Deep sequencing reveals a novel closterovirus associated with wild rose leaf rosette disease

YAN HE^{1,2,3,4}, ZUOKUN YANG^{1,2,3,4}, NI HONG^{1,2,3,4}, GUOPING WANG^{1,2,3,4}, GUOGUI NING⁵ AND WENXING XU^{1,2,3,4,*}

¹State Key Laboratory of Agricultural Microbiology, Wuhan, Hubei 430070, China

²College of Plant Science and Technology, Huazhong Agricultural University, Wuhan, Hubei 430070, China

³National Indoor Conservation Center of Virus-free Germplasms of Fruit Crops, Wuhan, Hubei 430070, China

⁴Key Laboratory of Plant Pathology of Hubei Province, Wuhan, Hubei 430070, China

⁵College of Horticulture and Forestry Sciences, Huazhong Agricultural University, Wuhan, Hubei 430070, China

SUMMARY

A bizarre virus-like symptom of a leaf rosette formed by dense small leaves on branches of wild roses (*Rosa multiflora* Thunb.), designated as 'wild rose leaf rosette disease' (WRLRD), was observed in China. To investigate the presumed causal virus, a wild rose sample affected by WRLRD was subjected to deep sequencing of small interfering RNAs (siRNAs) for a complete survey of the infecting viruses and viroids. The assembly of siRNAs led to the reconstruction of the complete genomes of three known viruses, namely *Apple stem grooving virus* (ASGV), *Blackberry chlorotic ringspot virus* (BCRV) and *Prunus necrotic ringspot virus* (PNRSV), and of a novel virus provisionally named 'rose leaf rosette-associated virus' (RLRaV). Phylogenetic analysis clearly placed RLRaV alongside members of the genus *Closterovirus*, family *Closteroviridae*. Genome organization of RLRaV RNA (17 653 nucleotides) showed 13 open reading frames (ORFs), except ORF1 and the quintuple gene block, most of which showed no significant similarities with known viral proteins, but, instead, had detectable identities to fungal or bacterial proteins. Additional novel molecular features indicated that RLRaV seems to be the most complex virus among the known genus members. To our knowledge, this is the first report of WRLRD and its associated closterovirus, as well as two ilarviruses and one capilovirus, infecting wild roses. Our findings present novel information about the closterovirus and the aetiology of this rose disease which should facilitate its control. More importantly, the novel features of RLRaV help to clarify the molecular and evolutionary features of the closterovirus.

Keywords: closterovirus, deep sequencing, rose, rose leaf rosette-associated virus, rose leaf rosette disease, siRNA.

INTRODUCTION

The rose (*Rosa rugosa* Thunb.) is one of the most important cut flowers and domestic flowers in the world, and is the top-selling flower in many countries (Milleza *et al.*, 2013). The wild rose (*Rosa multiflora* Thunb.) is commonly adopted as a rootstock to reinforce the resistance of rose cultivars against biotic and abiotic stresses. In some countries, virus-indexing procedures are often carried out to ensure that rootstocks are free of virus, but, despite this, viruses remain a problem for the industry. Viral diseases are seldom lethal, but can significantly reduce plant vigour, flower quality and production yield (Milleza *et al.*, 2013).

Recently, a severe virus-like disease, characterized by a leaf rosette or witches' broom symptom formed by dense small leaves on branches, was observed in China. The accompanying symptoms included remarkable decline, dieback and, eventually, plant death several years later. Here, we designate this disease as 'wild rose leaf rosette disease' (WRLRD). It is similar to the previously reported rose rosette disease (RRD), which is regarded as a serious problem by the nursery industry and many home gardeners, because it is lethal to the wild rose and potentially lethal to many ornamental rose species and cultivars. RRD is caused by *Rose rosette virus* (RRV) (Laney *et al.*, 2011). WRLRD differs from RRD in the following ways: first, the witches' broom symptom consists of the proliferation of only small leaves along the canes, instead of both leaves and twigs at the ends of canes; second, the leaves are not distorted and twisted; third, the thorns are not red-tinged or produced in excess.

Like RRD, WRLRD was initially considered to be a phytoplasma disease, but no phytoplasmas were detected using polymerase chain reaction (PCR) with universal primers specific for 16S ribosomal DNA (Lee *et al.*, 2004). Because the gardens in which the symptoms were observed had not been treated with herbicides, their involvement in the disease was also excluded. Deep sequencing of small RNAs (sRNAs), including the small interfering RNAs (siRNAs) and microRNAs (miRNAs), has been shown to be the most promising technique for the identification of RNA and DNA viruses and viroids without a priori knowledge (Adams *et al.*, 2009; Seguin *et al.*, 2014; Wu *et al.*, 2010).

*Correspondence: Email: xuwenxing@mail.hzau.edu.cn

In this study, we applied deep sequencing to characterize the viruses and/or viroids associated with WRLRD. The data revealed the presence of four viruses, namely *Apple stem grooving virus* (ASGV), *Blackberry chlorotic ringspot virus* (BCRV), *Prunus necrotic ringspot virus* (PNRSV) and a novel closterovirus (provisionally named 'rose leaf rosette-associated virus', RLRAV), in the WRLRD-affected wild roses. Molecular characterization of RLRAV showed that it has many novel features relative to known members of the genus *Closterovirus*. Further inspection of the co-infecting viruses in wild rose samples collected in different regions revealed that RLRAV is more likely than the other viruses to be responsible for WRLRD.

RESULTS

Raw deep sequencing data

A wild rose sample with WRLRD symptoms, including severe leaf rosette and dieback, was collected from a rose garden in Wuhan, Hubei Province, China (Fig. 1). A library, prepared from leaf sRNAs sequenced with the Solexa-Illumina platform, resulted in a total of 13 287 506 reads, and 10 154 346 clean reads were obtained. The majority of the reads were 20–25 nucleotides in length, with two peaks of 21 and 24 nucleotides (Fig. S1A, see Supporting Information). Analysis of the 5'-terminal nucleotide revealed a clear bias

Fig. 1 Symptoms of wild rose leaf rosette disease. (A) Overall view of wild rose leaf rosette disease. (B) A typical symptom of a leaf rosette. (C) A closer photograph of a leaf rosette. (D) Associated dieback symptoms.

depending on the size: U of >80% for the sRNAs of 19, 21 and 22 nucleotides, A of 70%, 62% and 98% for 18, 23 and 24 nucleotides, respectively, and G of 55% and 70% for 17 and 20 nucleotides, respectively (Fig. S1B).

Identification and characterization of a novel closterovirus in WRLRD-affected wild roses

De novo assembly of sRNAs generated 10 sequence contigs, with lengths ranging from 388 to 2466 nucleotides, having significant amino acid similarities to members of the family *Closteroviridae*. To obtain the complete genomic sequence of what appeared to be a novel member of this family, provisionally named 'rose leaf rosette-associated virus' (RLRaV) (being designated as isolate RLRaV-CWR.1), overlapping reverse transcription-polymerase chain reaction (RT-PCR) products covering the full genome were produced using primers based on the contigs obtained (Table S1, see Supporting Information). In general, the sequences obtained from Sanger sequencing were in agreement with those generated from sRNA assembly. After sequencing the 5' and 3' terminal regions, the complete genome was determined to be 17 653 nucleotides in length (GenBank Accession No. KJ748003). A BLASTn search of the full-length nucleotide

sequence indicated no significantly similar sequences in the National Center for Biotechnology Information (NCBI) database. Phylogenetic comparison of the amino acid sequence of the heat shock protein 70 homologue (HSP70h) of RLRaV with those of selected members in the family *Closteroviridae* clearly placed RLRaV alongside members of the genus *Closterovirus*, close to *Strawberry chlorotic fleck-associated virus* (SCFaV), *Raspberry leaf mottle virus* (RLMV) and *Citrus tristeza virus* (CTV) (Fig. 2).

Genomic organization of RLRaV RNA

RLRaV has untranslated terminal regions (UTRs) of 139 and 185 nucleotides at the 5' and 3' ends, respectively, and 13 open reading frames (ORFs) (Fig. 3A). The first ORF (ORF1a) codes for a multifunctional 309-kDa protein and includes two papain-like proteases (P-Pro), one methyltransferase (Mt) and one helicase (Hel) domain (Fig. 3A). The polyprotein shares a high amino acid identity with SCFaV, RLMV and CTV (32.1%, 29.5% and 23.9%, respectively) (Table S2, see Supporting Information). The two protease domains contain a conserved motif characteristic of members of the *Closteroviridae* and *Potyviridae* families (Fig. 3C), but having a low identity (27.5%) with each other. The first

Fig. 2 The phylogenetic tree based on the amino acid sequences of the HSP70h gene of the typical and selected members of *Closteroviridae*. The GenBank accession numbers, genera and virus acronyms are given directly in the dendrogram. The scale bar is 0.2 substitutions per site. Viruses included in this analysis are as follows: Beet yellows virus (BYV), Carrot yellow leaf virus (CYLV), Citrus tristeza virus (CTV), Grapevine leafroll-associated virus 2 (GLRaV-2), Mint virus 1 (MV-1), Raspberry leaf mottle virus (RLMV), Strawberry chlorotic fleck-associated virus (SCFaV), Grapevine leafroll-associated virus 3 (GLRaV-3), Plum bark necrosis stem pitting-associated virus (PBNSPaV), Pineapple mealybug wilt-associated virus 1 (PMWaV-1), Lettuce infectious yellows virus (LIYV), Bean yellow disorder virus (BYDV), Blackberry yellow vein-associated virus (BYVaV), Little cherry virus 1 (LChV-1), Grapevine leafroll-associated virus 7 (GLRaV-7) and rose leaf rosette-associated virus (RLRaV). *Arabidopsis thaliana* was used as the outgroup.

Fig. 3 Genome organization, frame shift and papain-like protease analysis of rose leaf rosette-associated virus (RLRaV). (A) Genomic organization of RLraV, showing the relative position of the open reading frames (ORFs) and their expression products. +1FS, putative ribosomal frame shift; P-Pro, papain-like protease; L1 and L2, P-Pro1 and P-Pro2, respectively; Mt, methyltransferase; Hel, helicase; RdRp, RNA-dependent RNA polymerase; HSP70h, heat shock protein homologue; CPm, minor coat protein; CP, coat protein. The line and numbers above the RLraV genome mark the start and end positions of the non-coding regions and ORFs along the genome. (B) Alignment of *Beet yellows virus* (BYV), *Citrus tristeza virus* (CTV) and RLraV nucleotide sequences (presented as DNA) in the vicinity of the proposed frame shift. Numbers below the virus name refer to the start and end positions of the nucleotide sequences. The nucleotides related to the frame shift are underlined. The encoded C-terminus of Hel and N-terminus of RdRp are presented below the nucleotide alignment. (C) Alignment of the C-termini of L-Pro using BLASTp. The arrowheads indicate the catalytic cysteine and histidine residues of the enzymes. L1 and L2, first and second L-Pro located at the C-terminus of the polyprotein, respectively. Codes after the virus names refer to the GenBank accession numbers; the number in brackets refers to the amount of amino acid that is not presented; the open triangle indicates a deletion. GRSLaV, *Grapevine rootstock stem lesion-associated virus* (Q8BEP9); GLRaV2, *Grapevine leafroll-associated virus 2* (O71209); CTV, *Citrus tristeza virus* (O10467); SCFaV, *Strawberry chlorotic fleck-associated virus* (YP_762676); BYV, *Beet yellows virus* (O89906).

protease domain (L1, amino acids 1–413) has high identities (32.5% and 40.3%, respectively) with those of *Beet yellows virus* (BYV) and CTV, whereas the second protease domain (L2, amino acids 414–618) shares lower identities (24% and 27.5%, respectively) with these viruses. It is worth noting that a deletion of a stretch of nine amino acids prior to the conserved histidine residue was found in L1 (Fig. 3C). The conserved Mt domain was identified between residues 679 and 1056 (Rozanov *et al.*, 1992), and the Hel domain was found at the C-terminus of the polyprotein (amino acids 2372–2645) (Marchler-Bauer *et al.*, 2003). A BLASTp search of the Hel domain showed insertions/deletions in six regions,

e.g. an insertion between amino acids 2406 and 2426 and a deletion between amino acids 2519 and 2520, when compared with other members in this family (Fig. S2, see Supporting Information).

Overall, the organization of ORF1a and ORF1b of the RLraV genome is similar to that of BYV (AF190581) and CTV (AF260651) (Agranovsky *et al.*, 1994; Dolja *et al.*, 2006; Karasev *et al.*, 1995), suggesting, by analogy, that ORF1b might be expressed via a +1 ribosomal frame shift resulting in the RNA-dependent RNA polymerase (RdRp) protein (Fig. 3B). The RdRp domain of the fusion protein encodes all eight conserved motifs identified previously

(Koonin, 1991) (data not shown) and shares a high amino acid identity of 57.9%–63.2% with those of the three closest viruses.

Downstream of the polymerase, an ORF codes for a 25-kDa putative protein (p25) that resembles the positions and sizes of p28 of SCFaV, p33 of CTV and p30 of *Beet yellow stunt virus* (BYSV), with no analogues in RLMV. No significant identities were found among p25, p28, p33 and p30, whereas a similar transmembrane region was found in p25. Two transmembrane domains were predicted at the C-terminus of p25, spanning either amino acids 214–231 (leaving the majority of the amino acids in the exterior and five residues in the cytosol) or amino acids 214–232 (alternatively leaving the majority of the amino acids in the cytosol and four residues in the exterior). The quintuple gene block (QGB) was found downstream of the p25 ORF and codes for a 7-kDa hydrophobic protein (p7a), a heat shock protein 70 homologue (HSP70h), a coat protein homologue (CPh), the minor coat protein (CPm) and the major coat protein (CP). BLASTp searches revealed that p7a has no significant similarity with viral proteins in the databases, but, instead, with proteins derived from various biological species, including a fungus, a marine animal and a bacterium (20.0%–21.5% identities) (Table S2). Four insertions (1–8 amino acids) and five deletions (1–15 amino acids) were found for HSP70h in comparison with members of the HSP70 protein family (Fig. S3, see Supporting Information), as well as seven insertions (1–8 amino acids) and four deletions (2–11 amino acids) for CPh in comparison with members in the *Closteroviridae* family (Fig. S4, see Supporting Information).

Downstream of the QGB, five ORFs code for putative proteins showing no significant identity with known viral proteins, except that p19 shares a low identity (15.5%) with p20 of SCFaV (Table S2), both of which do not have significant similarity with any viral proteins in the databases. Analysis of p19 revealed a putative zinc-finger domain (H2C2) between amino acids 65 and 86 (HADCEAIHDLRAAGDAVGVC). The ORF coding for a 17-kDa protein (p17) showed a conserved motif characteristic of the p20 superfamily of RNA silencing suppressors (RSSs) (Fig. S5, see Supporting Information). It is worth noting that each of the remaining five putative proteins shares detectable similarity with proteins from prokaryotic organisms, e.g. p25 shares 16.1% identity with proteins of *Mycobacterium gilvum*, and p19 shares 12.7% identity with a LuxR family transcriptional regulator of *Pseudomonas* sp. (Table S2). Of the remaining putative proteins, only p13 shares detectable similarity with a plant protein (15.8% identity), namely with a putative F-box protein of *Solanum tuberosum*. In addition, p17 and p12 also share detectable similarity with fungal proteins (Table S2).

Distribution of siRNA along the RLraV genome

Alignments of the siRNAs and RLraV sequences showed that the former fully cover the RLraV genome (Fig. S6B, see Supporting

Information). The majority of these reads were 20–22 nucleotides in length, with a dominant peak at 21 nucleotides (Fig. S6A). Analysis of the 5'-terminal nucleotide of siRNAs revealed a dominant U (40%–53%), an under-represented G (6%–11%) and moderate proportions of A (20%–33%) and C (23%–25%) (Fig. S1C). The frequency and distribution of siRNA coverage over the RLraV genome were almost even, with a slightly increased accumulation at the 3' terminal region regardless of the RNA polarities (the amount of siRNAs derived from the positive polarity was slightly higher than that from the negative polarity) (Fig. S6A). Examination of the siRNA profiles revealed a notable hotspot between positive polarity positions 15875 and 15895 within the gene region encoding p19 (Fig. S6B). The most conserved siRNA sequence for the hotspot was 5'-UCACGUGACUGUGAAGACGC-3'.

Identification of four co-infecting viruses in WRLRD-affected wild roses

Comparison of the virus-derived siRNAs through reference-guided assembly with known virus and viroid genomes also revealed complete genomes, without any gaps, of three reported viruses: ASGV, BCRV and PNRSV, which were designated as isolates ASGV-CWR.1, BCRV-CWR.1 and PNRSV-CWR.1, respectively.

The genome of ASGV-CWR.1 is 6496 nucleotides in length, similar to the known isolates. BLASTn searches revealed the highest identity (97%, 99% coverage and 0.0 e value) between ASGV-CWR.1 and another isolate from an apple in Japan (GenBank Accession No. D14995) (Yoshikawa *et al.*, 1992). The genomes of RNA1 to RNA3 of BCRV-CWR.1 are 3478, 2879 and 2290 nucleotides in length, respectively, and share the highest nucleotide sequence identity (96%–97%, 99%–100% coverage and 0.0 e value) with those of a BCRV isolate from *Rubus* sp. in the UK (DQ091193, DQ091194 and DQ091195) (Jones *et al.*, 2006). The genomes of RNA1 to RNA3 of PNRSV-CWR.1 are 3332, 2591 and 1957 nucleotides in length, respectively, and share the highest nucleotide sequence identity (97%–98%, 98%–100% coverage and 0.0 e value) with RNA1 and RNA2 (AF278534 and AF278535) of a cherry isolate and RNA3 (U57046) of a peach isolate in the USA (Di Terlizzi *et al.*, 2001; Scott *et al.*, 1998).

RT-PCR identification of the co-infected viruses

Twenty wild rose samples from different Wuhan regions in the autumn and spring of two subsequent years were collected for RT-PCR identification of RLraV using the primers listed in Table S3 (see Supporting Information). The results obtained indicated that 12 samples were positive for RLraV, which was confirmed by independently using the two pairs of primers, whereas only three samples were positive for BCRV, PNRSV or ASGV. Seven of the samples showing WRLRD or similar symptoms (small

leaves not so dense) were positive for RLraV, without co-infection by BCRV, ASGV and PRNSV (Fig. S7, see Supporting Information).

DISCUSSION

The bizarre symptoms of WRLRD are clearly different from those associated with known diseases, with no other viruses infecting roses or wild roses having been reported as responsible for WRLRD (Laney *et al.*, 2011; Milleza *et al.*, 2013; Moury *et al.*, 2001; Rivera and Engel, 2010). Therefore, to our knowledge, this is the first description of the disease. Deep sequencing of viral-derived siRNA revealed that four viruses, three known (PNRSV, BCRV and ASGV) and one reported here for the first time (RLraV), might induce the symptoms. Of these, PNRSV, from the genus *Ilarvirus*, family *Bromoviridae*, has been reported in many rose-growing regions world-wide (Moury *et al.*, 2001). PNRSV is either latent or induces line patterns, ring spots or yellow nets on rose leaves (Paduch-Cichal and Sala-Rejczak, 2011), but a leaf rosette symptom on roses or wild roses has never been reported in association with PNRSV. BCRV, also belonging to the genus *Ilarvirus*, has been reported in RRD-affected roses in the USA (Tzanetakis *et al.*, 2006a). ASGV, of the genus *Capillovirus*, family *Betaflexiviridae*, was the first reported virus to infect wild roses, but its biological impact on roses remains unclear. Most of the RLraV-infected samples were not co-infected by BCRV, ASGV or PRNSV, but showed WRLRD or similar symptoms, thus suggesting that RLraV is most likely to be responsible for WRLRD. To fulfil Koch's postulates, we tried to transmit the virus to *Nicotiana benthamiana* and cucumbers for virion purification, without success. As illustrated by SCFaV, it is quite difficult to purify the fragile flexuous virions for further inoculation into herbaceous hosts (Tzanetakis *et al.*, 2007). Recently, using an agroinoculation approach, transmission of CTV to *N. benthamiana* was accomplished. This approach, however, has been successful for only one specific CTV isolate so far (Ambrós *et al.*, 2011). We believe that the discovery of this novel virus is helpful for further clarification of the responsible aetiology and for the selection of sources and implementation of management practices for the control of this devastating disease.

More than 24 viruses belonging to 10 genera of eight families have been reported to infect roses or wild roses, but no closteroviruses or betaflexiviruses are known to infect *Rosa* plants (Milleza *et al.*, 2013). To our knowledge, this is the first report in this respect. A limited range of closteroviruses have been reported infecting rosaceous plants, i.e. SCFaV and *Strawberry pallidosis-associated virus* (SPaV) infecting strawberries, and RLMV and *Blackberry yellow vein-associated virus* (BYVaV) infecting raspberries and blackberries, respectively (Tzanetakis *et al.*, 2005, 2006b, 2007).

Considering that the full-length genome sequences of PRNSV, BCRV and ASGV can be assembled from their siRNAs and that they share high identities with sequences reported previously, we did not re-sequence their full genomes with conventional sequencing (although we corroborated their presence by RT-PCR of partial regions). However, for the novel virus RLraV, we performed overlapping RT-PCR to amplify the whole viral genome for conventional sequencing, so that a full-length genomic sequence independent of the siRNA-assembled contigs was obtained. Analysis of the overlapping RT-PCR fragments revealed a high-quality sequence, with clear-cut determination of each nucleotide, which was in good agreement with the sequence resulting from the siRNA-assembled contigs. Analysis of the siRNAs derived from the RLraV genome revealed that most siRNAs were 21 nucleotides in length, suggesting that their genesis is most likely mediated by DICER-like enzyme 4. This result differs from that obtained for CTV in Mexican lime, sour orange and sweet orange, in which both 21- and 22-nucleotide siRNAs were the most prevalent (Ruiz-Ruiz *et al.*, 2011). The bias found at the 5' termini of RLraV siRNAs is similar to the bias observed for CTV siRNAs (Ruiz-Ruiz *et al.*, 2011), but differs obviously from that of the total siRNAs in the sequenced wild rose sample (compare with Fig. S1B, C). This might be caused by different ARGONAUTES (AGOs) associated with their synthesis (Adams *et al.*, 2009; Mi *et al.*, 2008; Montgomery *et al.*, 2008), although technical reasons cannot be excluded at this stage. The even distribution of siRNA coverage over the RLraV genome is also different from CTV in the three citrus hosts, in which the siRNAs accumulate preferentially at the 3' terminal region (Ruiz-Ruiz *et al.*, 2011).

Comparative analysis of RLraV revealed two conserved gene blocks (ORF1 and QGB). The ORF1a–ORF1b encodes the replication-related proteins P-Pro, Mt, Hel and RdRp (Fig. 3A). Of the proteases, the L1 and L2 proteases play roles in replication and long-distance transport (Liu *et al.*, 2009; Peng *et al.*, 2001), whereas the others function in replication (Peremyslov *et al.*, 1998; Satyanarayana *et al.*, 1999) (Fig. 3A). The QGB, a hallmark of the family *Closteroviridae*, includes a homologue of HSP70 and three divergent copies of the CP gene (Fig. 3A). The proteins play roles in cell-to-cell movement and virion assembly, with Hsp70h (Napuli *et al.*, 2000; Peremyslov *et al.*, 1999, 2004; Satyanarayana *et al.*, 2004), CPh (Napuli *et al.*, 2003) and CPm (Agranovsky *et al.*, 1995; Alzhanova *et al.*, 2001, 2007; Satyanarayana *et al.*, 2004) being responsible for tail assembly, and CP for virion body assembly (Agranovsky *et al.*, 1995; Alzhanova *et al.*, 2000). The remaining ORFs showed no significant similarities with known viral proteins and are unique to RLraV, encoding the putative proteins p25, p7a, p7b, p13, p17 and p12, without detectable similarities to any viral proteins in the databases, and p19, which shared low similarity only with SCFaV p20. p19 contains a putative zinc-finger domain with two

coordinating histidines and two cysteines, similar to that found in CTV p23 (Flores *et al.*, 2013), suggesting its potential relationship to analogous proteins in CTV and other closteroviruses that are implicated in RNAi suppression (Chiba *et al.*, 2006; Lu *et al.*, 2004), and p17 contains a conserved motif characteristic of an RSS, suggesting that it is an analogue/homologue of BYV p21, CTV p20 and GLRaV-2 p24 (Chiba *et al.*, 2006; Lu *et al.*, 2004; Reed *et al.*, 2003). The remaining proteins, p7b, p13 and p12, also have no homologues in the genus *Closterovirus*, making the genomic organization of RLraV quite singular at the 3' terminus, because other genus members, e.g. SCFaV, RLMV, CTV and BYV, show very similar genomic organization, except that CTV contains the additional p18, p13 and p23 proteins (Fig. S8, see Supporting Information). It is worth noting that these unique putative proteins of RLraV share detectable similarity with fungal and bacterial proteins.

In summary, compared with the 8–12 ORFs of the known members of the genus *Closterovirus*, RLraV has 13 ORFs, some sharing similarities with bacterial or fungal proteins, two possible small-sized (17 and 19 kDa) RSSs, and a complex genome organization at the 3'-terminal half. In addition, some proteins, including L2, Mt, HSP70h and CPh, show amino acid insertions and deletions in their conserved motifs when compared with those of the members of the corresponding families. Considering the abundance of unique, rare and novel genes that appear to be picked at random by different closteroviruses (Dolja *et al.*, 2006), RLraV seems to be the most complex virus in its genus. In addition, unlike other members that contain two P-Pros in the genus, RLraV contains two P-Pros that share obviously different identities from the others, with L2 sharing a low identity (24.0%–27.5%) with the P-Pros of BYV and CTV, and L1 sharing a high identity (32.5%–40.3%) with them. Furthermore, the identities shared by L1 and other P-Pros are similar to those shared by the hallmark proteins (HSP70h and three divergent copies of CP) or other observed proteins and those of members of the genus, e.g. identities are 47.5%–52.2% for HSP70h, 32.3%–39.4% for CP, 34.2%–38.1% for CPh and 57.9%–63.2% for RdRp. These comparisons suggest that L1 of RLraV evolved together with MET-HEL-POL and QGB, whereas L2 was acquired independently, because they share a low identity (27.5%). The acquisition of the L1 gene might have facilitated RNA replication and conditioned the expansion of the progenitor's genome together with the replicase (Dolja *et al.*, 2006). The different evolution between L1 and L2 is consistent with the view that the last common ancestor of closteroviruses descended from a plant virus of the alphavirus-like superfamily that possessed a typical MET-HEL-POL replicase, a p6-like movement protein (MP) and a single CP, and this progenitor further evolved by acquiring several new genes (Dolja *et al.*, 2006). The finding of RLraV with novel features should aid in the understanding of the molecular and evolutionary features of the closterovirus.

EXPERIMENTAL PROCEDURES

Sample preparation and deep sequencing

Leaves were collected in a garden in Wuhan, Hubei Province, China in May 2013 from a wild rose sample showing small leaf rosettes on branches, dieback and severe decline symptoms (Fig. 1). Collected leaves were quickly frozen in liquid nitrogen, preserved in carbon dioxide ice blocks and shipped for 3–4 days to Biomarker Technologies Corporation, Beijing, China for deep sequencing. One microgram of total RNA was extracted, unique barcoded adaptors were added, RT-PCR was performed and the product was purified via polyacrylamide gel electrophoresis (PAGE) for sRNA library construction and sequencing in an Illumina HiSeq™ 2000 (Illumina, Inc., San Diego, CA, USA).

sRNA sequence processing, assembly and virus genome identification

Raw Illumina sRNA reads were trimmed and cleaned by removing the sequences smaller than 16 nucleotides or longer than 30 nucleotides, low-quality tags, polyA or N tags. Clean sRNAs were assembled *de novo* using Velvet with a k-mer value of 17 (Zerbino and Birney, 2008) and CAP3 with default values (Huang and Madan, 1999). Cleaned reads were aligned to the known virus and viroid genomes collected from the NCBI GenBank database using Bowtie software (Langmead *et al.*, 2009). Possible novel viroids were screened using a previously developed computational algorithm designed for the discovery of homologous independent circular RNAs (Wu *et al.*, 2010).

The distribution of viral small RNAs (vsRNAs) along the viral genome was determined using Bowtie software allowing up to three mismatches, and the base coverage at each position of the contigs was derived on the basis of the mpileup file generated from the alignments using SAMtools under default parameters (Li *et al.*, 2009); this information was used to correct potential base calling errors in the *de novo* assemblies by replacing them with the most frequent base appearing at that specific position.

Validation and completion of viral genome sequences with Sanger sequencing

Sequence gaps between clones were determined by RT-PCR using specific primers designed from the cDNA sequences obtained (Table S1). The 5' and 3' terminal sequences of the viral RNA were determined through rapid amplification of cDNA ends (RACE) with a kit (Invitrogen, Shanghai, China) following the manufacturer's instructions. Briefly, the RNA terminus was ligated with an adaptor primer using T4 RNAligase, and the oligonucleotide-ligated RNA was reverse transcribed with Moloney Murine Leukaemia Virus (M-MLV) reverse transcriptase and the primer complementary to the oligonucleotide used for the RNA ligation. The cDNA was subjected to nested amplification using the primers numbering 5' or 3' (F/R and Fn/Rn used for the first and second rounds of amplification, respectively) (Table S1). The authenticity of siRNA-assembled viral genome sequences was validated through Sanger sequencing of the RT-PCR products covering the full genome. The amplified PCR products were cloned into the pMD18-T vector (TaKaRa, Dalian, China) and trans-

formed into competent cells of *Escherichia coli* DH5 α . Sequencing was performed at Sangon Biotech (Shanghai) Co., Ltd., China, and each nucleotide was determined from at least three independent overlapping clones.

Sequence analysis

Sequence identity searches were performed using the NCBI databases with the BLAST program. Identity analyses were performed with the Molecular Evolutionary Genetic Analysis (MEGA; version 4) program (Tamura *et al.*, 2007). The tree was reconstructed by the neighbour-joining method with 1000 bootstrap replicates using MEGA5 (Tamura *et al.*, 2011). Bootstrap values are indicated at the branch points. ORFs were deduced using the DNAMAN DNA analysis software package (DNAMAN version 6.0; Lynnon Biosoft, Montreal, QC, Canada). Transmembrane domains were predicted at http://www.ch.embnet.org/software/TMPRED_form.html, and only scores above 500 were considered to be significant (Hofmann and Stoffel, 1993).

Virus detection with RT-PCR

Total RNA was extracted from wild rose leaves according to a previous method (Carra *et al.*, 2007). The extracted RNAs were subjected directly to reverse transcription. The first-strand cDNA was synthesized using M-MLV reverse transcriptase (Promega Corporation, Madison, WI, USA) employing random hexamer primers (TaKaRa Biotechnology Corporation, Dalian, China) at 37 °C for 1 h. Double-stranded DNA was amplified with *Taq* DNA polymerase (TaKaRa Biotechnology Corporation) with specific primers (Table S3) using the following reaction conditions: initial denaturation at 94 °C for 3 min, followed by 35 cycles of denaturation at 94 °C for 30 s, annealing at 56 °C (ASGV, BCRV, RLRAV) or 60 °C (PNRSV) for 30 s, extension at 72 °C for 45 s and a final extension at 72 °C for 10 min. PCR products were subjected to electrophoresis in 1% agarose gels, purified using an agarose gel DNA purification kit and ligated into the vector pMD18-T (TaKaRa Biotechnology Corporation), followed by transformation into *E. coli* DH5 α .

ACKNOWLEDGEMENTS

This research was financially supported by the National Natural Science Foundation of China (Nos. 31071662 and 30800746). The authors declare no competing interests. The authors thank Professor Ricardo Flores [Instituto de Biología Molecular y Celular de Plantas (UPV-CSIC), Valencia, Spain] for kindly reviewing the manuscript, and Professor Francesco Di Serio (Istituto di Virologia Vegetale, Consiglio Nazionale delle Ricerche, Strada delle Cacce 73, Torino, Italy) for kindly analysing possible novel viroids.

REFERENCES

- Adams, I., Glover, R., Monger, W., Mumford, R. and Jackeviciene, E. (2009) Next generation sequencing and metagenomic analysis: a universal diagnostic tool in plant virology. *Mol. Plant Pathol.* **10**, 537–545.
- Agranovsky, A.A., Koonin, E.V., Boyko, V.P., Maiss, E., Frötschl, R., Lunina, N.A. and Atabekov, J.G. (1994) Beet yellows closterovirus: complete genome structure and identification of a leader papain-like thiol protease. *Virology*, **198**, 311–324.
- Agranovsky, A.A., Lesemann, D.E., Maiss, E., Hull, R. and Atabekov, J.G. (1995) 'Rattlesnake' structure of a filamentous plant RNA virus built of two capsid proteins. *Proc. Natl. Acad. Sci. USA*, **92**, 2470–2473.
- Alzhanova, D.V., Hagiwara, Y., Peremyslov, V.V. and Dolja, V.V. (2000) Genetic analysis of the cell-to-cell movement of beet yellows closterovirus. *Virology*, **268**, 192–200.
- Alzhanova, D.V., Napuli, A.J., Creamer, R. and Dolja, V.V. (2001) Cell-to-cell movement and assembly of a plant closterovirus: roles for the capsid proteins and Hsp70 homolog. *EMBO J.* **20**, 6997–7007.
- Alzhanova, D.V., Prokhnevsky, A.I., Peremyslov, V.V. and Dolja, V.V. (2007) Virion tails of Beet yellows virus: coordinated assembly by three structural proteins. *Virology*, **359**, 220–226.
- Ambrós, S., El-Mohtar, C., Ruiz-Ruiz, S., Peña, L., Guerri, J., Dawson, W.O. and Moreno, P. (2011) Agroinoculation of *Citrus tristeza virus* causes systemic infection and symptoms in the presumed nonhost *Nicotiana benthamiana*. *Mol. Plant–Microbe Interact.* **24**, 1119–1131.
- Carra, A., Gambino, G. and Schubert, A. (2007) A cetyltrimethylammonium bromide-based method to extract low-molecular-weight RNA from polysaccharide-rich plant tissues. *Anal. Biochem.* **360**, 318–320.
- Chiba, M., Reed, J.C., Prokhnevsky, A.I., Chapman, E.J., Mawassi, M., Koonin, E.V., Carrington, J.C. and Dolja, V.V. (2006) Diverse suppressors of RNA silencing enhance agroinfection by a viral replicon. *Virology*, **346**, 7–14.
- Di Terlizzi, B., Skrzeczkowski, L.J., Mink, G.I., Scott, S.W. and Zimmerman, M.T. (2001) The RNA 5 of *Prunus necrotic ringspot virus* is a biologically inactive copy of the 3'-UTR of the genomic RNA 3. *Arch. Virol.* **146**, 825–833.
- Dolja, V.V., Kreuze, J.F. and Valkonen, J.P. (2006) Comparative and functional genomics of closteroviruses. *Virus Res.* **117**, 38–51.
- Flores, R., Ruiz-Ruiz, S., Soler, N., Sánchez-Navarro, J., Fagoaga, C., López, C., Navarro, L., Moreno, P. and Peña, L. (2013) *Citrus tristeza virus* p23: a unique protein mediating key virus–host interactions. *Front. Microbiol.* **98**, 1–9.
- Hofmann, K. and Stoffel, W. (1993) TMbase—a database of membrane spanning protein segments. *Biol. Chem. Hoppe Seyler*, **374**, 166.
- Huang, X. and Madan, A. (1999) CAP3: a DNA sequence assembly program. *Genome Res.* **9**, 868–877.
- Jones, A.T., McGavin, W.J., Gepp, V., Zimmerman, M.T. and Scott, S.W. (2006) Purification and properties of blackberry chlorotic ringspot, a new virus species in subgroup 1 of the genus *Ilarvirus* found naturally infecting blackberry in the UK. *Ann. Appl. Biol.* **149**, 125–135.
- Karasev, A.V., Boyko, V.P., Gowda, S., Nikolaeva, O.V., Hilf, M.E., Koonin, E.V., Niblett, C.L., Cline, K., Gumpf, D.J., Lee, R.F., Garnsey, S.M., Lewandowski, D.J. and Dawson, W.O. (1995) Complete sequence of the *Citrus tristeza virus* RNA genome. *Virology*, **208**, 511–520.
- Koonin, E.V. (1991) The phylogeny of RNA-dependent RNA polymerases of positive-strand RNA viruses. *J. Gen. Virol.* **72**, 2197–2206.
- Laney, A.G., Keller, K.E., Martin, R.R. and Tzanetakis, I.E. (2011) A discovery 70 years in the making: characterization of the rose rosette virus. *J. Gen. Virol.* **92**, 1727–1732.
- Langmead, B., Trapnell, C., Pop, M. and Salzberg, S.L. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25.
- Lee, I.M., Martini, M., Marccone, C. and Zhu, S.F. (2004) Classification of phytoplasma strains in the elm yellows group (16SrV) and proposal of 'Candidatus Phytoplasma ulmi' for the phytoplasma associated with elm yellows. *Int. J. Syst. Evol. Microbiol.* **54**, 337–347.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T. and Ruan, J. (2009) The sequence alignment/map format and SAMtools. *Bioinformatics*, **25**, 2078–2079.
- Liu, Y., Peremyslov, V.V., Medina, V. and Dolja, V.V. (2009) Tandem leader proteases of Grapevine leafroll-associated virus-2: host-specific functions in the infection cycle. *Virology*, **383**, 291–299.
- Lu, R., Folimonov, A.S., Shintaku, M., Li, W.X., Falk, B.W., Dawson, W.O. and Ding, S.W. (2004) Three distinct suppressors of RNA silencing encoded by a 20-kb viral RNA genome. *Proc. Natl. Acad. Sci. USA*, **101**, 15 742–15 747.
- Marchler-Bauer, A., Anderson, J.B., DeWeese-Scott, C., Fedorova, N.D., Geer, L.Y., He, S., Hurwitz, D.I., Jackson, J.D., Jacobs, A.R., Lanczycki, C.J., Liebert, C.A., Liu, C., Madej, T., Marchler, G.H., Mazumder, R., Nikolskaya, A.N., Panchenko, A.R., Rao, B.S., Shoemaker, B.A., Simonyan, V., Song, J.S., Thiessen, P.A., Vasudevan, S., Wang, Y., Yamashita, R.A., Yin, J.J. and Bryant, S.H. (2003) CDD, a curated Entrez database of conserved domain alignments. *Nucleic Acids Res.* **31**, 383–387.
- Mi, S., Cai, T., Hu, Y., Chen, Y., Hodges, E., Ni, F., Wu, L., Li, S., Zhou, H., Long, C., Chen, S., Hannon, G.J. and Qi, Y. (2008) Sorting of small RNAs into Arabidopsis

- argonaute complexes is directed by the 5' terminal nucleotide. *Cell*, **133**, 116–127.
- Milleza, E.J.M., Ward, L.I., Delmiglio, C., Tang, J.Z., Veerakone, S. and Perez-Egusquiza, Z. (2013) A survey of viruses infecting *Rosa* spp. in New Zealand. *Australas. Plant Pathol.* **42**, 313–320.
- Montgomery, T.A., Howell, M.D., Cuperus, J.T., Li, D., Hansen, J.E., Alexander, A.L., Chapman, E.J., Fahlgren, N., Allen, E. and Carrington, J.C. (2008) Specificity of ARGONAUTE7–miR390 interaction and dual functionality in TAS3 trans-acting siRNA formation. *Cell*, **133**, 128–141.
- Moury, B., Cardin, L., Onesto, J.P., Candresse, T. and Poupet, A. (2001) Survey of Prunus necrotic ringspot virus in rose and its variability in rose and *Prunus* spp. *Phytopathology*, **91**, 84–91.
- Napuli, A.J., Falk, B.W. and Dolja, V.V. (2000) Interaction between HSP70 homolog and filamentous virions of the Beet yellows virus. *Virology*, **274**, 232–239.
- Napuli, A.J., Alzhanova, D.V., Doneanu, C.E., Barofsky, D.F., Koonin, E.V. and Dolja, V.V. (2003) The 64-kDa capsid protein homolog of beet yellows virus is required for assembly of virion tails. *J. Virol.* **77**, 2377–2384.
- Paduch-Cichal, E. and Sala-Rejczak, K. (2011) Biological and molecular characterization of Prunus necrotic ringspot virus isolates from three rose cultivars. *Acta Physiol. Plant.* **33**, 2349–2354.
- Peng, C.W., Peremyslov, V.V., Mushegian, A.R., Dawson, W.O. and Dolja, V.V. (2001) Functional specialization and evolution of leader proteinases in the family Closteroviridae. *J. Virol.* **75**, 12 153–12 160.
- Peremyslov, V.V., Hagiwara, Y. and Dolja, V.V. (1998) Genes required for replication of the 15.5-kilobase RNA genome of a plant closterovirus. *J. Virol.* **72**, 5870–5876.
- Peremyslov, V.V., Hagiwara, Y. and Dolja, V.V. (1999) HSP70 homolog functions in cell-to-cell movement of a plant virus. *Proc. Natl. Acad. Sci. USA*, **96**, 14 771–14 776.
- Peremyslov, V.V., Andreev, I.A., Prokhnevsky, A.I., Duncan, G.H., Taliany, M.E. and Dolja, V.V. (2004) Complex molecular architecture of beet yellows virus particles. *Proc. Natl. Acad. Sci. USA*, **101**, 5030–5035.
- Reed, J.C., Kasschau, K.D., Prokhnevsky, A.I., Gopinath, K., Pogue, G.P., Carrington, J.C. and Dolja, V.V. (2003) Suppressor of RNA silencing encoded by Beet yellows virus. *Virology*, **306**, 203–209.
- Rivera, P.A. and Engel, E.A. (2010) Presence of rose spring dwarf-associated virus in Chile: partial genome sequence and detection in roses and their colonizing aphids. *Virus Genes*, **41**, 295–297.
- Rozanov, M.N., Koonin, E.V. and Gorbalenya, A.E. (1992) Conservation of the putative methyltransferase domain: a hallmark of the 'Sindbis-like' supergroup of positive-strand RNA viruses. *J. Gen. Virol.* **73**, 2129–2134.
- Ruiz-Ruiz, S., Navarro, B., Gisel, A., Pena, L., Navarro, L., Moreno, P., Di Serio, F. and Flores, R. (2011) *Citrus tristeza virus* infection induces the accumulation of viral small RNAs (21–24-nt) mapping preferentially at the 30-terminal region of the genomic RNA and affects the host small RNA profile. *Plant Mol. Biol.* **75**, 607–619.
- Satyanarayana, T., Gowda, S., Boyko, V.P., Albiach-Marti, M.R., Mawassi, M., Navas-Castillo, J., Karasev, A.V., Dolja, V., Hilf, M.E., Lewandowski, D.J., Moreno, P., Bar-Joseph, M., Garnsey, S.M. and Dawson, W.O. (1999) An engineered closterovirus RNA replicon and analysis of heterologous terminal sequences for replication. *Proc. Natl. Acad. Sci. USA*, **96**, 7433–7438.
- Satyanarayana, T., Gowda, S., Ayllon, M.A. and Dawson, W.O. (2004) Closterovirus bipolar virion: evidence for initiation of assembly by minor coat protein and its restriction to the genomic RNA 5' region. *Proc. Natl. Acad. Sci. USA*, **101**, 799–804.
- Scott, S.W., Zimmermann, M.T., Ge, X. and Mackenzie, D.J. (1998) The coat proteins and putative movement proteins of isolates of *Prunus necrotic ringspot virus* from different host species and geographic origins are extensively conserved. *Eur. J. Plant Pathol.* **104**, 155–161.
- Seguin, J., Rajeswaran, R., Malpica-López, N., Martin, R.R., Kasschau, K., Dolja, V.V., Otten, P., Farinelli, L. and Pooggin, M.M. (2014) De novo reconstruction of consensus master genomes of plant RNA and DNA viruses from siRNAs. *PLoS One*, **9**, e88513.
- Tamura, K., Dudley, J., Nei, M. and Kumar, S. (2007) MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* **24**, 1596–1599.
- Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M. and Kumar, S. (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* **28**, 2731–2739.
- Tzanetakis, I.E., Reed, J. and Martin, R.R. (2005) Nucleotide sequence, genome organization and phylogenetic analysis of Strawberry pallidosis associated virus, a new member of the genus crinivirus. *Arch. Virol.* **150**, 273–286.
- Tzanetakis, I.E., Gergerich, R.C. and Martin, R.R. (2006a) A new Ilarvirus found in rose. *Plant Pathol.* **55**, 568.
- Tzanetakis, I.E., Susaimuthu, J., Gergerich, R.C. and Martin, R.R. (2006b) Nucleotide sequence of Blackberry yellow vein associated virus, a novel member of the Closteroviridae. *Virus Res.* **116**, 196–200.
- Tzanetakis, I.E., Halgrena, A. and Martin, R.R. (2007) Strawberry chlorotic fleck: identification and characterization of a novel Closterovirus associated with the disease. *Virus Res.* **124**, 88–94.
- Wu, Q., Luo, Y., Lu, R., Lau, N., Lai, E.C., Li, W.X. and Ding, S.W. (2010) Virus discovery by deep sequencing and assembly of virus derived small silencing RNAs. *Proc. Natl. Acad. Sci. USA*, **107**, 1606–1611.
- Yoshikawa, N., Sasaki, E., Kato, M. and Takahashi, T. (1992) The nucleotide sequence of apple stem grooving capillovirus genome. *Virology*, **191**, 98–105.
- Zerbino, D.R. and Birney, E. (2008) Algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res.* **18**, 821–829.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Fig. S1 Graph of the percentages of variably sized small interfering RNAs (siRNAs) in clean reads in the wild rose sample affected by wild rose leaf rosette disease (WRLRD), and a percentage accumulation histogram for the frequency of the 5'-nucleotide of siRNAs. (A) Percentages of differently sized siRNAs. (B) Frequency of the 5'-nucleotide in the total clean siRNAs. (C) Frequency of the 5'-nucleotide in the rose leaf rosette-associated virus (RLRaV) siRNAs.

Fig. S2 The BLASTp search result for the RNA helicase of rose leaf rosette-associated virus (RLRaV), showing a motif conserved for the reverse transcriptase (RT, RNA-dependent DNA polymerase)-like family. ShMV, *Sunn-hemp mosaic virus* (P89202); HEV, *Hepatitis E virus* (P29324); CTV, *Citrus tristeza virus* (O10467); GLRaV2, *Grapevine leafroll-associated virus 2* (Q71209); LIYV, *Lettuce infectious yellows virus* (Q83044); TMGMV, *Tobacco mild green mosaic virus* (P18339); TMV, *Tobacco mosaic virus* (P03586); PDV, *Prune dwarf virus* (P89033); EMV, *Elm mottle virus* (Q98701). Filled and open triangles indicate an insertion or deletion, respectively, in the members obtained via the BLAST search.

Fig. S3 The BLASTp search result for HSP70h of rose leaf rosette-associated virus (RLRaV), showing a motif conserved for the HSP70 (70-kDa heat shock protein) family. Filled and open triangles indicate an insertion or deletion, respectively, in the members obtained via the BLAST search. The GenBank accession numbers are included in parentheses after the organism name as follows: *Saccharomyces cerevisiae* (3D2F_A), *Escherichia coli* (P0A6Y8), *Borrelia burgdorferi* (P0C922), *Bacillus megaterium* (P05646), *Clostridium acetobutylicum* (P30721), *Methanosarcina mazei* (P0CW13), *Pisum sativum* (Q02028), *Pavlova lutheri* (P30722) and *Kluyveromyces lactis* (P22010).

Fig. S4 The BLASTp search result for the coat protein homologue (CPh) of rose leaf rosette-associated virus (RLRaV), showing a motif conserved in a viral heat shock protein Hsp90 homologue. Filled and open triangles indicate an insertion or deletion, respec-

tively, in the representative sequences of the 18 known homologues. PMWaV1, *Pineapple mealybug wilt-associated virus* 1 (Q8V3A0); BYVaV, *Blackberry yellow vein-associated virus* (Q5ICV1); CYV, *Cucumber yellows virus* (Q806Z4); SPaV, *Strawberry pallidosis-associated virus* (Q6JGV9); LIYV, *Lettuce infectious yellows virus* (Q83048); LCV, *Little cherry virus* 1 (Q98160); CTV, *Citrus tristeza virus* (Q9YN83); MV1, *Mint virus* 1 (Q5G7G3); BYSV, *Beet yellow stunt virus* (Q65856).

Fig. S5 The BLASTp search result for p17 of rose leaf rosette-associated virus (RLRaV), showing a motif characteristic of the RSS_P20 superfamily, together with p20 or p23 of SCFaV, RLMV and CTV. SCFaV, *Strawberry chlorotic fleck-associated virus* (Q0GK46); RLMV, *Raspberry mottle virus* (YP_874193); CTV, *Citrus tristeza virus* (Q6X988).

Fig. S6 Size distribution of small RNAs (sRNAs). (A) Bar graph for size distribution of sRNAs in the library prepared from diseased wild rose leaves. (B) Distribution of all size classes (18–27 nucleotides) of viral small RNAs (vsRNAs) captured by high-throughput sequencing along the genome of rose leaf rosette-associated virus (RLRaV).

Fig. S7 Symptoms and infected viruses identified by reverse transcription-polymerase chain reaction (RT-PCR) of 20 samples collected in the Wuhan region. Sample names are shown under the

corresponding symptomatic photographs; more than one sample under the same photograph indicates that they show the same symptom. (+) and (–), positive and negative for rose leaf rosette-associated virus (RLRaV), respectively; 'BCRV+', 'PNRSV+' and 'ASGV+', positive for *Blackberry chlorotic ringspot virus* (BCRV), *Prunus necrotic ringspot virus* (PNRSV) and *Apple stem grooving virus* (ASGV), respectively; otherwise negative for BCRV, PNRSV and ASGV.

Fig. S8 Comparison of the genetic and structural complexity of rose leaf rosette-associated virus (RLRaV) with the closest members of *Strawberry chlorotic fleck-associated virus* (SCFaV), *Raspberry mottle virus* (RLMV), *Citrus tristeza virus* (CTV) and the type species *Beet yellows virus* (BYV). The full protein names are listed in Fig. 2.

Table S1 Primers used for reverse transcription-polymerase chain reaction (RT-PCR) to amplify the full-length genome of rose leaf rosette-associated virus (RLRaV).

Table S2 Summary of results of amino acid alignment for putative proteins of rose leaf rosette-associated virus (RLRaV).

Table S3 Primers used for detection of *Blackberry chlorotic ringspot virus* (BCRV), *Prunus necrotic ringspot virus* (PNRSV), *Apple stem grooving virus* (ASGV) and rose leaf rosette-associated virus (RLRaV).