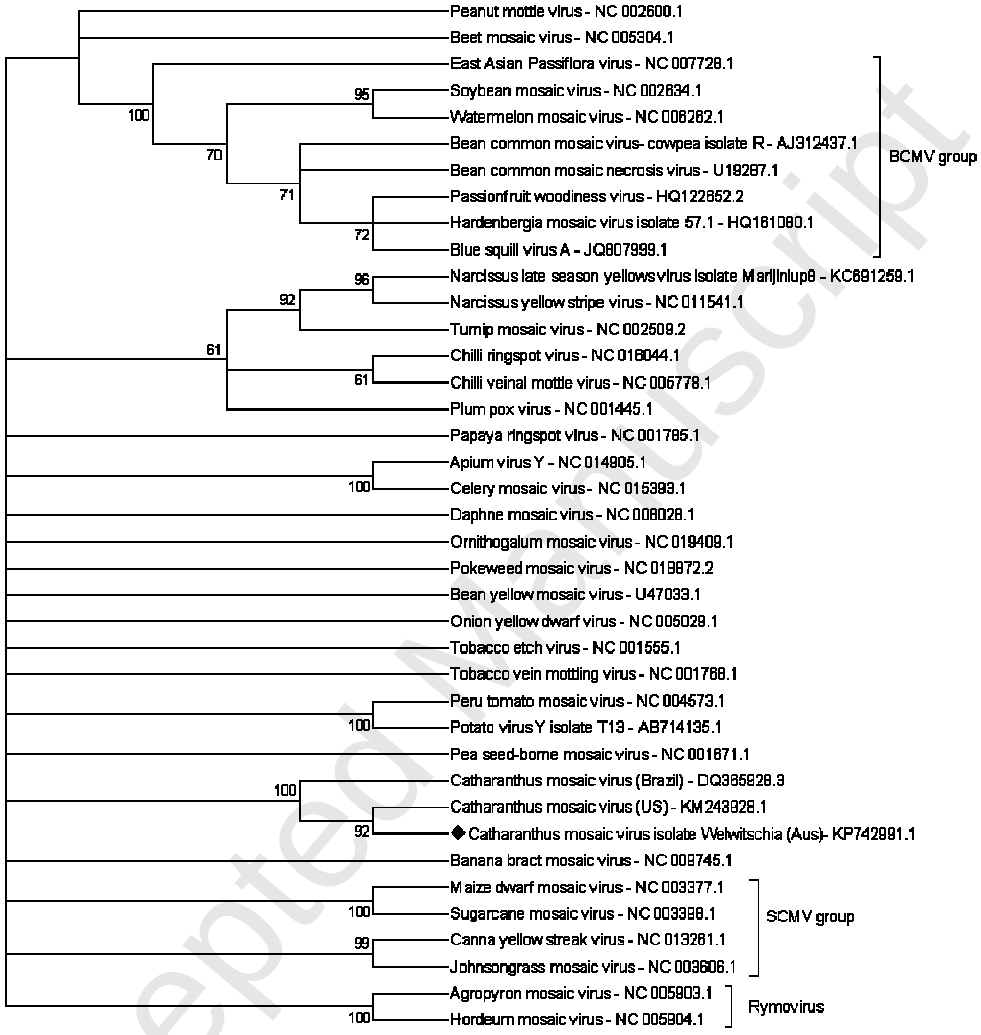

Figure 2 Condensed maximum likelihood tree inferred from amino acid sequences of 'coherently evolving coat protein' (cCP), made up of the CP coding region without the N terminal region, showing the position of CatMV-Welwitschia (black diamond). The bean common mosaic virus (BCMV) and sugarcane mosaic virus (SCMV) subgroups are shown. For branches with low statistical support (>50 % bootstrap confidence), they are condensed to form a multifurcating tree. The percentage of trees (>60%) in which the associated taxa clustered together is shown next to the branches. Sequences of agropyron mosaic virus and hordeum mosaic virus (genus *Rymovirus*, family *Potyviridae*) were used as outgroups.

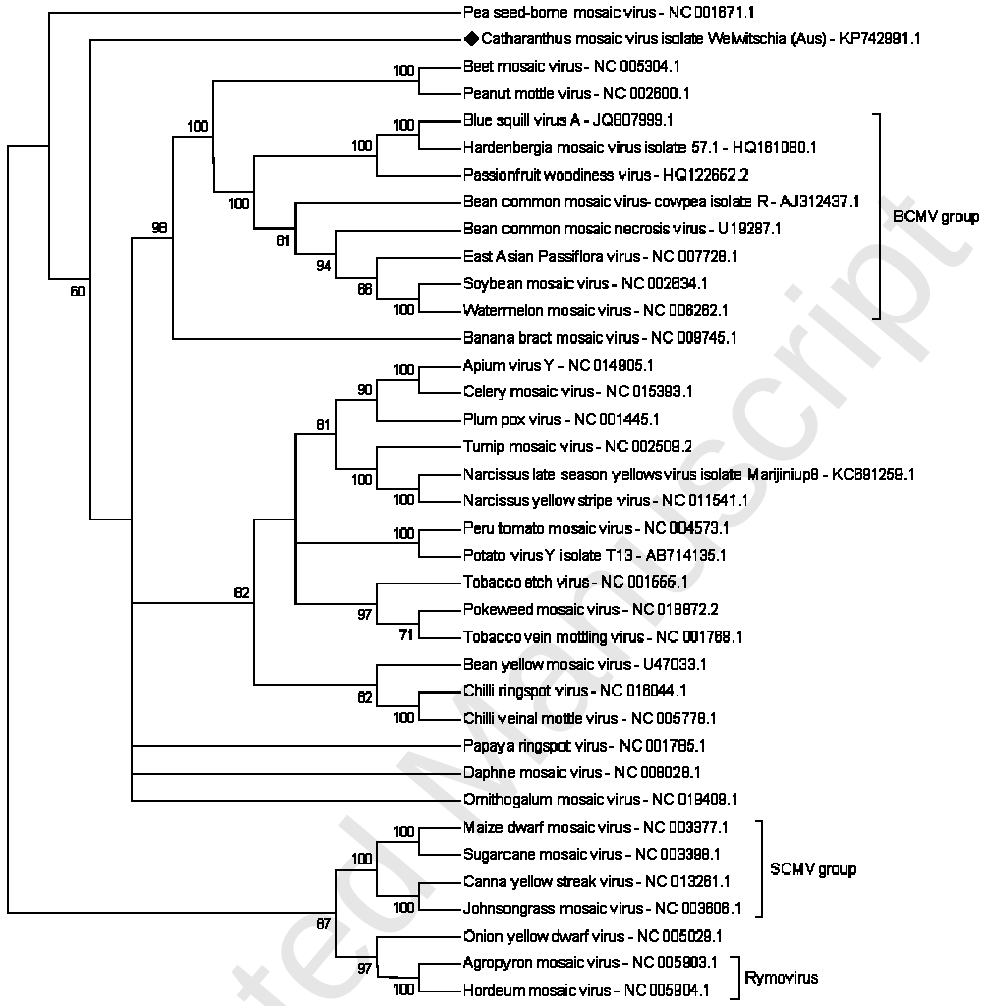

Figure 3 Condensed maximum likelihood tree inferred from amino acid sequences of complete polyproteins showing the position of catharanthus mosaic virus-isolate Welwitschia (marked with a black diamond). The bean common mosaic virus (BCMV) and sugarcane mosaic virus (SCMV) subgroups are shown. To form a condensed tree, branches were condensed if bootstrap confidence is >50 %. The percentage of trees (>60%) in which the associated taxa clustered together is shown next to the branches. Sequences of agropyron mosaic virus and hordeum mosaic virus (genus *Rymovirus*, family *Potyviridae*) were used as outgroups.

374 **Table 1** Comparison of percent identity of nucleotide (nt) and amino acid (aa) sequences of catharanthus mosaic virus isolate Welwitschia with
375 genome sequences and genes of other potyviruses. Regions analysed include P1, HC-Pro, P3, 6K1, CI, 6K2, VPg, NIa-Pro, NIb, CP, 'coherently
376 evolving coat protein' (cCP), made up of the CP coding region without the N terminal region, and PIPO.

Virus	Complete genome (nt/aa) (%)	P1 region (nt/aa) (%)	HC-Pro region (nt/aa) (%)	P3 region (nt/aa) (%)	6K1 region (nt/aa) (%)	CI region (nt/aa) (%)	6K2 region (nt/aa) (%)	VPg region (nt/aa) (%)	NIa-Pro region (nt/aa) (%)	NIb region (nt/aa) (%)	CP region (nt/aa) (%)	cCP region (nt/aa)	PIPO region (nt/aa) (%)
Catharanthus mosaic virus (DQ365928.3)					Sequence not available						81.7/89.5	82.9/90.6	Sequence not available
Catharanthus mosaic virus isolate Mandevilla – US (KM243928.1)					Sequence not available						97.2/97.7	98.2/98.3	Sequence not available
Bean yellow mosaic virus (U47033.1)	54.4/43.9	44.3/19.9	54.8/46.1	46.2/20.4	47.4/39.6	57.6/51.0	51.3/44.4	59.3/49.7	50.9/42.0	60.7/55.2	60.8/58.0	65.3/65.5	Not stated
East Asian Passiflora virus (NC_007728.1)	52.7/45.0	39.3/16.8	55.0/45.5	47.8/26.0	50.0/34.6	55.8/54.4	46.8/32.7	55.1/50.0	55.7/42.8	57.3/54.6	57.4/59.2	63.8/68.2	47.0/22.1
Maize dwarf mosaic virus (NC_003377.1)	53.9/45.0	43.7/12.3	53.2/43.3	46.7/22.3	47.5/32.8	55.9/51.5	49.7/35.2	58.6/57.3	54.4/41.1	58.9/56.1	61.2/56.1	68.1/66.9	38.3/19.6
Narcissus yellow stripe virus (NC_011541.1)	54.2/47.4	45.1/21.1	53.3/46.2	46.9/24.8	55.1/42.3	56.7/54.7	51.6/42.6	56.9/56.5	55.4/50.0	60.4/55.9	60.9/62.2	63.0/68.1	51.7/23.9
Onion yellow dwarf virus (NC_005029.1)	53.5/41.6	37.0/14.4	52.4/43.7	40.8/15.5	46.2/40.7	56.2/49.8	47.0/40.7	51.5/43.0	50.6/42.0	59.4/56.4	62.1/63.1	63.7/66.8	41.3/13.5
Ornithogalum mosaic virus (NC_019409.1)	51.4/46.3	40.5/19.2	54.2/46.0	43.7/23.7	50.3/32.7	57.6/55.7	49.5/42.6	58.2/56.7	53.1/43.6	58.4/57.1	64.7/60.8	67.5/64.3	Not stated
Pea seed-borne mosaic virus (NC_001671.1)	52.8/45.5	44.2/16.8	54.7/46.4	45.2/21.8	50.8/34.6	56.9/55.4	46.9/42.6	51.7/44.1	51.0/41.1	56.9/56.8	56.7/56.0	63.0/66.4	43.4/16.5
Plum pox virus (NC_001445.1)	54.4/47.8	45.5/23.6	53.5/45.3	46.2/27.5	55.1/44.2	58.4/58.3	51.9/55.2	56.7/56.9	57.0/44.4	58.4/60.6	53.6/49.8	64.7/68.1	43.9/20.8
Pokeweed mosaic virus (NC_018872.2)	53.8/46.8	47.2/21.8	56.6/47.6	44.2/21.7	46.5/33.3	57.2/56.6	43.0/39.7	55.8/53.7	51.1/45.1	61.2/58.0	57.9/58.8	62.1/66.8	Not stated
Potato virus Y isolate T13 (AB714135.1)	54.6/44.2	46.5/22.5	52.3/43.1	48.2/22.0	41.9/32.7	56.0/51.6	48.7/39.7	55.7/51.3	55.3/42.4	60.5/54.2	63.4/62.8	65.8/66.4	41.7/14.0
Sugarcane mosaic virus (NC_003398.1)	53.5/45.2	44.6/21.3	51.6/42.8	45.3/21.6	43.5/26.9	55.3/51.2	55.0/29.6	61.5/54.4	54.9/46.5	59.6/56.3	57.9/52.3	67.9/69.1	48.3/21.7

Turnip mosaic virus (NC_002509.2)	54.7/47.8	39.8/18.3	54.8/47.6	48.0/23.5	52.5/40.4	58.5/54.7	55.1/40.7	54.6/56.7	56.2/52.3	58.8/55.9	60.0/58.3	66.1/66.8	48.1/26.8

377

Highlights

- First record of catharanthus mosaic virus found in Australia
- First complete sequence of catharanthus mosaic virus determined
- First virus found from *Welwitschia mirabilis*, a long-lived gymnosperm originating from south-west Africa
- First potyvirus reported to infect Angiosperm and Gymnosperm hosts
- Illustrates how the international trade in live plants and propagules enables viruses to colonise new hosts and invade new lands