Identification of ambisense kolmioviruses

Mai Kishimoto[1,2,3,*], Sakiho Imai[3], Manabu Igarashi[4,5], Masayuki Horie[1,2,3,*]

[1] Laboratory of Veterinary Microbiology, Graduate School of Veterinary Science, Osaka Metropolitan University, Izumisano, Osaka, Japan

[2] Osaka International Research Center for Infectious Diseases, Osaka Metropolitan University, Osaka, Japan

[3] School of Veterinary Science, College of Life, Environment, and Advanced Science, Osaka Prefecture University, Izumisano, Osaka, Japan

[4] Division of Global Epidemiology, International Institute for Zoonosis Control, Hokkaido University, Sapporo, Hokkaido, Japan

[5] International Collaboration Unit, International Institute for Zoonosis Control, Hokkaido University, Sapporo, Hokkaido, Japan

[*]Corresponding authors

Mai Kishimoto, DVM, PhD

Email: m.kishimoto@omu.ac.jp

Masayuki Horie, DVM, PhD

E-mail: mhorie@omu.ac.jp

Abstract

Viruses in the family *Kolmioviridae* (KoVs) are unique among RNA viruses in that they have small, highly self-complementary, circular, negative-sense genomes and are satellite viruses requiring the envelope protein of co-infected helper viruses to form infectious virions. For approximately 40 years, only hepatitis D virus (HDV) had been identified, but recent studies have uncovered novel KoVs in diverse animal species. Serinus canaria-associated deltavirus (scKoV) was recently identified form a single RNA-seq data from canary skin. However, only the partial genome sequence corresponding to the transcript encoding DAg has been identified, leaving questions regarding its full genetic composition and origins. In this study, we examined publicly available RNA-seq data and identified multiple complete genome sequences of scKoV. Our results demonstrated that genetically diverse scKoVs infect multiple bird species with recent inter-order transmission. Additionally, scKoV was detected in various tissues including parenchymal organs, suggesting that scKoV is a genuine avian KoV. Notably, analysis of the genome structure and transcriptional pattern clearly revealed that scKoV transcribes a genomic-sense mRNA encoding a novel gene, in addition to the previously known antigenomic-sense DAg transcript. Furthermore, this newly identified gene appears to express two protein isoforms via alternative splicing. These findings mark the first KoV with an ambisense genome that utilizes alternative splicing, which may represent a strategy to increase gene-coding capacity despite the constraints imposed by its small and self-complementarity genome. Moreover, co-infection with avian bornaviruses was detected in most scKoV-positive RNA-seq datasets, suggesting that avian bornaviruses are candidate helper viruses for scKoV. In conclusion, our study on scKoV provides novel insights into the evolution and diversity of KoVs.

Keywords

46 kolmioviridae, deltavirus, ambisense, bird metagenome

47 **1. Introduction**

48 The family *Kolmioviridae* (KoVs), recently established in the realm *Ribozyviria*, consists of
49 RNA viruses with negative-stranded circular genomes of approximately 1.5-1.7 kb (1). A
50 representative KoV is human hepatitis D virus (HDV), which was first reported in 1977 (2).
51 After the discovery of HDV, no other related viruses were detected for approximately 40 years.
52 However, since 2018, many viruses related to HDV have been discovered in various vertebrates
53 and invertebrates, and currently, KoVs spanning 11 genera, along with additional unassigned
54 viruses, have been identified (1,3–8).

55 KoVs share unique features markedly different from other typical RNA viruses (RNA
56 viruses in the kingdom *Orthornavirae* of the realm *Riboviria*) in terms of their genomes,
57 replication, and transmission. The KoV genome is viroid-like, circular, and single-stranded
58 RNA that exhibits high self-complementarity, forming a rod-like structure (3–9). Unlike other
59 RNA viruses, KoVs only encode the delta antigen (DAg), a nucleocapsid-like protein, and lacks
60 RNA-dependent RNA polymerase, a hallmark of RNA viruses. Consequently, KoVs replicate in
61 the cell nucleus depending on the host DNA-dependent RNA polymerase (DdRp). This unique
62 replication occurs though a rolling-circle replication, utilizing the autocatalytic reaction of
63 ribozymes present in the KoV genome and antigenome. (10,11). As KoVs encode only DAg,
64 and lack proteins required for the formation of virus particles, they rely on the envelope proteins
65 of co-infecting "helper" viruses to produce infectious virions. Thus, KoVs are classified as
66 "satellite" viruses (5,12,13).

67 KoVs may utilize unique coding strategies to provide proteins essential for replication,
68 given the constraints of their small, highly self-complementary genomes. HDV expresses two
69 isoforms of the DAg, known as small and large DAg (S-HDAg and L-HDAg), through A-to-I
70 RNA editing, which is mediated by the host enzyme ADAR1 and modified the stop codon of S-

71 HDAg ORF (14–17). As a result, L-HDAg contains a 19-amino-acid extension at its C-terminus
72 that includes a nuclear export signal (18) and a farnesylation site which is required to interact
73 with helper virus envelope proteins (19). HDV utilizes these two isoforms to finely regulate
74 genome replication and virion formation. Although this RNA editing mechanism has only been
75 observed in HDV, other KoVs might employ different strategies to expand their coding capacity.
76 For example, Swiss snake colony virus 1 was reported to express two isoforms of DAg,
77 although the underlaying mechanisms remain unclear (20). Further, several KoVs, such as
78 DrDV-B, possess additional relatively long ORFs in their genomes (5,21). However, these
79 ORFs have yet to be shown to be expressed as mRNA, and there is no evidence that they are
80 functional viral genes.

81 The Serinus canaria-associated deltavirus (following the establishment of the family
82 *Kolmioviridae*, we designates it as Serinus canaria-associated kolmiovirus [scKoV]) was recently
83 identified through large-scale analysis of publicly available RNA-seq data, but its characteristics
84 are largely unclear (3). In the previous study, only partial sequences corresponding to the
85 transcript encoding DAg were identified. However, the rapid increase in publicly available
86 RNA-seq data offers an opportunity to identify the complete viral genome and conduct more
87 detailed analyses (22). Further, scKoV was suspected to have originated from contaminants in
88 the previous study, as it was only detected in a single RNA-seq dataset from canary skin and is
89 genetically distinct from viruses of the genus *Perithurisazvirus*, authentic avian KoVs (3) (Fig.
90 1). Thus, reanalyzing scKoV with the ever-increasing RNA-seq data may help clarify its origin.

91 In this study, we analyzed public RNA-seq data and identified multiple complete genome
92 sequences of scKoV. We provided strong evidence that scKoV is an authentic avian virus.
93 Further, we showed that scKoV transcribes a genomic-sense mRNA that encodes a novel gene
94 (ORF2) in addition to the well-known antigenomic-sense mRNA encoding DAg. We also

95 revealed that the ORF2 gene appears to express two protein isoforms via alternative splicing.
96 Our findings highlight an adaptive strategy by which KoVs maximize their gene-coding
97 capacity under the constraints of their small and highly self-complementary genomes.
98

99 **2. Results**

100 ***2.1. scKoV-derived reads were identified in multiple bird public RNA-seq datasets.***

101 As described above, the previous study detected scKoV in only one RNA-seq dataset. To further
102 identify scKoV-positive RNA-seq datasets, we mapped 15,452 publicly available bird RNA-seq
103 datasets to the partial sequence of scKoV (BR001664.1). A total of 66 RNA-seq datasets were
104 found to contain scKoV-derived reads, including the dataset in which scKoV was initially
105 identified in the previous study (Tables 1 and S1).

106 The scKoV-reads were found in four bird species: the common canary (*Serinus canaria*
107 [Linnaeus, 1758]), European serin (*Serinus serinus* [Linnaeus, 1766]), and house finch
108 (*Haemorhous mexicanus* [Statius Müller, 1776]) in the order *Passeriformes* and monk parakeet
109 (*Myiopsitta monachus* [Boddaert, 1783]) in the order *Psittaciformes*. The scKoV-positive RNA-
110 seq datasets were distributed across five BioProjects deposited by four independent research
111 groups (PRJNA300534 and PRJNA591356 were deposited by the same group), suggesting that
112 the scKoV-positive birds originated from diverse sources.

113 The scKoV-positive RNA-seq datasets were derived not only from skin samples but also in
114 various tissues including fat tissue, blood, brain, lung, and ovary. This strongly suggests that
115 scKoV is not a result of body-surface contamination but likely infects birds, potentially causing
116 systemic infection and/or viremia. These findings indicate that scKoV circulates among a broad
117 range of avian hosts, spanning different taxonomic orders.
118

119 ***2.2. The complete genome sequences of scKoV were identified.***

120 To obtain complete genome sequences of scKoV, each scKoV-positive RNA-seq dataset was
121 assembled. Among the resultant contigs, six contigs were identified as potential complete scKoV
122 genome sequences. Self-dotplot analyses revealed identical sequences at both ends of each contig
123 (Fig. S1), suggesting that they are circular contigs. We mapped the original RNA-seq datasets
124 back to the circularized contigs, confirming that multiple reads spanned the junctions
125 appropriately. These data strongly suggest that the contigs are indeed circular genomes.

126 To obtain additional complete scKoV genomes, we mapped scKoV-positive RNA-seq
127 datasets, from which no complete scKoV genomes had been obtained above, to the circularized
128 complete genomes of scKoV and extracted the consensus sequences. As a result, a total of 16
129 complete scKoV genomes were reconstructed from the RNA-seq datasets of *S. canaria*, *H.*
130 *mexicanus*, and *M. monachus* (Table 1). Notably, by this mapping-based method, we succesfully
131 recovered the complete scKoV genome from the RNA-seq datasets where only the partial
132 sequence had been initially reported in the previous study (3).

133

134 ***2.3. Genetically diverse scKoV circulate among multiple bird species.***

135 To elucidate the evolutionary relationships among scKoVs, we performed a phylogenetic analysis
136 using the complete nucleotide sequences of scKoV genomes. The resulting phylogenetic tree
137 showed three distinct clusters of scKoVs, designated as genotype 1, 2, and 3 (Fig. 2A). The
138 genotype 1 included scKoVs identified in all the three bird species (*S. canaria*, *H. mexicanus*, and
139 *M. monachus*), while the genotypes 2 and 3 were exclusively found in scKoVs from *S. canaria*
140 and *M. monachus*, respectively. Pairwise nucleotide sequence comparisons revealed high
141 nucleotide sequence identities (>95%) within each genotype, whereas inter-genotype nucleotide

142 sequence identities were <84% between genotypes 1 and 2, <83% between genotypes 1 and 3,
143 and <89% between genotypes 2 and 3 (Fig. 2B).

144 We also obtained data indicating that multiple genotypes of scKoV existed within individual
145 hosts. For example, all scKoV-positive RNA-seq datasets from BioProject PRJNA679636
146 originated from a single individual (23) , and both genotypes 1 and 3 were detected within these
147 datasets (Table 1). We thus investigated whether multiple genotypes of scKoV coexisted within
148 each scKoV-positive sample. The scKoV-positive RNA-seq datasets from BioProjects
149 PRJNA591356, PRJNA939570, and PRJNA679636 were mapped to the genomes of distinct
150 genotypes detected within the same BioProjects in a manner that ensures unique alignment to the
151 best match. The results indicated the coexistence of two genotypes of viruses within individual
152 RNA-seq data in varying ratios, with some data showing nearly a 1:1 ratio of mixed genotypes
153 (Fig. 2C). Apparently, these RNA-seq data were not derived from pooled samples of multiple
154 individuals based on the available metadata and literatures. Taken together, these findings suggest
155 that multiple genotypes of scKoV circulate among the birds and may occasionally cause mixed
156 infections.

157

158 ***2.4. scKoV potentially infects to various tissues.***

159 To understand the tissue tropism of scKoV, we analyzed the number of scKoV-derived reads
160 across the scKoV-positive RNA-seq datasets from BioProject PRJNA679636, which were
161 derived from various tissues of a single individual (23). The reads per million (RPM) of scKoV
162 was highest in the brain (264.1 RPM in the forebrain and 121.9 RPM in the cerebellum), followed
163 by the lung (164.1 RPM) and ovary (77.3 RPM), while only a small number of scKoV reads were
164 detected in the liver (0.2 RPM) and muscle (5.5 RPM) (Tables 1 and S1 and Fig. 2C). Moderate

165 amounts of scKoV reads were detected in the kidney (25.2 RPM), heart (22.7 RPM), and spleen
166 (21.8 RPM). These results suggest that scKoV infects a broad range of tissues.
167
168 ***2.5. scKoV exhibits ambisense gene coding and utilizes alternative splicing.***
169 We next characterized the genome structures of scKoV, revealing that the typical characteristics
170 of KoVs are conserved in scKoVs as follows. The scKoV genomes were approximately 1,670 bp
171 in length, with a GC content of approximately 55%, consistent with other KoV members (3–8).
172 The high self-complementarity of the circular genome, a typical feature of KoVs, was also
173 conserved in scKoV (Figs. 3A and S2). The relative positions of the DAg-encoding ORF, the
174 downstream poly-A signal, and both the genomic and antigenomic ribozymes were consistent
175 with those of other typical KoVs (Fig. 3B). Additionally, the predicted scKoV ribozymes exhibit
176 the conserved HDV ribozyme class, unlike toad and termite KoVs, which retain type III
177 hammerhead ribozymes (24) (Figs. 3C and S3). The ribozymes of scKoV contain sequence
178 variations, but these are expected to have minimal effects on their secondary structures (Fig. S3).
179 Notably, two additional ORFs (tentatively named ORF2 and ORF3) were identified on both
180 the genomic- and antigenomic-senses, respectively, which were located in the self-
181 complementary region of the DAg ORF (Figs. 3A and 3B). ORF2 is conserved in all scKoVs,
182 whereas ORF3 is absent in one virus (BR002468.1, assembled from SRR13480248) (Fig. S4).
183 Moreover, an additional poly-A signal was located immediately downstream of ORF2 (Fig. 3B).
184 To elucidate the transcriptional patterns of scKoVs, we mapped scKoV-positive RNA-seq
185 datasets to the corresponding viral genomes. The results exhibited a bimodal mapping pattern
186 (Fig. S5), suggesting the presence of a novel transcript in addition to the one encoding DAg. To
187 further characterize the putative novel transcript, we mapped strand-specific RNA-seq datasets
188 (BioProjects PRJNA591356 and PRJNA939570) to the corresponding viral genomes. The

189 mapping patterns clearly indicated the presence of a genomic-sense transcript encoding ORF2, as
190 well as the typical antigenomic-sense transcript encoding DAg (Figs. 3D and S5). The boundary
191 between the genomic and antigenomic transcripts was estimated to be located at the edge of the
192 rod-like structure of the scKoV genome, which is thought to function as a transcriptional promoter
193 (25). The edge of the rod in scKoV largely retained the conserved domains previously reported
194 in HDVs (Fig. S6) (25). Furthermore, based on the mapping patterns of the RNA-seq data in
195 BioProject PRJNA591356, the transcription start sites of the genomic- and antigenomic-sense
196 transcripts were predicted to be located between the internal and external bulges of the rod edge
197 (Fig. 3E). These data strongly suggest that scKoV expresses a genomic-sense transcript encoding
198 ORF2.

199 The mapping results also showed several putative splicing junctions. Among them, the one
200 located in the genomic-sense ORF2 transcript is prominent and reliable, as a large number of
201 reads consistently spanned the junction across all the RNA-seq data (Figs. 3D and S5). The
202 putative intron retains the typical features of major eukaryotic introns, including the GU-AG
203 motifs of donor/acceptor sites and a polypyrimidine tract present upstream of the AG motif at the
204 end of the intron (Fig. 3F) (26). Notably, the spliced ORF2 transcript encodes a protein that differs
205 from the unspliced ORF2 transcript in its C-terminal region (Fig. 3D and Table S2).

206 To investigate the function of the putative proteins translated from unspliced and spliced
207 ORF2 transcripts (designated ORF2a and ORF2b, respectively), we first performed sequence
208 similarity searches. However, neither protein showed sequence similarity to any known protein.
209 Additionally, no significant biases in hydrophobicity or polarity were observed (Figs. S7A and
210 S7B). We also carried out a series of in *silico* analyses, including functional domain searches,
211 predictions of topological structures, nuclear localization signals, and transmembrane domains
212 (Fig. S7C), but could not find any functional domains. An intrinsically disordered region was

213 identified exclusively in the C-terminal third of the ORF2b (Fig. S7D). Additionally, structural
214 predictions indicated that the N-terminus of the ORF2 protein forms a well-defined helical
215 structure, whereas the C-terminus of the ORF2 lacks a defined structure (Fig. S7E). These
216 findings suggest that scKoV expresses genomic-sense transcripts encoding proteins of unknown
217 function.
218
219 ***2.6. No evidence for RNA-editing was found on the stop codon of DAg.***
220 HDV utilizes the host A-to-I RNA editing machinery at the stop codon of DAg to express the
221 larger isoform of DAg protein. To investigate whether scKoV also employs A-to-I RNA editing,
222 we analyzed nucleotide variations at the stop codons of each ORF using the mapping data. Base
223 variations (A-to-G or T-to-C), potentially resulting from A-to-I editing of the genome or antigenome,
224 were observed in 11 out of the 16 mapping datasets (Table S3). However, except for SRR23747114,
225 the stop codons of the extended DAg ORFs, generated by the cancellation of the original stop
226 codon through A-to-I RNA editing, were located downstream of the poly-A signal (Fig. S8). These
227 results suggest that scKoV is unlikely to utilize host A-to-I RNA editing to produce DAg proteins
228 of different lengths.
229
230 ***2.7. Avian bornaviruses are the candidate helper virus for scKoV.***
231 To gain insights into potential helper viruses for scKoV, we searched for co-existing viruses in
232 the scKoV-positive RNA-seq datasets. BLAST searches using the assembled contigs revealed that
233 contigs derived from avian bornaviruses (ABVs), canary bornavirus 1 and 3 (CnBV-1 and -3) and
234 parrot bornavirus 2 (PaBV-2) (the genus *Orthobornavirus* of the family *Bornaviridae*) were
235 detected in 64 out of 66 scKoV-positive datasets (Figs. 4A and S9). CnBV-1 and 3 were detected
236 in the RNA-seq datasets from *S. canaria*, *S. serinus*, and *H. mexicanus* (the order *Passeriformes*),

237 while PaBV-2 was detected in the data from *M. monachus* (the order *Psittaciformes*). No other
238 enveloped viruses were identified. We also analyzed the correlation between their read counts. As
239 a result, the RPMs of scKoV and ABVs showed a weak positive correlation (Fig. 4B). These
240 results suggest that scKoV may utilize co-infecting ABVs as helper viruses.
241
242 ***2.8. Other possible ambisense kolmioviruses***
243 We investigated whether any known KoVs exhibit ambisense gene expression. Among the
244 KoVs with genome sequences deposited in GenBank or reported by the Serratus project (27),
245 six additional KoVs possessed genomic-sense ORFs upstream of poly-A signals (Figs 5A, 5B,
246 and S10). Of these, Amboli leaping frog (*Indirana chiravasi*) kolmiovirus (icKoV) and
247 Babylonia areolata KoV (baKoV) displayed clear bimodal transcription patterns (Figs.5C and
248 S10). Furthermore, for icKoV, the closest relative of scKoV, mapping of the strand-specific
249 RNA-seq dataset (SRR8954567) revealed transcripts derived from both the genome and
250 antigenome strands, like the pattern observed in scKoV (Fig. 5C). However, unlike scKoV,
251 icKoV lacks clear alternative splicing site. The icKoV ORF2 showed 36.6% amino acid
252 sequence identity with scKoV ORF2a, and sequence alignment revealed the N-terminal 80
253 residues are relatively well conserved (Fig. 5D).
254 For other potential ambisense KoVs, mapping patterns did not exhibit a clear bimodal
255 distribution (Fig. S10), suggesting that these ORFs may not be transcribed. As those RNA-seq
256 libraries were prepared via rRNA depletion rather than poly-A enrichment, we cannot exclude the
257 possibility that the data do not accurately reflect transcription. In addition, in DrDV-B, the ORF2
258 start codon is not conserved across all viral strains (Fig. S11A). These genomic-sense ORFs also
259 lack similarity in both secondary and 3D structure to scKoV ORF2 (Figs. S11B and S11C),

260 implying that they may not encode conserved genes. These results suggest that the ambisense
261 genome structure may be conserved among closely related species of scKoV.
262
263 **3. Discussion**
264 Although many novel KoVs have been identified recently, the diversity of KoVs remains unclear
265 since some KoVs have not yet been characterized well. We previously identified scKoV from a
266 single canary RNA-seq datasets, which is distantly related to the other avian KoVs, but only a
267 partial sequence corresponding to the DAg mRNA had been identified. In this study, we identified
268 the complete genome sequences of scKoV and revealed its unique genetic features among KoVs.
269 Notably, we showed a series of evidence strongly suggesting that scKoV expresses genomic-sense
270 transcripts encoding proteins of unknown function, designated ORF2, besides the antigenomic-
271 sense transcript encoding DAg (Figs. 3B and 3D). Although KoVs are currently classified as
272 negative-strand RNA viruses, our findings strongly suggest that scKoV is an ambisense virus.
273 Furthermore, our data indicate that the genomic-sense transcript undergoes alternative splicing,
274 potentially encoding two ORF2 protein isoforms. Therefore, scKoV employs a unique strategy to
275 maximize its coding capacity under the constraints of its small and self-complementary circular
276 genome.
277 Our results strongly suggest that the promoter of scKoV could initiate bidirectional
278 transcription. HDV transcribes mRNA using the host DdRp, with the rod-like edge of its circular
279 RNA genome functioning as a transcriptional promoter, and the promoter on the genomic-strand
280 initiate the transcription of antigenomic-strand mRNA (25). In scKoV, the mapping pattern clearly
281 revealed that the transcripts were expressed in both the genomic- and antigenomic-senses, with
282 the edge of the rod-like structure serving as the boundary (Figs. 3D and 3E). Some RNA-seq
283 datasets showed that the approximate initiation sites of both the genomic- and antigenomic-sense

284 transcripts are located between the internal and external bulges at edge of the rod-like structure,
285 which is close to the location previously identified in HDV (25). Although further experiments
286 are necessary, these observations suggest that both the genomic and antigenomic rod edges of
287 scKoV possess distinct promoter activity. The presence of KoVs with bidirectional and
288 monodirectional promoters provides valuable insights into their evolutionary processes and
289 difference in replication strategies.

290 Our mapping analyses revealed potential splicing events in the genomic transcripts of scKoV.
291 While the majority of RNA viruses replicate in the host cytoplasm, a few lineages, such as KoVs,
292 bornaviruses, orthomyxoviruses, and some rhabdoviruses, transcribe in the host cell nucleus and
293 employ the host splicing machinery, which occurs inside of the nucleus, to express multiple
294 proteins from the same genomic regions (28–31). However, evidence for alternative splicing in
295 KoVs has been lacking thus far. Here, we identified a putative splicing site in the scKoV ORF2
296 gene (Figs. 3D and 3F), which may lead to the expression of two isoforms of ORF2 proteins (Fig.
297 3D). Thus, this study provides the first evidence that KoV utilizes alternative splicing and
298 emphasizes the need for future characterization of other KoV strains, considering the potential
299 use of alternative splicing mechanisms.

300 The ambisense coding strategy of scKoV is analogous to that of other circular RNA viruses.
301 KoVs share the genome structure and the replication mechanism with circular RNA replicons,
302 including viroids, virusoids, and other unclassified mobile genetic elements (32,33). Recent
303 studies have revealed an unprecedented diversity of circular RNA replicons (27,34,35). Among
304 the newly identified circular RNA viruses, fungal viruses in the phylum *Ambiviricota* (the
305 kingdom *Orthornavirae* of the realm *Riboviria*) possess at least two non-overlapping ambisense
306 ORFs, one of which encodes a distinct viral RdRp, suggesting their role as evolutionary
307 intermediates (34,36). The genome of scKoV exhibits a similar organization to that of ambiviruses,

308 with the two ORFs (DAg and ORF2) arranged in an ambisense orientation within a self-
309 complementary region (Fig. 3). This suggests that ambisense coding strategies may be common
310 among certain circular RNA viruses.

311 Our findings strongly suggest that birds are the authentic host of scKoV. The previous study
312 proposed the possibility that scKoV does not infect canaries but rather originates from
313 contaminants, based on the following two points: (i) scKoV is distantly related to tgDeV, which
314 is an authentic avian virus found in a wide range of passerine birds; (ii) scKoV was detected only
315 in a single RNA-seq dataset obtained from canary skin (3). However, our study identified scKoV
316 in several passerine birds, including canaries, as well as in a parrot (Tables 1 and S1). Additionally,
317 scKoV was detected not only in the skin but also in multiple parenchymal organs, fat tissue, and
318 blood (Tables 1 and S1). Furthermore, the bimodal transcription patterns observed for scKoV
319 suggest active transcription of viral mRNA in those tissues (Fig. 3D). These findings strongly
320 suggest that scKoV is an authentic avian virus rather than a contaminant. The existence of avian
321 kolmioviruses that are phylogenetically distinct from the lineages of known reptilian, avian, and
322 mammalian KoVs (Fig. 1), suggests that KoVs infecting the above-mentioned animals are far
323 more genetically diverse than previously thought. Further studies on KoVs are required to
324 understand their host range and genetic diversity.

325 The identification of the complete scKoV genome will contribute to updating the taxonomy
326 of KoVs. The current genus demarcation criterion established by the International Committee on
327 Taxonomy of Viruses (ICTV) is DAg with ≤60% amino acid sequence identity (1). Thus, scKoV
328 is classified as a new genus because its DAg sequence identity to other KoVs is below 40% (3).
329 We further demonstrated the presence of genetic diversity within scKoVs. The ICTV species
330 demarcation criteria require ≤80% nucleotide genome identity and ≤70% S-DAg amino acid
331 identity (1). Based on these criteria, the three groups of scKoV are classified as different

332 genotypes of the same species (Fig. 2B). Given the limited information on viral genome sequences
333 for classifying KoVs into distinct species within the same genus, our findings on scKoV sequence
334 diversity may help refine future classification criteria for KoVs.

335 Our study revealed that scKoV has a broad host range and the potential for interspecies
336 transmission. Our findings suggest that scKoV, particularly genotype 1, is capable of interspecies
337 transmission between birds of different orders (Fig. 2A). It has been suggested that KoVs have
338 evolved through the multiple interspecies transmission (3,8). Thus, evidence of relatively recent
339 inter-order transmission of scKoV may provide insights into the evolutionary history of KoVs.
340 Further detailed analyses of the host specificity of scKoV are necessary.

341 We also reveal the mixed infections of different genotypes of scKoV in some individuals.
342 scKoVs of different genotypes were detected mixed within RNA-seq data from the same
343 individual (Fig. 2C). These mixed infections, observed across multiple BioProjects and in some
344 cases with nearly equal ratios and certain read numbers of different genotypes, suggest that they
345 are unlikely to be due to contamination and/or index hopping during sampling, library preparation,
346 and sequencing processes. HDV has been reported to undergo recombination between different
347 genotypes during coinfection (37). Our findings of mixed infection with scKoVs belonging to
348 different genotypes provide important insights into the epidemiology and evolution of KoVs.

349 Our study identified ABVs as candidate helper viruses for scKoV. The helper viruses for
350 non-HDV KoVs remain largely unknown except for snake deltavirus that utilizes the envelope
351 proteins of co-infected reptarenavirus or hartmanivirus in the family *Arenaviridae* (5,13). Our
352 previous study indicated co-existence of scKoV with CnBV in a single RNA-seq dataset
353 (SRR2915371); however, this RNA-seq dataset was derived from a pooled sample from multiple
354 individuals, and no definitive evidence of co-infection between scKoV and CnBV was obtained
355 (3). In this study, we identified CnBV-1, CnBV-3 and PaBV-2, which encode envelope

356 glycoproteins, in 64 out of 66 scKoV-positive RNA-seq datasets (Figs. 4A and S8). Notably,
357 except for the RNA-seq dataset in which scKoV was previously identified (3), none of these were
358 clearly from pooled samples. Although further experimental validation is needed, these results
359 suggest that ABVs are helper viruses for scKoV. Further, our data provide insights into the
360 hypothesis that KoVs have evolved with diverse satellite-helper relationships.

361 Co-infection with scKoV and ABV suggests that the pathogenicity of scKoV in infected
362 birds needs to be considered. Co-infection of HDV and HBV accelerates the pathogenic effects
363 of HBV, leading to severe or fulminant hepatitis and progression to hepatocellular carcinoma (38).
364 The pathogenicity of scKoV remains unknown due to the lack of metadata clarifying whether
365 scKoV-positive RNA-seq data were obtained from healthy birds. In contrast, the co-infecting
366 PaBV is a causative agent of proventricular dilatation disease (PDD), an immune-mediated
367 inflammatory disease (39,40). Similarly, CnBV was also reported to be associated with PDD in
368 canaries, but experimental infections using a CnBV isolate have not successfully reproduced the
369 pathology, suggesting that other factors may be involved in the onset of the disease (41).
370 Analogous to co-infection of HDV and HBV, scKoV may also contribute to or exacerbate PDD
371 through co-infection with CnBV. Further epidemiological studies of scKoV and CnBV is required
372 to clarify the etiology of PDD.

373 Our analysis suggests that scKoV may have the potential to infect various tissues. While
374 HDV is known for its high tropism for the liver (12), non-HDV KoVs have been detected in
375 multiple organs (3,7,8,13). Like other non-HDV KoVs, scKoV was found to predominantly target
376 the skin, fat tissue, blood, and some parenchymal organs, rather than the liver (Table 1 and Fig.
377 2C). ABVs, candidate helper viruses for scKoV, can also infect various organs (41,42). Our results
378 support the hypothesis that while the tissue tropism of scKoV and other KoVs is broad, it is likely
379 dependent on the tissue infectivity of the helper virus (3,7,13).

380 Taken together, we provide novel insights into the genomic features, replication strategies,
381 epidemiology, and satellite-helper relationships of KoVs through the comprehensive and detailed
382 analyses of the complete scKoV genomes. Further analyses are needed to explore the broader
383 diversity of KoVs.
384

385 **4. Materials and Methods**

386 ***4.1. Detection of scKoV-derived reads from public RNA-seq data***

387 RNA-seq data from birds (in the class *Aves* without *Gallus gallus* [Linnaeus, 1758]) were
388 downloaded from the NCBI Sequence Read Archive (43) (accessed on Jan 20, 2024) and mapped
389 to the partial scKoV genome (BR001664.1) using Magic-BLAST v1.7.1 (44) with the options "-
390 splice T -word_size 16 -lcase_masking". The mapped read numbers were counted using
391 SAMtools v1.19 (45). The mapping results were manually verified, and reads that were locally
392 aligned to sequences of less than 25 bases in the reference were excluded as non-specific mapping.
393

394 ***4.2. Identification of complete scKoV genomes***

395 The scKoV-positive RNA-seq data were downloaded and dumped using SRA Toolkit
396 (https://github.com/ncbi/sra-tools), which were preprocessed using fastp v0.23.4 (46) with the
397 options "-l 35 -x -y". The preprocessed reads were mapped to the host genome sequences using
398 HISAT2 v2.2.1 (47) with the default settings. The unmapped reads were extracted with SAMtools
399 v1.19 and subjected to *de novo* assembly using Trinity v2.15.1 (48) with the default settings. The
400 scKoV contigs were identified through a two-step sequence similarity search: BLASTn against
401 the partial scKoV genome and BLASTx against the clustered nr database (43), using BLAST+
402 v2.15.0 (49).

403 Self-dotplot analysis of the linear scKoV contigs was performed using the YASS program
404 with the default settings (50). Contigs with identical ends were manually circularized using
405 Geneious Prime v2024.0.5 (https://www.geneious.com). The circularities of scKoV contigs were
406 further verified by mapping short reads to the circularized scKoV contigs using Geneious and
407 manually checking the mapped reads across the circularized boundaries.
408 For RNA-seq data where circularized scKoV contigs could not be obtained by *de novo*
409 assembly, the scKoV-positive RNA-seq data were mapped to the circular scKoV contigs using
410 Geneious and extracted the circular consensus sequences.
411 For all the contigs, all positions were confirmed to be covered with at least three reads.
412
413 ***4.3. Phylogenetic analysis***
414 Phylogenetic relationship among all known KoVs were inferred as follows. The amino acid
415 sequences of KoV DAg were aligned by MAFFT v7.490 with the E-INS-I algorithm (51). The
416 phylogenetic tree was constructed using the Bayesian Markov chain Monte Carlo (MCMC)
417 method in MrBayes version 3.2.7a (52) with two independent runs, four chains per run, and the
418 JTT model of substitution. The analysis was run for 5 million iterations, with samples taken every
419 5,000 steps. The average standard deviation of split frequencies was 0.010.
420 For the nucleotide sequences of the full-length scKoV genomes, multiple sequence
421 alignment was performed using MAFFT v7.490 with the E-INS-I algorithm (51). Pairwise
422 sequence identities were calculated based on the alignment using Geneious. A maximum-
423 likelihood tree was constructed using the TVM+I+G4 model, selected as the best-fit model based
424 on the minimum AIC in RAxML v1.1.0 (53). The reliability of the tree was assessed by 1,000
425 bootstrap resampling with the transfer bootstrap expectation method (54). Both trees were
426 visualized and annotated in the Interactive Tree Of Life (iTOL) v6.9.1 (55).

427

428 ***4.4. Sequence Characterization***

429 The self-complementarities of scKoV genomes were analyzed using the Mfold web server (56)
430 with the default settings. Genomic and antigenomic ribozyme structures were initially inferred
431 using the IPknot web server (57), with the visualized output from IPknot serving as a guide for
432 manual ribozyme structure drawing. Poly-A signals (AAUAAA) were manually identified in both
433 strands of KoV genomes. ORFs were detected using the 'Find ORF' function in Geneious, with
434 a minimal threshold of 300 nucleotides. ORFs that overlap with the terminal rod-like structure of
435 the viral genome, where the putative transcriptional promoter is located, poly(A) signal, or the
436 predicted ribozyme regions were excluded.

437

438 ***4.5. Short read mapping to analyze mixed infection with scKoV of different genogroups.***

439 The short reads from scKoV-positive RNA-seq data belonging to BioProjects PRJNA591356,
440 PRJNA939570, and PRJNA679636 were mapped to scKoV sequences of different genogroups
441 detected from the same BioProject (scKoV from SRR10526804 and SRR10526851 for
442 PRJNA591356, scKoV from SRR23747122 and SRR23747110 for PRJNA939570, scKoV from
443 SRR13480248 and SRR13480240 for PRJNA679636). Mapping was performed using of HISAT
444 ver 2.2.1. with the `-k 1` option. The mapped read numbers were counted using SAMtools v1.19
445 (45). The data with RPM>10 were visualized.

446

447 ***4.6. Short read mapping to analyze transcriptional patterns***

448 The short reads from scKoV-positive RNA-seq data were preprocessed as described above (see
449 the section 4.2) and mapped to the corresponding scKoV genomes using HISAT2 v2.2.1 (47) with
450 the options "-k 2 --max-intronlen 1000 --rna-strandness RF". To accurately evaluate reads

451 spanning the circularized genome boundaries, 2x tandemly repeated genomes were used as the
452 references. The "-k 2" option in HISAT2 was selected to ensure precise mapping to the 2x
453 genomes. For strand-specific RNA-seq data, the "--rna-strandness RF" option was also used to
454 retain strand information. The mapped reads were counted using SAMtools v1.19 (45).
455
456 ***4.7. Characterization of ORF2 proteins***
457 The amino acid sequences of both ORF2a and ORF2b were analyzed using BLASTp, employing
458 the PSI-BLAST and the DELTA-BLAST algorithms with an e-value threshold of < 0.05.
459 Theoretical molecular weight and isoelectric point was calculated using ProtParam web server
460 (https://web.expasy.org/protparam/). Amino acid sequence alignments, colored according to
461 hydrophobicity and polarity, were generated using Geneious. Membrane topology of ORF2 proteins
462 was inferred using DeepTMHMM ver. 1.0 (58). Nuclear localization signal was inferred using the
463 cNLS mapper web server with a cut-off value of 3.0 (59). Intrinsically disordered region were
464 predicted using AIUPred web server (60). Secondary structures of ORF2 proteins were predicted using
465 JPred v4 (61).Protein structure predictions were performed by AlphaFold3 web server (62).
466
467 ***4.8. Detection of RNA-editing on the stop codon of DAg***
468 Sequence variations at the DAg stop codon were identified in the mapping data of section 4.5
469 using 'Find SNP/Variants' in Geneious.
470
471 ***4.9. Detection of co-existing viruses***
472 To identify co-existing viruses in scKoV-positive RNA-seq data, the assembled contigs (see the
473 section 4.2) were clustered using CD-HIT v 4.8.1 (63) with the 95 % threshold. The clustered
474 contigs of 500 nucleotides or longer were extracted using SeqKit v2.4.0 (64), which were

475 subjected to a two-step sequence similarity search. The first search was conducted by MMseqs2
476 (version c48da9d781b81804727b5cccfed7f97cfcc20c9d) (65) with the default settings against a
477 custom database consisting of viral protein sequences with a threshold E-value of 0.001. The
478 custom database was created as follows. Viral protein sequences were downloaded from the NCBI
479 nr database on June 22, 2023, which was clustered by using CD-HIT v4.8.1 (63) with a threshold
480 of 0.9. Contigs alignable to viral sequences other than retroviruses and environmental viruses
481 were extracted. The second BLASTx search was then performed using the extracted contigs as
482 queries against the NCBI clustered nr database (43) with the options "-e-value < 1e-4 -word_size
483 2 -max_target_seqs 10" on Aug 26, 2024. The contigs with the best hit to viruses were extracted
484 and manually analyzed. Contigs with the best hit to bornaviruses (the family *Bornaviridae*) were
485 further subjected to a BLASTn search against a database of CnBVs and PaBVs to accurately
486 identify the species/genotypes.

487 To quantify ABVs-derived reads in scKoV-positive RNA-seq data, preprocessed reads were
488 mapped using HISAT2 (47) with the option "-k 1" against CnBV-1 (NC_030690.1), CnBV-3
489 (NC_024296.1), and newly obtained PaBV contig assembled from SRR13480248
490 (Supplementary Materials) because the genome sequence of co-existed PaBV is slightly different
491 from the ones deposited in INSDC.

492

493

494 **Acknowledgements**

495 The super-computing resources were provided by Human Genome Center, the Institute of
496 Medical Science, the University of Tokyo.

This study was supported by KAKENHI grant numbers 21H01199 (MH), 22K19234 (MI and MH), 23K20902 (MH), 24K21922 (MH), and 24K18455 (MK) and the 2024 Osaka Metropolitan University (OMU) Strategic Research Promotion Project (Young Researcher) (MK).

501 **References**

502 1. Kuhn JH, Babaian A, Bergner LM, Dény P, Glebe D, Horie M, et al. ICTV
503 Virus Taxonomy Profile: Kolmioviridae 2024. J Gen Virol. 2024;105(2):001963.
504 2. Rizzetto M, Canese MG, Aricò S, Crivelli O, Trepo C, Bonino F, et al.
505 Immunofluorescence detection of new antigen-antibody system (delta/anti-delta)
506 associated to hepatitis B virus in liver and in serum of HBsAg carriers. Gut. 1977 Dec
507 1;18(12):997–1003.
508 3. Iwamoto M, Shibata Y, Kawasaki J, Kojima S, Li YT, Iwami S, et al.
509 Identification of novel avian and mammalian deltaviruses provides new insights into
510 deltavirus evolution. Virus Evol. 2021 Jan 20;7(1):veab003.
511 4. Wille M, Netter HJ, Littlejohn M, Yuen L, Shi M, Eden JS, et al. A Divergent
512 Hepatitis D-Like Agent in Birds. Viruses. 2018 Dec;10(12):720.
513 5. Hetzel U, Szirovicza L, Smura T, Prähauser B, Vapalahti O, Kipar A, et al.
514 Identification of a Novel Deltavirus in Boa Constrictors. mBio. 2019 Apr
515 2;10(2):10.1128/mbio.00014-19.
516 6. Chang WS, Pettersson JHO, Le Lay C, Shi M, Lo N, Wille M, et al. Novel
517 hepatitis D-like agents in vertebrates and invertebrates. Virus Evol. 2019 Jul
518 1;5(2):vez021.
519 7. Paraskevopoulou S, Pirzer F, Goldmann N, Schmid J, Corman VM, Gottula
520 LT, et al. Mammalian deltavirus without hepadnavirus coinfection in the neotropical
521 rodent Proechimys semispinosus. Proc Natl Acad Sci. 2020 Jul 28;117(30):17977–83.
522 8. Bergner LM, Orton RJ, Broos A, Tello C, Becker DJ, Carrera JE, et al.
523 Diversification of mammalian deltaviruses by host shifting. Proc Natl Acad Sci. 2021
524 Jan 19;118(3):e2019907118.
525 9. Wang KS, Choo QL, Weiner AJ, Ou JH, Najarian RC, Thayer RM, et al.
526 Structure, sequence and expression of the hepatitis delta (δ) viral genome. Nature. 1986
527 Oct;323(6088):508–14.
528 10. Macnaughton TB, Beard MR, Chao M, Gowans EJ, Lai MMC. Endogenous
529 Promoters Can Direct the Transcription of Hepatitis Delta Virus RNA from a
530 Recircularized cDNA Template. Virology. 1993 Oct 1;196(2):629–36.
531 11. Jeng KS, Daniel A, Lai MM. A pseudoknot ribozyme structure is active in vivo
532 and required for hepatitis delta virus RNA replication. J Virol. 1996 Apr;70(4):2403–
533 10.
534 12. Asselah T, Rizzetto M. Hepatitis D Virus Infection. N Engl J Med. 2023 Jul
535 5;389(1):58–70.

536 13. Szirovicza L, Hetzel U, Kipar A, Martinez-Sobrido L, Vapalahti O, Hepojoki J.
537 Snake Deltavirus Utilizes Envelope Proteins of Different Viruses To Generate
538 Infectious Particles. mBio. 2020 Mar 17;11(2):e03250-19.
539 14. Bergmann KF, Gerin JL. Antigens of Hepatitis Delta Virus in the Liver and
540 Serum of Humans and Animals. J Infect Dis. 1986 Oct 1;154(4):702–6.
541 15. Bonino F, Heermann KH, Rizzetto M, Gerlich WH. Hepatitis delta virus:
542 protein composition of delta antigen and its hepatitis B virus-derived envelope. J Virol.
543 1986 Jun;58(3):945–50.
544 16. Wong SK, Lazinski DW. Replicating hepatitis delta virus RNA is edited in the
545 nucleus by the small form of ADAR1. Proc Natl Acad Sci. 2002 Nov 12;99(23):15118–
546 23.
547 17. Taylor JM. Infection by Hepatitis Delta Virus. Viruses. 2020 Jun;12(6):648.
548 18. Lee CH, Chang SC, Wu CHH, Chang MF. A Novel Chromosome Region
549 Maintenance 1-independent Nuclear Export Signal of the Large Form of Hepatitis Delta
550 Antigen That Is Required for the Viral Assembly *. J Biol Chem. 2001 Mar
551 1;276(11):8142–8.
552 19. O'Malley B, Lazinski DW. Roles of Carboxyl-Terminal and Farnesylated
553 Residues in the Functions of the Large Hepatitis Delta Antigen. J Virol. 2005 Jan
554 15;79(2):1142–53.
555 20. Khalfi P, Denis Z, McKellar J, Merolla G, Chavey C, Ursic-Bedoya J, et al.
556 Comparative analysis of human, rodent and snake deltavirus replication. PLOS Pathog.
557 2024 Mar 5;20(3):e1012060.
558 21. Khalfi P, Denis Z, Merolla G, Chavey C, Ursic-Bedoya J, Soppa L, et al.
559 Unraveling human, rodent and snake Kolmioviridae replication to anticipate cross-
560 species transmission [Internet]. bioRxiv; 2023 [cited 2023 Oct 20]. p.
561 2023.05.17.541162. Available from:
562 https://www.biorxiv.org/content/10.1101/2023.05.17.541162v1
563 22. Kawasaki J, Kojima S, Tomonaga K, Horie M. Hidden Viral Sequences in
564 Public Sequencing Data and Warning for Future Emerging Diseases. mBio. 2021 Aug
565 17;12(4):10.1128/mbio.01638-21.
566 23. Huang Z, De O. Furo I, Liu J, Peona V, Gomes AJB, Cen W, et al. Recurrent
567 chromosome reshuffling and the evolution of neo-sex chromosomes in parrots. Nat
568 Commun. 2022 Feb 17;13(1):944.
569 24. de la Peña M, Ceprián R, Casey JL, Cervera A. Hepatitis delta virus-like
570 circular RNAs from diverse metazoans encode conserved hammerhead ribozymes.
571 Virus Evol. 2021 Jan 20;7(1):veab016.

572 25. Beard MR, MacNaughton TB, Gowans EJ. Identification and characterization
573 of a hepatitis delta virus RNA transcriptional promoter. J Virol. 1996 Aug;70(8):4986–
574 95.
575 26. Wilkinson ME, Charenton C, Nagai K. RNA Splicing by the Spliceosome.
576 Annu Rev Biochem. 2020 Jun 20;89(Volume 89, 2020):359–88.
577 27. Edgar RC, Taylor J, Lin V, Altman T, Barbera P, Meleshko D, et al. Petabase-
578 scale sequence alignment catalyses viral discovery. Nature. 2022 Feb;602(7895):142–7.
579 28. Dubois J, Terrier O, Rosa-Calatrava M. Influenza Viruses and mRNA Splicing:
580 Doing More with Less. mBio. 2014 May 13;5(3):10.1128/mbio.00070-14.
581 29. Schneider PA, Schneemann A, Lipkin WI. RNA splicing in Borna disease
582 virus, a nonsegmented, negative-strand RNA virus. J Virol. 1994 Aug;68(8):5007–12.
583 30. Cubitt B, Oldstone C, Valcarcel J, de la Torre JC. RNA splicing contributes to
584 the generation of mature mRNAs of Borna disease virus, a non-segmented negative
585 strand RNA virus. Virus Res. 1994 Oct 1;34(1):69–79.
586 31. Kuwata R, Isawa H, Hoshino K, Tsuda Y, Yanase T, Sasaki T, et al. RNA
587 Splicing in a New Rhabdovirus from Culex Mosquitoes. J Virol. 2011 Jul;85(13):6185–
588 96.
589 32. Elena SF, Dopazo J, Flores R, Diener TO, Moya A. Phylogeny of viroids,
590 viroidlike satellite RNAs, and the viroidlike domain of hepatitis delta virus RNA. Proc
591 Natl Acad Sci. 1991 Jul;88(13):5631–4.
592 33. Jenkins GM, Woelk CH, Rambaut A, Holmes EC. Testing the Extent of
593 Sequence Similarity Among Viroids, Satellite RNAs, and Hepatitis Delta Virus. J Mol
594 Evol. 2000 Jan 1;50(1):98–102.
595 34. Forgia M, Navarro B, Daghino S, Cervera A, Gisel A, Perotto S, et al. Hybrids
596 of RNA viruses and viroid-like elements replicate in fungi. Nat Commun. 2023 May
597 5;14(1):2591.
598 35. Zheludev IN, Edgar RC, Lopez-Galiano MJ, Peña M de la, Babaian A, Bhatt
599 AS, et al. Viroid-like colonists of human microbiomes [Internet]. bioRxiv; 2024 [cited
600 2024 Aug 29]. p. 2024.01.20.576352. Available from:
601 https://www.biorxiv.org/content/10.1101/2024.01.20.576352v1
602 36. Kuhn JH, Botella L, de la Peña M, Vainio EJ, Krupovic M, Lee BD, et al.
603 Ambiviricota, a novel ribovirian phylum for viruses with viroid-like properties. J Virol.
604 2024 Jun 10;98(7):e00831-24.
605 37. Wang TC, Chao M. RNA Recombination of Hepatitis Delta Virus in Natural
606 Mixed-Genotype Infection and Transfected Cultured Cells. J Virol. 2005 Feb
607 15;79(4):2221–9.

608 38. Rizzetto M. The adventure of delta. Liver Int. 2016;36(S1):135–40.
609 39. Honkavuori KS, Shivaprasad HL, Williams BL, Quan PL, Hornig M, Street C,
610 et al. Novel Borna Virus in Psittacine Birds with Proventricular Dilatation Disease.
611 Emerg Infect Dis. 2008 Dec;14(12):1883–6.
612 40. Kistler AL, Gancz A, Clubb S, Skewes-Cox P, Fischer K, Sorber K, et al.
613 Recovery of divergent avian bornaviruses from cases of proventricular dilatation
614 disease: Identification of a candidate etiologic agent. Virol J. 2008 Jul 31;5(1):88.
615 41. Rubbenstroth D, Rinder M, Stein M, Höper D, Kaspers B, Brosinski K, et al.
616 Avian bornaviruses are widely distributed in canary birds (*Serinus canaria* f.
617 domestica). Vet Microbiol. 2013 Aug 30;165(3):287–95.
618 42. Gancz AY, Kistler AL, Greninger AL, Farnoushi Y, Mechani S, Perl S, et al.
619 Experimental induction of proventricular dilatation disease in cockatiels (Nymphicus
620 hollandicus) inoculated with brain homogenates containing avian bornavirus 4. Virol J.
621 2009 Jul 9;6(1):100.
622 43. Sayers EW, Bolton EE, Brister JR, Canese K, Chan J, Comeau DC, et al.
623 Database resources of the National Center for Biotechnology Information in 2023.
624 Nucleic Acids Res. 2023 Jan 6;51(D1):D29–38.
625 44. Boratyn GM, Thierry-Mieg J, Thierry-Mieg D, Busby B, Madden TL. Magic-
626 BLAST, an accurate RNA-seq aligner for long and short reads. BMC Bioinformatics.
627 2019 Jul 25;20(1):405.
628 45. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, et al.
629 Twelve years of SAMtools and BCFtools. GigaScience. 2021 Feb 1;10(2):giab008.
630 46. Chen S. Ultrafast one-pass FASTQ data preprocessing, quality control, and
631 deduplication using fastp. iMeta. 2023;2(2):e107.
632 47. Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. Graph-based genome
633 alignment and genotyping with HISAT2 and HISAT-genotype. Nat Biotechnol. 2019
634 Aug;37(8):907–15.
635 48. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, et al.
636 Full-length transcriptome assembly from RNA-Seq data without a reference genome.
637 Nat Biotechnol. 2011 Jul;29(7):644–52.
638 49. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al.
639 BLAST+: architecture and applications. BMC Bioinformatics. 2009 Dec 15;10(1):421.
640 50. Noé L, Kucherov G. YASS: enhancing the sensitivity of DNA similarity
641 search. Nucleic Acids Res. 2005 Jul 1;33(suppl_2):W540–3.

642 51. Katoh K, Standley DM. MAFFT Multiple Sequence Alignment Software
643 Version 7: Improvements in Performance and Usability. Mol Biol Evol. 2013 Apr
644 1;30(4):772–80.
645 52. Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, et al.
646 MrBayes 3.2: Efficient Bayesian Phylogenetic Inference and Model Choice Across a
647 Large Model Space. Syst Biol. 2012 May 1;61(3):539–42.
648 53. Kozlov AM, Darriba D, Flouri T, Morel B, Stamatakis A. RAxML-NG: a fast,
649 scalable and user-friendly tool for maximum likelihood phylogenetic inference.
650 Bioinformatics. 2019 Nov 1;35(21):4453–5.
651 54. Lemoine F, Domelevo Entfellner JB, Wilkinson E, Correia D, Dávila Felipe M,
652 De Oliveira T, et al. Renewing Felsenstein's phylogenetic bootstrap in the era of big
653 data. Nature. 2018 Apr;556(7702):452–6.
654 55. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v5: an online tool for
655 phylogenetic tree display and annotation. Nucleic Acids Res. 2021 Jul 2;49(W1):293–6.
656 56. Zuker M. Mfold web server for nucleic acid folding and hybridization
657 prediction. Nucleic Acids Res. 2003 Jul 1;31(13):3406–15.
658 57. Sato K, Kato Y, Hamada M, Akutsu T, Asai K. IPknot: fast and accurate
659 prediction of RNA secondary structures with pseudoknots using integer programming.
660 Bioinformatics. 2011 Jul 1;27(13):i85–93.
661 58. Hallgren J, Tsirigos KD, Pedersen MD, Armenteros JJA, Marcatili P, Nielsen
662 H, et al. DeepTMHMM predicts alpha and beta transmembrane proteins using deep
663 neural networks [Internet]. bioRxiv; 2022 [cited 2024 Sep 20]. p. 2022.04.08.487609.
664 Available from: https://www.biorxiv.org/content/10.1101/2022.04.08.487609v1
665 59. Kosugi S, Hasebe M, Tomita M, Yanagawa H. Systematic identification of cell
666 cycle-dependent yeast nucleocytoplasmic shuttling proteins by prediction of composite
667 motifs. Proc Natl Acad Sci. 2009 Jun 23;106(25):10171–6.
668 60. Erdős G, Dosztányi Z. AIUPred: combining energy estimation with deep
669 learning for the enhanced prediction of protein disorder. Nucleic Acids Res. 2024 Jul
670 5;52(W1):W176–81.
671 61. Drozdetskiy A, Cole C, Procter J, Barton GJ. JPred4: a protein secondary
672 structure prediction server. Nucleic Acids Res. 2015 Jul 1;43(W1):W389–94.
673 62. Abramson J, Adler J, Dunger J, Evans R, Green T, Pritzel A, et al. Accurate
674 structure prediction of biomolecular interactions with AlphaFold 3. Nature. 2024
675 Jun;630(8016):493–500.
676 63. Fu L, Niu B, Zhu Z, Wu S, Li W. CD-HIT: accelerated for clustering the next-
677 generation sequencing data. Bioinformatics. 2012 Dec 1;28(23):3150–2.

678 64. Shen W, Le S, Li Y, Hu F. SeqKit: A Cross-Platform and Ultrafast Toolkit for
679 FASTA/Q File Manipulation. PLOS ONE. 2016 Oct 5;11(10):e0163962.
680 65. Steinegger M, Söding J. MMseqs2 enables sensitive protein sequence
681 searching for the analysis of massive data sets. Nat Biotechnol. 2017 Nov;35(11):1026–
682 8.
683

684 **Figure legends**

685 **Fig. 1 Phylogenetic relationships of kolmioviruses.** The phylogenetic tree was inferred by the
686 Bayesian Markov chain Monte Carlo method based on an amino acid sequence alignment of DAg.
687 Posterior probabilities are shown for all nodes. Figure arts were created with BioRender.com or
688 downloaded from https://www.phylopic.org/. The scale bar indicates the number of amino acid
689 substitutions per site.

690

691 **Fig. 2 Phylogenetic analysis based on nucleotide sequences of complete scKoV genomes.** (**A**)
692 The phylogenetic tree was inferred by the maximum likelihood method using nucleotide
693 sequences of complete scKoV genomes. Bootstrap values are indicated at each branch. The
694 BioProjects of each SRA data that scKoV sequences originated are indicated with dark blue, blue,
695 light blue, and red circles, respectively. The scale bar indicates the number of nucleotide
696 substitutions per site. (**B**) Heat map of pairwise identities of scKoV variants based on nucleotide
697 sequences of complete genomes (lower left) and amino acid sequences of DAg (upper right). (**C**)
698 Detection of mixed infections with scKoV of different genogroups. For BioProject
699 PRJNA679636, their tissue of origin is indicated.

700

701 **Fig. 3 Genome organization and transcriptional pattern of scKoV.** (**A**) Self-
702 complementarities of scKoV. The predicted RNA structures were visualized using the Mfold web
703 server (56). (B) Genome organization of scKoV. Colored arrows indicate annotations (ORFs,
704 ribozymes, and poly-A signals). The numbers indicate nucleotide positions. (**C**) Predicted
705 genomic and antigenomic structures of scKoV. Gray numbers indicate nucleotide positions of
706 each predicted ribozyme. The names of secondary structural elements of ribozymes are indicated
707 in blue. Red triangles indicate catalytic sites. Information of SRR10526804 is indicated as a

708 representative. Information about other variants is provided in Figs. S2, S3, and S4. (**D**) Strand-
709 specific mapping coverage of original short reads of scKoV full-genome sequence. Colored
710 arrows indicate annotations (ORFs, ribozymes, poly-A signals). Light pink box indicates putative
711 intron. Green triangle indicates edge of the rod-like structure of circulated scKoV genome.
712 Information of SRR10526804 is indicated as a representative. Information about other variants is
713 provided in Fig. S5. (**E**) Secondary structures and expanded mapping patterns of the putative
714 promoter regions of scKoV. Conserved domains were annotated as previously reported (25). (**F**)
715 Nucleotide alignment of predicted intron region using representative variants of scKoV in each
716 genotype. Dots indicate common bases in the alignment. Gray and blue box indicate conserved
717 GU-AG boundary and polypyrimidine tract, respectively. Gray numbers indicate nucleotide
718 position in alignment of full-genome sequences.
719
720 **Fig. 4 Co-existing viruses in scKoV-positive RNA-seq data.** (**A**) Co-existing avian bornaviruses
721 (ABVs) (left panel) and scKoV-derived read number (right panel) detected from each RNA-seq
722 dataset are shown. Black indicates the presence of a viral contig, while white indicates its absence.
723 Read per million (RPM) of scKoV is shown as a heatmap. Results from RNA-seq data with
724 scKoV RPM $\geq$ 5 are shown. Full information is available in Fig. S8. (**B**) Scatter plot showing
725 RPM of scKoV and ABVs of each RNA-seq data. The linear prediction is shown as a dotted line.
726 The correlation coefficient (R^2) of the linear prediction is shown on the graph.
727
728 **Fig. 5 identification of other potential ambisense KoV.** (A) Phylogenetic tree of all known
729 KoVs. The tree was inferred by the Bayesian Markov chain Monte Carlo method based on an
730 amino acid sequence alignment of DAg. Posterior probabilities are shown for all nodes. The scale
731 bar indicates the number of amino acid substitutions per site. KoVs with potential ambisense

genome structures are shown in blue, and those with confirmed ambisense transcription patterns are shown in red. (**B**) Genome organization of Amboli leaping frog KoV (icKoV). Colored arrows indicate annotations (ORFs, ribozymes, and poly-A signals). The numbers indicate nucleotide positions. (**C**) Strand-specific mapping coverage of original short reads of icKoV. Colored arrows indicate annotations (ORFs, ribozymes, poly-A signals). Green triangle indicates edge of the rod-like structure of circulated scKoV genome. (**D**) Amino acid alignment of scKoV ORF2a and icKoV ORF2. Identical amino acid residues are highlighted in black, similar substitutions in grey, and gaps are shown as dashes.

Table 1. Detection of scDeV reads in public RNA-seq data

Bio Project	SRA	Host				RPM (read per million)[a]	Tissue	full-length scDeV genome		
		Common name	Taxonomy					DDBJ accession	Length (bp)	GC content (%)
			Order	Family	Species					
PRJNA300534[b]	SRR2915371[c]	Common canary	*Passeriformes*	*Fringillidae*	*Serinus canaria*	10.17	Skin	BR002460	1,672	54.7
PRJNA939570	SRR23747102					41.63	Fat tissue	BR002469	1,672	
	SRR23747110					96.01	Fat tissue	BR002470	1,674	
	SRR23747114					5.26	Fat tissue	BR002471	1,676	
	SRR23747122					57.30	Fat tissue	BR002472	1,672	
PRJNA591356[b]	SRR10526804					758.69	Skin	BR002461[d]	1,672	54.5
	SRR10526806					570.30	Skin	BR002464[d]	1,672	54.6
	SRR10526807					525.41	Skin	BR002462	1,672	54.7
	SRR10526821					153.89	Skin	BR002465[d]	1,676	55.5
	SRR10526846					809.00	Skin	BR002461[d]	1,672	54.5
	SRR10526848					685.58	Skin	BR002464[d]	1,672	54.6
	SRR10526849					488.29	Skin			
	SRR10526851					194.19	Skin	BR002465[d]	1,676	55.5
	SRR10526845	European serin	*Passeriformes*	*Fringillidae*	*Serinus serinus*	5.52	Skin			
	SRR10526817	House finch	*Passeriformes*	*Fringillidae*	*Haemorhous mexicanus*	20.35	Skin			
	SRR10526824					16.80	Skin	BR002463	1,672	54.0
PRJNA679636	SRR13480239	Monk parakeet	*Psittaciformes*	*Psittacidae*	*Myiopsitta monachus*	16.38	Brain	BR002466	1,671	54.6
	SRR13480240					38.42	Brain	BR002467	1,671	54.6
	SRR13480246					22.48	Lung			
	SRR13480248					11.49	Ovary	BR002468	1,671	54.9

[a]This table only shows the SRA data wirh RPM > 5. The full version of the table is available as Supplementary Table S1.

[b]PRJNA300534 and PRJNA591356 were deposited by the same group

[c]Identified in the previous study (Iwamoto et al. Virus Evol, 2021)

[d]Identical accession numbers indicate that the genome sequences obtained are identical.

Fig. 1

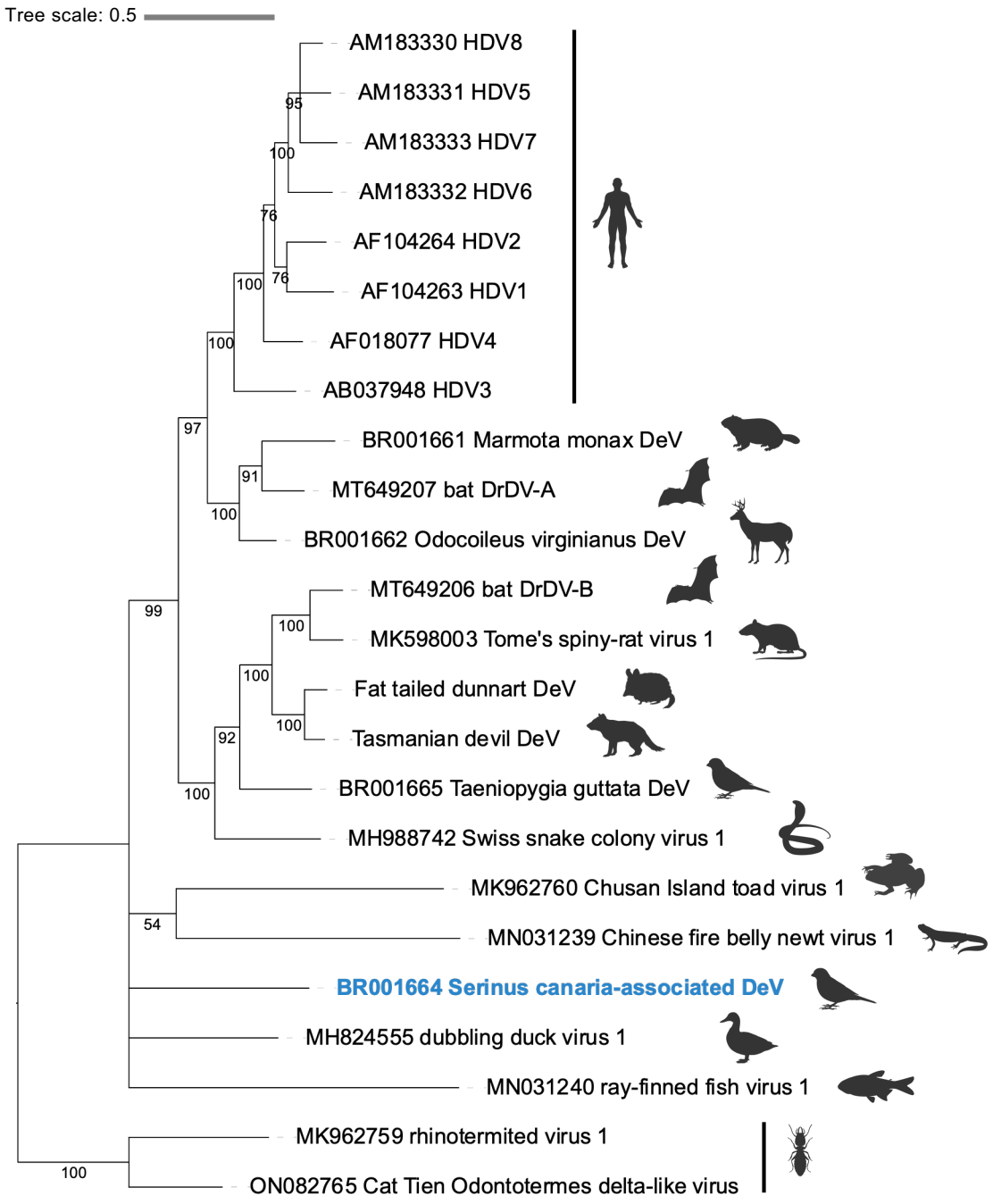

Tree scale: 0.5
AM183330 HDV8
AM183331 HDV5
AM183333 HDV7
AM183332 HDV6
AF104264 HDV2
AF104263 HDV1
AF018077 HDV4
AB037948 HDV3
BR001661 Marmota monax DeV
MT649207 bat DrDV-A
BR001662 Odocoileus virginianus DeV
MT649206 bat DrDV-B
MK598003 Tome's spiny-rat virus 1
Fat tailed dunnart DeV
Tasmanian devil DeV
BR001665 Taeniopygia guttata DeV
MH988742 Swiss snake colony virus 1
MK962760 Chusan Island toad virus 1
MN031239 Chinese fire belly newt virus 1
BR001664 Serinus canaria-associated DeV
MH824555 dubbling duck virus 1
MN031240 ray-finned fish virus 1
MK962759 rhinotermited virus 1
ON082765 Cat Tien Odontotermes delta-like virus
95
100
76
76
100
100
97
91
100
99
100
100
100
92
100
54
100

Fig. 2

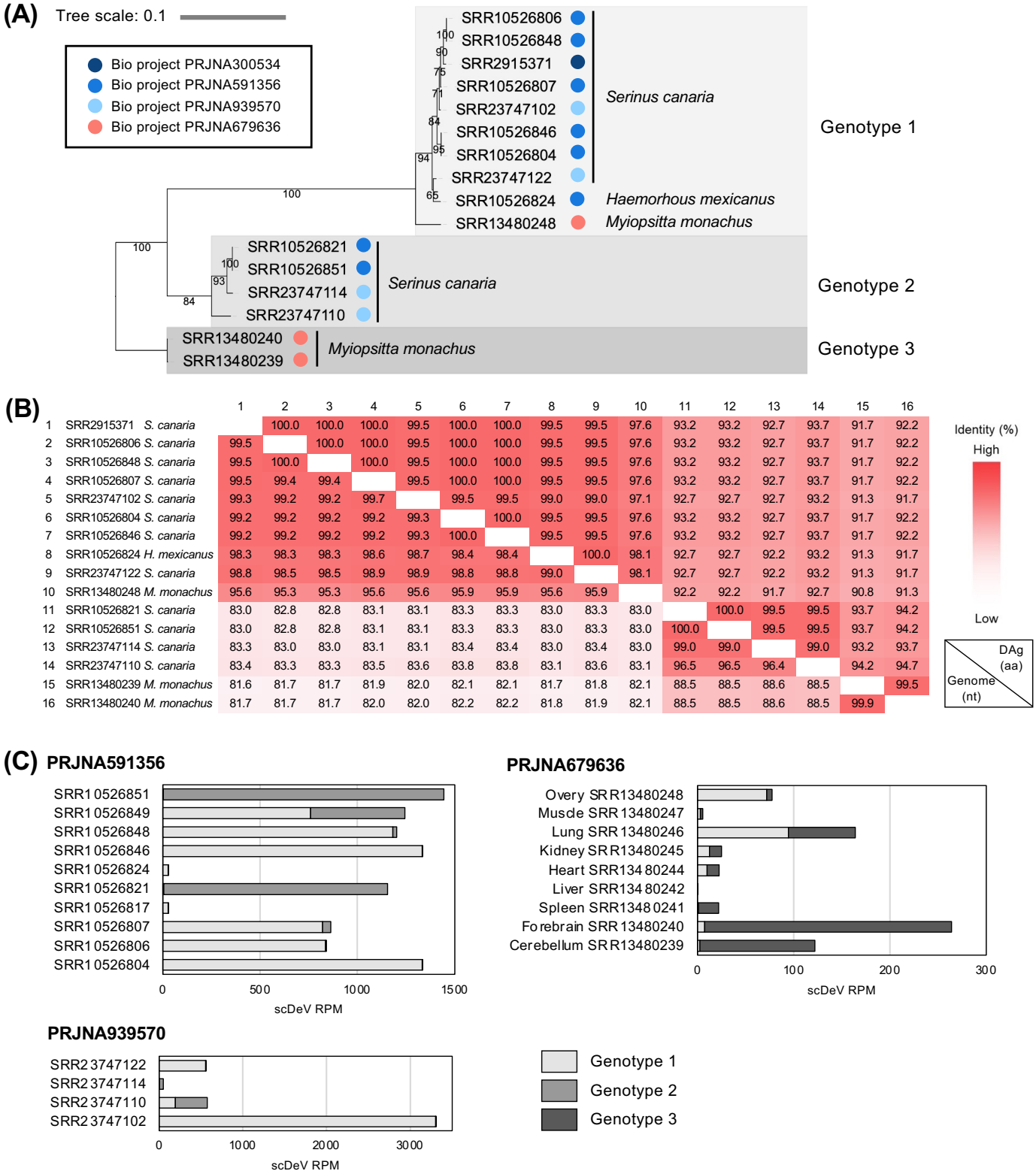

(A)
Tree scale: 0.1
Bio project PRJNA300534
Bio project PRJNA591356
Bio project PRJNA939570
Bio project PRJNA679636
SRR10526806
SRR10526848
SRR2915371
SRR10526807
SRR23747102
SRR10526846
SRR10526804
SRR23747122
SRR10526824
SRR13480248
Serinus canaria
Haemorhous mexicanus
Myiopsitta monachus
Genotype 1
SRR10526821
SRR10526851
SRR23747114
SRR23747110
Serinus canaria
Genotype 2
SRR13480240
SRR13480239
Myiopsitta monachus
Genotype 3
(B)
Identity (%)
High
Low
DAg (aa)
Genome (nt)
(C)
PRJNA591356
SRR10526851
SRR10526849
SRR10526848
SRR10526846
SRR10526824
SRR10526821
SRR10526817
SRR10526807
SRR10526806
SRR10526804
scDeV RPM
PRJNA679636
Overy SRR13480248
Muscle SRR13480247
Lung SRR13480246
Kidney SRR13480245
Heart SRR13480244
Liver SRR13480242
Spleen SRR13480241
Forebrain SRR13480240
Cerebellum SRR13480239
scDeV RPM
PRJNA939570
SRR23747122
SRR23747114
SRR23747110
SRR23747102
scDeV RPM
Genotype 1
Genotype 2
Genotype 3

Fig. 3

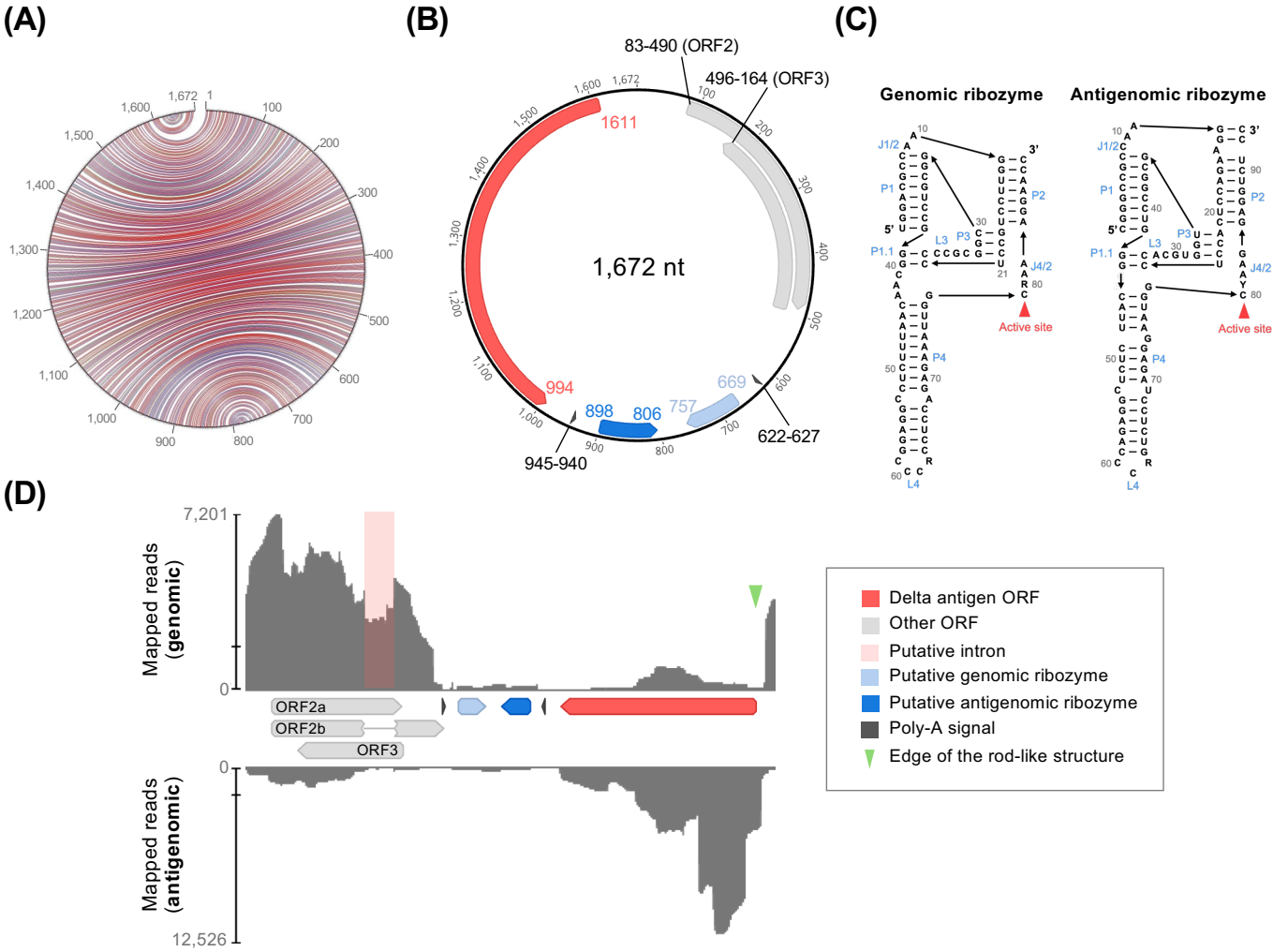

(A)
(B)
(C)
83-490 (ORF2)
496-164 (ORF3)
1611
1,672 nt
994
898
806
757
669
622-627
945-940
Genomic ribozyme
Antigenomic ribozyme
Active site
(D)
Mapped reads (genomic)
Mapped reads (antigenomic)
7,201
12,526
ORF2a
ORF2b
ORF3
Delta antigen ORF
Other ORF
Putative intron
Putative genomic ribozyme
Putative antigenomic ribozyme
Poly-A signal
Edge of the rod-like structure
(E)
Genotype 1 (SRR10526804)
Genotype 2 (SRR10526851)
Turnover
External bulge
Internal bulge
Start codon of DAg ORF
GC-rich
3,452
4,444
1,488
4,181
(F)
SRR10526804 (genotype 1)
SRR10526851 (genotype 2)
SRR13480240 (genotype 3)
375
468
CUCCG GUGAGCACGCAGCUGCGCGA ··· AGUCGCCGGGAGUUCCCUCUCUUCUGGCAG GGGCG
Doner
Polypyrimidine tract
Acceptor
Intron

Fig. 4

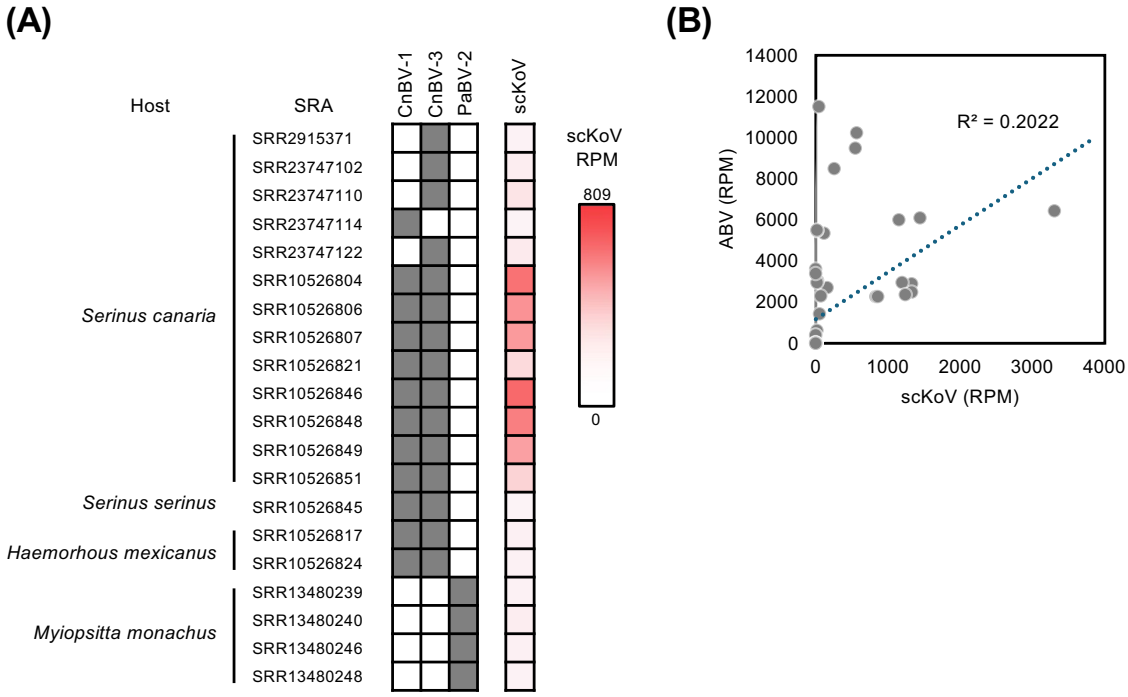
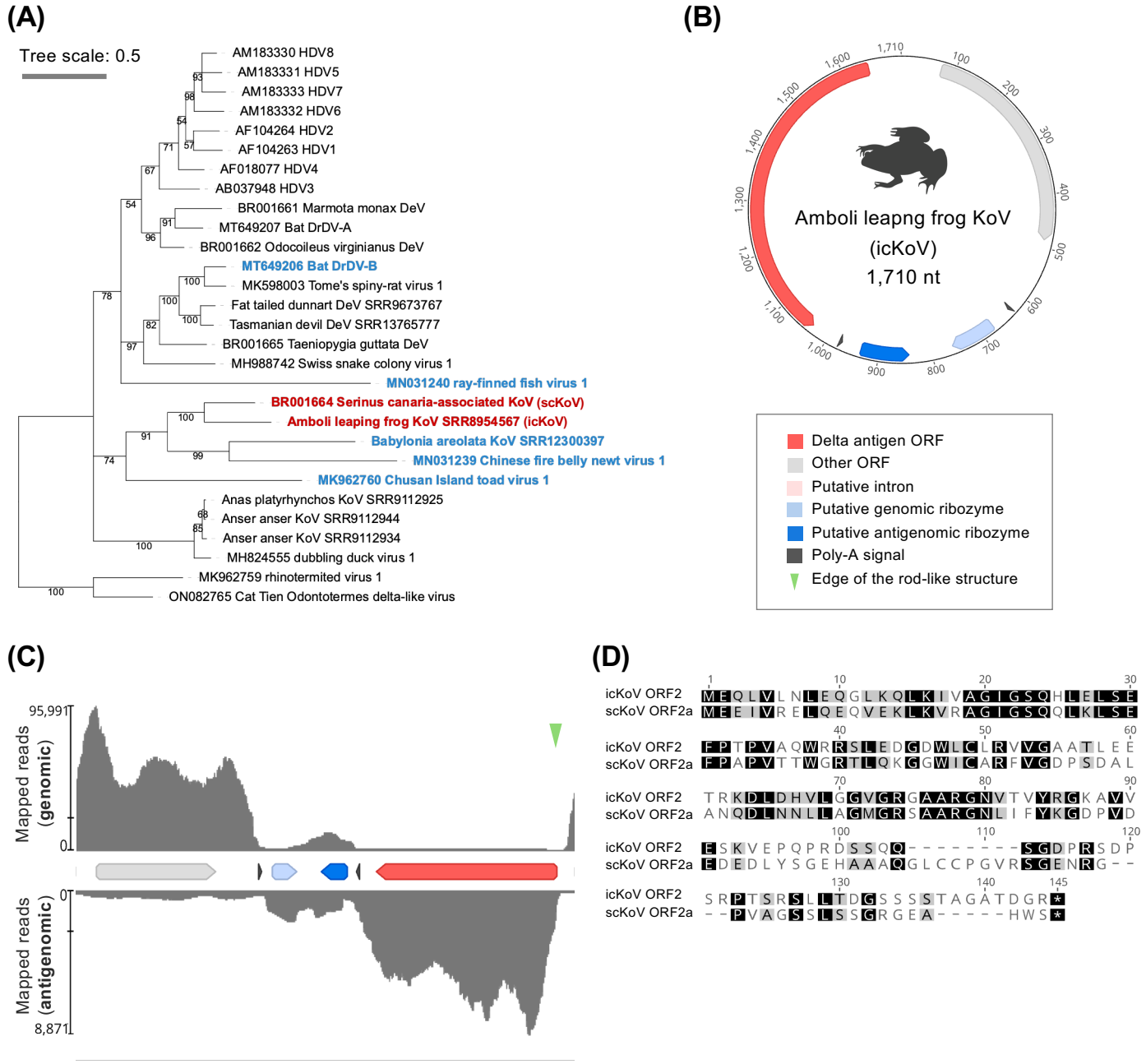

(A)
Host
SRA
CnBV-1
CnBV-3
PaBV-2
scKoV
scKoV RPM
809
0
Serinus canaria
Serinus serinus
Haemorhous mexicanus
Myiopsitta monachus
SRR2915371
SRR23747102
SRR23747110
SRR23747114
SRR23747122
SRR10526804
SRR10526806
SRR10526807
SRR10526821
SRR10526846
SRR10526848
SRR10526849
SRR10526851
SRR10526845
SRR10526817
SRR10526824
SRR13480239
SRR13480240
SRR13480246
SRR13480248
(B)
R² = 0.2022
ABV (RPM)
scKoV (RPM)
0
2000
4000
6000
8000
10000
12000
14000
0
1000
2000
3000
4000

Fig. 5


(A)
Tree scale: 0.5
AM183330 HDV8
AM183331 HDV5
AM183333 HDV7
AM183332 HDV6
AF104264 HDV2
AF104263 HDV1
AF018077 HDV4
AB037948 HDV3
BR001661 Marmota monax DeV
MT649207 Bat DrDV-A
BR001662 Odocoileus virginianus DeV
MT649206 Bat DrDV-B
MK598003 Tome's spiny-rat virus 1
Fat tailed dunnart DeV SRR9673767
Tasmanian devil DeV SRR13765777
BR001665 Taeniopygia guttata DeV
MH988742 Swiss snake colony virus 1
MN031240 ray-finned fish virus 1
BR001664 Serinus canaria-associated KoV (scKoV)
Amboli leaping frog KoV SRR8954567 (icKoV)
Babylonia areolata KoV SRR12300397
MN031239 Chinese fire belly newt virus 1
MK962760 Chusan Island toad virus 1
Anas platyrhynchos KoV SRR9112925
Anser anser KoV SRR9112944
Anser anser KoV SRR9112934
MH824555 dubbling duck virus 1
MK962759 rhinotermited virus 1
ON082765 Cat Tien Odontotermes delta-like virus
(B)
Amboli leapng frog KoV
(icKoV)
1,710 nt
Delta antigen ORF
Other ORF
Putative intron
Putative genomic ribozyme
Putative antigenomic ribozyme
Poly-A signal
Edge of the rod-like structure
(C)
Mapped reads (genomic)
95,991
0
Mapped reads (antigenomic)
0
8,871
(D)
icKoV ORF2
scKoV ORF2a