Data

Data S1. (separate file)

Motifs identified in the RdRp domains of novel RNA virus phyla inferred in this work. The conserved motifs G [GxS], F [long motif enriched in basic amino acids], A [DxxxxD], B [(S/T)Gxxx(T/G)xxxN], C [(S/G)DD], D[GxxxK], and E [FL] were manually inspected in multiple sequence alignments of the domain protein sequences, and their identity was supported by HHPred search. Sections of the alignments not containing motifs are not shown. Note that motif A of "*Taraviricota*" is DxxxxE instead of the canonical DxxxxD. Motif G was not found in most of the cases.

Data S2. (separate file)

Motifs identified in the RdRp domains of the novel RNA virus classes inferred in this work. Note the motif permutation (C-A-B, instead of A-B-C) in the class 42 (phylum *Kitrinoviricota*), and the unusual motif C [IDD] in the classes 66 and 67 (negarnaviricot subphylum *Haploviricotina*). Asterisks indicate domain motifs with less confidence. Legend follows **Data S1**.

Data S3. (separate file)

Cluster-specific Maximum-likelihood phylogenetic trees. The phylogenetic trees were based on the amino-acid sequences of the RdRp domain. Tips are labelled with GenBank accession number, virus name and associated virus family if previously established/assigned. Viruses identified in this study are highlighted in red, whereas previously known sequences are highlighted in black or grey. The tree scale represents one substitution per amino acid. The multiple sequence alignments and phylogenetic trees are available online for download (see **Data and materials availability** section).

Data S4. (separate file)

Sequence alignment used for the Global RdRp phylogenetic. Aligned sequences were derived from both individual megataxa and consensus sequences. The colors coding scheme represents amino-acid properties per column in the alignment. The multiple sequence alignment and phylogenetic tree are available online for download (see **Data and materials availability** section).