eLIFE
elifesciences.org

Figures and figure supplements

Unprecedented genomic diversity of RNA viruses in arthropods reveals the ancestry of negative-sense RNA viruses

Ci-Xiu Li, et al.

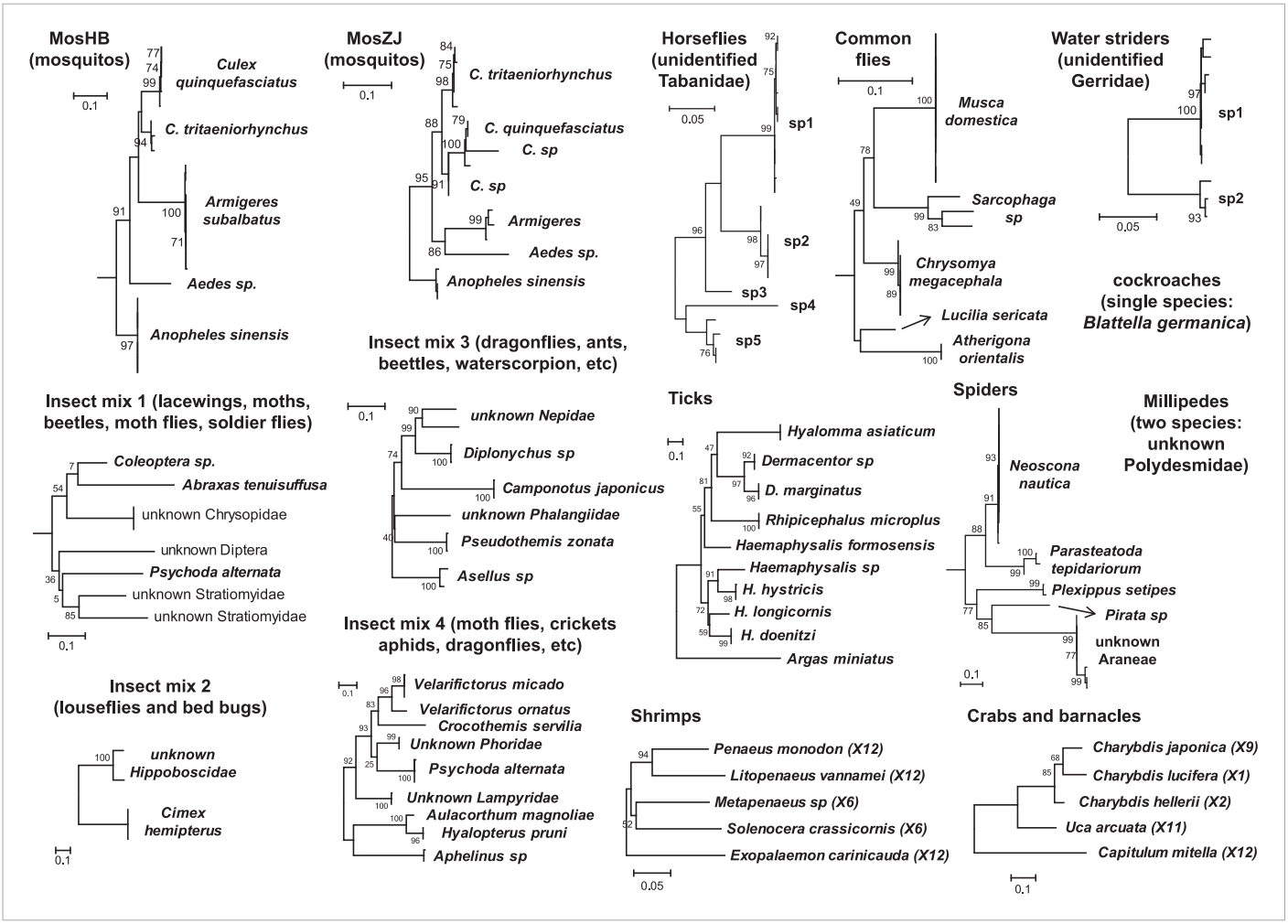

Figure 1. Host component of each pool used in the RNA-seq library construction and sequencing. The taxonomic units in the tree correspond to the unit samples used in the RNA extraction. Species or genus information is marked to the left of the tree.

DOI: 10.7554/eLife.05378.004

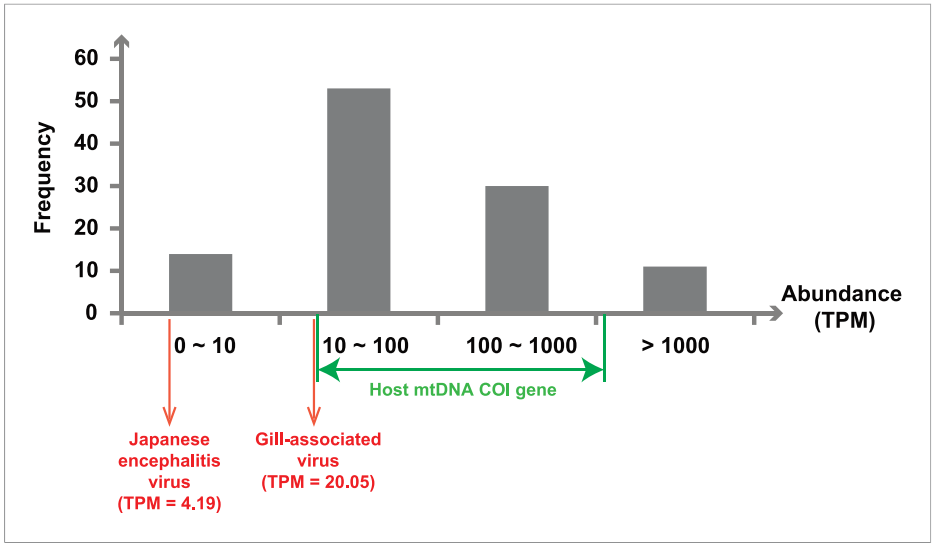

Figure 2. Abundance level (transcripts per million—TPM) of the RdRp genes from the negative-sense RNA viruses detected in this study. Abundance is calculated after the removal of ribosomal RNA reads. As a comparison, we show the abundance of the two well characterized (positive-sense) RNA viruses: Japanese encephalitis virus and Gill-associated virus found in the Mosquito-Hubei and Shrimp libraries, respectively, as well as the range of abundance of host mitochondrial COI genes in these same multi-host libraries.

DOI: 10.7554/eLife.05378.008

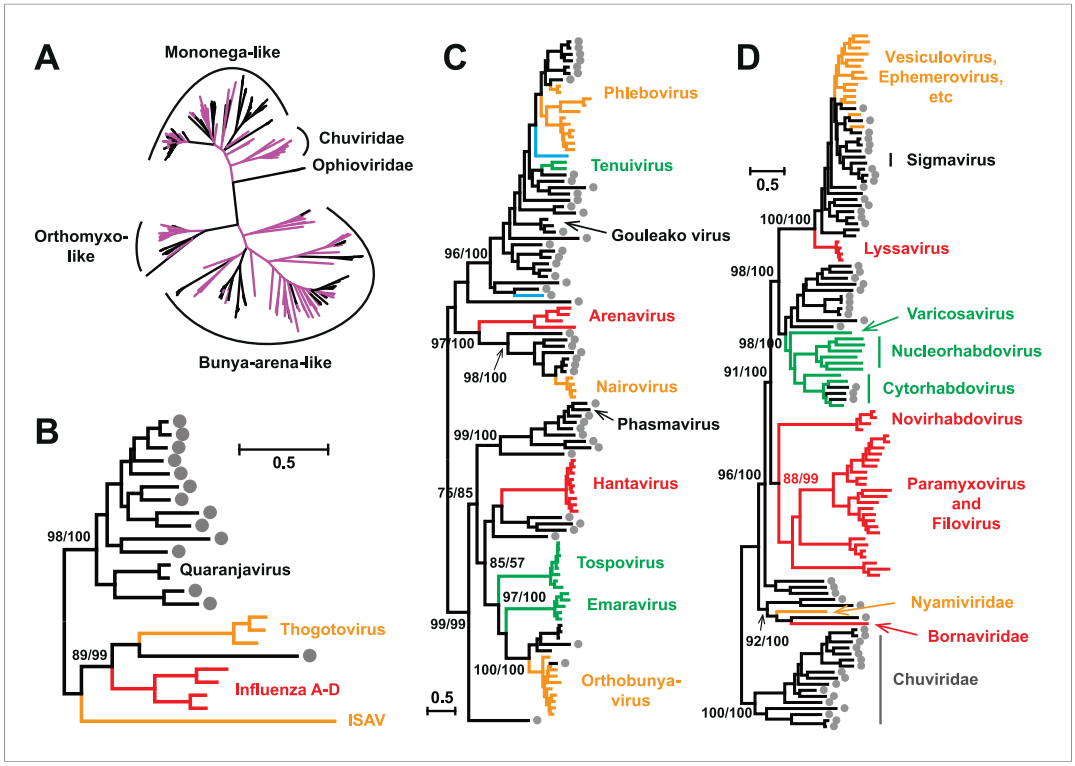

Figure 3. Evolutionary history of negative-sense RNA viruses based on RdRp. This is initially displayed in an unrooted maximum likelihood (ML) tree including all major groups of negative-sense RNA viruses (**A**). Separate and more detailed ML phylogenies are then shown for the Orthomyxoviridae-like (**B**), Bunya-Arenaviridae-like (**C**), and Mononegavirales-like viruses (**D**). In all the phylogenies, the RdRp sequences described here from arthropods are either shaded purple or marked with solid gray circles. The names of previously defined genera/families are labeled to the right of the phylogenies. Based on their host types, the branches are shaded red (vertebrate-specific), yellow (vertebrate and arthropod), green (plant and arthropod), blue (non-arthropod invertebrates), or black (arthropod only). For clarity, statistical supports (i.e., approximate likelihood-ratio test (aLRT) with Shimodaira–Hasegawa-like procedure/posterior probabilities) are shown for key internal nodes only.

DOI: 10.7554/eLife.05378.009

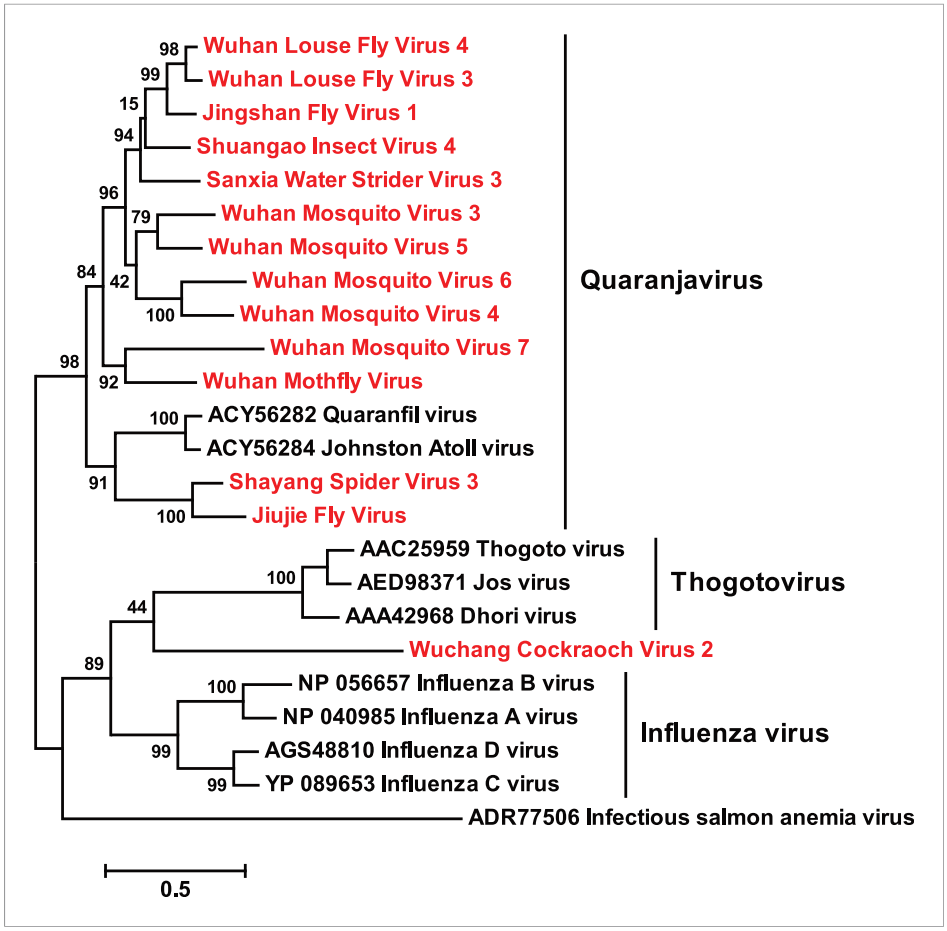

Figure 3—figure supplement 1. A fully labeled ML phylogeny for Orthomyxoviridae-like viruses. The phylogeny is reconstructed using RdRp alignments. Statistical support from the approximate likelihood-ratio test (aLRT) is shown on each node of the tree. The names of the viruses discovered in this study are shown in red. The names of reference sequences, which contain both the GenBank accession number and the virus species name, are shown in black. The names of previously defined genera/families are shown to the right of the phylogenies.

DOI: 10.7554/eLife.05378.010

Figure 3—figure supplement 2. A fully labeled ML phylogeny for Bunya-Arenaviridae-like viruses. The phylogeny is reconstructed using RdRp alignments. Statistical support from the aLRT is shown on each node of the tree. The names of the viruses discovered in this study are shown in red. The names of reference sequences, which contain both the GenBank accession number and the virus species name, are shown in black. The names of previously defined genera/families are shown to the right of the phylogenies.

DOI: 10.7554/eLife.05378.011

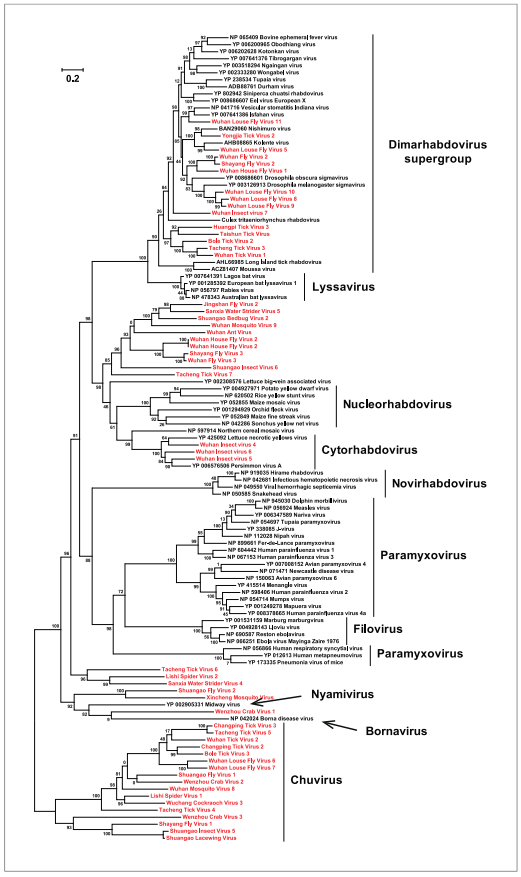

Figure 3—figure supplement 3. A fully labeled ML phylogeny for Mononegavirales-like viruses. The phylogeny is reconstructed using RdRp alignments. Statistical support from the aLRT is shown on each node of the tree. The names of the viruses discovered in this study are shown in red. The names of reference sequences, which contain both the GenBank accession number and the virus species name, are shown in black. The names of previously defined genera/families are shown to the right of the phylogenies.

DOI: 10.7554/eLife.05378.012

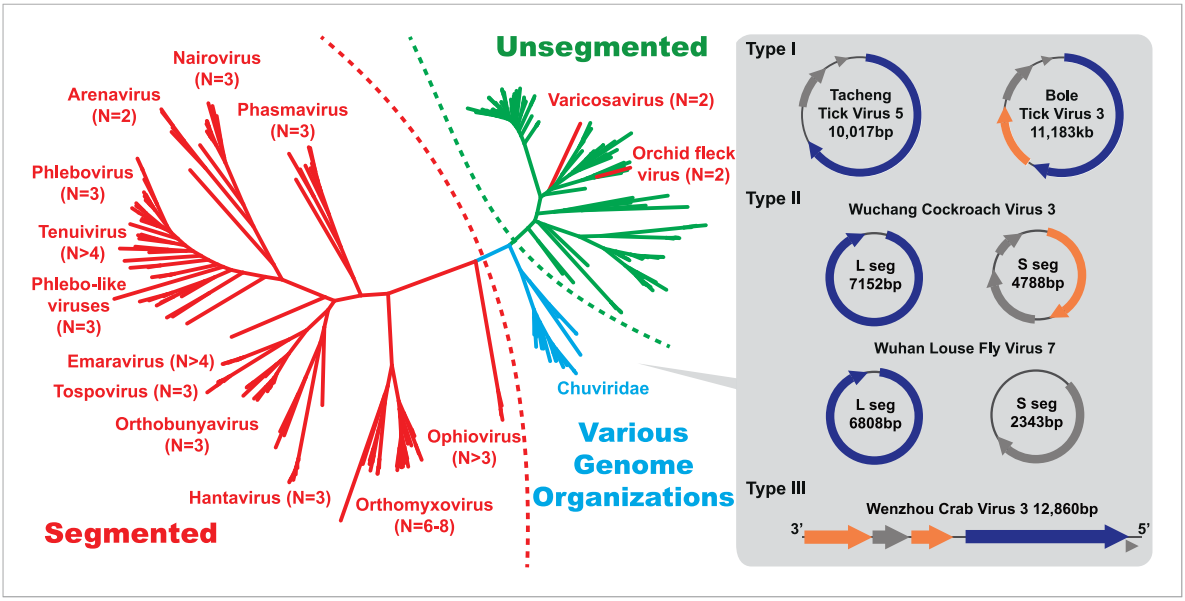

Figure 4. The unrooted ML phylogeny based on RdRp showing the topological position of segmented viruses within the genetic diversity of negative-sense RNA viruses. The segmented viruses are labeled with segment numbers and shaded red. The unsegmented viruses are shaded green. The Chuviridae, which exhibit a wide variety of genome organizations, are shaded cyan. Three major types of putative chuvirus genomes (circular, circular and segmented, and linear) are shown in the right panel and are annotated with predicted ORFs: putative RdRp genes are shaded blue, putative glycoprotein genes are shaded orange, and the remaining ORFs are shaded gray.

DOI: 10.7554/eLife.05378.013

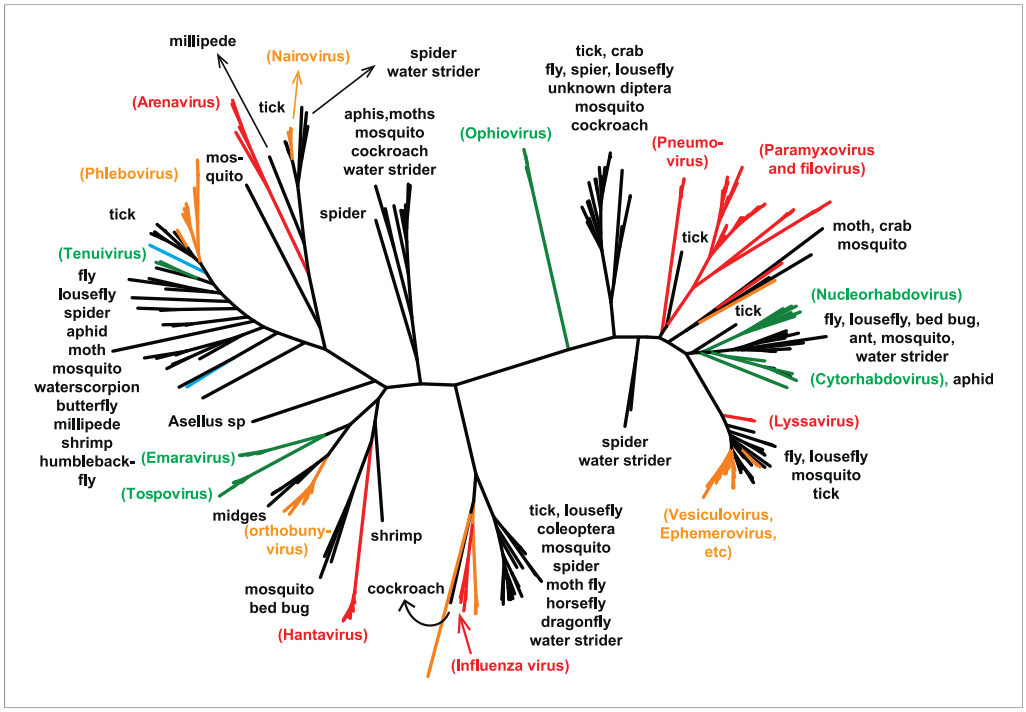

Figure 5. The unrooted ML phylogeny of negative-sense RNA viruses (RdRp) with the common names of the principle arthropod hosts analyzed in this study indicated. Vertebrate-specific viruses are shaded red, those infecting both vertebrates and arthropods (or with unknown vectors) are shaded yellow, those infecting both plants and arthropods are shaded green, those infecting non-arthropod invertebrates are shaded blue, and the remainder (arthropod only) are shaded black.

DOI: 10.7554/eLife.05378.014

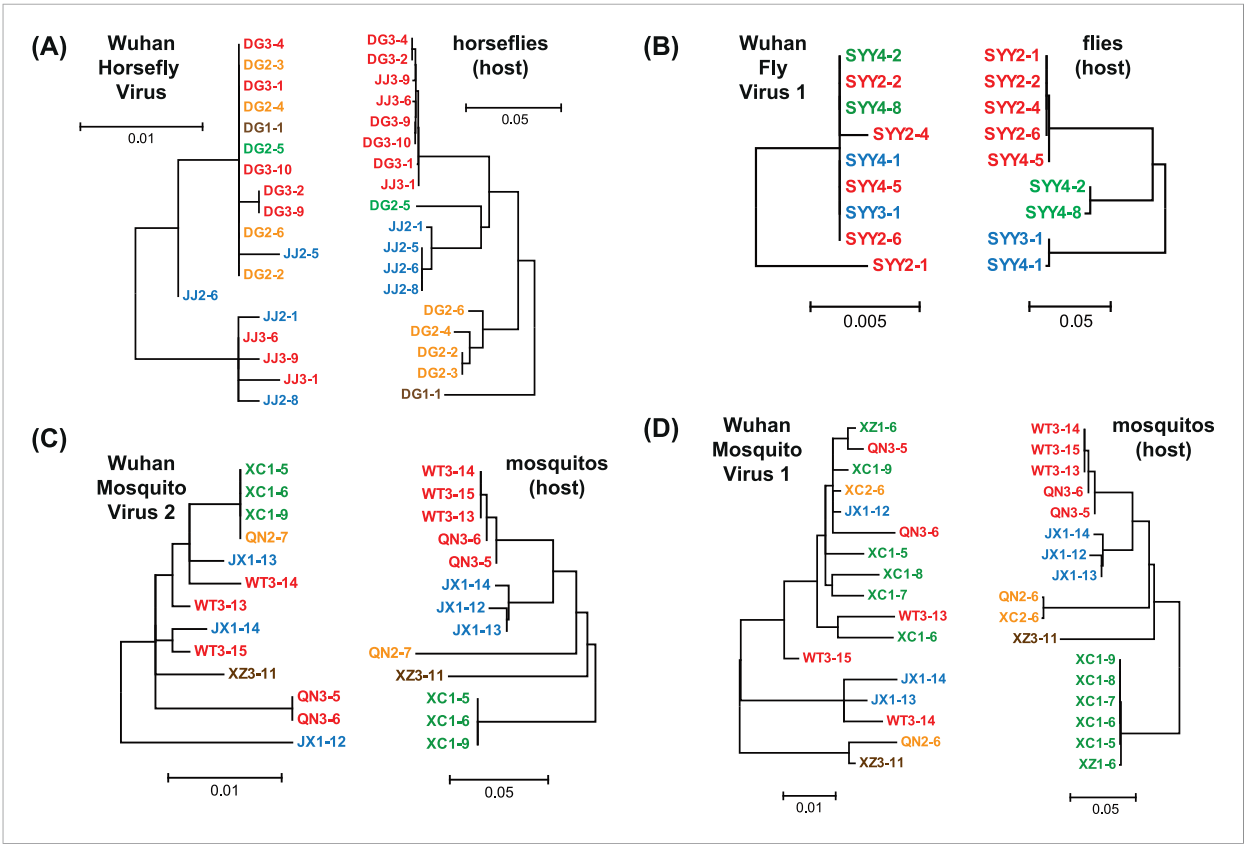

Figure 6. Phylogenetic congruence between viruses (M segments) and hosts, including (**A**) Wuhan Horsefly Virus, (**B**) Wuhan Fly Virus 1, (**C**) Wuhan Mosquito Virus 2, and (**D**) Wuhan Mosquito Virus 1. Different host species/genera are distinguished with different colors, which are then mapped onto virus phylogeny to assess the phylogenetic congruence. ML phylogenetic trees were inferred in all cases.

DOI: 10.7554/eLife.05378.015

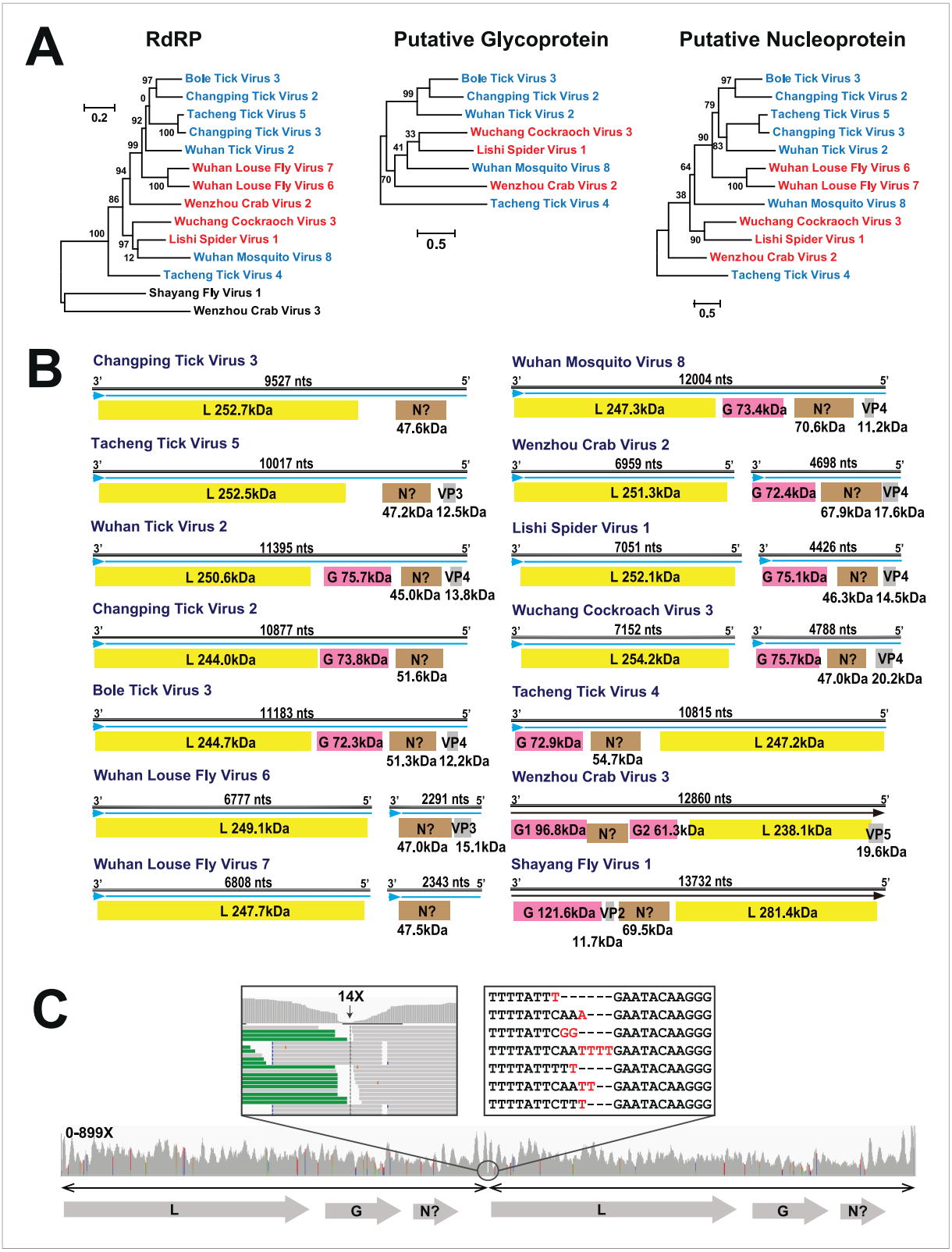

Figure 7. The differing genome organizations in the Chuviridae. (**A**) ML trees of three main putative proteins conserved among the chuviruses. Viruses with circular genomes (Type I) are shaded blue, while those with segmented genomes (Type II) are shaded red. (**B**) Structures of all complete chuvirus genomes. Circular genomes are indicated with the arrow (blue) situated at the 3′ end, and the genome is drawn in a linear form for ease of comparison only, being broken at the region of variable sequence (refer to the 'Materials and methods'). (**C**) An example showing mapping of sequencing reads to the

Figure 7. continued on next page

Figure 7. Continued

circular chuvirus genome. The template for mapping contains two genomes connected head-to-tail. The two boxes magnify the genomic region containing abundant sequence variation.

DOI: 10.7554/eLife.05378.016

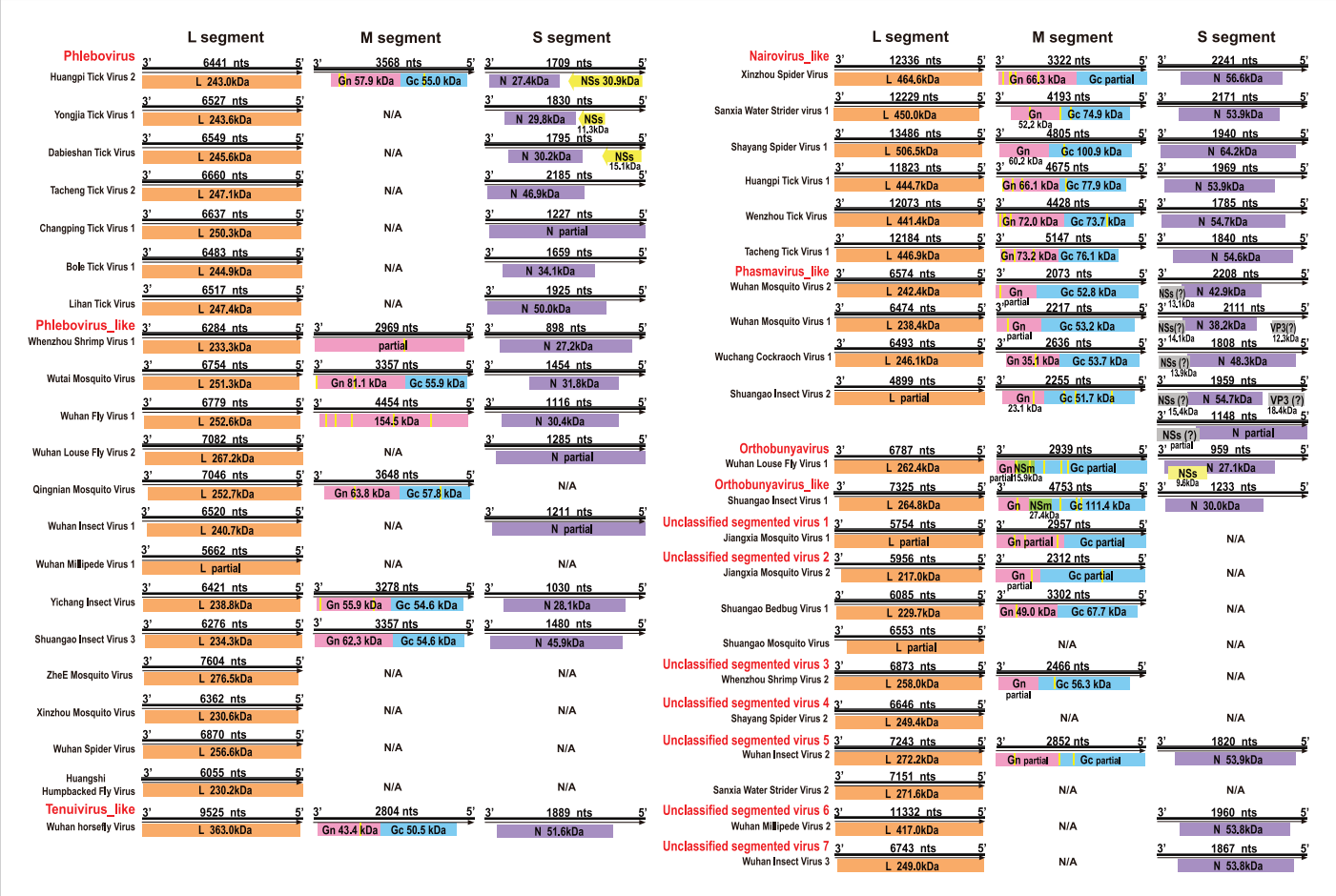

Figure 8. Genome structures of segmented negative-sense RNA viruses. Predicted viral proteins homologous to known viral proteins are shown and colored according to their putative functions. The numbers below each ORF box give the predicted molecular mass.

DOI: 10.7554/eLife.05378.017

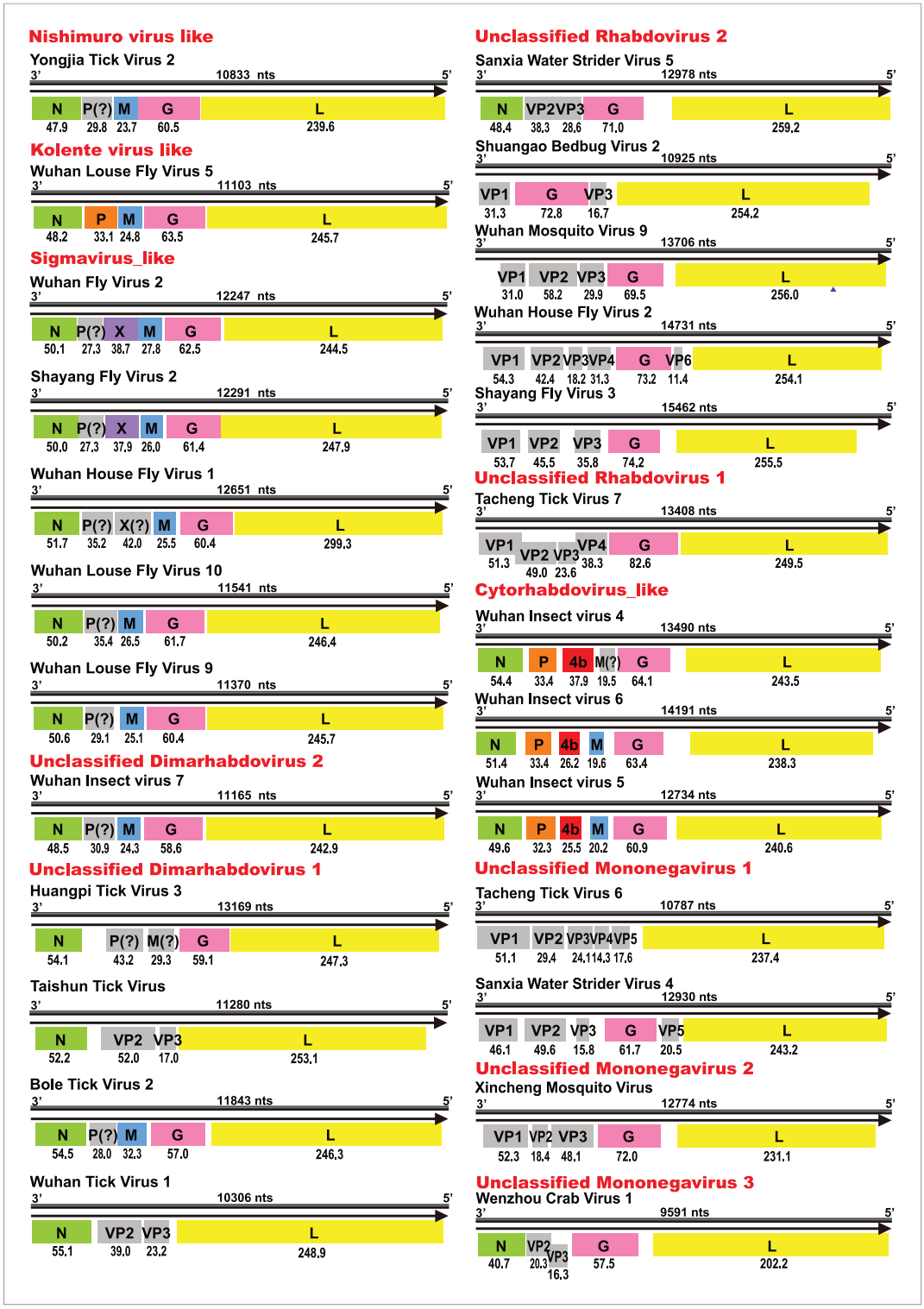

Figure 9. Genome structures of unsegmented negative-sense RNA viruses. Predicted ORFs encoding viral proteins with >10 kDa molecular mass are shown and colored according to their putative functions. The numbers below each ORF box give the predicted molecular mass.

DOI: 10.7554/eLife.05378.018

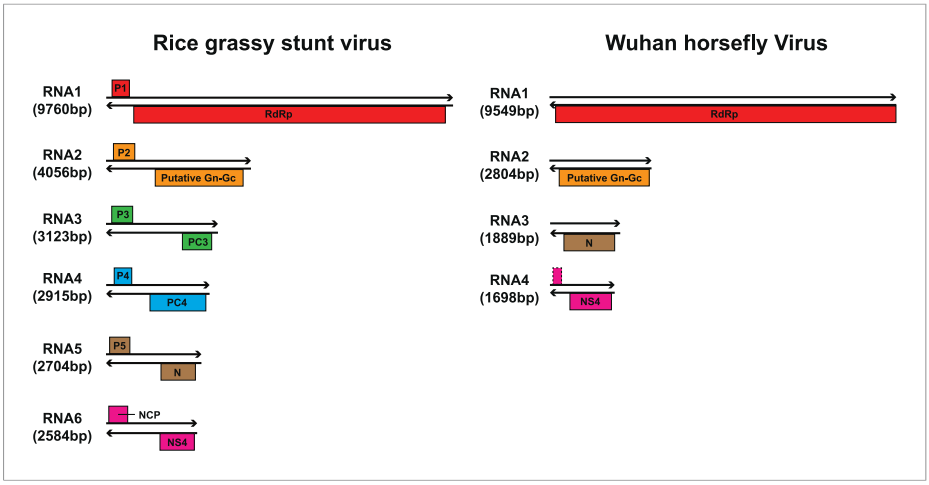

Figure 10. Comparison of the genome structure of a potential tenui-like virus from horsefly with a prototype tenuivirus (Rice grassy stunt virus) genome.

DOI: 10.7554/eLife.05378.019

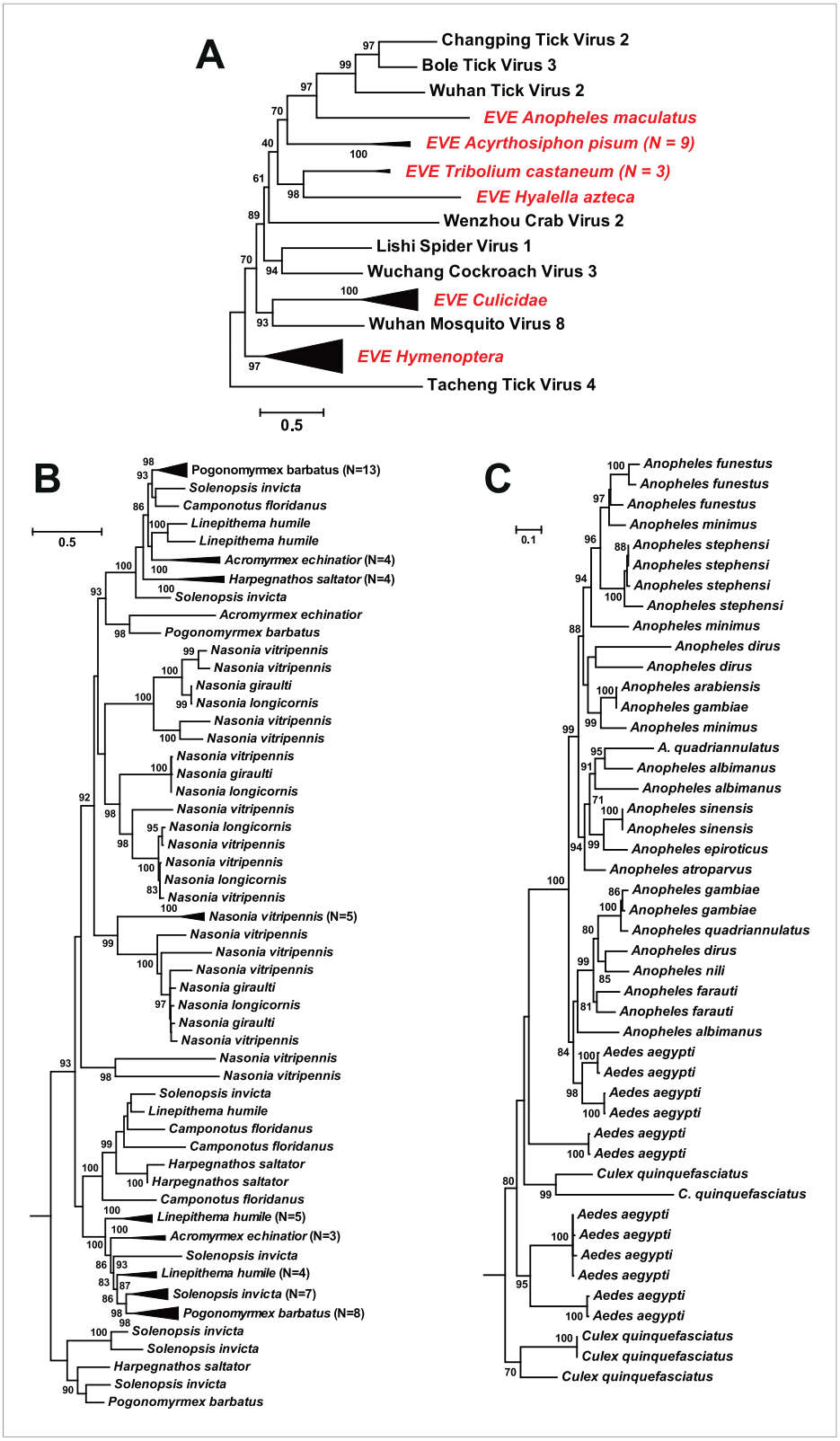

Figure 11. ML phylogeny of EVEs based on the glycoprotein of chuviruses in the context of exogenous members of this family (**A**), with subtrees magnified for (**B**) the Culicidae clade and (**C**) the Hymenoptera clade. The EVEs used in the phylogeny covered the complete or near complete length of the glycoprotein gene and are shown in red and labeled according to host taxonomy in the overall tree. For clarity, monophyletic groups are collapsed based on the

Figure 11. continued on next page

Figure 11. Continued

host taxonomy. Only bootstrap values >70% are shown.

DOI: 10.7554/eLife.05378.021

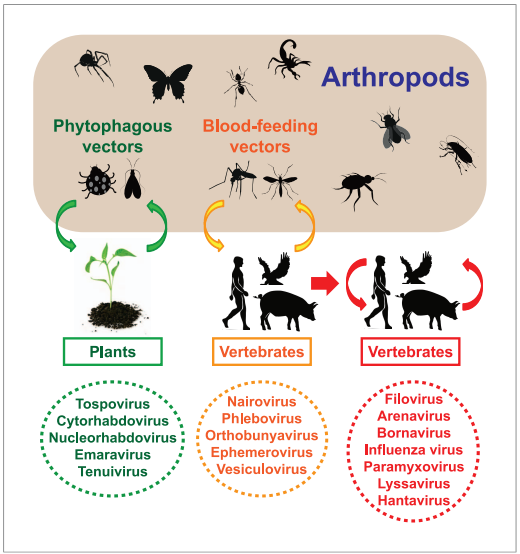

Figure 12. Transmission of negative-sense RNA viruses in arthropods and non-arthropods. Three types of transmission cycle are shown: (i) those between arthropods and plants are shaded green; (ii) those between arthropods and vertebrates are shaded yellow; and (iii) those that are vertebrate-only are shaded red. Viruses associated with each transmission type are also indicated.

DOI: 10.7554/eLife.05378.022