Integrating structural modeling and divergence dating of RNA-dependent RNA polymerases to resolve the evolutionary history of plant and fungal viruses: from sobemoviruses to sobelivirads

Jérôme Gracy,[1,‡] Mahan Ghafari,[2,§] Gilles Labesse,[1, ‡‡] Denis Fargette,[3,§§] Eugénie Hébrard,[3,*,†]

[1]Centre de Biochimie Structurale, Montpellier, France

[2]Department of Biology, University of Oxford, UK

[3]PHIM Plant Health Institute, Univ Montpellier, IRD, CIRAD, INRAE, Institut Agro, Montpellier, France

[‡]https://orcid.org/0000-0002-9778-0327

[§]https://orcid.org/0000-0002-4123-3287

[‡‡]https://orcid.org/0000-0002-6861-3300

[§§]https://orcid.org/0000-0003-2078-8141

[†]https://orcid.org/0000-0002-2128-1302

[*]Corresponding author: E-mail: eugenie.hebrard@ird.fr

Abstract

RNA-dependent RNA polymerases (RdRps) are crucial for RNA virus replication and serve as key marker genes for defining deep taxonomic ranks and for understanding viral evolutionary history. Despite their conserved functions and motifs, the high genetic diversity of RdRps complicates precise sequence comparisons across viral families, hindering accurate taxonomic classification of new species. Three-dimensional (3D) RdRp structures can help overcome these challenges through structure-guided alignments. However, such data are rare for myco- and phytoviruses, limiting investigation of their ecological and evolutionary links. In this study, we focused on the highly divergent order *Sobelivirales*, which includes sobemoviruses infecting plants — a taxon known for its ancient origin — and barnaviruses infecting fungi, which remain relatively unknown. Using deep-learning structural modeling, we generated reliable 3D models for 44 sobemoviral and sobelivirad species. Structure-guided alignments, together with new barnaviral and relevant outgroup sequences, enabled robust phylogenetic reconstruction, allowing us to propose revisions of existing viral families and suggest new evolutionary scenarios. Divergence dates were estimated for the first time at this taxonomical rank using the Prisoner of War model, which revealed a divergence of plant and fungal sobelivirads 27.0 ± 10 million years ago — much more recent than the separation of their respective hosts. This result suggests that cross-kingdom host shifts have contributed more likely to the evolutionary history of *Sobelivirales* than strict virus-host codivergence. Based on an extended dataset of 127 species, structure and sequence conservation analyses identified molecular signatures of sobeliviral families. These conserved and extented RdRp motifs will facilitate future taxonomic assignments and the development of diagnostic tools. Our interdisciplinary approach, integrating structure modeling and divergence dating, offers new insights into the evolutionary divergence of plant and fungal viruses, with potential applications to other viral orders and families.

Keywords : polymerase, evolutionary history, divergence date, protein structure, modeling, taxonomy, sobemovirus, polerovirus, barnavirus, sobelivirales.

Author summary

RNA-dependent RNA polymerases (RdRps) are essential for RNA virus replication and serve as important markers for classifying viruses and understanding their evolution. However, with metagenomic studies that rapidly expand the known diversity of RNA viruses, it is increasingly difficult to compare highly divergent RdRps and accurately classify new species. When available, 3D structures of RdRps can help overcome this challenge through structure-guided alignments.

Here, we focused on the highly divergent order *Sobelivirales*, which groups phytoviruses and mycoviruses. Deep-learning models were used to generate reliable 3D structures for 44 representative viral species. Structure-guided alignments combined with the identification of new barnaviruses and outgroups allowed us to build a more accurate viral phylogeny. Based on this finding, we proposed updates to existing viral families and genera within the order *Sobelivirales*. We estimated divergence dates using a model that previously uncovered the ancient origins of sobemovirus. Notably, we provided the first estimate of when these plant and fungal viruses diverged ~27.0 ± 10 million years ago, suggesting cross-kingdom host shifts rather than strict virus-host codivergence. We also identified molecular signatures that are useful for future virus classification and diagnosis, with potential applications to other viral groups.

62 **Introduction**

63 Recent advances in viral metagenomics have revealed how extensive and unexplored
64 viral diversity is, and have contributed significantly to a better understanding of viral evolution.
65 Major adaptations of viral taxonomic classification have been made to incorporate numerous
66 new lineages and their evolutionary links [1,2]. This includes the creation of novel upper ranks
67 and subdivisions such as classes and phyla [3]. The identification of viral hallmark genes
68 (VHG) facilitated the assignment of viral taxa into these novel groups [4,5]. Polymerases
69 especially RNA-dependent RNA polymerases (RdRp) represent one of the major groups of
70 VHG due to their crucial function for realm *Riboviria* (RNA viruses) [6]. RdRp comparative
71 analyses revealed unexpected relationships between some viral groups, prompting taxonomic
72 reassessments and modifications. For example, the family *Luteoviridae* had initially grouped
73 the genera *Luteovirus* and *Polerovirus*, primarly due to their similar circulative transmission
74 mode mediated by aphids and the related readthrough domain of their capsid protein (CP).
75 However, based on the low sequence similarity of their RdRps, the family *Luteoviridae* was
76 abolished by the International Committee on Taxonomy of Viruses (ICTV) and the two genera
77 were subsequently reassigned to the families *Tombusviridae* and *Solemoviridae*, respectively
78 [7].

79 Viral RdRps are monomeric proteins responsible for synthesizing complementary RNA
80 strands from a template RNA. They exhibit an overall 3D structure often likened to a cupped
81 human right hand in which the fingers and thumb domains interact to form a characteristic
82 central tunnel (for review [8–11]). The palm domain is formed by a β-sheet with three to six β-
83 strands above two α-helices, the fingers domain is a mixed α/β structure and the thumb domain
84 is highly variable with a predominantly helical structure. The fingers and thumb domains
85 contribute to the RdRp overall stability and to proper positioning of the RNA template and

86 nascent RNA strand. The palm domain is essential for the accurate and efficient RNA chain
87 elongation. It allows (i) the positioning of the primer 3'OH, or the nucleotidylated viral protein
88 genome-linked (VPg) depending on the viral taxon, (ii) the binding of incoming nucleotides
89 and (iii) their incorporation to the RNA chain using two metal ions as cofactors. Comparisons
90 of RdRp sequences of diverse viral groups has identified short linear motifs of highly conserved
91 amino acids [12–14]. Five motifs (A-E) were reported in the palm domain, two motifs (F and
92 G) in the fingers domain and one putative and less conserved motif (H) in the thumb domain.
93 High-throughput taxonomic assignments were often based on the ABC motifs only [15–17].
94 Analyzing the 3D structures of RdRps later identified larger protein segments referred to as
95 homomorphs, determined by structural similarities rather than sequence composition [18].
96 Enlarging the dataset of available 3D structures and releasing sequence conservation constraints
97 allowed the identification of a common structural core and ancient phylogenetic relationships
98 between distants taxa [19]. However, for some viral phyla such as those of plant and fungus
99 viruses, the absence of 3D structures poses problems for comparing RdRps and revealing their
100 evolutionary history. To address this challenge, multiple protein alignments were restricted to
101 the central RdRp palm domain, using classical algorithms combined with manual refinements
102 based on available 3D structures from taxonomically distant related viruses, such as animal or
103 human viruses. These approaches led to the creation of the kingdom *Orthornavirae* and its
104 associated phyla [4,6]. However, these approaches lack the precision needed to analyze in detail
105 phylogenetic relationships between lower taxonomic ranks such as classes, orders and families.

106 Although the monophyly of viruses in the kingdom *Orthornavirae* has been established
107 based on comparison of their RdRps, reconstruction of their evolutionary history remains
108 difficult due to their strong, probably ancient, evolutionary divergence and the central role of
109 horizontal transfer of other viral genes in plant viruses [20]. In the picorna-like supergroup, now
110 referred as the phylum *Pisuviricota*, viral species share common structural and functional

111 features in RdRp. On the other hand, they also exhibit contrasting biological characteristics in
112 terms of host range (animals, insects, plants, fungi, protists), as well as variations in genomic
113 molecule (positive single-stranded RNA or double-stranded RNA) and size (ranging from the
114 largest to the smallest RNA genomes) (S1 Table). Phylogenetic analyses of viral taxa, such as
115 the phylum *Pisuviricota* or its main class *Pisoniviricetes,* can reveal evolutionary relationships
116 that help formulate hypotheses about the origins of shared or distinct traits among these viruses.

117 Viruses of fungi also named mycoviruses have long been understudied, but advances of
118 high-throughput sequencing technologies have revealed the large and unexpected diversity of
119 fungal viruses [21,22]. Moreover, recent discoveries, such as cross-kingdom transmission
120 events, highlight ecological and evolutionary links with plant viruses [20,22–24]. In this study,
121 we compared the RdRps inside the order *Sobelivirales*, as markers of their common history to
122 gain insights into host range evolution. This order is one of the six orders of positive-sense
123 single-stranded RNA (+ssRNA) mycoviruses. It comprises families *Solemoviridae,*
124 *Barnaviridae* and *Alvernaviridae*, which contains plant, fungus and micro-algae viruses,
125 respectively. Solemovirids (members of the family *Solemoviridae*) includes four plant viral
126 genera (*Sobemovirus*, *Polerovirus*, *Enamovirus* and *Polemovirus*), while barnavirids and
127 alvernavirids consist of single genera, *Barnavirus* and *Dinornavirus*, each composed of only
128 one approved species infecting fungus or dinoflagellates, respectively (Table 1). Apart from the
129 alvernavirids which are not yet well known, these families share the same characteristic
130 genomic arrangement of their main polyprotein such as chymotrypsine-like protease / VPg /
131 RdRp [25,26], but their sequence conservation is restricted to very short linear motifs.
132 Similarities between RdRp active sites of sobemovirus and polero-/enamovirus were noted as
133 early as 1991 [12] — leading finally to their taxonomic grouping in 2021 [27]. Although RdRp
134 homologies between sobemoviruses and barnaviruses were proposed in 2008 [10], they have
135 not yet resulted in taxonomic revisions. Furthermore, a newly proposed genus of fungal virus,

136 *Hubsclerovirus*, has recently been assigned to the phytoviral family *Solemoviridae* instead of
137 the mycoviral family *Barnaviridae* [28]. Thus, comparing more precisely the RdRps from these
138 viral families would provide valuable insights into the complex history of plant and fungus
139 viruses and would contribute to revise their taxonomy.

140 Our focus on sobelivirads (members of the order *Sobelivirales*) enabled the application
141 of modern methodologies to reconstruct their deep evolutionary histories. Notably, the genus
142 *Sobemovirus* exemplifies a unique case where the impact of time-dependent rate phenomenon
143 (TDRP) on underestimation of divergence times using standard molecular clock methods has
144 been considered using the Prisoner of War model (PoW) [29,30]. The PoW model was
145 developed to account for the time-dependent rate phenomenon in viral evolution [31], whereby
146 substitution rate estimates inferred over short timescales tend to decline when extrapolated over
147 deeper evolutionary timescales due to factors like epistasis, nucleotide biases, and substitution
148 saturation. Under this framework, observed genetic divergence is interpreted as arising from a
149 spectrum of effective rate classes : rapidly evolving sites dominate short-term divergence,
150 whereas over longer timescales these sites become increasingly saturated or purged by
151 selection, and more slowly evolving sites contribute proportionally more to the observed
152 divergence. PoW model therefore provides a distance-to-time transformation that maps genetic
153 distances at internal nodes of a tree to calendar time without relying on fixed-node calibration
154 information. In the case of sobelivirads, this framework is especially relevant because standard
155 time-invariant molecular-clock extrapolations using short-term rates underestimate the very
156 ancient origin of sobemoviruses (over four million years) [30,32].

157 The PoW model application needs three key prerequisites: the detection of a temporal
158 signal in time-stamped sequences, identification of non-recombinant genomic regions, and a
159 robust phylogenetic signal in the data. Here, to overcome the last and more difficult issue at the

160 high taxonomic rank of the order *Sobelivirales*, we leveraged both the structural protein
161 conservation and the advancements in deep-learning algorithms. Indeed, the development of
162 structural phylogenetics, recently accelerated with the efficiency of Alphafold (AF) predictions
163 [33], have demonstrated that 3D data can be used to build more accurate sequence alignments
164 [34] and to reconstruct relationships even between highly divergent proteins in the 'twilight
165 zone' of sequence similarity (for review [35]). This approach has been increasingly applied to
166 large collections of viral metagenomic data in order to infer more precise phylogenetic
167 relationships [17,36–38]. By combining it with the discriminant taxonomical criteria for
168 sobelivirads, i.e. polyprotein arrangement, we were able to discard proteins suspected of
169 resulting from evolutionary convergence, to find new candidate barnaviruses from the NCBI
170 Virus database and to select strong and accurate outgroups to reconstruct the sobelivirad
171 phylogeny.

172 In this paper, we hypothesized that structural modeling would reveal conserved motifs
173 undetectable by sequence alignment alone, and that divergence dates would support ancient
174 horizontal transfers between plant and fungal viruses. We followed an integrative and multi-
175 step approach. Once we had assessed the reliability of the 3D model of RYMV RdRp, we
176 extended our analysis at an intrageneric level revealing the structural conservation among 26
177 sobemoviral polymerases, including in the divergent sequence domains of the index and thumb.
178 Then, the 3D models of 19 sobelivirad RdRps were generated to produce a structure-guided
179 sequence alignement and to identify the structural features of each viral genera. The RdRp
180 features combined with the characteristic polyprotein arrangement of sobelivirads allowed the
181 identification of new candidate barnaviruses and outgroups. These sequences contributed to
182 generate a new structure-guided alignement and a robust phylogenetic topology. Hence,
183 divergence dates between the two main Sobarna- and Polena-clades and between
184 sobemoviruses infecting plants and barnaviruses infecting fungi were estimated. Finally,

185 molecular signatures of sobelivirads and of each clades were identified refining the definitions
186 of the sobelivirad RdRp conserved motifs. Potential applications of our results including
187 taxonomical revisions, metagenomic sequence assignments, diagnostic tool development and
188 risks of cross-kingdom host jumps were further discussed.

189 **Results**

190 **Structural features of RYMV RdRp among the picornavirus supergroup**

191 The diversity and evolution of rice yellow mottle virus (*Sobemovirus RYMV*) have been
192 extensively studied [39,40], the estimation of its short-term substitution rate was the starting
193 point of deep evolutionary analyses of sobemoviruses [30,32]. So, we firstly modeled the
194 RYMV RdRp (Fig 1A) using Alphafold v2.2 algorithm [33], which has been shown to predict
195 protein structures with near-experimental accuracy. Predicted 3D model showed local distance
196 difference test (plDDT) scores ranging from 92 to 98, indicating highly reliable structural model
197 (maximum score = 100) (Fig 1A).

198 Up to now, the closest available experimental structures of RdRp belong to poliovirus
199 (PV) and foot-and-mouth disease virus (FMDV) [41], two well-studied species of the order
200 *Picornavirales* in the phylum *Pisuviricota* infecting humans or animals, respectively. The 3D
201 model of RYMV RdRp was aligned with them using mTM-align algorithm [42], a tool that
202 performs global superposition of protein backbones based on TM-scores and RMSD values.
203 The characteristic fingers-palm-thumb organization was observed in all three structures (Fig
204 1B). Although amino acid conservation between RYMV and PV/FMDV sequences was low
205 (<10% identity) except in the RdRp motifs, the nature and 3D orientation of most secondary
206 structural elements (α-helices and β-sheets) were preserved across all three structures (Fig 1C
207 & 1D). The central part of the protein, which forms the conserved palm domain, showed high
208 local structural similarity, with root mean square deviations (RMSD) below 1 angstrom (Å),

209 validating the approach. One of the major conserved feature of picornavirales RdRps was
210 observed, *i.e.* the index finger at the amino-terminal region is position to interact with the top
211 of the thumb, trapping the ring finger beneath it (Fig 1B). Additionally, the orientation of the
212 so-called pinky finger helix, which has been previously described to distinguish two
213 picornavirad groups including FMDV or PV [41], would bring the predicted structure of RYMV
214 closer to that of FMDV (Fig 1B-D).

215 The structure-guided alignment generated from the RYMV RdRp model facilitated
216 precise amino acid sequence comparison, in particular for the variable N-terminal domain and
217 the poorly conserved motif H located at the C-terminal extremity (Fig 1E). Globally, only 27
218 out of 560 sites (4.8%) accross the sequence alignment were strictly identical in the three
219 RdRps. Less than 50% of these conserved residues were located in the ABC motifs, ~80% in
220 all the 7 RdRp well-known motifs and ~95% in the entire homomorphs, showing the added
221 value of considering the structural domains rather than the sequence motifs only. However, as
222 expected, the phylogenetic distance between picornavirads and sobelivirads resulted in
223 significant structural differences. In the RdRp core domain, there was a larger number of β-
224 strand elements which are longer in RYMV compared to the two picornaviruses. In contrast,
225 the thumb domain, formed by α-helices at the carboxy-terminal part of RdRps, was smaller in
226 the predicted RYMV structure. Comparison with experimental structures, rare and
227 phylogenetically distant, was informative but insufficient ; it must be complemented by
228 modeling of more closely related proteins.

229 **Structural conservation and phylogeny of sobemoviral RdRps**

230 The RdRp 3D models of the 26 ICTV-approved sobemoviral species (S2 Table) were
231 generated using the same method to assess the structural conservation at the intrageneric scale.
232 They all showed high reliability, with plDDT scores comprised between 90 and 98 (S3 Fig).

233 Three dimensional superpositions of sobemovirus RdRps revealed a high degree of structural
234 similarity in the alpha-carbon traces, with RMSD to RYMV below 1Å (Fig 2A). Only the length
235 of two external loops varied from two to six residues depending the viral species. These loops
236 located at the thumb joint and at the back of the fingers, are named here as « orange and dark
237 green loops », respectively, based on the protein rainbow color code (from N-terminus dark
238 blue to C-terminus red) (Fig 2A).

239 The structure-based alignment of the 26 sobemovirus RdRps computed by mTM-align
240 has a mean length of 446 residues across 472 sites (S4 Fig). With a mean pairwise sequence
241 identity values among sobemovirus RdRps at 54.3% ± 3.2%, the resulting amino acid
242 conservation in the whole mTM-alignment was relatively low, with 120 out of 446 residues
243 being strictly conserved (27% of the mean length). The strictly conserved residues were located
244 in RdRp motifs (31%) or more extended in homomorphs (69%) but 21% belonged to the RdRp
245 extremities upstream and dowstream of the motifs G and D, respectively. Noteworthy, the 75
246 C-terminal residues including motifs E and H have not been considered in previous sobemoviral
247 alignments (Ghafari et al., 2024). Including this C-terminal part increased alignment length by
248 18%, and revealed 9 new conserved residues (7.5%).

249 The RdRp structural alignment was converted to a multiple sequence alignment of
250 nucleic acids by replacing each amino acid by its corresponding codon from their respective
251 viral genomes. The structure-guided alignment was used to reconstruct the maximum-
252 likelihood phylogenetic tree of sobemoviruses using SeaView v5.05 (Fig 2B). Phylogenetic
253 relationships between the viral species confirmed the previously reported lineages and their
254 clustering by plant family hosts [30]. RYMV belong to one of the two lineages infecting
255 monocots host plants of the *Poaceae* family. The other sobemovirus species were distributed in
256 one and two lineages infecting plants of the *Solanaceae* and *Fabaceae* families, respectively,

257 or to different and unique plant families. Hence, the structure-guided alignment of all
258 sobemovirus RdRps recapitulates known phylogenetic relationships, confirming that our
259 approach preserves evolutionary signals even at fine taxonomic scales.

260 **Structural features and structure-guided alignment of sobelivirad RdRps**

261 The order *Sobelivirales* comprised 122 viral species unevenly distributed into seven
262 genera and three families (Table 1). The genera *Barnavirus, Dinornavirus, Husbsclerovirus* and
263 *Polemovirus* contain only one member recognized by ICTV whereas poleroviruses,
264 sobemoviruses and enamoviruses contain 77, 26 and 15 species, respectively. A subset of 20
265 RdRps including maximum six taxons in each genera, including sobemoviruses, were selected
266 (Table 1). Given the amino acid sequence similarity between the RdRps of the unique
267 polemovirus and the selected poleroviruses (% ID = 64% ± 0.9%), they were grouped together
268 as « poler/moviruses » for the remainder of the analyses. Indeed, *Polemovirus PNLV* originated
269 from a recombination event between a polerovirus as major parent and a sobemovirus on its CP
270 gene [43].

271 The RdRp 3D models were generated in a view to guide the sequence alignement at the
272 intergeneric scale. The RdRp models showed plDDT scores ranging mostly from 90 to 98,
273 indicating highly reliable structural models as for the sobemoviral ones (Fig 3A-D and S5 Fig).
274 However, the RdRp 3D model of the Heterocapsa circularisquama RNA virus 01 (HcRNAV),
275 the unique dinornavirus, showed low-confidence plDDT scores (< 60) for 10% of its residues.
276 As a result, dinornaviruses were excluded from dowstream analyses to avoid bias in structural
277 alignments. As expected based on the amino acid sequence conservation (% ID >50%) at the
278 intrageneric scale, comparison of RdRp 3D models confirmed a high structure conservation at
279 this scale (RMSD <1 Å) (Table 2). Sequence and structure conservations of sobemoviral RdRps
280 was significatively lower from those of poler/moviruses.

281 The 3D superimposition of the 19 RdRp models obtained after structural alignment
282 confirmed the global strutural conservation of sobelivirad RdRps (1.9 Å < RMSD < 2.5 Å) (Fig
283 4A). While the RdRp core was highly conserved as expected, few additional structural features
284 were observed in some external loops depending on the viral taxons. The « orange loop », found
285 to be highly reliable (plDDT> 80-90) but variable in length among the sobemoviral RdRps, was
286 identified as characteristic of this genera (Fig 2A). By contrast the « dark green loop » was
287 rather stable in most of the sobelivirad RdRps analyzed here, except in those from certain
288 sobemoviruses, and therefore not considered in the dowstream analyses. In the polem/rovirus
289 and enamovirus RdRps, an extra « green helix » was identified at the thumb junction in addition
290 to a longer « red loop » located at the thumb tip. These two specific secondary elements showed
291 a conserved length and high plDDT scores (Fig 3A & 3E). Similarly, high scores were observed
292 for a « light blue helix » located at the edge of the palm in the hubsclerovirus RdRp (Fig 3D &
293 3H). Interestingly, this extended domain was also observed in the barnavirus RdRp but rather
294 as an unfolded loop (Fig 3C & 3G). However, this « light blue loop » and a « yellow helix »
295 found in the barnavirus RdRp showed low plDDT values (<70) indicating that their 3D
296 conformation remained uncertain. At this step, we cannot exclude that these three structural
297 features were restricted to the unique RdRp sequences of hubsclerovirus and barnavirus species
298 available.

299 The 3D modeling facilitated sequence alignment and optimization of gap positions and
300 lengths. The structure-guided alignment of the 19 sobelivirad RdRps has a mean length of 456
301 residues across 581 sites (Fig 4B & S6A-B Fig). Amino acid conservation was notably low,
302 with less than 12% of residues being strictly conserved. As expected, the solvent accessibility
303 scores and predicted secondary structure elements were well conserved along the linear
304 sequences in sobelivirad RdRps. Due to their low sequence conservation, the RdRp extremities
305 are generally not considered in the sequence alignment. However, they play a crucial structural

306 role in the protein structure of picorna-like RdRps (thumb/index finger interaction). The
307 structure-guided approach has enabled us to explicitely include these RdRp extremities in the
308 sequence alignment, for the first time, at this intergeneric taxonomic scale (Fig 4A & 4B).

309 Intermediate to large insertions and deletions (indels) to account for additional
310 secondary structure elements were heterogeneously distributed along the linear alignment of
311 the RdRps (Fig 4B and S6A-B Fig). They correspond to the secondary structure elements
312 already identified through the global superimposition of the 3D models and could be considered
313 as candidate taxonomical features due to their high intrageneric conservation. For instance,
314 insertions at positions 334-347 and 538-552 were characteristic to the « green helix » and « red
315 loop » in both poler/mo- and enamoviruses and insertions at positions 490-503 to the « orange
316 loop » in sobemoviruses. To a lesser extend due to unique species in the genera and to lower
317 plDDT scores, insertions at positions 120-163 and 443-455 might be characteristic to the « light
318 blue helix or loop » found in the barnavirus or husclerovirus and to the « yellow loop » in the
319 barnavirus. Notably, all these features were located outside the ABC motifs, mostly at the
320 protein extremities. Based on highly reliable structural models, our RdRp sequence alignment
321 contributed to define accurately indel lengths and locations all along the protein sequence and
322 to identify candidate structural features of each sobelivirad genera.

323 **Identification of new barnaviruses and outgroup RdRp sequences**

324 To balance the underrepresented genera in our RdRps dataset and to confirm the
325 sobelivirad structural features, RdRp sequences of new candidate sobelivirad species were
326 searched in the NCBI Virus database. Ten sequences tagged as husbscleroviruses and
327 polemoviruses were not further considered because they belong to the known HuSRV and
328 PnLV species. By contrast, 69 sequences tagged as classified or unclassified barnavirids were
329 extracted from the database. To avoid false positives due to structural evolutive convergence,

330 we firstly filtered the candidate sequences based on the specific arrangement of domains in the
331 polyprotein (protease / VPg / polymerase) — considered as a major taxonomical criteria of
332 sobelivirads. Thirthy one sequences were identified based on the conserved triad H-D-
333 Gx(S/C)G of the protease and the ABC motif of the polymerase (S7 Table). To avoid
334 redondancy, only one MBV isolate (accession number NC001633) out of three was selected.
335 Based on the high conservation of indels at the intrageneric scale observed in this study, we
336 discarded all the RdRp sequences containing long insertions between the 8 RdRp conserved
337 motifs. Under this stringent filter, five sequences were identified and considered as barnavirus
338 tentative species (Table 3). Three of them, namely CoBV, RsBV and TuBV, were originated
339 from phytoparasitic fungi [44–46], and two, referred here as tBV1 and tBV2, from
340 metagenomic samples [47].

341 To reconstruct the evolutionary history of sobelivirad taxons by phylogenetic analyses,
342 a careful outgroup selection is needed. As candidate outgroups must share a common ancestor
343 with sobelivirads, they should also share a similar genomic organisation. Polyprotein sequences
344 tagged as unclassified sobeli-, solemo- and sobemo-like were extracted from the NCBI virus
345 database. From 204 initial sequences, 55 sequences were identified based on the conserved
346 protease and RdRp motifs (S8 Table). Noteworthy, sobemoviruses recently approved by ICTV,
347 such as XYDV and PLSV (accession numbers ON828429 and ON137710), still appeared as
348 unclassified in the NCBI Virus database. After removal of sequences with long insertions and
349 based on basal phylogenetic position, Soybean thrips sobemo-like virus 5 (SoTV5,
350 MW023867), Scaphoideus titanus sobemo-like virus 1 (ScTV1, MN982389), and Hypera
351 postica associated sobemovirus 1 (HyPV1, MW676135) were selected as outgroups for the
352 following analyses (Table 3). They all originated from phytophageous insects [48–50].

353 The 3D models of the 5 candidate barnaviral and the 3 outgroup RdRps were generated
354 using Alphafold (S9 Fig). They showed plDDT scores ranging mostly from 90 to 98, indicating
355 highly reliable structural models. The low plDDT score of the « light blue loop » of barnavirus
356 RdRp was observed only for MBV but not for the new barnaviruses. The candidate structural
357 features of barnaviruses, the « light blue and yellow loops », showed length variability similarly
358 to the « orange loop » of sobemoviruses. In comparaison to the new barnaviruses, the three
359 outgroup models showed more structural conservation among them. However, the 3D
360 superimposition of the total dataset revealed several specific features in external loops of the 3
361 outgroups (S10 Fig). Both a longer and a shorter loops were identified in their « dark green »
362 region.

363 **Sobelivirad phylogeny and divergence date estimation**

364 The structure-guided alignment of the 27 RdRps including candidate barnaviruses and
365 outgroups was generated with mTM-align (Fig 5A & 5B). Following the same method than for
366 the sobemovirus phylogeny, the structural alignment was converted to a multiple sequence
367 alignment of nucleic acids by replacing each amino acid by its corresponding codon, and, then,
368 it was used to construct the maximum-likelihood phylogenetic tree of *Sobelivirales* with
369 SeaView (Fig 6A). The three supplemental sequences were confirmed as outgroups and formed
370 a well-supported cluster. They contributed to reveal two main clades with high bootstrap
371 supports. On the one hand, phylogenetic relationships between poler/moviruses and
372 enamoviruses were confirmed, this clade was referred to as Polena-clade. On the other hand,
373 plant sobemoviruses and fungal barna-/hubsclerovirus clustered together, establishing the
374 monophyly of this group for the first time, referred to as Sobarna-clade. The structure-guided
375 alignment of our relevant dataset resulted in the first accurate phylogeny of sobelivirads — a
376 starting point to propose a taxonomical revision of sobelivirad families.

377 To assess the robustness of the inferred topology to the alignment strategy, we also
378 examined an alternative structure-informed alignment based on the 3Di structural alphabet
379 [51,52]. Apart from the position of HuSRV, which appeared distantly related to the
380 sobemoviruses rather than the barnaviruses and the absence of posterior support values, this
381 analysis recovered a broadly similar tree topology (S11 Fig). We therefore retained the mTM-
382 align/codon-based tree for the main analyses, as it enabled downstream nucleotide phylogenetic
383 inference and PoW-based divergence dating. According to this analysis, the most recent
384 common ancestor (TMRCA) of sobemoviruses dates back to 2.2 million years before present
385 (BP) (95% highest posterior density (HPD): 1.0-3.8 million years BP) (Fig 6B), aligning closely
386 with a recent estimate [30]. The divergence time of the Polena-clade was estimated to 5.1
387 million years BP (95% HPD: 2.6–8.3). The findings here provides the first estimate of the
388 TMRCA for fungal (barna)viruses, which dates back to 4.6 million years BP (95% HPD: 2.0-
389 8.1), and for the Sobarna-clade to 18.7 million years BP (95% HPD: 11.1– 27.1). Furthermore,
390 this analysis offers the first estimation of the TMRCA of the *Sobelivirales* order, reaching back
391 to 27.0 (95% HPD: 18.1-37.5) million years BP, which is nearly ten times older than the origins
392 of sobemoviruses.

393 **Integrating structural and phylogenetic analyses of sobelivirad RdRps**

394 In this study, 44 sobelivirad polymerases (including 26 sobemoviral RdRps) were
395 modeled and compared. In total, only 43 residues were strictly conserved at the intergeneric
396 scale (less than 9% of the mean sequence length) (Fig 4 & 5). Despite significant divergence in
397 their protein sequences, the RdRp core within the central tunnel of sobelivirads were found to
398 be highly conserved in structure across all genera in agreement with strong functional restraints.
399 A fourth of the conserved residues are located within the ABC motifs, half in the well-
400 characterized RdRp motifs and >80% in homomorphs. Excluding the residues shared with the

401 picornavirus PV and FMDV, 29 out of these 43 conserved amino acids could represent
402 interesting sobelivirad markers for taxonomical assignement, in particular nine of them located
403 in RdRp motifs including 382-SG-383, 288-PH-289, 545-FCS-547 in the motifs A, F and E,
404 respectively (Fig 7).

405 In order to identify markers of the two major clades identified with the phylogenetic
406 analyses of sobelivirads, the amino acid sequence conservation inside the RdRp core was
407 compared between the Polena- versus Sobarna-clades. All ICTV-recognized poleroviruses and
408 enamoviruses were included in the RdRp sequence dataset in addition to the known
409 sobemoviruses and the candidate barnaviruses (S12 Table). This dataset of 127 sequences
410 allowed the identification of 10 clade-specific residues, *i.e.* amino acids conserved in one clade
411 and distinct in the other clade (Fig 7). Six of them were localized at the same positions 206,
412 381, and 499 in motifs G, A and C, respectively (referring to the alignment numbering in Fig
413 5A & 5B). Indeed, the RdRps of the sobarna-clade harbour one isoleucine and two serines at
414 these positions whereas a cysteine, an alanine and a glycine were observed in the polena-clade.
415 Three other residues were specific of the sobarna-clade : 207P in motif G, 520H and 523K, the
416 last two occupying notable 3D positions in motif D. Indeed, The lysine at position 523 pointed
417 into the entry of the central tunnel, while the histidine at position 520 participated to the
418 stabilization of the region with its interaction with the residue 292E (Fig 7). Notably, the
419 sobemoviral « orange loop » between Motifs D and E was also validated in the large dataset. In
420 the Polena-clade, a strictly conserved lysine at position 290 within Motif F on the picky finger
421 domain also pointed into the central tunnel, alongside the conserved sobelivirad lysine at
422 position 293 (Fig 7). In addition, the polena- « green helix » upstream Motif A and the polena-
423 « red loop » dowstream Motif H provided robust indel signatures. Integrating structural and
424 phylogenetic analyses of sobelivirad RdRps revealed molecular signatures of these deeply
425 divergent clades, refining the definitions of the conserved motifs (Fig 7).

426 **Discussion**

427 Our study aligns with recent efforts to leverage structural phylogenomics for resolving
428 deep evolutionary relationships in highly divergent proteins [35]. We combined AlphaFold,
429 mTM-align and PoW dating to produce a robust and extensive sequence alignement of RdRps
430 and to investigate the evolutionary history of sobelivirads. By contrast to the alignments based
431 on structural alphabets such as FoldSeek 3Di [51], our methodology based on global backbone
432 superpositions offers the advantages to capture extended motifs and to be less sensitive to local
433 inaccuracies in flexible regions. Protein modeling successfully provided high-confidence
434 structural elements to complement anchor points based on the classical functional motifs. Most
435 of these structural features are located outside the ABC motif, often in the RdRp sequence
436 extremities generally not considered for the alignment despite their functional role. For
437 instance, in the NLM's conserved domain database (CDD) [53], the barnavirid RdRp domain
438 (cd23184) was restricted to the 259 amino acids in the palm domain (corresponding to aa 170-
439 429 in Fig 5A & 5B). However, we showed that the conserved motifs G and H in the index
440 finger and thumb domains were very well predicted by Alphafold (plDDT > 90), and showing
441 high structural similarity (RMSD < 2Å). Hence, our structure-guided approach doubled the
442 alignment length of the sobelivirad RdRp and increased its quality considering now the
443 extremities of RdRp and its intra-molecular interactions. Such an extension can result in a more
444 precise phylogenetic analysis.

445 As for other mycovirus taxons, barnavirid diversity remains largely unknown. Up to
446 now, the family *Barnaviridae* contains only one viral species recognized by ICTV (*Mushroom*
447 *bacilliform virus*, MBV). However, the NCBI Virus database contains 69 sequences tagged as
448 « Barnaviridae » and the NCBI Taxonomy browser inventories 13 « unclassified Barnavirus »
449 and 4 « unclassified Barnaviridae ». Moreover, the number of metagenomic publications

450 describing partial sequences related to barnavirids or referred to as « barna-like » viruses is
451 increasing [44–47,54–57]. In the CDD database, besides rhizoctonia solani barnavirus 1
452 (RsBV) and tulasnella barnavirus 1 (TuBV), apple barna-like 1 (ApBV) is also considered and
453 included in the short protein alignment of the RdRp palm domain. In addition to another
454 independent taxonomical criteria (the polyprotein domain arrangment), our study provides
455 stringent structural criteria of RdRps to assign sequences to the order *Sobelivirales*. We
456 considered that, except the five sequences used in our dataset (RsBV, TuBV, CoBV, tBV1,
457 tBV2), all the other metagenomic sequences cited above should not be considered as barnavirids
458 in absence of other taxonomical evidences. In the near future, new metagenomic datasets
459 targeting fungi could increase the number of barnavirid sequences.

460 The same approach combining polyprotein domain arrangement and RdRp structural
461 features allowed the identification of three outgroups from metagenomic data based on their
462 genetic distance from the sobelivirad dataset. These outgroups enhanced the quality of the
463 phylogenetic tree and contributed to reveal two main clades with high bootstrap supports. This
464 phylogenic analysis led to a new taxonomic proposal : the creation of two new families
465 Polenaviridae (clustering the genera *Polerovirus*, *Enamovirus* and *Polemovirus*) and
466 Sobarnaviridae (clustering genera *Sobemovirus*, *Barnavirus* and *Hubsclerovirus*), replacing the
467 *Solemoviridae* and *Barnaviridae*. By assessing the conservation of the structure-guided
468 alignment of the 127 RdRp sobelivirid sequences, we enriched the common conserved motifs,
469 identified molecular signatures of the two new proposed families, and revealed structural
470 features that can be used as marker of viral genera. Interestingly, newly identified putative
471 sobemo-like and sobeli-like viruses often represent a significant portion of newly discovered
472 plant and fungus taxa within environnemental viromes [17,47,58–60]. Extending this approach
473 in the near future will accelerate their taxonomic classification, a priority for virus research

474 [61]. Additionally, these RdRp features could be useful for the development of molecular tools
475 for diagnostic and detection such as PCR primers.

476 At a higher taxonomical rank, in the class *Pisoniviricetes,* the well-known
477 picornaviruses show both structural and functional similarities with sobelivirids, such as the use
478 of a Viral Protein genome-linked (VPg) to prime RNA replication by the RdRp. However,
479 picornaviral VPg are very small (~20 residues) and are cleaved from the RdRp by a viral
480 protease located downtream of the VPg domain. In contrast, the sobarnavirids and polenavirids
481 share a conserved genomic organization (protease/VPg/polymerase) and have large VPgs (80-
482 150 residues), which likely result in or induce specific molecular mechanisms. While
483 interactions between picornaviral VPg and RdRps and their structure-function relationships
484 have been extensively studied in human and animal viruses, very little data exist for plant
485 viruses. Our approach could help to fill this knowledge gap, or at least provide a better
486 understanding of plant RdRp features through *in silico* comparative analyses.

487 For example, the comparative analysis of Motif E, unique to RdRps compared to other
488 polymerases, offers a good starting point for better understanding of some of the plant RdRp
489 features. The amino acid composition of this motif is directly linked to the viral replication
490 strategy, whether it is based on protein priming (VPg) or not. Described as a « primer grip »,
491 this motif has been shown to help to correctly position the 3' OH end of the primer strand [9,18].
492 In picornaviral RdRps, the conserved phenylalanine in Motif E, which faces Motif C, is
493 followed by two residues with relatively long side chains (leucine-lysine, LK), while in
494 flaviviral RdRps, the central channel is narrower, and these residues have rather short side
495 chains (cysteine-serine, CS). Interestingly, plant and fungus sobelivirad RdRps share the same
496 conserved Motif E as flavivirids. Additionaly, the homomorph that contains Motif E forms part
497 of the NTP entry tunnel and can vary in length depending on the viral species. Sobemoviruses

498 exhibit a longer « orange loop » at the thumb joint compared to other members of *Sobelivirales*,
499 another unexpected similarity with flaviviral hepatitis virus C. However, at this taxonomic rank,
500 evolutive convergence should not be discarded.

501 At the sobelivirad scale, taxonomical factors such as genomic arrangement limit the
502 convergence risk. The common evolutionary history of plant and fungus sobelivirad RdRps
503 could be illustrated by Motif D. This motif influences the rate of nucleotide addition to the
504 nascent RNA strand. Characterized by a glycine hinge and a positively charged lysine, both
505 located in similar positions across available RdRp crystal structures, the distance inside the
506 linear sequence between these two conserved residues varies across different viral families [9].
507 In the Sobarnaviridae clade, lysine (or its equivalent arginine) is conserved at position N+4, i.e.
508 position 523 (referring to Fig 7). In the Polenaviridae clade, our initial analysis was too stringent
509 to detect the conserved lysine. However, a lysine was found at position N+2 – i.e. position 521,
510 in a large part of the sequences, with its orientation differing from the Sobernaviridae lysine in
511 the 3D structural model. Motif D has been shown to influence not only incorporation speed but
512 also fidelity, both of which are closely linked to viral pathogenesis and virulence. For example,
513 increased fidelity in a poliovirus mutant has been associated with viral attenuation [62]. The
514 putative role of lysine 523 could be tested via site-directed mutagenesis in model sobemoviruses
515 like RYMV.

516 The structural differences and evolutionary distances between RdRps of sobemoviruses
517 and poleroviruses (two phytovirus genera in the order *Sobelivirales*) were found to be greater
518 than those between sobemoviruses and barnaviruses (a mycovirus genus in the family
519 *Barnaviridae*) as well as hubscleroviruses (a mycovirus genus in the family *Solemoviridae*).
520 Given the role of RdRp as a major marker of viral evolutionary history, our results confirmed
521 that horizontal viral transfert(s) (HVT) between plant/fungal viruses occured in the Order

522 *Sobelivirales*. The direction of the HVT(s) cannot yet be determined without further data on the
523 hosts of metagenomic sequences we used as outgroups.

524 Using an interdisciplinary approach, we provided the first divergence time estimation
525 between these plant and fungal viruses, dating back to ~27 million years BP. The TDRP-
526 corrected evolutionary history of *Sobelivirales* was extended five thousand times, far beyond
527 previous estimates [32]. However, this timeframe should be interpreted cautiously. Our current
528 application of the PoW model is based on a short-term substitution rate estimated for the RYMV
529 polymerase gene and assumes that this rate provides a reasonable proxy across the broader
530 sobelivirad phylogeny. If short-term rates vary substantially among plant- and fungal-infecting
531 clades, some node ages may be shifted to older or younger than their true values. More
532 generally, while PoW accounts for time-dependent rate decay across timescales, it does not
533 fully resolve lineage-specific branch-rate heterogeneity. These estimates should therefore be
534 viewed as a first unified temporal framework for sobelivirad evolution rather than as precise
535 point estimates for all clades. Nevertheless, even with our current conservative estimates, the
536 estimate time favour a model in which the present-day distribution of sobelivirads across plants
537 and fungi reflects past host shifts across kingdoms rather than uninterrupted codivergence with
538 their current hosts over hundreds of millions of years [63]. Other lineages of mycoviruses either
539 with single- or double-stranded RNA genomes also showed close phylogenetic relationships
540 with plant viruses at different taxonomical ranks [20,24,64], suggesting that taxonomy of these
541 viruses should not be considered independently but rather as a continuum. An interdisciplinary
542 approach at the interface of biophysics, bioinformatics, and virology should be encouraged to
543 better understand evolution, interactions and biological mechanisms of myco- and
544 phytoviruses.

545

546 **Material and methods**

547 Sequence datasets

548 The sobemoviral dataset used in this study contained all the RdRp sequences from
549 RYMV to the other 25 ICTV-approved sobemoviral species (S2 Table). The sobelivirad dataset
550 contained a subset of representative RdRp sequences from the viral order (Table 1). Six
551 poleroviruses, six sobemoviruses and four enamoviruses were selected to maximize the RdRp
552 sequence diversity for each genera based on their phylogeny. For each of the genera *Barnavirus,*
553 *Husbsclerovirus* and *Polemovirus,* only one ICTV-approved member was available. The unique
554 member of *Dinornavirus* (HcRNAV) was excluded from the dataset due to low-confidence
555 plDDT scores (see above).

556 Five putative barnaviral sequences were added to the 19 sobelivirad datased based on a
557 three-step methodology. Firstly, polyprotein sequences tagged as « Barnaviridae » or
558 « unclassified Barnaviridae » were extracted from the NCBI Virus database and aligned with
559 MUSCLE [65] in SeaView v5.05 [66]. The N-terminal extremities of the RdRps often
560 incomplete were manually refine. Secondly, the presence of the conserved triad H-D-Gx(S/C)G
561 of the protease upstream the ABC motif of the polymerase were checked (S7 Table). Finally,
562 sequences with long insertions between the 8 RdRp conserved motifs were discarded (Table 3).
563 In addition, three outgroup sequences tagged as unclassified sobeli-, solemo- and sobemo-like
564 were selected and added to result in the 27 RdRp dataset using the same process (S8 Table,
565 Table 3). A total sobelivirad dataset made of 127 RdRp sequences was also constitued adding
566 all ICTV-recognized poleroviruses and enamoviruses (S12 Table) to the 26 sobemoviral and
567 the 27 sobelivirad datasets described above.

568 3D model prediction, evaluation and structural alignment

569 The 3D structures of the RYMV, the 25 other sobemoviral and the 27 sobelivirad RdRps
570 were predicted using a locally installed version of Alphafold v2.2 [33]. For each sequence, 5

571 models were generated using default Alphafold parameters and the highest-ranked model was
572 conserved. The pLDDT reliability scores were available in the B-factor field of each Alphafold
573 model. Models showing 10% of their residues having low-confidence plDDT scores (< 60)
574 were not considered in this study.

575 The RYMV 3D model was superposed to the closest experimental structures available,
576 PDB 1RA6 and 1U09 of poliovirus (PV) and foot-and-mouth disease virus (FMDV) RdRps,
577 respectively, using mTM-align [42]. Pairwise RMSD scores were calculated using ProFit
578 (Martin A.C.R., http://www.bioinf.org.uk/software/profit/). The aligned sequences and their
579 corresponding RMSD scores in Angstroms (identifier suffixed with d3.RMSD_1ra6A an
580 .d3.RMSD_1u09A) are interfaced with the 3D models using Jsmol applet and script [67]
581 (https://pat.cbs.cnrs.fr/sobemo/rmsd.html). The Cα deviation colors can be projected onto one
582 of the displayed structures by cliking one of these identifiers. The pairwise sequence identity
583 between the proteins was calculated with Ident and Sim software [68]. Descriptive statistics
584 (mean and standard deviation), one-way analysis of variance (ANOVA) and Tukey's honest
585 significant difference (HSD) post-hoc test were performed to identify specific pairwise
586 differences ($p < 0.05$). The strictly conserved and the similar residues were visualized onto the
587 sequence alignment using Color Align Conservation [68]. Homomorphs were identifed based
588 on [18].

589 The 3D models of the 26 sobemoviral and the 27 sobelivirad RdRps were superposed
590 using mTM-align and were interfaced with their sequences using JSmol applet and script
591 (https://pat.cbs.cnrs.fr/sobemo/sobemo_v7/sobemo and https://pat.cbs.cnrs.fr/sobemo/sobemo
592 _v7/sobeli, respectively). By default, the 3D models were colored using standard rainbow code
593 (from N-terminus, dark blue, to C-terminus, red). Selected Alphafold models can be displayed
594 and customized by using the toggle buttons and the popup menus from the right bottom panel
595 and the colors of the amino acids of one aligned sequence from the top panel can be projected

596 on the selected 3D structure(s) by clicking on the corresponding protein identifier in the top left
597 panel. We carefully delineated the polymerase domain shared by the proteins by inspecting both
598 the positional conservation of the aligned sequences, the superposition of their predicted
599 secondary structures and the solvent accessibility. The secondary structure and the solvent
600 accessibility of each sobelivirad model residue was calculated using STRIDE [69].

601 Phylogenetic analysis

602 The mTM-align structural alignments of sobemoviral and sobelivirad RdRps were
603 translated into multiple sequence alignments of nucleic acids by replacing each amino acid by
604 their corresponding codon from the viral genomes using PAL2NAL [70]. JModelTest 2 [71]
605 was used to identify the best nucleotide substitution model for phylogenetic analysis based on
606 the Akaike Information Criterion. The Maximum-Likelihood tree was then constructed with
607 the General Time Reversible Model with an optimised rate variation and class for invariant sites
608 (GTR + G + I) using PhyML implemented in SeaView v 5.05 [66]. The reliability of the
609 phylogenetic tree nodes was estimated via a bootstrap resampling procedure with 100
610 replicates. The sobemoviral and sobelivirad phylogenetic tree were midpoint rooted and rooted
611 using the outgroup branch, respectively. The mTM-align tree was compared to an alternative
612 one based on the 3Di structural alphabet using Foldtree powered by Foldseek [51,52].

613 Time-tree reconstruction using the Prisoner of War model

614 We used the Prisoner of War (PoW) model of virus evolution to account for the power-
615 law decay in the substitution rate estimates of sobelivirads, as described in Ghafari et al. [30],
616 where we also established the presence of a robust and representative molecular clock signal
617 for the RdRp locus for sobemoviruses. Briefly, the short-term substitution rate for the RYMV
618 polymerase gene, inferred from a standard molecular clock model was used as the baseline rate.
619 RYMV provides the only available sobelivirad lineage for which a robust short-term rate
620 estimate had previously been established. PoW model was then applied to root-to-node genetic

621 distances in an ultrametric distance tree. For each internal node, the genetic node height was
622 transformed into calendar time using the PoW distance-to-time mapping, parameterised by the
623 short-term rate distribution estimated from RYMV and the fastest-evolving rate class.
624 Calendar-time branch durations were then obtained as the differences between transformed
625 parent and child node ages. The maximum clade credibility tree was constructed after 100
626 iterations using TreeAnnotator v.1.10.4 and visualized through a lineages-through-time plot
627 using the phytools library in R [72].

628 Polymerase evolution

629 The 44 sequences of sobemoviral and sobelivirad RdRps were compared alone and with
630 those of PV and FDMV using Color Align Conservation to identify candidate sobelivirad
631 markers. Then, the total sobelivirad dataset made of 127 RdRp proteins were aligned using
632 MAFFT [73] to identify RdRp core markers of the Polena- versus Sobarna- clades. The three
633 outgroups were not considered in this analysis. At each site of the 8 RdRp conserved motifs,
634 amino acids present in all sequences of one clade (or that were similar based on BLOSUM score
635 ≥ 0) and absent from sequences of the other clade were considered as clade-specific residue.
636 The location of these sobelivirad and clade-specific marker residues were visualized and
637 colored on the 3D models of RYMV and PLRV as sticks on Cα trace using JSmol. Motif
638 visualization was performed with WebLogo v2.8.2 [74]. Each sequence logo consists of stacks
639 of symbols, one stack for each position in the sequence. The overall height of the stack indicates
640 the sequence conservation at that position, while the height of symbols within the stack
641 indicates the relative frequency of each amino at that position.

642 **Acknowledgements**

643 This work benefited from discussions with Aurore Comte and Nils Poulicard, other members
644 of Virus/Cereal Interactions in Tropical Agroecosystems group and colleagues of Plant Health

Institute Montpellier. This work was supported by the French Infrastructure for Integrated Structural Biology (FRISBI) ANR-10-INBS-0005 and a Wellcome Trust Early-Career Award (grant 309205/Z/24/Z) to M.G.

Data availability

The sequence accession numbers used in this study and their alignments were given in Tables, Figures and Supporting data. The 3D models generated using Alphafold were available online.

Funding

None.

Conflict of interest : None declared.

References

1. Simmonds P, Adams MJ, Benkő M, Breitbart M, Brister JR, Carstens EB, et al. Virus taxonomy in the age of metagenomics. Nat Rev Microbiol. mars 2017;15(3):161-8. doi:10.1038/nrmicro.2016.177

2. Gorbalenya AE. Increasing the number of available ranks in virus taxonomy from five to ten and adopting the Baltimore classes as taxa at the basal rank. Arch Virol. oct 2018;163(10):2933-6. doi:10.1007/s00705-018-3915-6

3. Siddell SG, Walker PJ, Lefkowitz EJ, Mushegian AR, Adams MJ, Dutilh BE, et al. Additional changes to taxonomy ratified in a special vote by the International Committee on Taxonomy of Viruses (October 2018). Arch Virol. mars 2019;164(3):943-6. doi:10.1007/s00705-018-04136-2

4. Koonin EV, Dolja VV, Krupovic M, Varsani A, Wolf YI, Yutin N, et al. Global Organization and Proposed Megataxonomy of the Virus World. Microbiol Mol Biol Rev. 20 mai 2020;84(2):e00061-19. doi:10.1128/MMBR.00061-19

5. Wolf YI, Kazlauskas D, Iranzo J, Lucía-Sanz A, Kuhn JH, Krupovic M, et al. Origins and Evolution of the Global RNA Virome. Racaniello VR, éditeur. mBio. 21 déc 2018;9(6):e02329-18. doi:10.1128/mBio.02329-18

6. Simmonds P, Adriaenssens EM, Zerbini FM, Abrescia NGA, Aiewsakun P, Alfenas-Zerbini P, et al. Four principles to establish a universal virus taxonomy. PLOS Biol. 13 févr 2023;21(2):e3001922. doi:10.1371/journal.pbio.3001922

675 7. Walker PJ, Siddell SG, Lefkowitz EJ, Mushegian AR, Adriaenssens EM, Alfenas-Zerbini P, et al.
676 Changes to virus taxonomy and to the International Code of Virus Classification and
677 Nomenclature ratified by the International Committee on Taxonomy of Viruses (2021). Arch
678 Virol. sept 2021;166(9):2633-48. doi:10.1007/s00705-021-05156-1

679 8. Ferrero D, Ferrer-Orta C, Verdaguer N. Viral RNA-Dependent RNA Polymerases: A Structural
680 Overview. In: Harris JR, Bhella D, éditeurs. Virus Protein and Nucleoprotein Complexes [Internet].
681 Singapore: Springer Singapore; 2018 [cité 9 oct 2024]. p. 39-71. (Subcellular Biochemistry).
682 Disponible sur: https://link.springer.com/10.1007/978-981-10-8456-0_3 doi:10.1007/978-981-
683 10-8456-0_3

684 9. Jácome R, Becerra A, Ponce de León S, Lazcano A. Structural Analysis of Monomeric RNA-
685 Dependent Polymerases: Evolutionary and Therapeutic Implications. Kuhn JH, éditeur. PLOS
686 ONE. 23 sept 2015;10(9):e0139001. doi:10.1371/journal.pone.0139001

687 10. Koonin EV, Wolf YI, Nagasaki K, Dolja VV. The Big Bang of picorna-like virus evolution antedates
688 the radiation of eukaryotic supergroups. Nat Rev Microbiol. déc 2008;6(12):925-39.
689 doi:10.1038/nrmicro2030

690 11. te Velthuis AJW. Common and unique features of viral RNA-dependent polymerases. Cell Mol
691 Life Sci. nov 2014;71(22):4403-20. doi:10.1007/s00018-014-1695-z

692 12. Koonin EV. The phylogeny of RNA-dependent RNA polymerases of positive-strand RNA viruses. J
693 Gen Virol. 1 sept 1991;72(9):2197-206. doi:10.1099/0022-1317-72-9-2197

694 13. Poch O, Sauvaget I, Delarue M, Tordo N. Identification of four conserved motifs among the RNA-
695 dependent polymerase encoding elements. EMBO J. déc 1989;8(12):3867-74.
696 doi:10.1002/j.1460-2075.1989.tb08565.x

697 14. Gorbalenya AE, Pringle FM, Zeddam JL, Luke BT, Cameron CE, Kalmakoff J, et al. The Palm
698 Subdomain-based Active Site is Internally Permuted in Viral RNA-dependent RNA Polymerases of
699 an Ancient Lineage. J Mol Biol. nov 2002;324(1):47-62. doi:10.1016/S0022-2836(02)01033-1

700 15. Charon J, Buchmann JP, Sadiq S, Holmes EC. RdRp-scan: A bioinformatic resource to identify and
701 annotate divergent RNA viruses in metagenomic sequence data. Virus Evol. 20 oct
702 2022;8(2):veac082. doi:10.1093/ve/veac082

703 16. Edgar RC, Taylor B, Lin V, Altman T, Barbera P, Meleshko D, et al. Petabase-scale sequence
704 alignment catalyses viral discovery. Nature. 3 févr 2022;602(7895):142-7. doi:10.1038/s41586-
705 021-04332-2

706 17. Hou X, He Y, Fang P, Mei SQ, Xu Z, Wu WC, et al. Using artificial intelligence to document the
707 hidden RNA virosphere. Cell. oct 2024;S0092867424010857. doi:10.1016/j.cell.2024.09.027

708 18. Lang DM, Zemla AT, Zhou CLE. Highly similar structural frames link the template tunnel and NTP
709 entry tunnel to the exterior surface in RNA-dependent RNA polymerases. Nucleic Acids Res. févr
710 2013;41(3):1464-82. doi:10.1093/nar/gks1251

711 19. Mönttinen HAM, Ravantti JJ, Poranen MM. Structure Unveils Relationships between RNA Virus
712 Polymerases. Viruses. 17 févr 2021;13(2):313. doi:10.3390/v13020313

713 20. Dolja VV, Krupovic M, Koonin EV. Deep Roots and Splendid Boughs of the Global Plant Virome.
714 Annu Rev Phytopathol. 25 août 2020;58(1):23-53. doi:10.1146/annurev-phyto-030320-041346

715 21. Kondo H, Botella L, Suzuki N. Mycovirus Diversity and Evolution Revealed/Inferred from Recent
716 Studies. Annu Rev Phytopathol. 26 août 2022;60(1):307-36. doi:10.1146/annurev-phyto-021621-
717 122122

718 22. Xie J, Jiang D. Understanding the Diversity, Evolution, Ecology, and Applications of Mycoviruses.
719 Annu Rev Microbiol. 10 nov 2024;78(1):595-620. doi:10.1146/annurev-micro-041522-105358

720 23. Andika IB, Tian M, Bian R, Cao X, Luo M, Kondo H, et al. Cross-Kingdom Interactions Between
721 Plant and Fungal Viruses. Annu Rev Virol. 29 sept 2023;10(1):119-38. doi:10.1146/annurev-
722 virology-111821-122539

723 24. Roossinck MJ. Evolutionary and ecological links between plant and fungal viruses. New Phytol.
724 janv 2019;221(1):86-92. doi:10.1111/nph.15364

725 25. Gorbalenya AE, Donchenko AP, Blinov VM, Koonin EV. Cysteine proteases of positive strand RNA
726 viruses and chymotrypsin-like serine proteases: A distinct protein superfamily with a common
727 structural fold. FEBS Lett. 30 janv 1989;243(2):103-14. doi:10.1016/0014-5793(89)80109-7

728 26. Revill PA, Davidson AD, Wright PJ. Mushroom Bacilliform Virus RNA: The Initiation of Translation
729 at the 5' End of the Genome and Identification of the VPg. Virology. sept 1998;249(2):231-7.
730 doi:10.1006/viro.1998.9345

731 27. Sõmera M, Fargette D, Hébrard E, Sarmiento C, ICTV Report Consortium. ICTV Virus Taxonomy
732 Profile: Solemoviridae 2021: This article is part of the ICTV Virus Taxonomy Profiles collection. J
733 Gen Virol. 9 déc 2021;102(12). doi:10.1099/jgv.0.001707

734 28. Simmonds P, Adriaenssens EM, Lefkowitz EJ, Oksanen HM, Siddell SG, Zerbini FM, et al. Changes
735 to virus taxonomy and the ICTV Statutes ratified by the International Committee on Taxonomy of
736 Viruses (2024). Arch Virol. nov 2024;169(11):236. doi:10.1007/s00705-024-06143-y

737 29. Ghafari M, Simmonds P, Pybus OG, Katzourakis A. A mechanistic evolutionary model explains the
738 time-dependent pattern of substitution rates in viruses. Curr Biol. nov 2021;31(21):4689-
739 4696.e5. doi:10.1016/j.cub.2021.08.020

740 30. Ghafari M, Sõmera M, Sarmiento C, Niehl A, Hébrard E, Tsoleridis T, et al. Revisiting the origins of
741 the Sobemovirus genus: A case for ancient origins of plant viruses. Nagy PD, éditeur. PLOS
742 Pathog. 11 janv 2024;20(1):e1011911. doi:10.1371/journal.ppat.1011911

743 31. Aiewsakun P, Katzourakis A. Time-Dependent Rate Phenomenon in Viruses. Ross SR, éditeur. J
744 Virol. 15 août 2016;90(16):7184-95. doi:10.1128/JVI.00593-16

745 32. Fargette D, Pinel A, Rakotomalala M, Sangu E, Traoré O, Sérémé D, et al. *Rice Yellow Mottle Virus*
746 , an RNA Plant Virus, Evolves as Rapidly as Most RNA Animal Viruses. J Virol. avr
747 2008;82(7):3584-9. doi:10.1128/JVI.02506-07

748 33. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein
749 structure prediction with AlphaFold. Nature. 26 août 2021;596(7873):583-9.
750 doi:10.1038/s41586-021-03819-2

751 34. Carpentier M, Chomilier J. Protein multiple alignments: sequence-based versus structure-based
752 programs. Valencia A, éditeur. Bioinformatics. 15 oct 2019;35(20):3970-80.
753 doi:10.1093/bioinformatics/btz236

754 35. Puente-Lelievre C, Malik A, Douglas J. Protein Structural Phylogenetics. Alvarez-Ponce D, éditeur.
755 Genome Biol Evol. 30 juill 2025;17(8):evaf139. doi:10.1093/gbe/evaf139

756 36. Gutnik D, Evseev P, Miroshnikov K, Shneider M. Using AlphaFold Predictions in Viral Research.
757 Curr Issues Mol Biol. 21 avr 2023;45(4):3705-32. doi:10.3390/cimb45040240

758 37. Mifsud JCO, Lytras S, Oliver MR, Toon K, Costa VA, Holmes EC, et al. Mapping glycoprotein
759 structure reveals Flaviviridae evolutionary history. Nature. 19 sept 2024;633(8030):695-703.
760 doi:10.1038/s41586-024-07899-8

761 38. Nino Barreat JG, Katzourakis A. Deep mining reveals the diversity of endogenous viral elements
762 in vertebrate genomes. Nat Microbiol. 22 oct 2024;9(11):3013-24. doi:10.1038/s41564-024-
763 01825-4

764 39. Ndikumana I, Onaga G, Pinel-Galzi A, Rocu P, Hubert J, Wéré HK, et al. Grains, trade and war in
765 the multimodal transmission of Rice yellow mottle virus: An historical and phylogeographical
766 retrospective. Malmstrom, CM, éditeur. PLOS Pathog. 17 juin 2025;21(6):e1013168.
767 doi:10.1371/journal.ppat.1013168

768 40. Hébrard E, Poulicard N, Rakotomalala M. Rice Yellow Mottle Virus (Solemoviridae). In:
769 Encyclopedia of Virology [Internet]. Elsevier; 2021 [cité 18 févr 2026]. p. 675-80. Disponible sur:
770 https://linkinghub.elsevier.com/retrieve/pii/B9780128096338212442 doi:10.1016/B978-0-12-
771 809633-8.21244-2

772 41. Peersen OB. Picornaviral polymerase structure, function, and fidelity modulation. Virus Res. avr
773 2017;234:4-20. doi:10.1016/j.virusres.2017.01.026

774 42. Dong R, Peng Z, Zhang Y, Yang J. mTM-align: an algorithm for fast and accurate multiple protein
775 structure alignment. Valencia A, éditeur. Bioinformatics. 15 mai 2018;34(10):1719-25.
776 doi:10.1093/bioinformatics/btx828

777 43. Aus Dem Siepen M, Pohl JO, Koo BJ, Wege C, Jeske H. Poinsettia latent virus is not a cryptic virus,
778 but a natural polerovirus–sobemovirus hybrid. Virology. juin 2005;336(2):240-50.
779 doi:10.1016/j.virol.2005.03.020

780 44. Nibert ML, Manny AR, Debat HJ, Firth AE, Bertini L, Caruso C. A barnavirus sequence mined from
781 a transcriptome of the Antarctic pearlwort Colobanthus quitensis. Arch Virol. juill
782 2018;163(7):1921-6. doi:10.1007/s00705-018-3794-x

783 45. Marzano SYL, Nelson BD, Ajayi-Oyetunde O, Bradley CA, Hughes TJ, Hartman GL, et al.
784 Identification of Diverse Mycoviruses through Metatranscriptomics Characterization of the
785 Viromes of Five Major Fungal Plant Pathogens. Simon A, éditeur. J Virol. août
786 2016;90(15):6846-63. doi:10.1128/JVI.00357-16

787 46. Sutela S, Forgia M, Vainio EJ, Chiapello M, Daghino S, Vallino M, et al. The virome from a
788 collection of endomycorrhizal fungi reveals new viral taxa with unprecedented genome
789 organization. Virus Evol. 1 juill 2020;6(2):veaa076. doi:10.1093/ve/veaa076

790 47. Chen YM, Sadiq S, Tian JH, Chen X, Lin XD, Shen JJ, et al. RNA viromes from terrestrial sites across
791 China expand environmental viral diversity. Nat Microbiol. 28 juill 2022;7(8):1312-23.
792 doi:10.1038/s41564-022-01180-2

48. Thekke-Veetil T, Lagos-Kutz D, McCoppin NK, Hartman GL, Ju HK, Lim HS, et al. Soybean Thrips (Thysanoptera: Thripidae) Harbor Highly Diverse Populations of Arthropod, Fungal and Plant Viruses. Viruses. 1 déc 2020;12(12):1376. doi:10.3390/v12121376

49. Ottati S, Chiapello M, Galetto L, Bosco D, Marzachì C, Abbà S. New Viral Sequences Identified in the Flavescence Dorée Phytoplasma Vector Scaphoideus titanus. Viruses. 6 mars 2020;12(3):287. doi:10.3390/v12030287

50. François S, Antoine-Lorquin A, Kulikowski M, Frayssinet M, Filloux D, Fernandez E, et al. Characterisation of the Viral Community Associated with the Alfalfa Weevil (Hypera postica) and Its Host Plant, Alfalfa (Medicago sativa). Viruses. 28 avr 2021;13(5):791. doi:10.3390/v13050791

51. Van Kempen M, Kim SS, Tumescheit C, Mirdita M, Lee J, Gilchrist CLM, et al. Fast and accurate protein structure search with Foldseek. Nat Biotechnol. févr 2024;42(2):243-6. doi:10.1038/s41587-023-01773-0

52. Moi D, Bernard C, Steinegger M, Nevers Y, Langleib M, Dessimoz C. Structural phylogenetics unravels the evolutionary diversification of communication systems in gram-positive bacteria and their viruses. Nat Struct Mol Biol. déc 2025;32(12):2492-502. doi:10.1038/s41594-025-01649-8

53. Wang J, Chitsaz F, Derbyshire MK, Gonzales NR, Gwadz M, Lu S, et al. The conserved domain database in 2023. Nucleic Acids Res. 6 janv 2023;51(D1):D384-8. doi:10.1093/nar/gkac1096

54. Reyes-Proaño E, Griffin AM, Duarte A, Sheng H, Schroeder BK, Murray TD, et al. Afrina barna-like Virus, a Novel Virus Associated with Afrina sporoboliae, the Drop Seed Gall-Forming Nematode. Viruses. 23 juill 2025;17(8):1032. doi:10.3390/v17081032

55. Wright AA, Cross AR, Harper SJ. A bushel of viruses: Identification of seventeen novel putative viruses by RNA-seq in six apple trees. Melcher U, éditeur. PLOS ONE. 13 janv 2020;15(1):e0227669. doi:10.1371/journal.pone.0227669

56. De Miccolis Angelini RM, Raguseo C, Rotolo C, Gerin D, Faretra F, Pollastro S. The Mycovirome in a Worldwide Collection of the Brown Rot Fungus Monilinia fructicola. J Fungi. 6 mai 2022;8(5):481. doi:10.3390/jof8050481

57. Zhou K, Zhang F, Deng Y. Comparative Analysis of Viromes Identified in Multiple Macrofungi. Viruses. 12 avr 2024;16(4):597. doi:10.3390/v16040597

58. Duarte MF, De Andrade IA, Silva JMF, De Melo FL, Machado AM, Inoue-Nagata AK, et al. Metagenomic analyses of plant virus sequences in sewage water for plant viruses monitoring. Trop Plant Pathol. 28 avr 2023;48(4):408-16. doi:10.1007/s40858-023-00575-8

59. Maksimović O, Bačnik K, Rivarez MPS, Vučurović A, Mehle N, Ravnikar M, et al. Virome analysis of irrigation water sources provides extensive insights into the diversity and distribution of plant viruses in agroecosystems. Water Res. févr 2024;249:120712. doi:10.1016/j.watres.2023.120712

60. Shi M, Lin XD, Tian JH, Chen LJ, Chen X, Li CX, et al. Redefining the invertebrate RNA virosphere. Nature. déc 2016;540(7634):539-43. doi:10.1038/nature20167

61. Gorbalenya AE, Siddell SG. Recognizing species as a new focus of virus research. Morrison TE, éditeur. PLOS Pathog. 4 mars 2021;17(3):e1009318. doi:10.1371/journal.ppat.1009318

832 62. Castro C, Smidansky ED, Arnold JJ, Maksimchuk KR, Moustafa I, Uchida A, et al. Nucleic acid
833 polymerases use a general acid for nucleotidyl transfer. Nat Struct Mol Biol. févr
834 2009;16(2):212-8. doi:10.1038/nsmb.1540

835 63. Selosse, Marc-André, Strullu-Derrien, Christine, Martin, Francis M., Kamoun, Sophien, Kenrick,
836 Paul. Plants, fungi and oomycetes: a 400-million year affair that shapes the biosphere. New
837 Phytol. 2015;206(2):501-6. Located at: JSTOR

838 64. Ghabrial SA, Castón JR, Jiang D, Nibert ML, Suzuki N. 50-plus years of fungal viruses. Virology. mai
839 2015;479-480:356-68. doi:10.1016/j.virol.2015.02.034

840 65. Edgar RC. MUSCLE: a multiple sequence alignment method with reduced time and space
841 complexity. BMC Bioinformatics. 19 août 2004;5(1):113. doi:10.1186/1471-2105-5-113

842 66. Guindon S, Gascuel O. A Simple, Fast, and Accurate Algorithm to Estimate Large Phylogenies by
843 Maximum Likelihood. Rannala B, éditeur. Syst Biol. 1 oct 2003;52(5):696-704.
844 doi:10.1080/10635150390235520

845 67. Hanson RM. *Jmol* – a paradigm shift in crystallographic visualization. J Appl Crystallogr. 1 oct
846 2010;43(5):1250-60. doi:10.1107/S0021889810030256

847 68. Stothard P. The Sequence Manipulation Suite: JavaScript Programs for Analyzing and Formatting
848 Protein and DNA Sequences. BioTechniques. juin 2000;28(6):1102-4. doi:10.2144/00286ir01

849 69. Frishman D, Argos P. Knowledge-based protein secondary structure assignment. Proteins Struct
850 Funct Bioinforma. déc 1995;23(4):566-79. doi:10.1002/prot.340230412

851 70. Suyama M, Torrents D, Bork P. PAL2NAL: robust conversion of protein sequence alignments into
852 the corresponding codon alignments. Nucleic Acids Res. 1 juill 2006;34(Web Server):W609-12.
853 doi:10.1093/nar/gkl315

854 71. Darriba D, Taboada GL, Doallo R, Posada D. jModelTest 2: more models, new heuristics and
855 parallel computing. Nat Methods. août 2012;9(8):772-772. doi:10.1038/nmeth.2109

856 72. Revell LJ. phytools: an R package for phylogenetic comparative biology (and other things).
857 Methods Ecol Evol. avr 2012;3(2):217-23. doi:10.1111/j.2041-210X.2011.00169.x

858 73. Katoh K, Standley DM. MAFFT Multiple Sequence Alignment Software Version 7: Improvements
859 in Performance and Usability. Mol Biol Evol. 1 avr 2013;30(4):772-80.
860 doi:10.1093/molbev/mst010

861 74. Crooks GE, Hon G, Chandonia JM, Brenner SE. WebLogo: A Sequence Logo Generator: Figure 1.
862 Genome Res. juin 2004;14(6):1188-90. doi:10.1101/gr.849004

863

Tables

Table 1: Composition of the order *Sobelivirales* and species analyzed in detail in this study. Names of viral families and genera were indicated with number of species into brackets. For the species selected in the dataset, virus name, abbreviations and accession numbers were indicated. Viral species used as representative member in this study were in bold.

Family	Genus (N species)	Virus name	Abbr.	Genbank accession number	Uniprot accession nb
Solemoviridae	*Sobemovirus* (26)	Papaya lethal yellowing virus	PLYV	NC018449	J7GBL4
		Rice yellow mottle virus	**RYMV**	AJ608207	Q709L0
		Solanum nodiflorum mottle virus	SNMoV	NC033706	A0A1P7XK80
		Southern bean mosaic virus	SBMV	AF055887	O72157
		Turnip rosette virus	TRoV	NC004553	Q80QX5
		Xufa yellow dwarf virus	XYDV	ON828429	A0A977KCX8
	Polerovirus (77)	Beet chlorosis virus	BChV	NC002766	A0A385L367
		Beet mild yellowing virus	BMYV	NC003491	Q45GB9
		Cereal yellow dwarf virus RPS	CYDV-RPS	AF235168	Q9IZF5
		Cucurbit aphid-borne virus	CaBYV	NC003688	Q65969
		Potato leafroll virus	**PLRV**	AF453394	Q84837
		Turnip yellows virus	TuYV	NC003743	P09507
	Enamovirus (15)	Alfalfa enamovirus 1	AEV	KU297983	A0A166J1Q2
		Citrus vein enation virus	CVEV	HF679486	S0EQ55
		Grapevine enamovirus 1	GEV	KX645875	A0A1B3B596
		Pea enation mosaic virus 1	**PeMV**	MN497823	A0A6M8PM86
	Polemovirus (1)	Poinsettia latent virus	PnLV	AJ867490	Q5NDM9
	Hubsclerovirus (1)	Hubei sclerotinia RNA virus 1	HuSRV	MK889164	A0A649X039
Barnaviridae	*Barnavirus* (1)	Mushroom bacilliform virus	MBV	U07551	Q9YPD4
Alvernaviridae	*Dinornavirus* (1)	Heterocapsa circularisquama RNA virus	HcRNAV	AB218608	Q3C207

871 **Table 2** : Structure and sequence conservations of sobelivirad RdRps selected in this study.

Viral genera	Sample number	Structural deviation (mean RMSD, Å)*	Sequence conservation (mean pairwise % ID)*
Sobemovirus	6	0.72 ± 0.12 (a)	54.3 ± 3.2 (a)
Poler/movirus	7	0.56 ± 0.16 (b)	66.3 ± 4.4 (b)
Enamovirus	4	0.62 ± 0.23 (a, b)	62.8 ± 15.8 (b)
Barnavirus	1	-	-
Hubsclerovirus	1	-	-

872 * Letters into brackets (a, b) indicate significative differences between groups (Tukey test, $p < 0.05$).

873 ANOVA : F(Structure) = 18.19, $p < 0.001$; F(Sequence) = 10.67, $p < 0.001$.

874

875 **Table 3**: Virus name, accession numbers and references of the five candidate barnaviruses and

876 the three outgroup viruses used for the phylogenetic analysis of sobelivirads

Viral species	Abbr.	NCBI access nb	Uniprot accession nb	Reference
Colobanthus quitensis associated barnavirus 1	CoBV	MG686618	A0A2P1JMX2	(Nibert et al 2018)
Rhizoctonia solani barnavirus 1	RsBV	KP900904	A0A0M4KGX8	(Marzano et al 2016)
Tulasnella barnavirus 1	TuBV	MN738552	QNN89176	(Sutella et al 2020)
Barnaviridae sp.	tBV1	MZ218174	UHS71527	(Chen et al 2022)
Barnaviridae sp.	tBV2	MZ218185	UHS71731	(Chen et al 2022)
Soybean thrips sobemo-like virus 5	SoTV5	MW023867	QPZ88400	[48]
Scaphoideus titanus sobemo-like virus 1	ScTV1	MN982389	QIJ56913	[49]
Hypera postica associated sobemovirus 1	HyPV1	MW676135	QUS52861	[50]

877

878 **Figure captions**

879

880 **Figure 1:** (A) RdRp 3D model of RYMV generated using Alphafold 2.2 colored by per-residue

881 model confidence score (local distance difference test, plDDT). The regions modeled with high

882 confidence (plDDT>90) are shown dark blue. The regions colored light blue (plDDT>70) were

883 modeled with confidence. (B) RYMV RdRp domains. The palm and thumb domains are colored
884 on RYMV model, grey and pink. The index, picky and other fingers are colored magenta, deep
885 purple and white, respectively. (C & D) RdRp structures of picornaviruses (PDB 1RA6,
886 Poliovirus and PDB 1U09, Foot-and-mouth disease virus) are compared to RYMV model and
887 colored based on root mean square deviations (RMSD) to RYMV RdRp model. The regions
888 with high local structural similarity (RMSD<1Å) are shown dark blue. The regions colored
889 light blue (1<RMSD<2) and green (2<RMSD<3) were similar whereas colored other colors,
890 similarity was low. The picornavirus feature, which was the index finger interacted with the top
891 of the thumb trapping the ring finger beneath it, was conserved (black arrow) whereas the α-
892 helix of picornaviruses located at the carboxy-terminal extremity (thumb domain) was lacking
893 in RYMV predicted structure (shown red). The orientation of the pinky finger in the RYMV
894 RdRp model was closer to FMDV rather than PV RdRp structure (grey arrows). Dynamic view
895 and alignment available at https://pat.cbs.cnrs.fr/sobemo/rmsd.html (E) Structure-guided
896 alignment of RdRp sequences of RYMV and the two picornaviruses PV and FMDV using
897 mTM-align. Strictly conserved and similar residues are highlighted (black and grey,
898 respectively), RdRp conserved motifs and homomorphs are outlined (black and grey,
899 respectively). The thumb, index, picky finger domains are outlined pink, magenta and blue
900 violet, respectively.

901 **Figure 2** : (A) Superimposition of sobemoviral RdRp 3D models (26 species) colored using
902 standard rainbow code (from N-terminus, dark blue, to C-terminus, red). The models showed
903 high reliability (plDDT scores >90) and high local structural similarity (RMSD to RYMV <1Å).
904 Major variable segments at the intra-genera scale are pointed out (grey arrows). Dynamic view
905 and alignment available at https://pat.cbs.cnrs.fr/sobemo/sobemo_v7/sobemo B) Sobemoviral
906 maximum-likelihood phylogenetic tree obtained using the structure-guided mTM-alignment
907 and the GTR+G+I substitution model in SeaView v5.05. Viral species are colored when they

908 belong to lineages with common plant host families as reported in Ghafari et al 2023 (orange,
909 *Fabaceae*, green, *Poaceae,* blue, *Solanaceae* and pale green, *Poales*). Bootstrap supports based
910 on 100 replicates were indicated if up or egal to 75%.

911 **Figure 3 :** (A-D) RdRp 3D models of genus type members colored by per-residue model
912 confidence score (local distance difference test, plDDT). The regions modeled with high
913 confidence (plDDT>90) are shown dark blue. The regions colored light blue (plDDT>70were
914 modeled with confidence whereas yellow (plDDT>50) and orange (plDDT<50), confidence
915 was low to very low. PLRV (Polerovirus), PeMV (Enamovirus), MBV (Barnavirus), HuSRV
916 (Hubsclerovirus). (E-H) Superimposition of RdRp 3D models by genera colored using standard
917 rainbow code (from N-terminus, dark blue, to C-terminus, red). Poler/movirus (5 species),
918 Enamovirus (4 species), Barnavirus (1 species), Hubsclerovirus (1 species). Structural features
919 at the intergeneric level are pointed out (black arrows see Fig 4). Dynamic view and alignment
920 available at https://pat.cbs.cnrs.fr/sobemo/sobemo_v7/sobeli

921 **Figure 4** : (A) Inter-genera superimposition of RdRp 3D models. For clarity reason, only one
922 model per genera was represented (sobemovirus RYMV, poler/movirus PLRV, enamovirus
923 PeMV, barnavirus MBV, hubsclerovirus HuSRV). The thumb, index, picky finger domains
924 were colored pink, magenta and blue violet, respectively. The conserved motifs A-B-C
925 localized at the center of the channel were colored black. Structural features of the different
926 genera were colored using the rainbow code (as reported in Fig 2 and 3) and pointed out with
927 black arrows. Dynamic view and alignment available at
928 https://pat.cbs.cnrs.fr/sobemo/sobemo_v7/sobeli/ (B) Structure-guided alignment of the 5
929 sobelivirad RdRp sequences using mTM-align. Strictly conserved and similar residues are
930 highlighted (black and grey, respectively), RdRp conserved motifs are outlined in black. The
931 thumb, index, picky finger domains are outlined pink, magenta and blue violet, respectively.

932 Color code of the structural features of each genera was the same as above. Alignment available
933 at https://pat.cbs.cnrs.fr/sobemo/sobemo_v7/ali19.html

934 **Figure 5** : (A-B) Structure-guided alignment of the 27 sobelivirad RdRp sequences using mTm-
935 align. Strictly conserved and similar residues are highlighted (black and grey, respectively),
936 RdRp conserved motifs are outlined in black. The thumb, index, picky finger domains are
937 outlined pink, magenta and blue violet, respectively. Color code of the structural features of
938 each genera was the same as above. Dynamic view and alignment available at
939 https://pat.cbs.cnrs.fr/sobemo/sobemo_v7/sobeli

940 **Figure 6 :** (A) Sobelivirad maximum-likelihood phylogenetic tree obtained using the structure-
941 guided mTm-alignment and the GTR+G+I substitution model in SeaView v5.05. Viral species
942 are colored according to the five genera and the outgroups. Two main clades, namely Sobarna-
943 vs. Polena-clades, were revealed clustering sobemo- and barnaviruses vs. poler/mo- and
944 enamoviruses, respectively. Bootstrap supports based on 100 replicates were indicated if up or
945 egal to 75%. (B) PoW-transformed time tree of the sobelivirads. The horizontal blue bars show
946 95% highest posterior density of the node age.

947 **Figure 7 :** (A) Sobelivirad-specific and clade-specific residues of RdRp motifs located on 3D
948 models. Remarkable residues were figured as sticks on Cα trace of one representative species
949 (sobarna, RYMV and polena-, AlEV, respectively). Sobelivirad-specific, Sobarna- and polena-
950 specific residues were colored magenta, red & blue, respectively, except clade-specific
951 polymorphic positions colored yellow (arrows). Remarkable residues at positions 263, 489 or
952 in interaction 365-486 were indicated by arrows or circled in dotted line (see text). (B)
953 Optimized motifs of sobelivirad RdRps. (a) For each conserved motif, RdRp conserved residues
954 were colored (black), additional sobelivirad-specific residues (magenta), sobarna-specific
955 residues (red) and polena-specific residues (blue). Common clade-specific positions were

956 highlighted (yellow open boxes). Each sequence logo consists of stacks of symbols, one stack
957 for each position in the sequence. The overall height of the stack indicates the sequence
958 conservation at that position, while the height of symbols within the stack indicates the relative
959 frequency of each amino at that position. Position numbering referred to the alignment with 19
960 sequences.

961 **Supporting data**

962 **S1 Table :** Taxonomy and biological characteristics in the phylum Pisuviricota.

963 **S2 Table :** Dataset of the 26 sobemoviral sequences.

964 **S3 Figure** : RdRp 3D models of sobemoviruses colored by per-residue model confidence score
965 (local distance difference test, plDDT).

966 **S4 Figure** : Structure-guided alignment of RdRp sequences of the 26 sobemoviruses using
967 mTm-align.

968 **S5 Figure** : RdRp 3D model of pole/moviruses, enamoviruses and dinornavirus colored by per-
969 residue model confidence score (local distance difference test, plDDT).

970 **S6 Figure** : (A-B) Structure-guided alignment of RdRp sequences of the 19 sobelivirads using
971 mTm-align.

972 **S7 Table** : Dataset of 31 sequences from NCBI Virus database, tagged as Barnaviridae and
973 containing both protease and sobelivirad RdRp conserved motifs.

974 **S8 Table** : Dataset of 55 sequences from NCBI Virus database, tagged as unclassified sobeli-,
975 solemo- and sobemo-like and containing both protease and RdRp conserved motifs

976 **S9 Figure** : (A-B) RdRp 3D models of candidate barnaviruses and outgroups colored by per-
977 residue model confidence score (local distance difference test, plDDT). (C-D) Superimposition
978 of RdRp 3D models of barnaviruses and outgroups.

979 **S10 Figure** : Structure-guided alignment of RdRp sequences of the 6 barnaviruses, 3 outgroups
980 and 4 sobelivirads representative of each other genera.

S11 Figure : Phylogenetic tree obtained using Foldtree.

S12 Table : Additionnal 99 sequences (genus, name species and accession numbers).

Figure 1

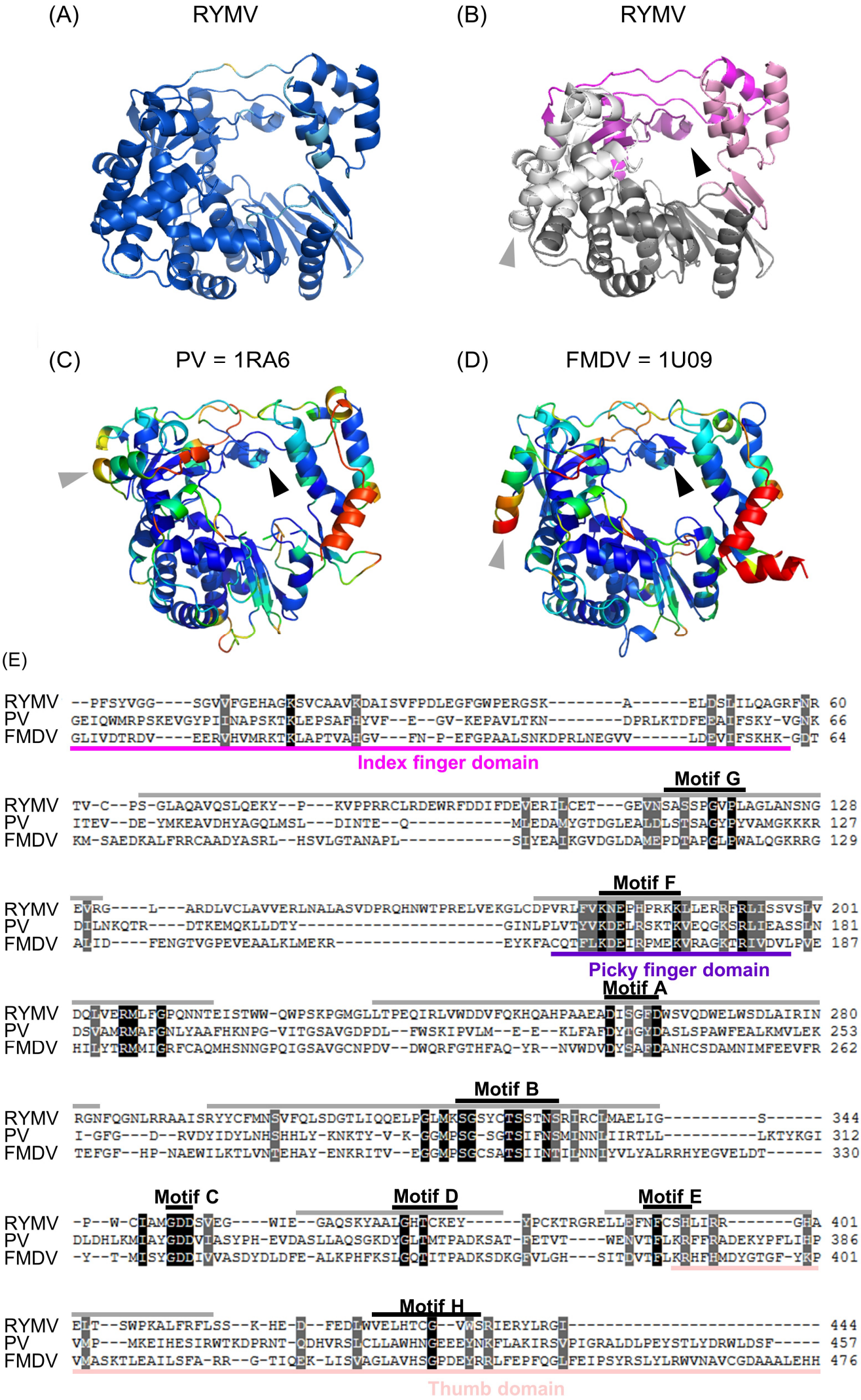

(A)
RYMV
(B)
RYMV
(C)
PV = 1RA6
(D)
FMDV = 1U09
(E)
RYMV --PFSYVGG----SGVVFGEHAGKSVCAAVKDAISVFPDLEGFGWPERGSK--------A------ELDSLILQAGRFNR 60
PV GEIQWMRPSKEVGYPIINAPSKTKLEPSAFHYVF--E--GV-KEPAVLTKN--------DPRLKTDFEEAIFSKY-VGNK 66
FMDV GLIVDTRDV----EERVHVMRKTKLAPTVAHGV---FN-P-EFGPAALSNKDPRLNEGVV------LDEVIFSKHK-GDT 64
Index finger domain
Motif G
RYMV TV-C--PS-GLAQAVQSLQEKY--P---KVPPRRCLRDEWRFDDIFDEVERILCET---GEVNSASSPGVPLAGLANSNG 128
PV ITEV--DE-YMKEAVDHYAGQLMSL---DINTE--Q----------MLEDAMYGTDGLEALDLSTSAGYPYVAMGKKKR 127
FMDV KM-SAEDKALFRRCAADYASRL--HSVLGTANAPL----------SIYEAIKGVDGLDAMEPDTAPGLPWALQGKRRG 129
Motif F
RYMV EVRG----L---ARDLVCLAVVERLNALASVDPRQHNWTPRELVEKGLCDPVRLFVKNEPHPRKKLLERRFRLISSVSLV 201
PV DILNKQTR----DTKEMQKLLDTY----------GINLPLVTYVKDELRSKTKVEQGKSRLIEASSLN 181
FMDV ALID----FENGTVGPEVEAALKLMEKR----------EYKFACQTFLKDEIRPMEKVRAGKTRIVDVLPVE 187
Picky finger domain
Motif A
RYMV DQLVERMLFGPQNNTEISTWW-QWPSKPGMGLLTPEQIRLVWDDVFQKHQAHPAAEADISGFDWSVQDWELWSDLAIRIN 280
PV DSVAMRMAFGNLYAAFHKNPG-VITGSAVGDPDL--FWSKIPVLM--E-E--KLFAFDYTGYDASLSPAWFEALKMVLEK 253
FMDV HILYTRMMIGRFCAQMHSNNGPQIGSAVGCNPDV--DWQRFGTHFAQ-YR--NVWDVDYSAFDANHCSDAMNIMFEEVFR 262
Motif B
RYMV RGNFQGNLRRAAISRYYCFMNSVFQLSDGTLIQQELPGLMKSGSYCTSSTNSRIRCLMAELIG----------S------ 344
PV I-GFG---D--RVDYIDYLNHSHHLY-KNKTY-V-K-GGMPSG-SGTSIFNSMINNLIIRTLL----------LKTYKGI 312
FMDV TEFGF--HP-NAEWILKTLVNTEHAY-ENKRITV--EGGMPSGCSATSIINTILNNIYVLYALRRHYEGVELDT------ 330
Motif C
Motif D
Motif E
RYMV -P--W-CIAMGDDSVEG----WIE--GAQSKYAALGHTCKEY------YPCKTRGRELLEFNFCSHLIRR-------GHA 401
PV DLDHLKMIAYGDDVIASYPH-EVDASLLAQSGKDYGLTMTPADKSAT-FETVT----WENVTFLKRFFRADEKYPFLIHP 386
FMDV -Y--T-MISYGDDIVVASDYDLDFE-ALKPHFKSLGQTITPADKSDKGFVLGH---SITDVTFLKRHFHMDYGTGF-YKP 401
Motif H
RYMV ELT--SWPKALFRFLSS--K-HE-D--FEDLWVELHTCG--VWSRIERYLRGI---------- 444
PV VMP---MKEIHESIRWTKDPRNT-QDHVRSLCLLAWHNGEEEYNKFLAKIRSVPIGRALDLPEYSTLYDRWLDSF----- 457
FMDV VMASKTLEAILSFA-RR--G-TIQEK-LISVAGLAVHSGPDEYRRLFEPFQGLFEIPSYRSLYLRWVNAVCGDAAALEHH 476
Thumb domain

Figure 2

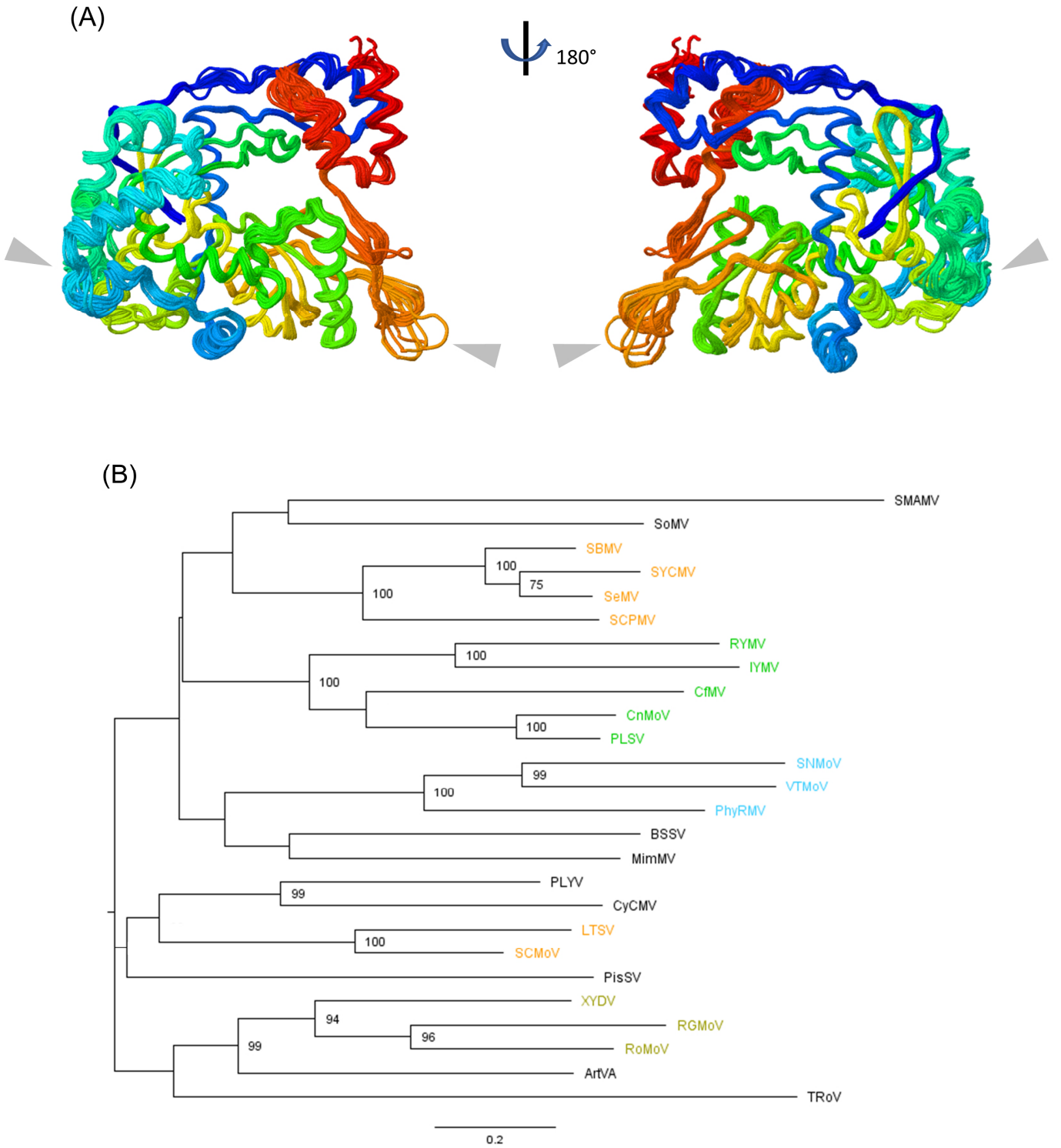

(A)
180°
(B)
SMAMV
SoMV
SBMV
SYCMV
SeMV
SCPMV
RYMV
IYMV
CfMV
CnMoV
PLSV
SNMoV
VTMoV
PhyRMV
BSSV
MimMV
PLYV
CyCMV
LTSV
SCMoV
PisSV
XYDV
RGMoV
RoMoV
ArtVA
TRoV
100
75
100
100
100
100
99
100
99
100
94
96
99
0.2

Figure 3

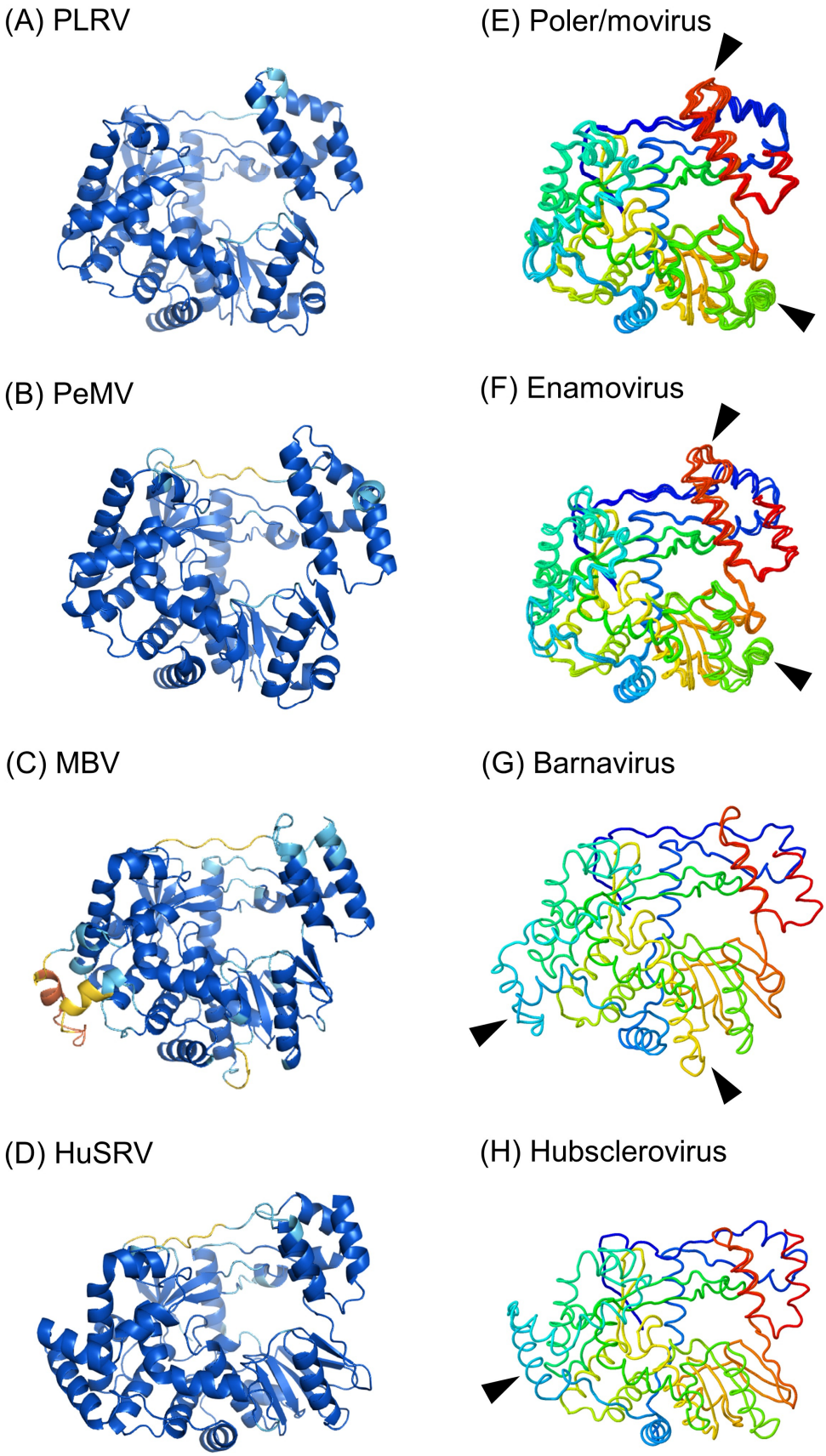

(A) PLRV
(E) Poler/movirus
(B) PeMV
(F) Enamovirus
(C) MBV
(G) Barnavirus
(D) HuSRV
(H) Hubsclerovirus

Figure 4

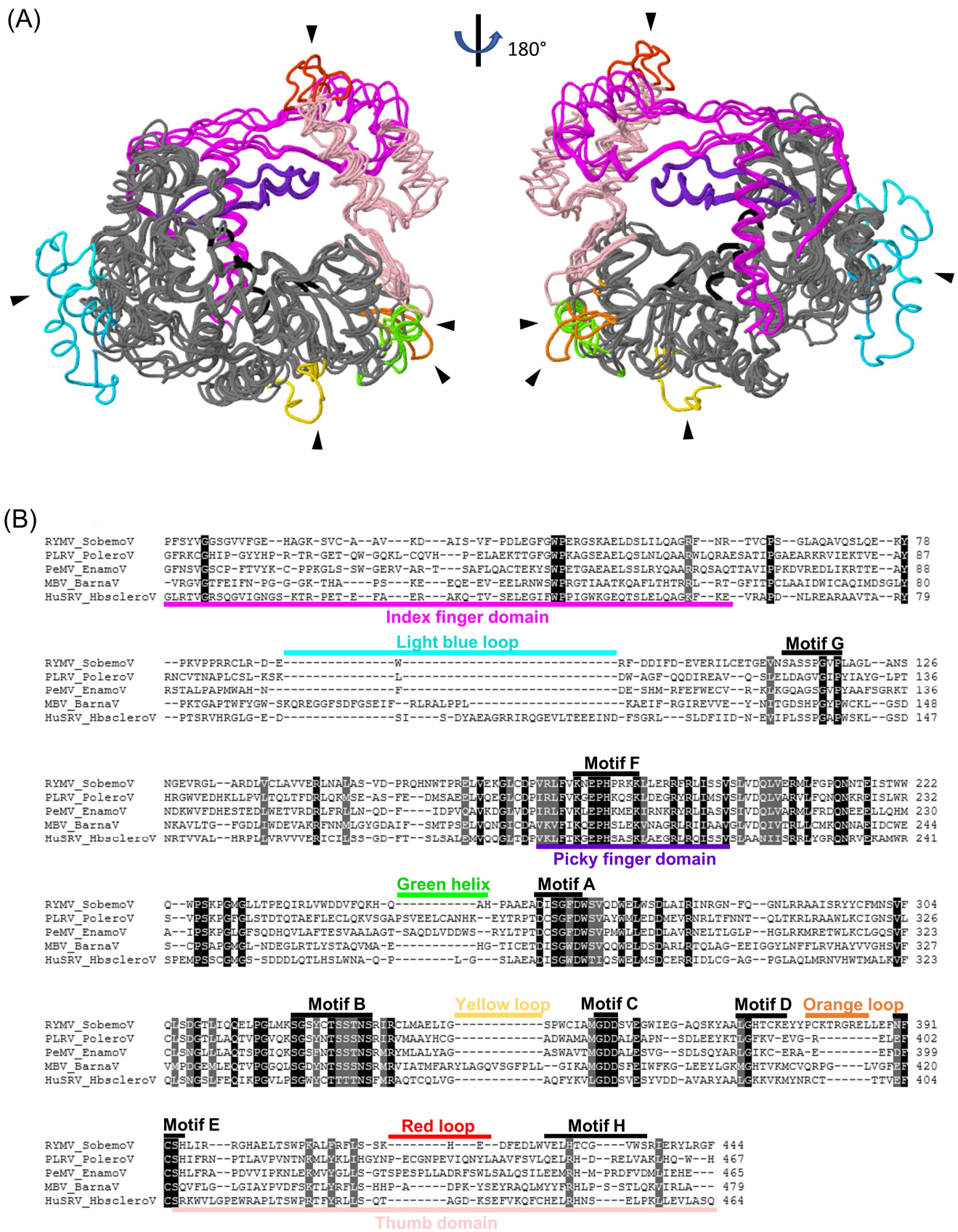

(A)
180°
(B)
Index finger domain
Light blue loop
Motif G
Motif F
Picky finger domain
Green helix
Motif A
Motif B
Yellow loop
Motif C
Motif D
Orange loop
Motif E
Red loop
Motif H
Thumb domain
RYMV_SobemoV
PLRV_PoleroV
PeMV_EnamoV
MBV_BarnaV
HuSRV_HbscleroV

Figure 5A

```
tBV2_BarnaV      NLKFVGKFGRSAG----GRFQ--GHG--Q----VD---GA---DRL-AY-PQLNDWGWPSRDGEAARTSFGIHAGKF--KRE--REVPCDAQRKQAIDAV 76
RsBV_BarnaV      YLRCVGSFKSPTG----GNQG--SMP--C----PR---GV---ERR-IY-DKLSNWRWPRKENAAVKKSLEIHAEGL--RREEGRRVPSRTEREAVLKVI 78
TuBV_BarnaV      ALSFVGEFCRGVD----SRKG--KHG--K----PT---HY---ERS-CH-PGLNNWWWPERNAGATNRALHLHADAL--AVK-STRDPTDDEIEAVVQAI 77
tBV1_BarnaV      PLAKVGKFSAATG----KGGQ--HKH--A----PT---WT---EKR-KY-RALFNWAWPTQTAEGVYHSLVLHSRGL--GK--GSRRPTAAEEEAVTQAL 76
MBV_BarnaV       -VRGVGTFEIFNP-----GGG--KTH--A----PS---KE---EQE-EV-EELRNWSWPRGTIAATKQAFLTHTRRL--RTG--FITPCLAAIDWICAQI 74
CoBV_BarnaV      NLRWSGKFKGAKG----GAKQ--KSL--G----PD---KE---ERK-EH-PGLLNWSWVPGFDQHQRRSLGLHADRL--GR--ATWIPNQEFRNVLARGL 76
HuSRV_HbscleroV  GLRTVGRSQGVIGNGSK------TR-PET-E--FA---ER---AKQ-TV-SELEGIFWPPIGWKGEQTSLELQAGKF--KE---VRAPD--NLREARAAV 75
TroV_SobemoV     GLKEVGKSGVTFS----E-P-KKK--GTK-A--IT---EA---FSR-EI-PEVANYAWPQRGSKAERKSLYLQASRF--RR---TEEPN--GIGEVVERL 74
XYDV_SobemoV     PLFARGRSGVKFR----E-S-GRT--VVG-S--AV---RD---AIQ-HF-PTLSEFDWPERGTKAEIGSLLLQASRF--RP---TEAPS--NLPQATAEL 74
RYMV_SobemoV     PFSYVGSGVVFG----E-H-AGK--SVC-A--AV---KD---AIS-VF-PDLEGFGWPERGSKAELDSLILQAGRF--NR---TVCPS--GLAQAVQSL 74
PLYV_SobemoV     PLTYVGRSGVKFR----E-T-GRS--PID-Q--RV---TN---ALK-HL-PELSEFGWPERGSEAELKSLKIQAERF--RQ---TGSPE--GLEEACRRL 74
SNMoV_SobemoV    LFNSVGISTCKFR----E-GSK-HAS--R-G--DV---AA---AQR-VF-SELNEYGWPERGSRAEKHSLFLQASRH--RR---VPKPK--GLEEACREL 74
SBMV_SobemoV     PLRFVGQSACKFR----E-L-CRK--DTP-D--EV---LR---ATR-VF-PELSDFSWPERGSKAELHSLLLQAGKF--NP---TGIPR--NLEGACQNL 74
PnLV_PolemoV     GFRPCGKIPKFYR----P-R-K-R--RVS-EW-GQNL-ARKH---S-SLGEITQGFGWPEAGAEAELRSLRLQAQRWLERSK-SSVIPSAIEREIVISRL 83
CYDV_PoleroV     GFQRVGKLSPQYY----P-R-A-R--STT-AW-GERL-CAEH---P-LLGEKTKGFGWPAVGATAELTSLRLQAARWLERSE-SAKIPSDAARKNVIDRT 83
PLRV_PoleroV     GFRKCGHIPGYYH----P-R-T-R--GET-QW-GQKL-CQVH---P-ELAEKTTGFGWPKAGSEAELQSLNLQAARWLQRAE-SATIPGAEARKRVIEKT 83
TuYV_PoleroV     GFRHCGKLPQYYH----P-K-Q-K--EES-SW-GKTL-VGNH---P-ALGEKTSGFGWPKFGPEAELKSLRLQASRWLERAQ-SAEIPSDAERERVIQKT 83
BChV_PoleroV     GFEEVGSLPKFYH----S-K-Q-R--LNS-EY-GVRV-AEEY---S-ELKELTAGFGWPKFGAQAELTSLRLQADRWLQRME-SAQIPSTEGEESVINRL 83
BMYV_PoleroV     GFIKVGELPKFYF----S-K-Q-K--GCS-DW-GTKL-TSLH---P-ELEEKTRGFGWPKFGPAAELKSLRLQAARWLERAE-QVKIPSTEERERVIEKC 83
CaBYV_PoleroV    GFDKVGSLPQFYH----A-K-Q-K--KSS-NW-GDKI-CKQH---P-EMGDLTKGFGWPQFGAKAELKSLRLQAARWLERAQ-SVKIPSSEEREHVIERC 83
GREnV_EnamoV     GFKWAGRCKANFW----P-A-Q-GT-DWS-EW-GARK-LE-A---SEWLRRKVEGYGWPSFGAEAELRSLRLQAARRAARSK-SAVLPSNEKRERVIKET 84
CVEV_EnamoV      GFCKVGSCDFTKW----K-S-P-TK-ELS-EW-GFHR-AQ-A---STLLQACQESYGWPDTGAEAELSSLRYQAAKRAAAQT-KALTPSKAARWSVIQRT 84
AlEV_EnamoV      GFNAVGSCPFTVY----K-C-P-PK-GLS-RW-GERV-AR-A---SSFLQACTEKYSWPETGAEAELSSLRYQAARRQSAQT-TAVIPPKDVREDLIKRT 84
PeMV_EnamoV      GFNSVGSCPFTVY----K-C-P-PK-GLS-SW-GERV-AR-T---SAFLQACTEKYSWPETGAEAELSSLRYQAARRQSAQT-TAVIPPKDVREDLIKRT 84

ScTV1_OutGroup   YFTFEGQESPIPV----F-G-G-E--PVEN-DR---IP--AEYR-D-TLKKVSSGFGWPKFGHNAELISLLRHGSLRKDL---RSDGPSAQVRTLIFMAM 80
SoTV5_OutGroup   YVDFLAKESRIYV----R-K-N-F--PERY-P----WN--KD-V-S-LMEDLAKEYDWPDLGHVAEIRSLTKHLPLRDSV---AVDTIDPDQRVIVEEEV 78
HyPV1_OutGroup   -FEPVGKKPGLFQ----R-V-S-Q--PVEF-P----YP--EAIK-A-ELRDMSAGLTWPNNDYGTEMLSFLTHARIRQQV---RDQPMDQGLKSRLIEVL 78
```

Index finger domain

Light blue loop

```
tBV2_BarnaV      CRQGS-EY--PKTRTPT-W-LTRV-CG--DRR--------WAE-------G-EVKF----------SRQM-F-EGVGGLREKIE--R-K 126
RsBV_BarnaV      RE--GYEY--PNTRVPA-G-FS---RG--EVEG-------W--A--E-G-R--YEV----------TV-EDF-E---SVKDSIQ--R-D 123
TuBV_BarnaV      TK--V--Y--PKTRVPS-S-FL---NG--ESCRG------P--E--EYG-N--HFF----------TQ-ESF-E---PIKSLIE--K-D 122
tBV1_BarnaV      LS--SGEY--PITSVPH-L-FL---GG--EA-G-------Y--E------G--YQL----------KG--VP-E---VTRRAVE--K-D 117
MBV_BarnaV       MD--SGLY--PKTGAPT-W-FYG--WSKQREGGFSDFGS----EIFRLRALP--PL----------KA-EIF-R---GIREVVE--Y-N 131
CoBV_BarnaV      LD--SQQY--PKTTPPG-D-VE---WK-------------T------------GRT----------R--IL-A---SRRLTVE--R-D 111
HuSRV_HbscleroV  TA--R--Y--PTSRVHR-G-LG---E--D----------SI------------SDYAEAGRRIRQGEVLTEEEIND-FS-GRL------SLDFIID-N-E 130
TroV_SobemoV     VK--E--Y--PTTNCPP-E-FK---G--R----------W-----------------------------------DY--QDIFE----YVTDVAR-SSD 108
XYDV_SobemoV     VK--E--Y--PKIRQRS-I-LQ---HR-E----------W-----------------------------------DK--EEVQK----AISEIAQ-R-Q 108
RYMV_SobemoV     QE--K--Y--PKVPPRR-C-LR---D--E----------W-----------------------------------RF--DDIFD----EVERILCETGE 109
PLYV_SobemoV     KD--R--Y--PPARPRR-C-LR---G--E----------W-----------------------------------DE--RETAT----EVAKICE-S-S 107
SNMoV_SobemoV    EK--L--Y--PRSRAKQ-C-FR--KQT-F----------L-----------------------------------DD--ELLLK----AIKETAF-SPE 110
SBMV_SobemoV     LE--R--Y--PASKSCY-C-LR---GE-A----------W-----------------------------------SF--DAVYE----EVCKKAQ-SAE 109
PnLV_PolemoV     VE--S--YKICRSEAPL-C-SS---GS-D----------L-----------------------------------SW--KGF-L---EDFREAV--S-S 118
CYDV_PoleroV     VQ--A--YSNCKTNVPR-C-TR---D--Q----------L-----------------------------------NW--DDF-R---IDFLEAI--K-S 117
PLRV_PoleroV     VE--A--YRNCVTNAPL-C-SL---KS-K----------L-----------------------------------DW--AGF-Q---QDIREAV--Q-S 118
TuYV_PoleroV     AD--V--YHPCQTNGPA-A-TR---GG-T----------L-----------------------------------TW--NNF-M---IDFKQAV--F-S 118
BChV_PoleroV     VE--A--YSCARTNAPT-S-TA---GN-S----------L-----------------------------------TW--EGF-I---EDIKEAV--S-S 118
BMYV_PoleroV     VE--A--RSPTQTRGPM-A-TR---GS-K----------L-----------------------------------SW--NNF-L---EDFKTAV--F-S 118
CaBYV_PoleroV    CR--A--YQAAKTNGPM-A-TR---GD-R----------L-----------------------------------SW--DNF-L---QDFKQAV--L-S 118
GREnV_EnamoV     VS--E--YAHTAQRSPD-W-IR--NG-------------L-----------------------------------NW--RLA-R---LYFGECL--N-S 118
CVEV_EnamoV      CV--A--YGVTRRHAPR-W-TR---T-------------M-----------------------------------NR--ADA-R---FYFCEAV--R-S 117
AlEV_EnamoV      TE--A--YRSTALPAPM-W-AH---N-------------F-----------------------------------DE--SHM-R---FEFWECV--R-K 117
PeMV_EnamoV      TE--A--YRSTALPAPM-W-AH---N-------------F-----------------------------------DE--SHM-R---FEFWECV--R-K 117

ScTV1_OutGroup   AK--W--YERARWSIPDDFM-----------------------------------------------------------Q-F---SHFQRAV--A-R 107
SoTV5_OutGroup   TK--L--ASAALTVLPHDFM-----------------------------------------------------------S-R---SHFDRVI--A-D 105
HyPV1_OutGroup   LE--R--LNPIRYKIPDDFL-----------------------------------------------------------E-Y---THFKRVI--H-G 105
```

Motif G

Motif F

```
tBV2_BarnaV      IVKDSTPGYEYAVL--GANNALVL-E------R--YGNTHWDIVCDRLNRMVEYTGNY----H-DLTPVELVKLGFCDPVKVFWKQEPHNKKKLVEGRVR 210
RsBV_BarnaV      INKDSTPGYEYVAL--GGKNKQVM-E------N--YGSFHWNTVAQSFNNALRLGD--KV--F-SMSPSELVKAGVCDVVKVFIKQEPHSLKKIESGKLR 207
TuBV_BarnaV      VNKDSSPGIELMTL--GSTNKIVM-E------K--YNSEHWQSVVDQFNAALHYGE--DV--F-TMSPQELIQKGICHPVRAFWKQEPHSAKKITSGKLR 206
tBV1_BarnaV      INMDSTPGYEHVAL--GSSNEKVL-E------K--YGNVHWDTVLEQFNNALRLGE--EV--F-SLSPTELIQFGICDPVKVFWKQEPHSRAKLDAGKLR 201
MBV_BarnaV       ITGDSHPGYEWCKL--GSDNKAVL-T------G--FGDLHWDEVAKRFNNMLGYGD--AI--F-SMTPSELVQNGICDAVKVFIKQEPHSLEKVNAGRLR 215
CoBV_BarnaV      VTKDSSPGYEWNLL--GDTNAEVM-T------T--NGTLVWSEVERRMNYIIDNFE--AV--M-EMGPEELVRNGACDPVKLFIKQEPHSQKKIVQGFLR 195
HuSRV_HbscleroV  VIPLSSPGAEWSKL--GSDNRTVV-A------L--HRPLLVRVVVERICILSSGD---F---T-SLSALEMVQQGLTDPVKLFTKQEPHSASKLAEGRLR 212
TroV_SobemoV     INGKASPGVELSSI--AAKNEVLV-T------R--HLDFLVHAVVQRLFHLSDEI---LP--E-KPSPEWLVSQGYCDPVRVFVKQEPHPLRKLDEGRVR 191
XYDV_SobemoV     VNRDASPGVELAIL--GTKNGVVL-D------Q--HHDLLVEAVWARLEALAQAD---L---SPQQDPVSLVVNGLCDPVRLFVKQEPHPKRKLREGRFR 191
RYMV_SobemoV     VNSASSPGVELAGL--ANSNGEVR-G------L--ARDLLVCLAVVERLNALASVD---PRQH--NWTPRELVEKGLCDPVRLFVKNEPHPRKKLLERRFR 193
PLYV_SobemoV     VNQRASPGVELSVF--GSSNGAVL-D------K--HRDLLVTAVVARLNLLSEKI---I-----DYDPVQLVQFGYCDPVRLFVKQEPHSLKKITEGRFR 188
SNMoV_SobemoV    INDKASPGSEWSKL--AERNGLLL-N------S--FGDFLVEAVRQRILKLAAWDPT-DL--L-MLKPSAMVRMGLVDPARVFVKQEPHTNKKLSEGRYR 195
SBMV_SobemoV     INEKASPGVELSRL--ASTNKDLL-K------R--HLELLVALCVTERLFLLSEAE---DL--L-DESPVDLVRRGLCDPVRLFVKQEPHASRKVREGRFR 192
PnLV_PolemoV     LELDAGIGVEYIGYG-YPTHRGWV-E------NPRLLPVLSRLVYARLQRLATL----SV--D-GKTPEELVRDGLVDPVRVFVKQEPHKQSKLDEGRYR 203
CYDV_PoleroV     LQLDAGVGIEMITAG-LPTHRGWV-E------DPDLLPVLARLTFDRLLTMSKA----SL--E-TRSPEQLVKENLCDPIRLFVKQEPHKQSKLDEGRYR 202
PLRV_PoleroV     LELDAGVGIEYIAYG-LPTHRGWV-E------DHKLLPVLTQLTFDRLQKMSEA----SF--E-DMSAEELVQEGLCDPIRLFVKGEPHKQSKLDEGRYR 203
TuYV_PoleroV     LEFDAGIELEYIAYG-KPTHRGWV-E------DQKLLPILAQLTFFRLQKMLEV----NF--E-DMGPEELVRNGLCDPIRLFVKGEPHKQAKLDEGRYR 203
BChV_PoleroV     LELDAGVGVEYIAYG-TRTHRDWV-F------NQELLPVLTRLTFNRLQKMLEV----NS--D-DLNAEQLVQHGLCDPIRVFVKGEPHKVSKLEEGRYR 203
BMYV_PoleroV     LELEAGVGVEYVAYG-RRTHRGWI-E------DPDLLPVLARFTFDRLQKLSEA----KF--E-HMSPEQLVQEGLCDPIRLFVKGEPHKQSKLDEGRYR 203
CaBYV_PoleroV    LEMDAGIGVEYIAYG-KPTHRGWV-E------DKKLLPILARLTFNRLQKMLEV----RY--V-DLSPAELVRRGLCDPIRVFVKGEPHKQAKLDEGRYR 203
GREnV_EnamoV     MTMDSGAGVEYASFLNRKINQKWL-D------SEDKVEAVWDLVRARLERMLTY----SW-----VSPEQAVLDGVCDPVRVFVKGEPHKVAKLHEGRFR 202
CVEV_EnamoV      LKPDSGSGLEFAAFYNRTTHSDWC-Y------DDASFEAVFDLVWARLGRLRSG----SF-----RNPVQAVQDGLCDPIRLFVKLEPHKRAKIEQKRFR 201
AlEV_EnamoV      LKGQAGSGVEYAAFSNRRTNDKWV-F------DHESTEDVWETVRDRLFRLLNG----EF-----IDPVQAVKDGLVDPIRLFVKLEPHKMEKIRNKRYR 201
PeMV_EnamoV      LKGQAGSGVEYAAFSGRKTNDKWV-F------DHESTEDVWETVRDRLFRLLNQ----DF-----IDPVQAVKDGLVDPIRLFVKLEPHKMEKIRNKRYR 201

ScTV1_OutGroup   LDWTSSPGYEYML-R-HVNNGQFFSCKD-GVPSQERLDAVWSMVQQRLLKG------------------------DADPIRLFVKPEAHSEKKLKSEKYR 180
SoTV5_OutGroup   LDMNSSPGFEYLY-E-YTTNKEFFGARD-GVKDEARVEAVWSKVLSRIQAR------------------------DSDPIRLFIKPEPHKKKLVEGRLR 178
HyPV1_OutGroup   INRQASPGYEYLL-Q-YTTNEQMFCIGEDGKIPEYICKEYYDMIMERIQDE------------------------TYDPIRVFIKPEAITKKLYYQHRER 179
```

Picky finger domain

Figure 5B

Green helix
Motif A

```
tBV2_BarnaV      LISSVSLVDQIIPRMLHTTQNEWPIEHWPT--CASKPGMGL-HDDGMKVLSSNIQELLD------------LSGENMCTDISGWDWSVPEWLLHDDMECR 295
RsBV_BarnaV      LISSVSLVDQIKTRLLCRTQNLNPIESWES--CPSKPGLGL-HDDGLQVIAANIKRFLS------------Q-GVVAEADVSGWDWSVQDWELEFDAECR 291
TuBV_BarnaV      IHAAVSLIEQLKTRLLCSTQNKMPIACWQS--CPSKPGLGL-HDEGLFVIQQNMKEILE------------H-GELASTDIGSWDWSVKMWELNADAERR 290
tBV1_BarnaV      IISSVSLVDQIKTRLLCSQQNKAPINSWQT--CPSKPGLGL-HDDGLASIADTIKDFLS------------H-GKVIETDVSGWDWSVADWELRVDAECR 285
MBV_BarnaV       IHAAVGLVDQIVTRLLCMKQNNAPIDCWES--CPSAPGMGL-NDEGLRTLYSTAQVMAE------------H-GTICETDISGWDWSVQQWELDSDARLR 299
CoBV_BarnaV      IHASVSLCDQLLTRTLSARQNKTPIMNWET--CPSKPGMGL-HDEGLAKLSENARRILT------------E-GRIMATDLSSWDWTVQEWELEVDAKRR 279
HuSRV_HbscleroV  QISSVSLAANIISRRLYGRQNRVPKAMWRSPEMPSSCQMG--SSDDDLQTLHSLWNA-QP---------L-G---SLAEADISGWDWTIQSWELMSDCERR 297
TroV_SobemoV     LISSVSLVDQLVPRVLFGRQNRKPITQWKS--IPSKPGMGLSLTEQMKSVFEQVSKLAA------------S-REAAEADISGFDWSVQEWELEMDLEVR 276
XYDV_SobemoV     LISSVSLIDQLVPRVLFGYQNSLPIATWKT--CPSKPGMGLSVPEQAKAIWEYAEYRHT------------R-APAAEADISGFDWSVQVWELMADVEMR 276
RYMV_SobemoV     LISSVSLVDQLVPRMLFGPQNNTPISTWWQ--WPSKPGMGLLTPEQIRLVWDDVFQKHQ------------A-HPAAEADISGFDWSVQDWELWSDLAIR 278
PLYV_SobemoV     LISSVSLVDQLVPRMLFGPQNELPITRWRN--IPSKPGMGLSLEEQAQTIWSDLSYKHL------------K-HPAAEADISGFDWSVQDWELWADLKMR 273
SNMoV_SobemoV    LISSVSLIVDQIIPRLLFGPQNQLPIATWST--IPSKPGMGLSLRSQAELLWNDLKHKHR------------L-SPASEADISGFDWSVQEWELWSDLSMR 280
SBMV_SobemoV     LISSVSLVDQLVPRMLFGPQNQLPIAEWEH--IPSKPGMGLSLRQQAKSLFDDLRVKHS------------R-CPAAEADISGFDWSVQDWELWADVEMR 277
PnLV_PolemoV     LIMSVSLIDQLVARVLFQKQNKLPLLLWRS--IPSKPGFGLSTVEQVEEFIDHLARVVDV-KSDDLLENWR--ELMVPTDCSGFDWSVSDWMLKDDMEVR 298
CYDV_PoleroV     LIMSVSLIDQLVARVLFQAQNKSPIALWSA--IPSKPGFGLSTEDQVESFINVLADTAGA-RPEEICDKWR--DLLVPTDCSGFDWSVSDWMLADDMEVR 297
PLRV_PoleroV     LIMSVSLVDQLVARVLFQNQNKRPISLWRS--VPSKPGFGLSTDTQTAEFLECLQKVSGAPSVEELCANHK--EYTRPTDCSGFDWSVAYWMLEDDMEVR 299
TuYV_PoleroV     LIMSVSLVDQLVARVLFQNQNKRPIALWRA--IPSKPGFGLSTDEQVLDFVESLARQVGT-TTTEVVANWK--NYLTPTDCSGFDWSVADWMLHDDMIVR 298
BChV_PoleroV     LIMSVSLVDQLVARVLFQNQNKRPIELWRA--VPSKPGFGLSTDDQIEDFVKVLASQLGE-EPQEVFNNWS--TKLIPTDCSGFDWSVADWMLEDDMEVR 298
BMYV_PoleroV     LIMSVSLVDQLVARVLFQNQNKRPIALWRA--IPSKPGFGLSTDGQVVDFMQALSAQVGV-NTAELLQDWK--SHLIPTDCSGFDWSVSDWLLEDDMEVR 298
CaBYV_PoleroV    LIMSVSLVDQLVARVLFQNQNKRPIALWRV--VPSKPGFGLSTDEQVAEFMQILSAQVGL-TPSDLITEWR--AHMIATDCSGFDWSVSDWLLEDDMEVR 298
GREnV_EnamoV     IHASVSMVDQLVARMLFQSQNKQPLRMWFS--IPSKPGMGLSTDDQVREFLDSLVAVSGAPSADALVADWQ--KWCVPTDCSGFDWSVPMWLLEDDMECR 298
CVEV_EnamoV      LHASVSLADQLVARMLFQEQNQAPLDMYMY--IPSKPGLGFSKDSQVLEFTKSIAYLAHT-TPEDLVNNWS--EHLIPTDCSGFDWSVPMWLLEDDLAVR 296
AlEV_EnamoV      LHASVSLVDQLVARMLFREQNQEEPLLQHMA--IPSKPGLGFSQDRQVLAFTESVAALAGT-TADDLVENWS--RYLTPTDCSGFDWSVPMWLLEDDLAVR 296
PeMV_EnamoV      LHASVSLVDQLVARMLFRDQNEEPLLQHMA--IPSKPGLGFSQDHQVLAFTESVAALAGT-SAQDLVDDWS--RYLTPTDCSGFDWSVPMWLLEDDLAVR 296

ScTV1_OutGroup   LISSVSVVDQIIDHMLFGDLNDVMIENWIH--IPNKPGWAPFG-GGWR-FI--P------------K--ET----WVATDASSWDWTVQPWLLELALEFR 256
SoTV5_OutGroup   LISSVSIIDQIIDQMLFGDQNEAFLQAYHY--TPVKVGWSWMN-GGWR-EV--P------------R--QN----MLACDKSSWDWTVSAWLLDLELMIR 254
HyPV1_OutGroup   LISSVSVIDQLIDHLLFDEKQNIMIGNWDK--IPPKVGWTPIV-GGWK-LV--P------------Q--VN----VQAADKSAWDWTMQGWIVEILLEIT 255
```

Motif B
Yellow loop
Motif C

```
tBV2_BarnaV      IKLAC--ADDKSLYAHLCRINAMVVS-RSVYVDPDGFCWAQTAYGIQNSGRYCTSSSNSRMRVLLTVHCRLSAGK--PAL--VK--G-RLGIISMGDDSD 385
RsBV_BarnaV      AALAG--ADPKGVFAFLLRVHAYCVA-NSVYVLPNGEMYEQTVPGAQLSGDYNTSSTNSRMRVCASLFSRLWAGK--PLL--VD--G-RIPVSAMGDDSF 381
TuBV_BarnaV      RRLAG--ATKAGVFNFLLRVHAYCVG-HSVFMTPGGSLWAQTIAGGQLSGDYNTSSSNSRMRVCAALMARQFAGQ--PLL--VG--G-KIPVCAMGDDAF 380
tBV1_BarnaV      RLLAR--ARKGGVFDFLLRVHAHCVA-NSVFVTPDGEMYAQTIPGGQLSGDYNTSSTNSRMRVVASMMARLWAT-GS-PL-VN---G-CIGIAAMGDDSF 375
MBV_BarnaV       TQLAG--EEIGGYLNFFLRVHAYVVG-HSVFVMPDGEMLEQTVPGGQLSGDYNTSSSNSRMRVIATMFARYLAGQ-VS---GF---P-LLGIKAMGDDSF 388
CoBV_BarnaV      ILLAG--IDEDHPMAKLLMAHAHIVA-NSVYVDSAGNMFAQTIPGGQLSGDYNTGSSNSAMRVIATQAARLKVN-QELYL-ER-ISAKELLVCSMGDDTL 373
HuSRV_HbscleroV  IDLCG--AG--PGLAQLMRNVHWTMA-LKVFQLSNGSLFEQIKPGVLPSGWYCTTTNSFMRAQTCQLVG----------AQ--------FYKVMGDDSV 374
TroV_SobemoV     LRLGN--FP--PKLELAARNRFKCFM-NSVFQLSNGELISQVSPGLMKSGSYCTSSTNSRIRVAMAYLIG----------SP--------WCIAMGDDSV 353
XYDV_SobemoV     IQLGD--FS--PKVERAARNRFICLA-NSVFQLSDGTLISQGLPGLMKSGSYCTSSSNSRIRCLMAKLIG----------SE--------WAMAMGDDSV 353
RYMV_SobemoV     INRGN--FQ--GNLRRAAISRYYCFM-NSVFQLSDGTLIQQELPGLMKSGSYCTSSTNSRIRCLMAELIG----------SP--------WCIAMGDDSV 355
PLYV_SobemoV     IELGN--ME--GRLKIAAINRFTCFM-NSVFQLSDGSLIAQGKPGLMKSGSYCTSSTNSRIRCLMAELIG----------AP--------WCIAMGDDSV 350
SNMoV_SobemoV    VNLCDD-MH--EGLKRLMINRFRCFM-FSVFQLSNGELLEQTSPGIMKSGSYCTSSTNSRIRCLMGVLIG----------SP--------WIVAMGDDSV 358
SBMV_SobemoV     IVLGG--FG--HKLAKAAQNRFSCFM-NSVFQLSDGTLIEQQLPGIMKSGSYCTSSTNSRIRCLMAELIG----------SP--------WCIAMGDDSV 354
PnLV_PolemoV     NRLTIN-CN--DLTRRLRNSWLYCLS-NSDLALSDGSLLAQEVPGVQKSGSYNTSSTNSRIRVMAAYFAG----------AS--------WAVAIGDDAL 376
CYDV_PoleroV     NRLTID-CN--ELTRHLRAVWLQGIS-NSVLCLSDGTMLSQVKPGVQKSGSYNTSSTNSRIRVMAAYHCG----------AS--------WAIAMGDDAL 375
PLRV_PoleroV     NRLTFN-NT--QLTKRLRAAWLKCIG-NSVLCLSDGTLLAQTVPGVQKSGSYNTSSSNSRIRVMAAYHCG----------AD--------WAMAMGDDAL 377
TuYV_PoleroV     NRLTID-LN--PATERLRSCWLRCIS-NSVLCLSDGTLLAQIHPGVQKSGSYNTSSSNSRIRVMAAFHTG----------AI--------WAMAMGDDAL 376
BChV_PoleroV     NRLTRN-NN--HTTKRLRSVWLKCIS-NSVLCLSDGCLLSQRVPGVQKSGSYNTSSSNSRIRVMAAYHCG----------AS--------WAMAMGDDAL 376
BMYV_PoleroV     NRLTLD-IN--DLTRRLRAGWLKCLA-NSVLCLSDGTLLSQQVPGVQKSGSYNTSSSNSRIRVMAAYHSG----------AS--------WAIAMGDDAL 376
CaBYV_PoleroV    NRLTLD-LN--ETTRRLRSAWLYCIS-NSVLCLSDGTLLAQRVPGVQKSGSYNTSSSNSRIRVMAAYHCG----------AE--------WAMAMGDDAL 376
GREnV_EnamoV     NRLTAE-IT--QDLITLRSVWLKCIG-NSLFSLSDGCLLAQTTPGIQKSGSYNTSSSNSRMRVLMSKHAG----------AK--------WCVAMGDDAI 376
CVEV_EnamoV      NELTID-CS--DDLIKMRSEWLRCLG-QSVLVLSNGMMVAQLEPGIQKSGSFNTSSTNSRMRYMASLYAG----------AS--------WAVTMGDDAL 374
AlEV_EnamoV      NELTLG-LP--YGLRKMRETWLKCLG-QSVFCLSNGLLLAQTSPGIQKSGSFNTSSTNSRMRYMLALYAG----------AD--------WAVTMGDDAL 374
PeMV_EnamoV      NELTLG-LP--HGLRKMRETWLKCLG-QSVFCLSNGLLLAQTSPGIQKSGSFNTSSTNSRMRYMLALYAG----------AS--------WAVTMGDDAL 374

ScTV1_OutGroup   KRQCVN-LT--RRWEELANWRYACLFRDPLFVTTAGHMFRQKQPGVMKSGCVNTIADNSVMQVLLHLRVCAELGQ--K--MT--------YLFTMGDDRL 341
SoTV5_OutGroup   SNLIVSEFK--PYWLELARWRYKCLFVDNIFVTSGGLVFRVKKAGLMKSGCVNTIVTNSLLQLVLHVRVSLELGR--E--IT--------NIWAMGDDTL 340
HyPV1_OutGroup   LRLCDN-PT--NNWKNLVCRRYKQLFSDAIFILSNGCMYKQKDKGLMKSGSVMTIMDNSLAQYVMHMRIVLELGL--E--EG--------ILWSMGDDTL 340
```

Motif D
Orange loop
Motif E
Red loop
Motif H

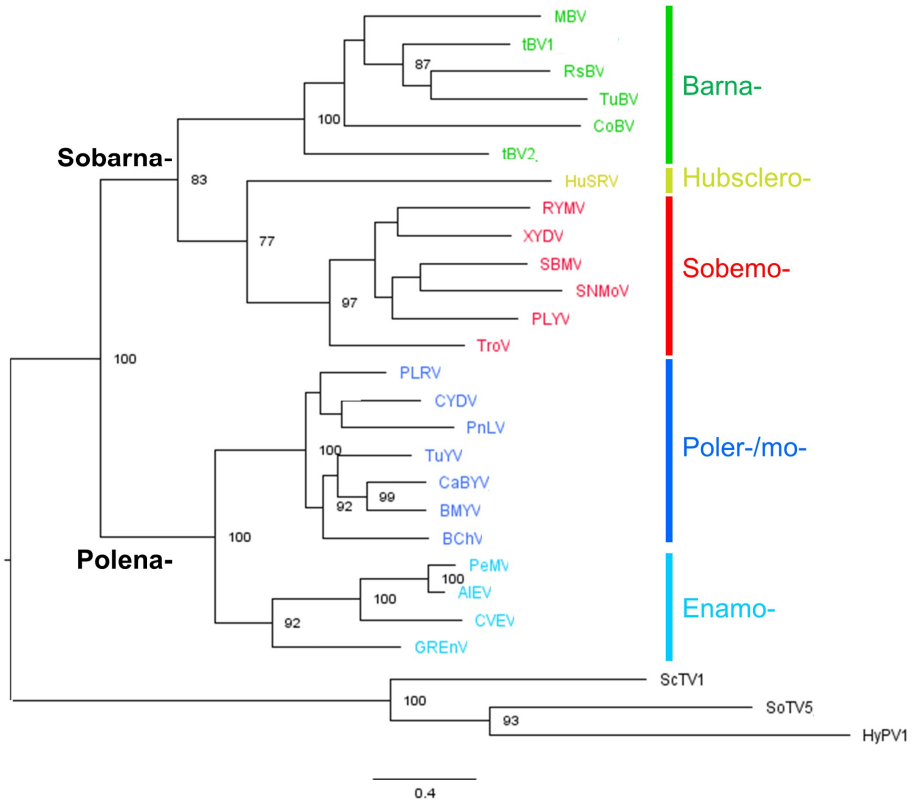
```
tBV2_BarnaV      EVAF--R-D-ILDWFGRLGFTVKMVEYN-T--------TVAGSEFCSQRFL-G-NGFAYPQAPAGTVFRFLS-RPP-T-----TEE-LPE--LWSQLSWY LRHLSGE-E-KEVIG--KLA 476
RsBV_BarnaV      ELDF--P-E-LQENMRSIGHNVRFVKRS-T--------SLQGVEFCSQVFD-E-DGFAAPADPSTTVYRFLS-HKV-T-----FSE-YPE--LWAQLSWY LRHQKKGDE-RDIIS--GLG 473
TuBV_BarnaV      ELAY--P-G-MAEHIRSLGHVVKFVELN-D--------KIQGAKFCSQEFD-E-NGTAYPEDPTTTLYRFLS-HNK-A-----SDD-MVM--LQVQLAWY FRHLPRE-L-FREIM--TTA 471
tBV1_BarnaV      EMEA--P-G-IGEGLQLMGHSVKFLKEN-A--------TLDGVNFCSQVFT-S-EGTAYPEDPSTTVFRFLS-HKH-G-----EQS-YLE--LWVQLNWY LRHLPHSVI-SQTIQ--ELA 467
MBV_BarnaV       EIWF--K-G-LEEYLGKMGHTVKMCVQR-P--------GLVGFEFCSQVFL-G-LGIAYPVDFSTTLYRFLS-HHP-A-----DPK-YSE--YRAQLMYY FRHLPSS-T-LQKVI--RLA 479
CoBV_BarnaV      ELEV--T-G-VPEVLEEIGHIVKDVEFH-E--------TLQDVEFCSQRFN-A-DGSAYPVDYNTTLYRFLS-HAP-D-----DVE-YSS--YKAQLMYI FRHMNQE-A-RDKIE--HTA 464
HuSRV_HbscleroV  ESYV--D-D-AVARYAALGKKVKMYNRCT-----------TTVEFCSRKWVLGPEWRAPLTSWPTTFYRLLS-QT--------AGD-KSE--FVKQFCHE LRHNS---E-LPKLL--EVL 461
TroV_SobemoV     EGYV--T-N-AREKYESLGHICKDYLVCQKK-K----GELDGFNFCSHWIS---RSHSYLDSVGTLYRFLE-SS--------N-------EELEILEAE LGTHP---R-WREIV--SKL 439
XYDV_SobemoV     EGWV--E-N-AKEKYLALGHTCKDYIACETE-ITRKGRKLRRFNFCSHEIS---SQGAFLTTWETTLFRFLE-SD--------H-------ESFDNLKFE LSTSP---K-WPSIL--RYL 443
RYMV_SobemoV     EGWI--E-G-AQSKYAALGHTCKEYYPCKTR-G----RELLEFNFCSHLIR---RGHAELTSWPTALFRFLS-SK--------H-------EDFEDLWVE LHTCG---V-WSRIE--RYL 441
PLYV_SobemoV     EGYV--P-N-ARRKYEELGHVCKDYMPCKVDSE----GNLEKVNFCSHELS---EGKFWLTSWPTTVFRYLN-SK--------A-------PEIGDLKAE LWGNP---H-WGRIY--EVV 437
SNMoV_SobemoV    EGWV--T-G-AKEKYLELGHVCKEYRPCETTRD----GELSKVNFCSHELS---EGRFWLTSWATTLYKFLD-SP--------N------ESH-VELERE LSSGP---M-WPRVK--RYL 445
SBMV_SobemoV     EGWV--D-G-AKDKYMRLGHTCKDYKPCATTIS----GRLYEVEFCSHVIR---EDRCWLASWPTTLFKYLS-EG--------K-------WFFEDLERD VSSSP---H-WPRIR--HYV 441
PnLV_PolemoV     ESID-----TTLAVYKSLGFKV-EVSE--------------DLEFCSHIFK-T-RSLAIPVNTSRMLYRLIYGYEP-ECGNLDVLRNYLC--ALASVLHE LRH-D-----QDL-VQ-N-L 462
CYDV_PoleroV     EAPD-----TDLSKYKDLGFKV-EVSK--------------ELEFCSHIFK-S-PTLAIPVNANRMLYRLIHGYNP-ECGNAEVIVNYLN--AASSVLHE LRH-D-----QEL-CA-L-L 461
PLRV_PoleroV     EAPN-----SDLEEYKTLGFKV-EVGR--------------ELEFCSHIFR-N-PTLAVPVNTNRMLYKLIHGYNP-ECGNPEVIQNYLA--AVFSVLQE LRH-D-----REL-VA-K-L 463
TuYV_PoleroV     ESNP-----ADLAAYKKLGFKV-EVSG--------------QLEFCSHIFR-A-PDLALPVNENRMIYKLIYGYNP-GSGNAEVVSNYLA--ACFSVLNE LRH-D-----PAS-VE-L-L 462
BChV_PoleroV     ESVD-----SNLTEYKKLGFKV-EVAK--------------QLEFCSHIFK-N-ERLALPLNVKRMLYKLIYGYNP-DSGNLEAIKNYLD--ACHSIVNE IRH-D-----ESL-VQ-K-I 462
BMYV_PoleroV     ESVD-----ADLSRYSSLGFKV-EVSS--------------QLEFCSHIFE-E-ENLAVPVNKARMLYKLIHGYEP-ECGNLEVLTNYLA--ACFSILNE LRS-D-----PEL-VA-P-L 462
CaBYV_PoleroV    ESAC-----SNLERYKSLGFKV-EESS--------------KLEFCSHIFE-K-EDLAIPVNKARMLYKLIHGYEP-ECGNVEVLINYLA--ACFSILNE LRS-D-----PSL-VE-T-L 462
GREnV_EnamoV     ESTD-----TNLSAYEEMGFKC-EAAD--------------KFDFCSHIFH-S-AKVALPTNVGRMLVGPLC-GVSPESQSQQDRIRWLM--SVGSILQE LRH-L----PPED-LAEL-H 464
CVEV_EnamoV      ESVD-----SDLSQYALLGLIC-ERAE--------------EFDFCSHIFK-A-PSVVVPKNIHRMIFGLLS-GVSPISPVQEARIQWLQ--AFQSISEE MRH-M----PAEF-WEEF-R 462
AlEV_EnamoV      ESVN-----SDLSQYARLGIKC-ERAE--------------EFDFCSHLFR-A-PDVIIPKNLERMVYGLLS-GTSPESPLLADRFCWLA--SLQSILEE MRH-M----PQEF-VDML-I 462
PeMV_EnamoV      ESVG-----SDLSQYARLGIKC-ERAE--------------EFDFCSHLFR-A-PDVVIPKNLERMVYGLLS-GTSPESPLLADRFSWLS--ALQSILEE MRH-M----PRDF-VDML-I 462

ScTV1_OutGroup   QEPVEDPKS-YYEMTSQF-CILKSVAH--------------VNEFAGFRFR-G--KWVEPVHKARHAYNILH-MED-------S--VVVP--LAHSYMLNYHRSQ---F-KQF-ME-D-L 423
SoTV5_OutGroup   QPDVAWGDD-YVRRLRKY-CILKETNR--------------VSEFAGFRWQ-G--NFIEPLYRARHAFNILH-VKP-------K--DFDV--FALSYNLLYWRSS---L-LGE-IQ-R-L 422
HyPV1_OutGroup   QTRQ--PKE-YFEKLAQY-CILKQVDT--------------ISEFAGMRFK-G--MQVTPSYLSRHAYKLLH-LDP-------E--LKDE--VALAYYLLYHRSS---N-KGK-IA-K-V 420
```

Thumb domain

Figure 6

(A)


MBV
tBV1
RsBV
TuBV
CoBV
tBV2
HuSRV
RYMV
XYDV
SBMV
SNMoV
PLYV
TroV
PLRV
CYDV
PnLV
TuYV
CaBYV
BMYV
BChV
PeMV
AlEV
CVEV
GREnV
ScTV1
SoTV5
HyPV1
Barna-
Hubsclero-
Sobemo-
Poler-/mo-
Enamo-
Sobarna-
Polena-
87
100
83
77
97
100
100
99
92
100
100
100
92
100
93
0.4

(B)

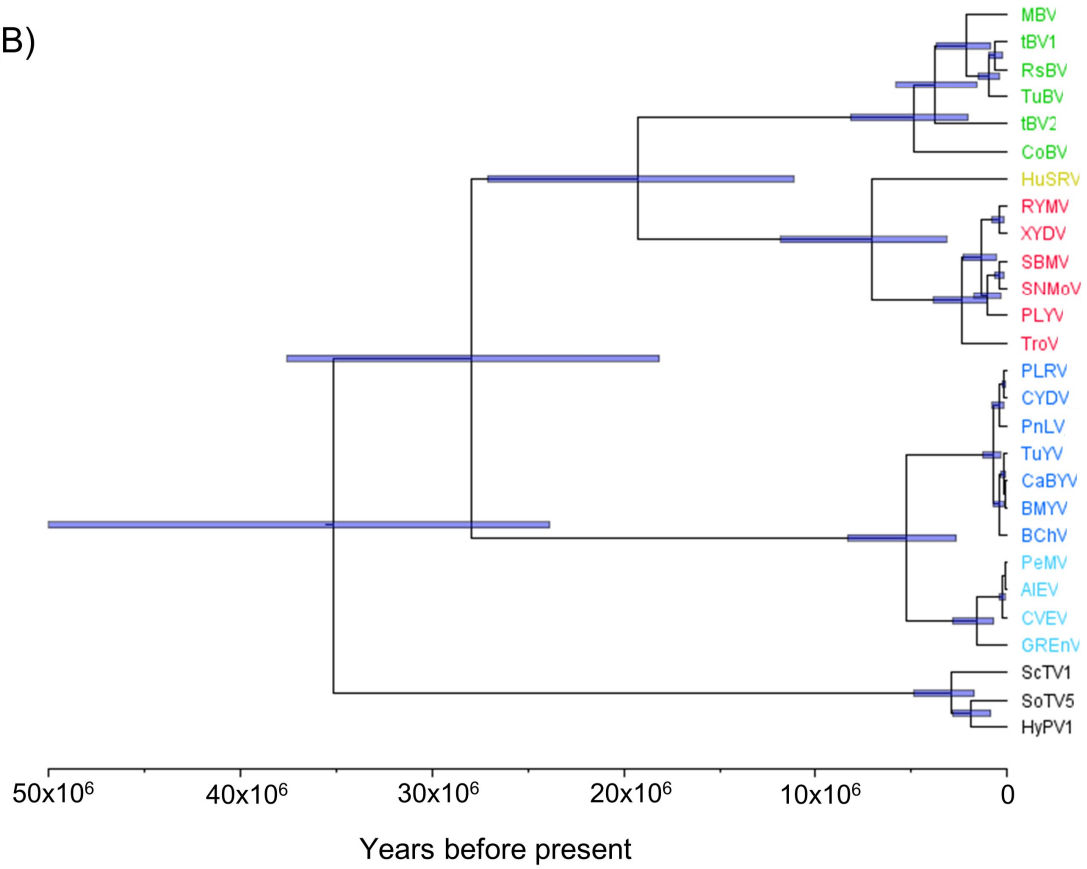
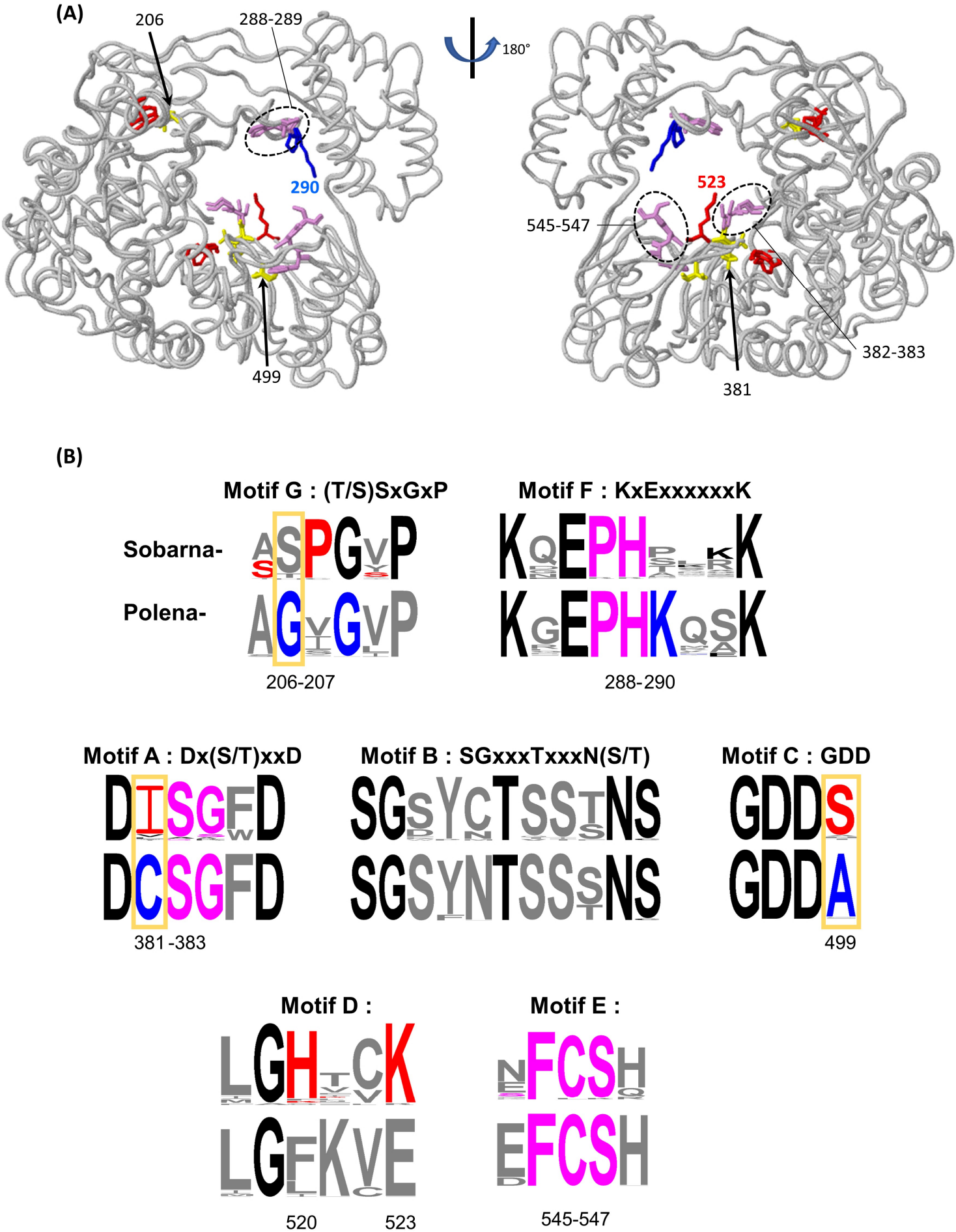

MBV
tBV1
RsBV
TuBV
tBV2
CoBV
HuSRV
RYMV
XYDV
SBMV
SNMoV
PLYV
TroV
PLRV
CYDV
PnLV
TuYV
CaBYV
BMYV
BChV
PeMV
AlEV
CVEV
GREnV
ScTV1
SoTV5
HyPV1
50x10^6
40x10^6
30x10^6
20x10^6
10x10^6
0
Years before present

Figure 7


(A)
206
288-289
290
499
180°
523
545-547
382-383
381
(B)
Motif G : (T/S)SxGxP
Motif F : KxExxxxxxK
Sobarna-
Polena-
206-207
288-290
Motif A : Dx(S/T)xxD
Motif B : SGxxxTxxxN(S/T)
Motif C : GDD
381-383
499
Motif D :
Motif E :
520
523
545-547