Discovery of novel non-retroviral endogenous viral elements reveals their long-term integration history in spiders

Hengrui Hu[1], Just M. Vlak[2], Zhihong Hu[1]*, Manli Wang[1]*

1. State Key Laboratory of Virology and Biosafety, Wuhan Institute of Virology, Chinese Academy of Science, Wuhan, China.

2. Laboratory of Virology, Wageningen University, Binnenhaven 11, 6709 PD, Wageningen, The Netherlands.

*Corresponding authors

Zhihong Hu, Dr., Prof.

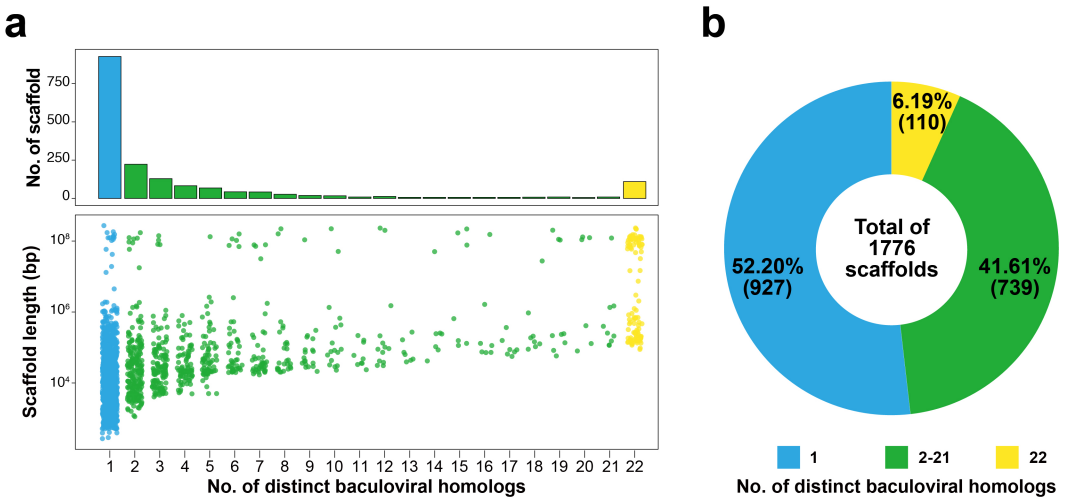
Mailing address: Wuhan Institute of Virology, Chinese Academy of Sciences, Wuhan 430071, P.R. China. Email: huzh@wh.iov.cn. Tel/Fax: +86-27-87998916

Manli Wang, Dr. Prof.

Mailing address: Wuhan Institute of Virology, Chinese Academy of Sciences, Wuhan 430071, P.R. China. Email: wangml@wh.iov.cn Tel/Fax: +86-27-87998086

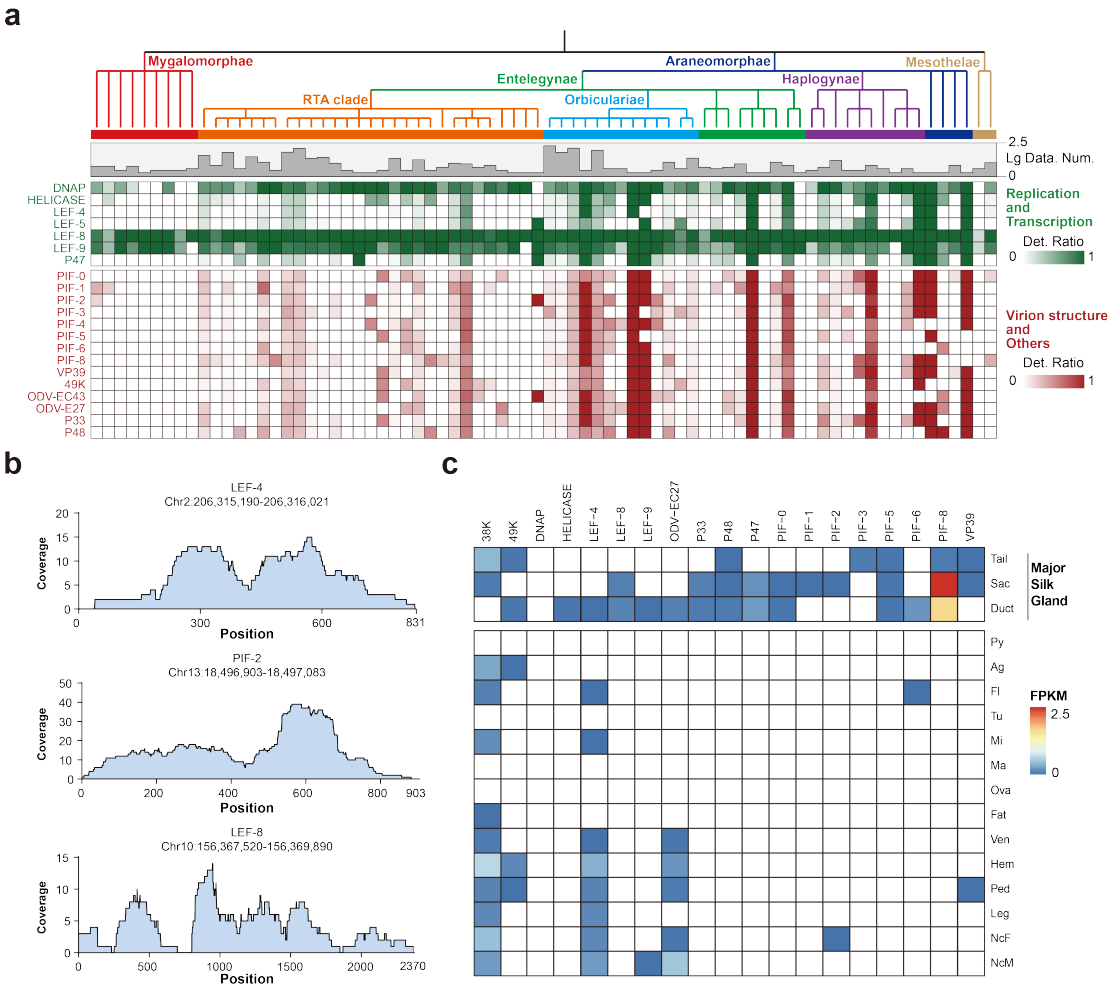
Supplementary Figures

Supplementary Fig. 1 Summary of the distribution of homologs of baculoviral core proteins in spider scaffolds



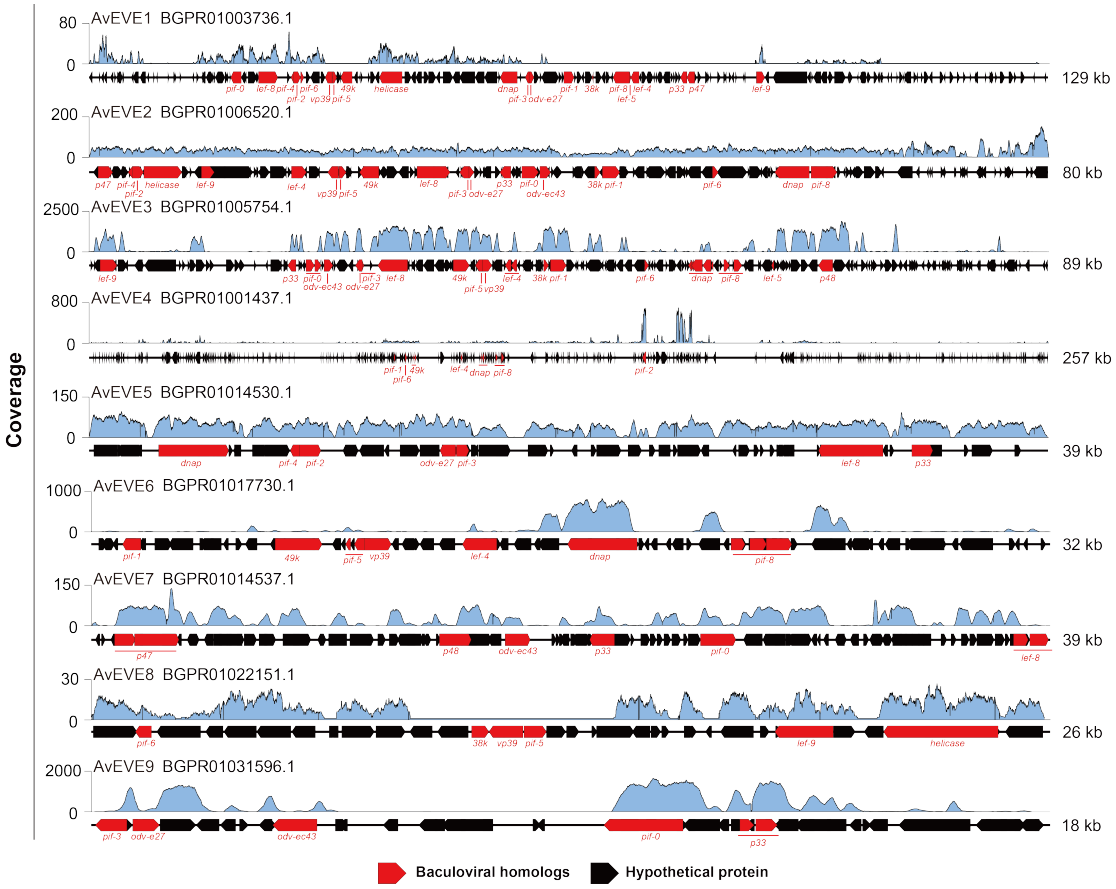
a. The bottom dot plot illustrates the correlation between scaffold length and the number of distinct homologs of baculoviral core proteins found within the WGS databases for Araneae species and SpiderDB. The top bar plot depicts the total count of scaffolds categorized by the number of distinct baculoviral homologs they contain. b. The pie chart provides a visual breakdown, representing the proportion of scaffolds that harbor 1, 2-21, and all 22 distinct homologs of baculoviral core proteins, respectively.

Supplementary Fig. 2 Baculoviral homologs found in spider TSA database and *T. clavata* transcriptomic data



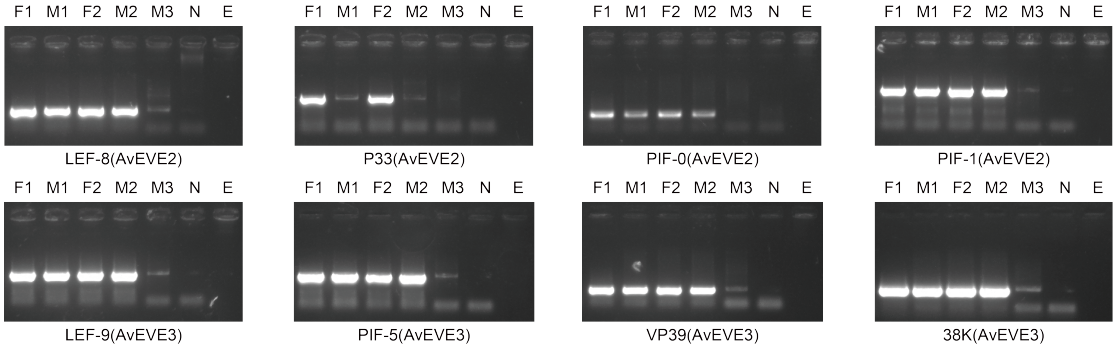
a. A heatmap of the baculoviral homologs (BHs) detection ratio in different spider families, the corresponding common tree, and the dataset number for each family are plotted above the heatmap. b. mRNA read coverage of three representative BHs (LEF-4, PIF-2, and LEF-8) in *T. clavata* transcriptome sequence data. c. Heatmap of BHs' FPKM values from different tissues in *T. clavata* transcriptome sequence data. Py: Pyriform gland; Ag: Aggregate gland; Fl: Flagelliform gland; Tu: Tubuliform gland; Mi: Minor ampullate gland; Ma: Major ampullate gland; Ova: Ovary Fat: Fat body; Ven: Venom gland; Hem: Hemolymph; Ped: Pedipalp; NcF: Female whole body; NcM: Male whole body.

Supplementary Fig. 3 Schematic diagrams of top nine scaffolds containing highest number of distinct baculovirus-related genes in *A. ventricosus* (AvEVE1-9)



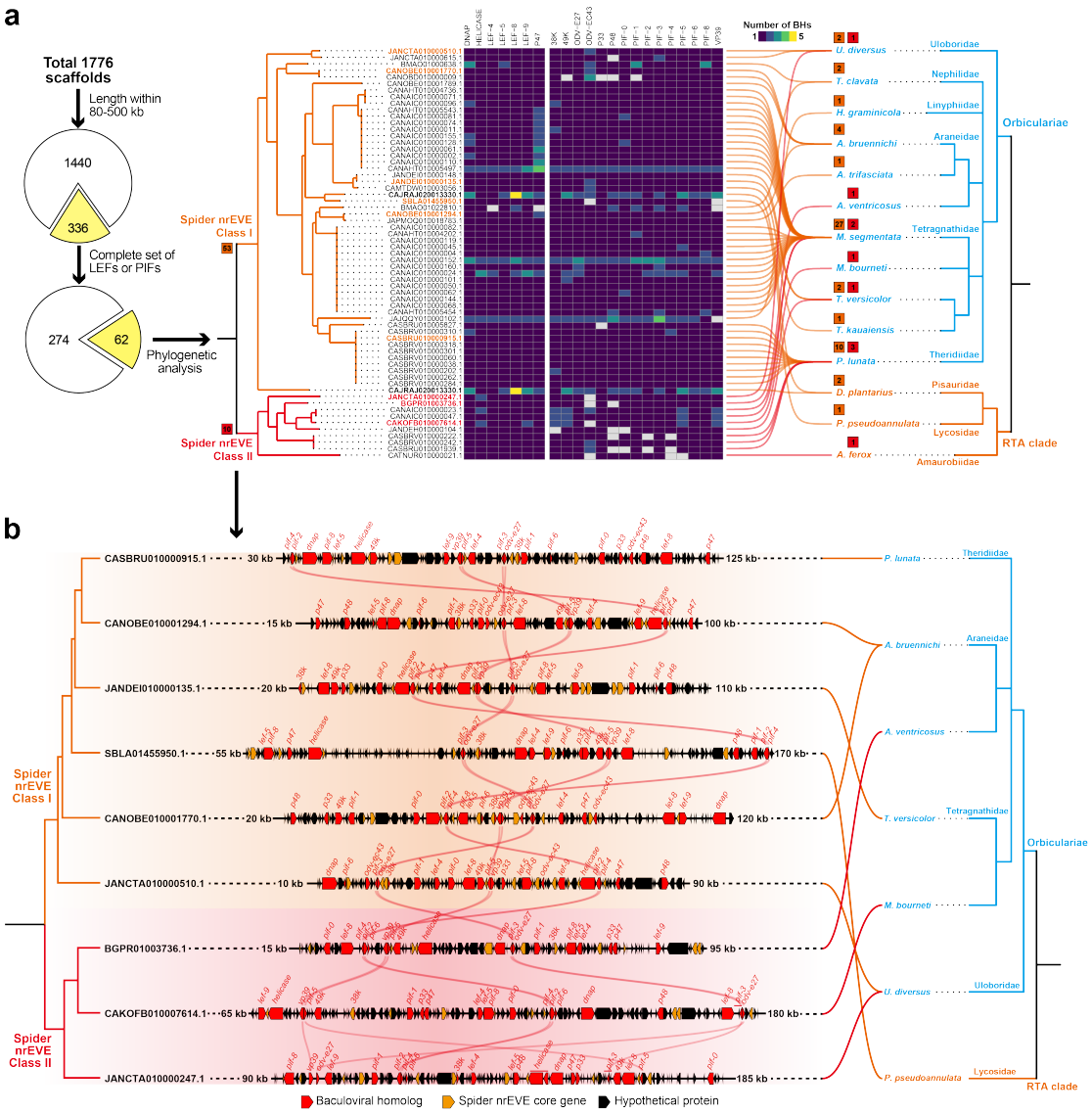
The reads of NGS data from an adult female *A. ventricosus* were aligned to NCBI WGS scaffolds (BGPR01) containing spider nrEVEs *via* bowtie2. The reads coverage was extracted *via* samtools and presented above the schematic diagrams of AvEVE1-9. The scaffold schematic diagrams were generated *via* R package 'gggenome'. The genes with red lines indicated pseudogenes.

Supplementary Fig. 4 PCR results demonstrate the presence of spider nrEVEs in different individuals of *A. ventricosus*



PCR results of baculoviral homologs in *A. ventricosus* samples from two female (F1 and F2) and three males (M1, M2, and M3). Lane N represents the negative control, and Lane E represents the blank control. The experiments were repeated for twice with similar results.

Supplementary Fig. 5 Schematic diagrams of nine spider nrEVEs with nearly complete viral genome



a. The black arrows represent the workflow for screening nine spider nrEVEs with nearly complete viral genomes, while the pie charts illustrate the specific number of scaffolds during processing. The heatmap depicts the distribution of baculoviral homologs across different scaffolds. The number of scaffolds within the tree branch was presented in the box, with orange and red color represent Class I and II. The scaffolds selected for further analysis were marked in orange and red. b. The diagram shows the nrEVE regions of nine representative scaffolds containing nearly complete putative NALDV genome. Three pair of conserve arranged ORFs (*pif-2* and *pif-4*, *vp39* and *pif-5*, *pif-3* and *odv-e27*) from different scaffolds were connected with lines to

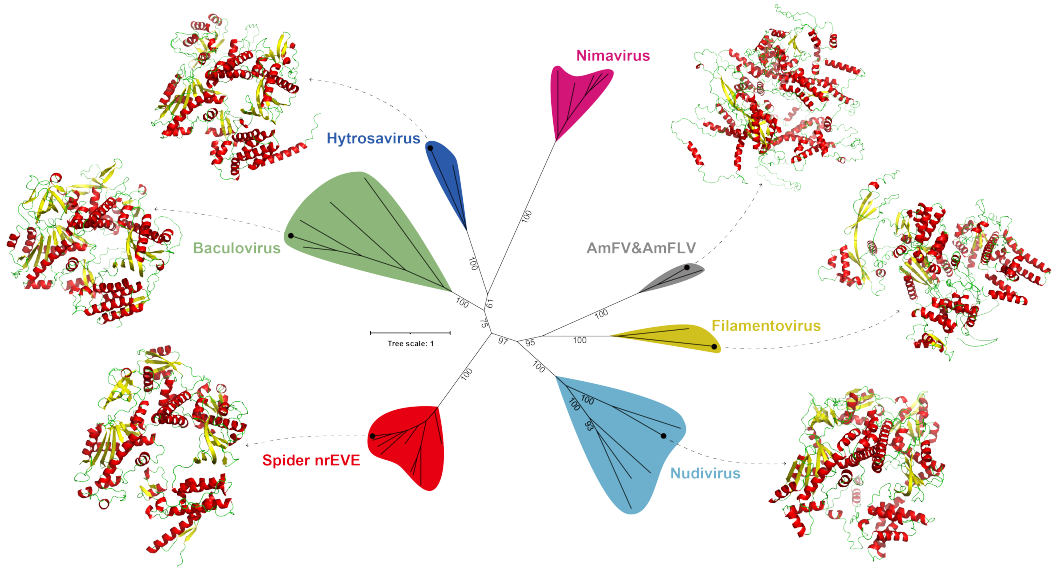
display as examples of rearrangements of genome organization. The hypothetical ORFs were plotted *via* R package ‘gggenome’. The trees of scaffolds were generated using LEF-8 sequences and the nrEVE classes were indicated, with orange and red branches representing Classes I and II. The trees of spider on the right are derived from the common tree of NCBI taxonomy. The scaffolds were linked to the corresponding spiders in the species tree and the line color is consistent with the trees of scaffolds.

Supplementary Fig. 6 Alignments and phylogenetic trees of different proteins from representative NALDVs and spider nrEVEs

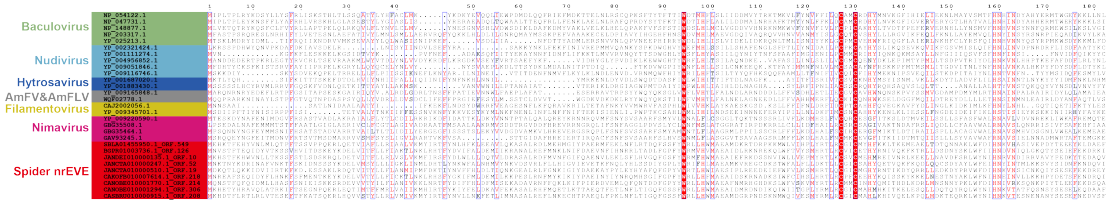
The sequence alignments and phylogenetic trees of ten proteins (DNAP, P33, PIF-0, PIF-1, PIF-2, PIF-3, PIF-5, LEF-4, LEF-8, and LEF-9) from representative NALDVs and spider nrEVEs. The sequences were firstly aligned *via* mafft and then trimmed *via* trimal with '-gt 0.8'. The phylogenetic trees were generated *via* IQ-TREE2 with using ModelFinder Plus to determine the best-fit tree model.

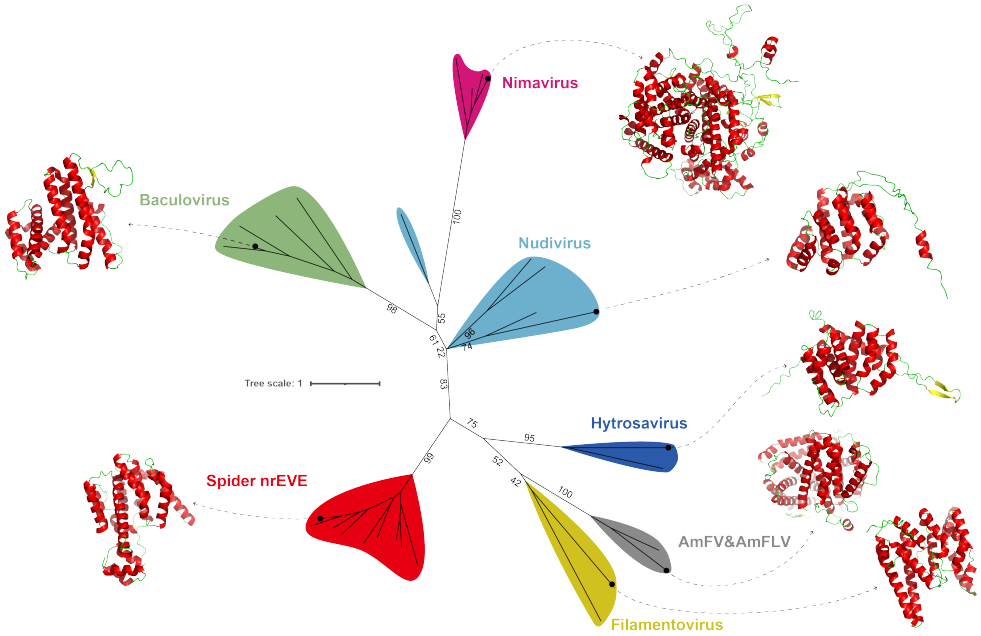
DNAP



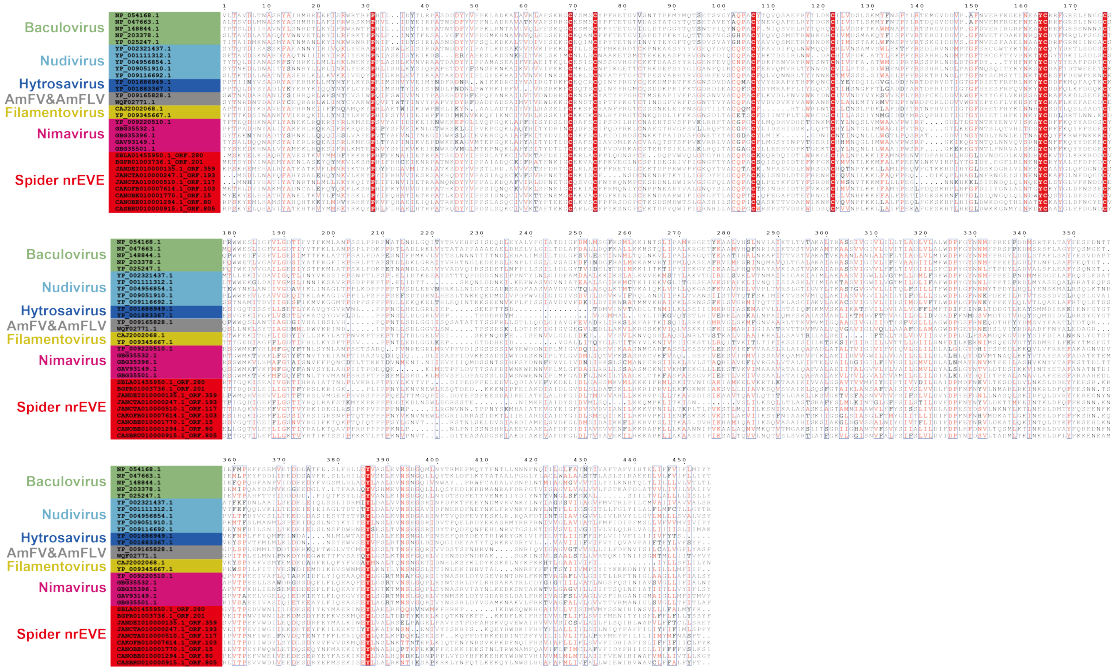


P33


Baculovirus
Nudivirus
Hytrosavirus
AmFV&AmFLV
Filamentovirus
Nimavirus
Spider nrEVE

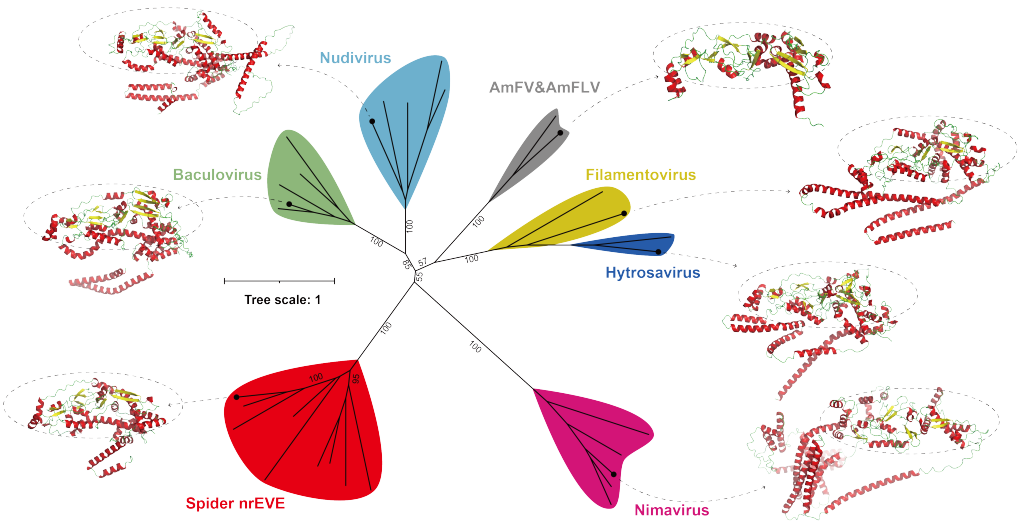

Nimavirus
Baculovirus
Nudivirus
Tree scale: 1
Hytrosavirus
Spider nrEVE
AmFV&AmFLV
Filamentovirus

PIF-0

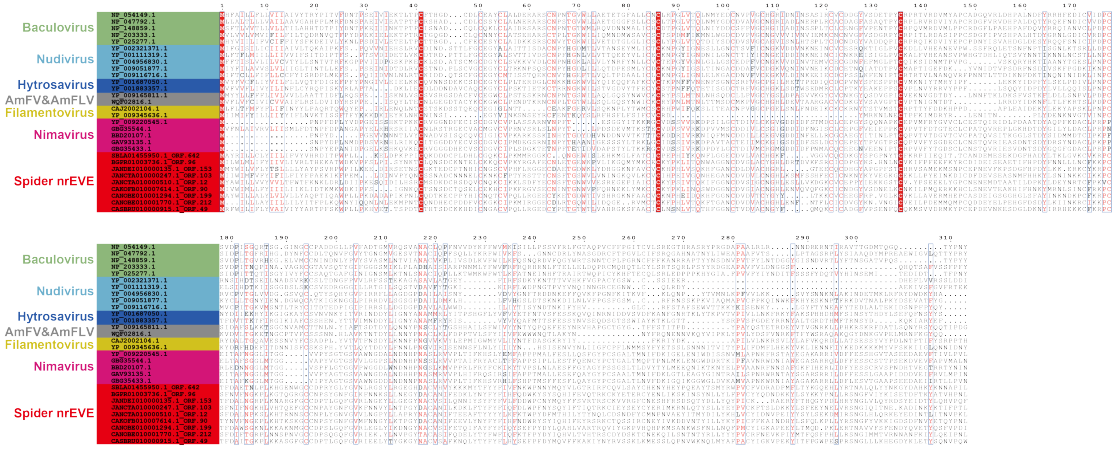

Baculovirus
Nudivirus
Hytrosavirus
AmFV&AmFLV
Filamentovirus
Nimavirus
Spider nrEVE

Baculovirus
Nudivirus
Hytrosavirus
AmFV&AmFLV
Filamentovirus
Nimavirus
Spider nrEVE

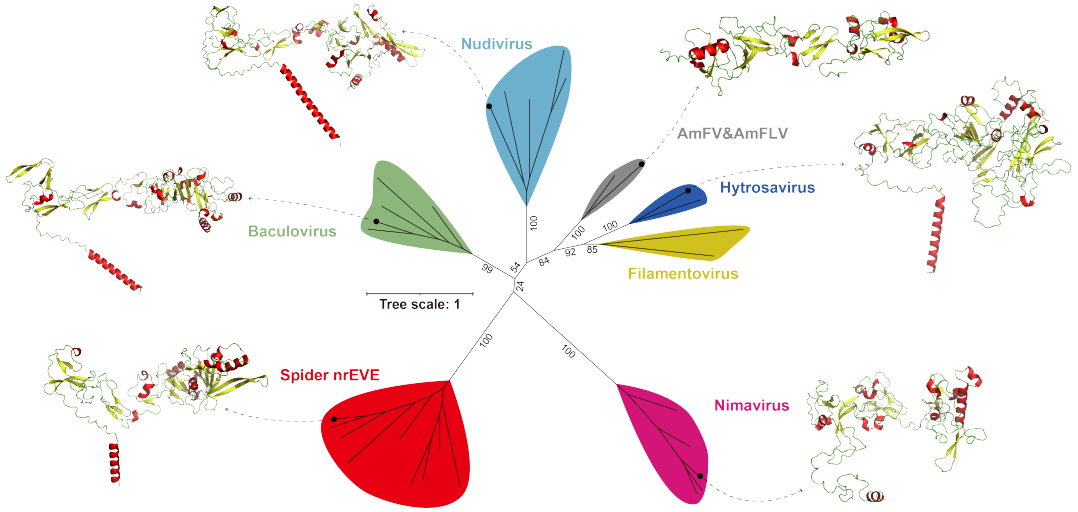
Baculovirus
Nudivirus
Hytrosavirus
AmFV&AmFLV
Filamentovirus
Nimavirus
Spider nrEVE


Nudivirus
AmFV&AmFLV
Baculovirus
Filamentovirus
Hytrosavirus
Tree scale: 1
Spider nrEVE
Nimavirus

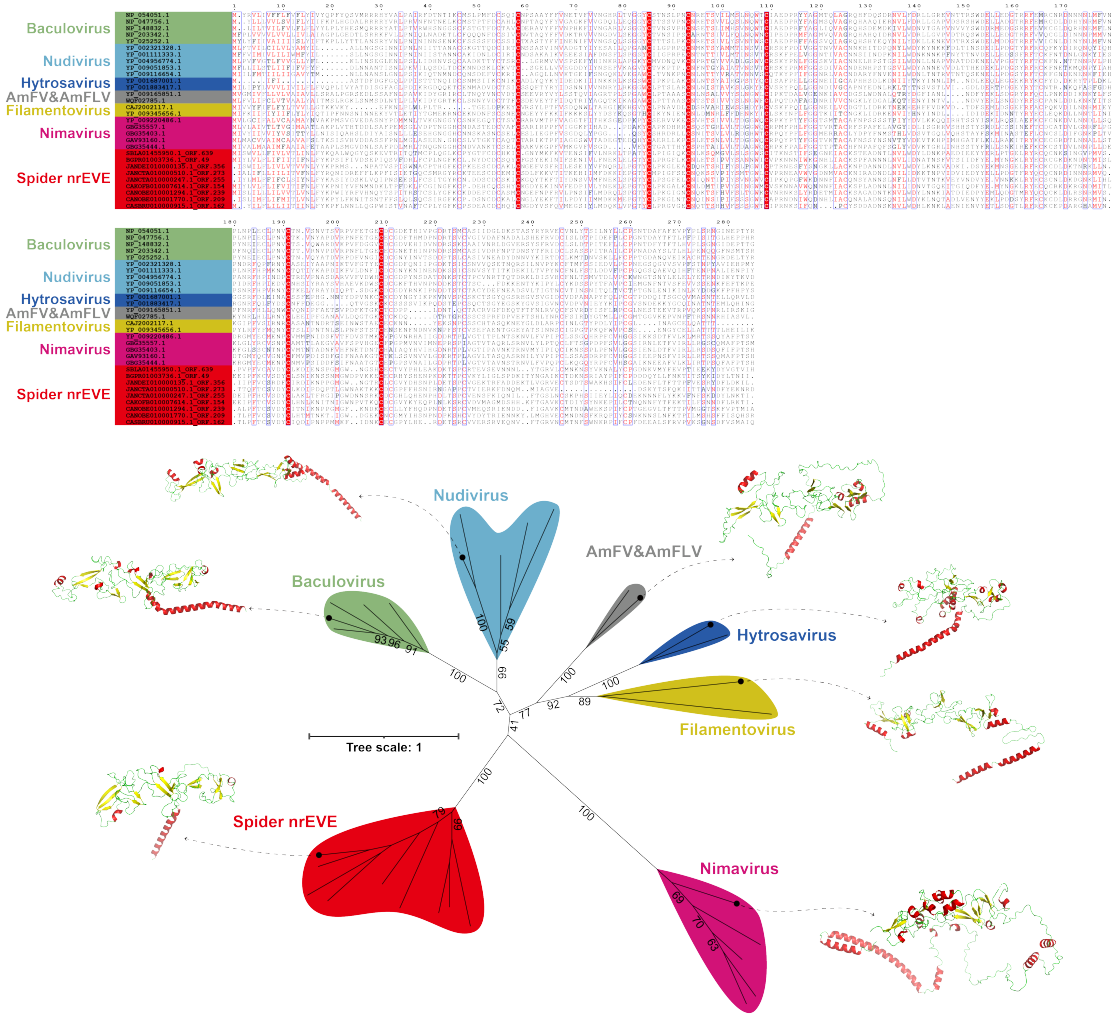
PIF-1


Baculovirus
Nudivirus
Hytrosavirus
AmFV&AmFLV
Filamentovirus
Nimavirus
Spider nrEVE

Baculovirus
Nudivirus
Hytrosavirus
AmFV&AmFLV
Filamentovirus
Nimavirus
Spider nrEVE

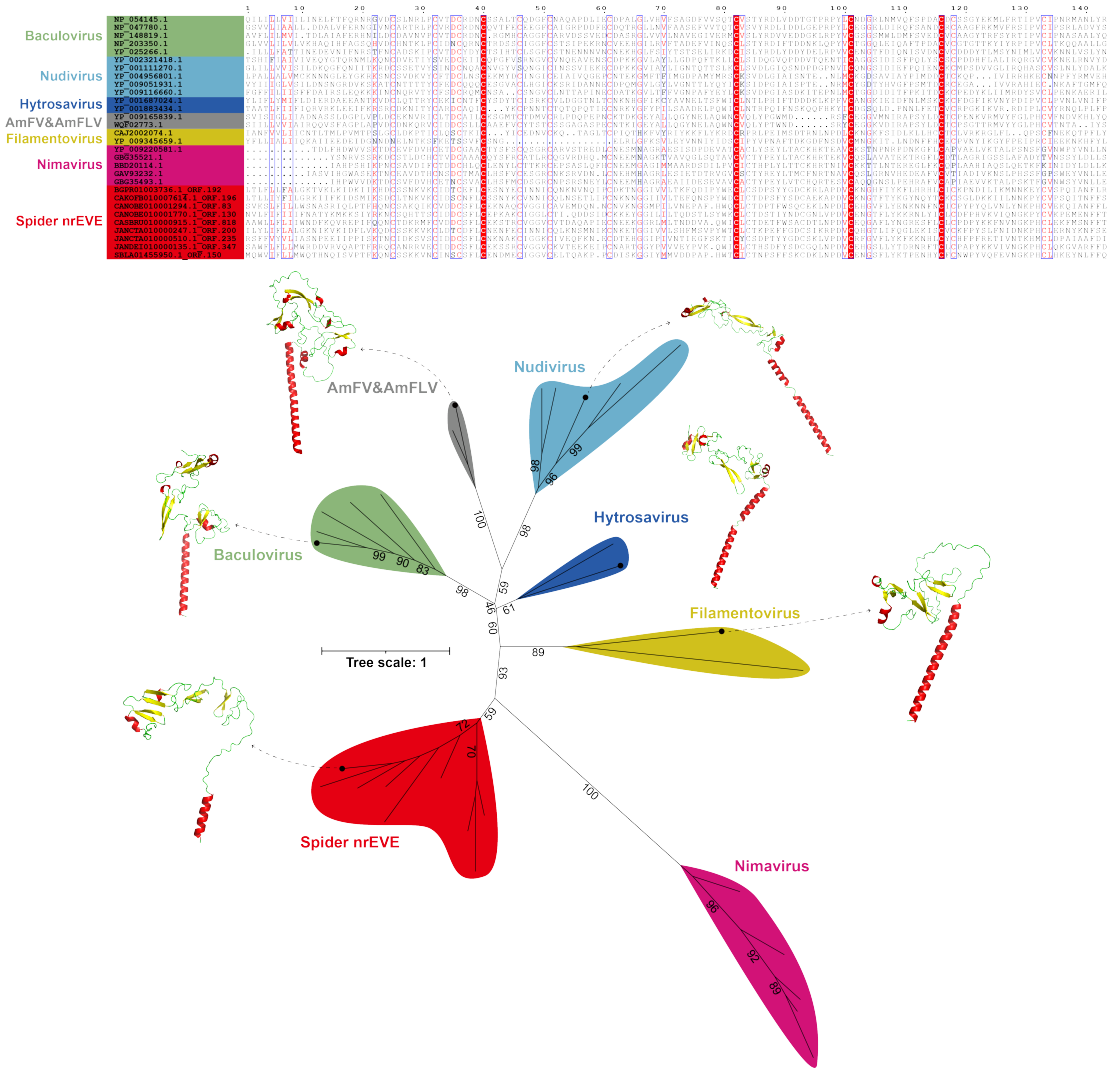

Nudivirus
AmFV&AmFLV
Hytrosavirus
Baculovirus
Filamentovirus
Tree scale: 1
Spider nrEVE
Nimavirus

PIF-2

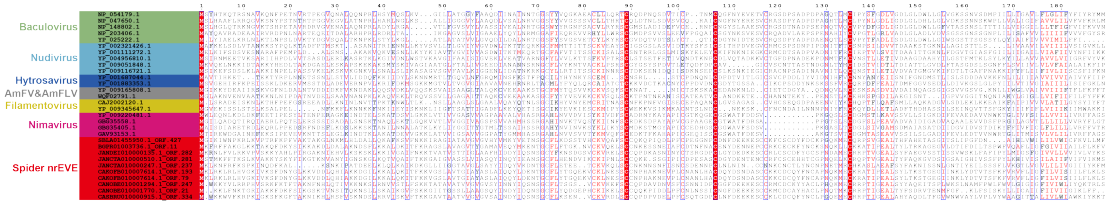

Baculovirus
Nudivirus
Hytrosavirus
AmFV&AmFLV
Filamentovirus
Nimavirus
Spider nrEVE

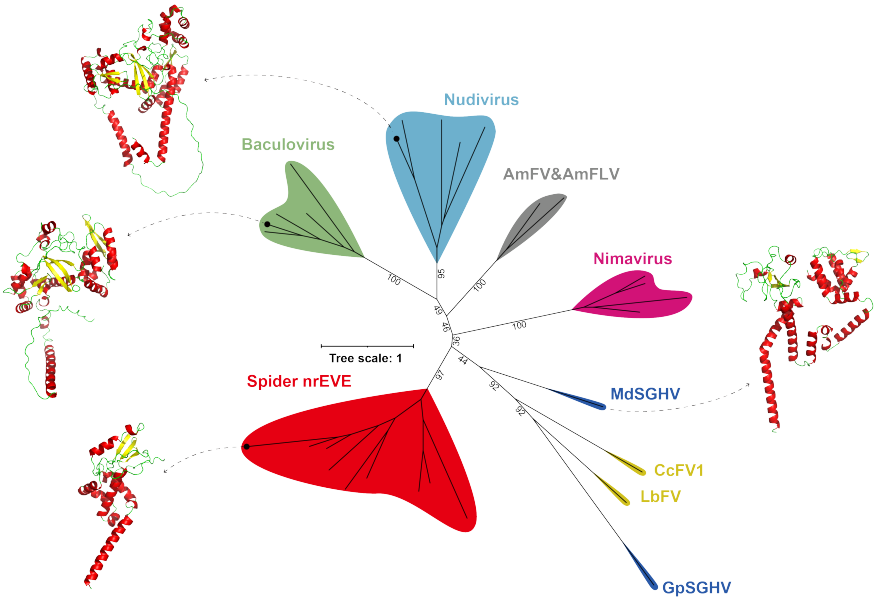
Nudivirus
AmFV&AmFLV
Baculovirus
Hytrosavirus
Filamentovirus
Tree scale: 1
Spider nrEVE
Nimavirus

PIF-4

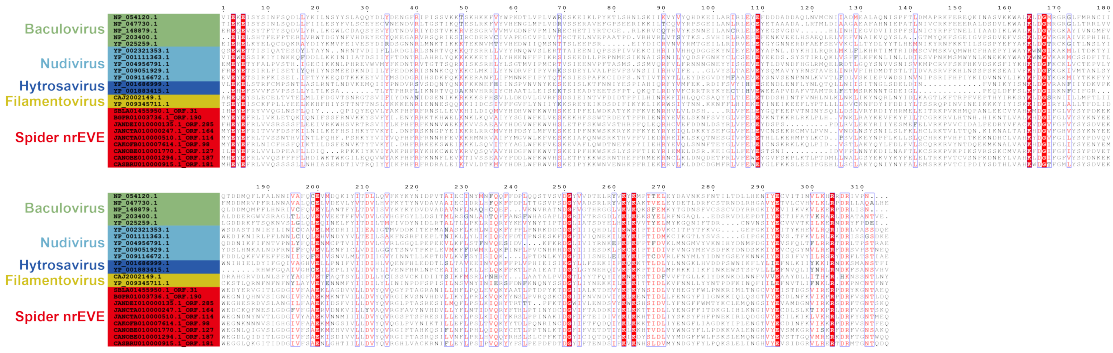

Baculovirus
Nudivirus
Hytrosavirus
AmFV&AmFLV
Filamentovirus
Nimavirus
Spider nrEVE
AmFV&AmFLV
Nudivirus
Hytrosavirus
Baculovirus
Filamentovirus
Spider nrEVE
Nimavirus
Tree scale: 1

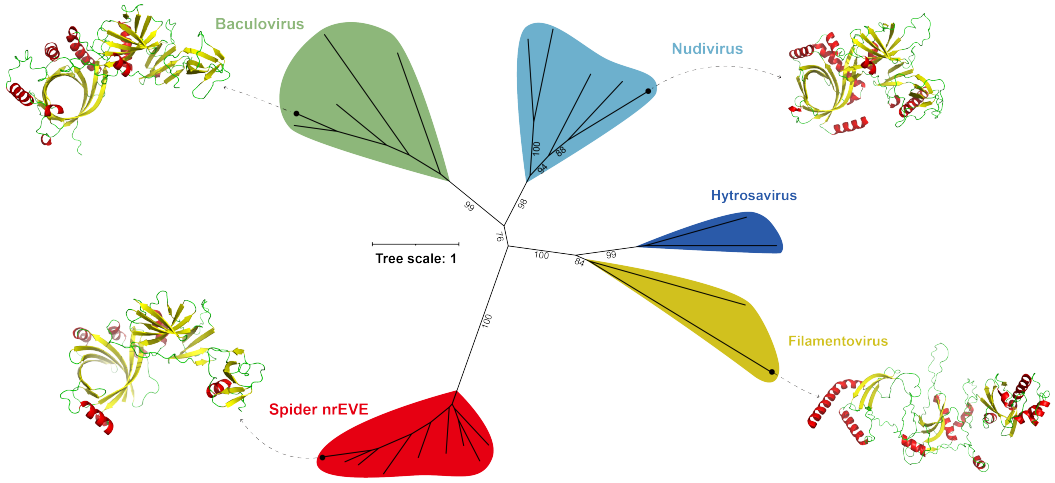
PIF-5


Baculovirus
Nudivirus
Hytrosavirus
AmFV&AmFLV
Filamentovirus
Nimavirus
Spider nrEVE

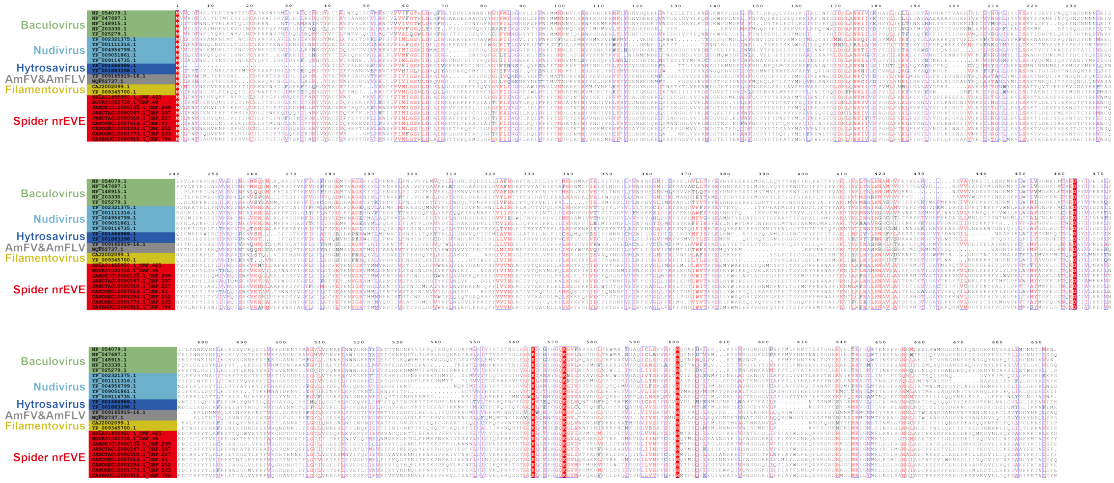

Nudivirus
Baculovirus
AmFV&AmFLV
Nimavirus
Tree scale: 1
Spider nrEVE
MdSGHV
CcFV1
LbFV
GpSGHV

LEF-4


Baculovirus
Nudivirus
Hytrosavirus
Filamentovirus
Spider nrEVE

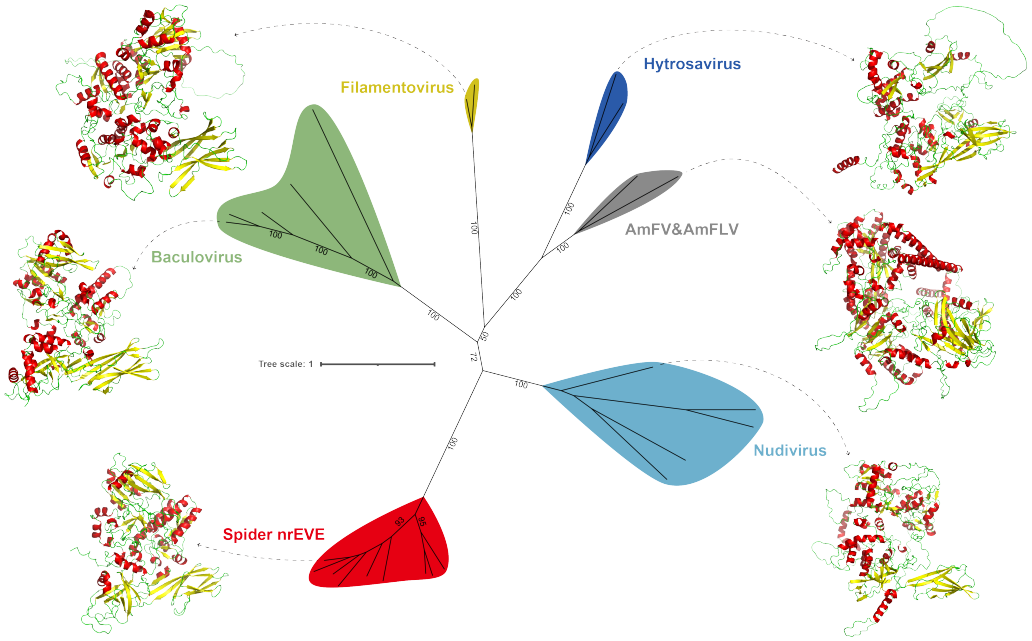

Baculovirus
Nudivirus
Hytrosavirus
Tree scale: 1
Filamentovirus
Spider nrEVE

LEF-8

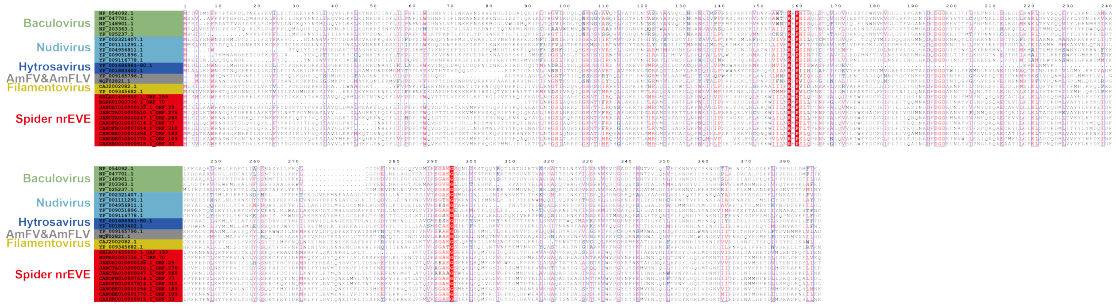

Baculovirus
Nudivirus
Hytrosavirus
AmFV&AmFLV
Filamentovirus
Spider nrEVE

Baculovirus
Nudivirus
Hytrosavirus
AmFV&AmFLV
Filamentovirus
Spider nrEVE

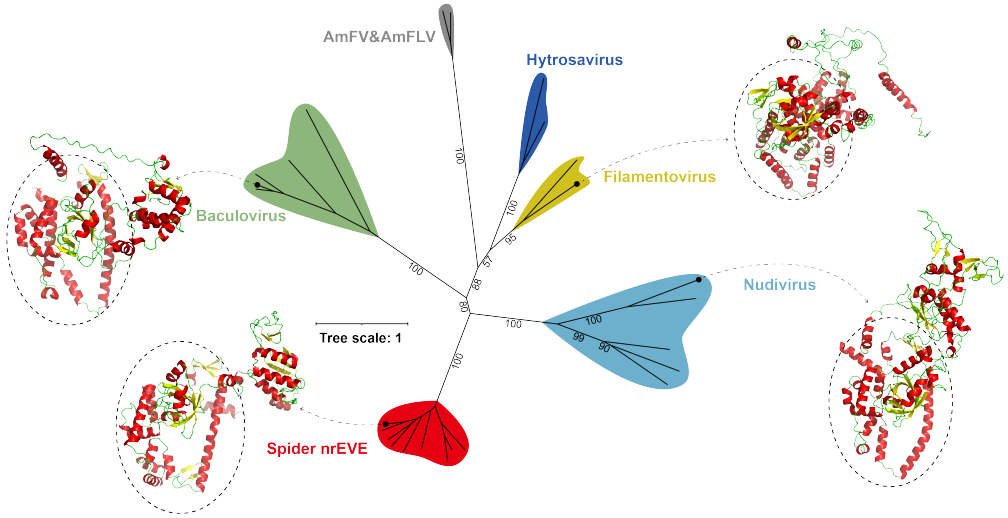
Baculovirus
Nudivirus
Hytrosavirus
AmFV&AmFLV
Filamentovirus
Spider nrEVE


Filamentovirus
Hytrosavirus
AmFV&AmFLV
Baculovirus
Tree scale: 1
Nudivirus
Spider nrEVE

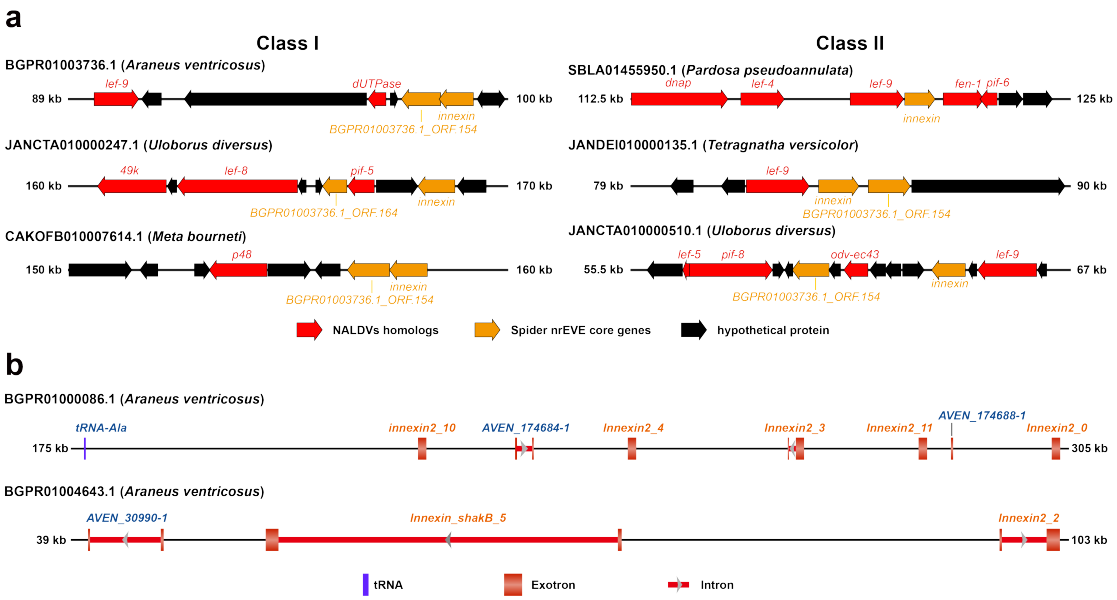
LEF-9


Baculovirus
Nudivirus
Hytrosavirus
AmFV&AmFLV
Filamentovirus
Spider nrEVE

Baculovirus
Nudivirus
Hytrosavirus
AmFV&AmFLV
Filamentovirus
Spider nrEVE

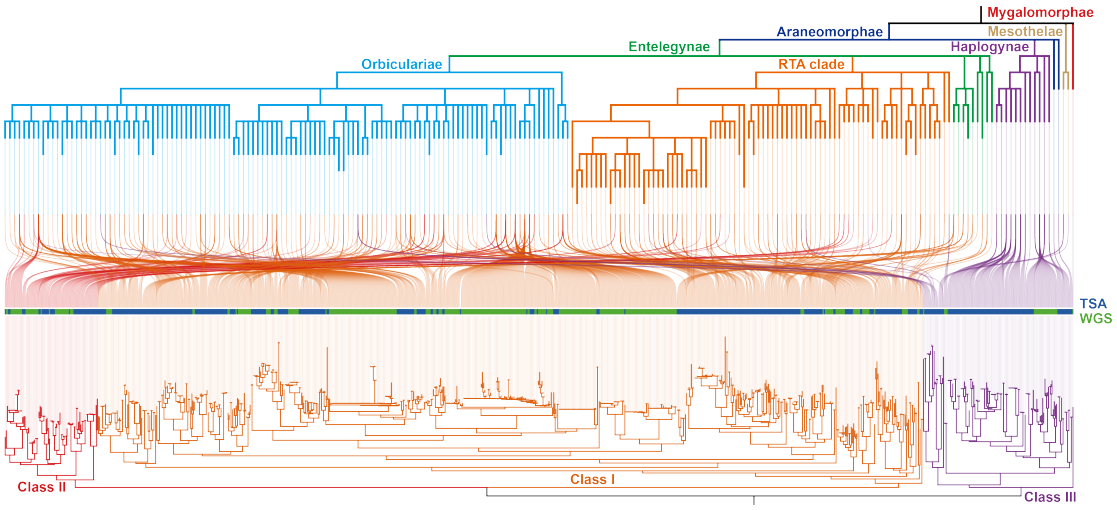

AmFV&AmFLV
Hytrosavirus
Filamentovirus
Baculovirus
Nudivirus
Spider nrEVE
Tree scale: 1

Supplementary Fig. 7 Schematic diagrams of spider scaffolds harboring virus-derived innexins or spider-innexins.



a. Scaffolds containing virus-derived innexin. The spider nrEVEs belonging to Class I were shown on the left panel, and those of Class II were shown on the right panel. b. Two scaffolds containing spider-innexins.

Supplementary Fig. 8 Co-phylogenetic tree of spider nrEVEs and their spider hosts



The phylogenetic tree of spider nrEVEs was a maximum-likelihood phylogenetic tree built using PIF-0 sequences from the WGS and TSA databases. The phylogenetic tree of the spiders was based on the common tree from the NCBI taxonomy. The co-phylogenetic analysis was conducted by R package ‘phytools’ (v2.1-1).