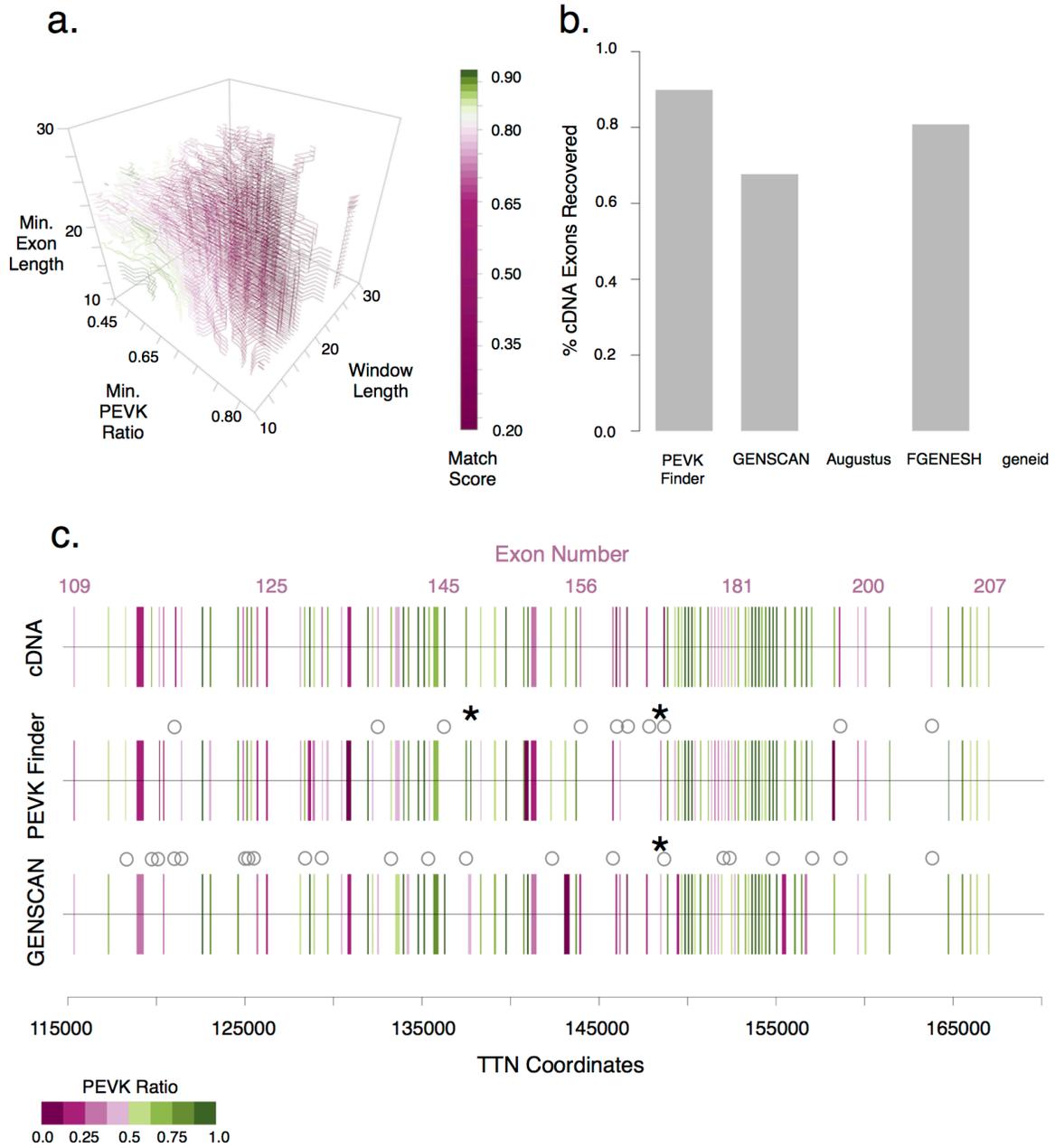
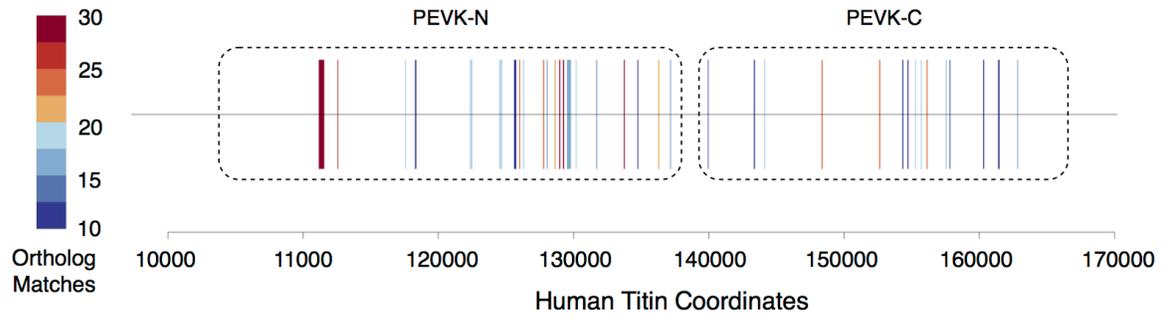


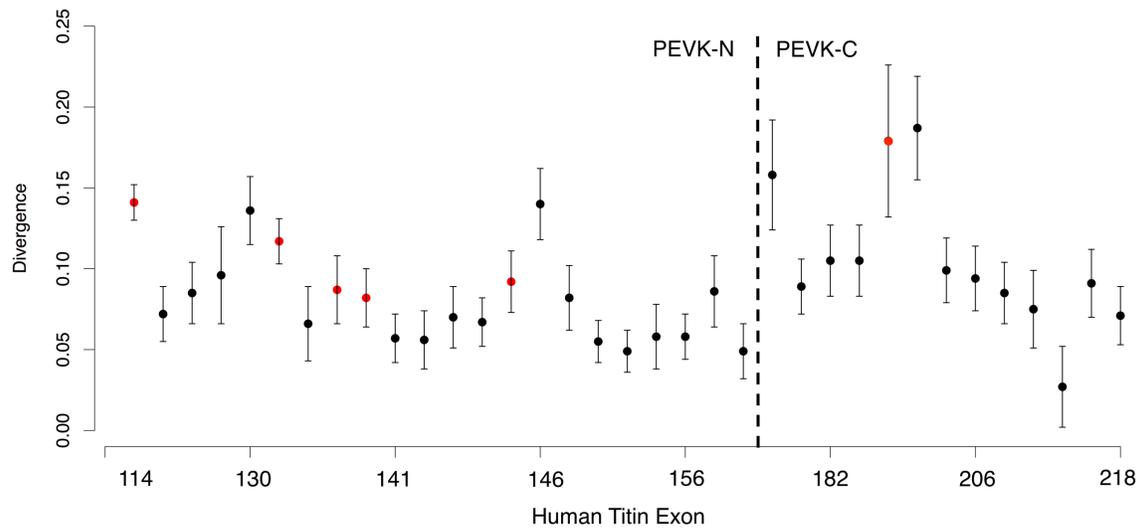
**Figure S1.**



**Figure S2.**



**Figure S3.**



**Figure S4.**

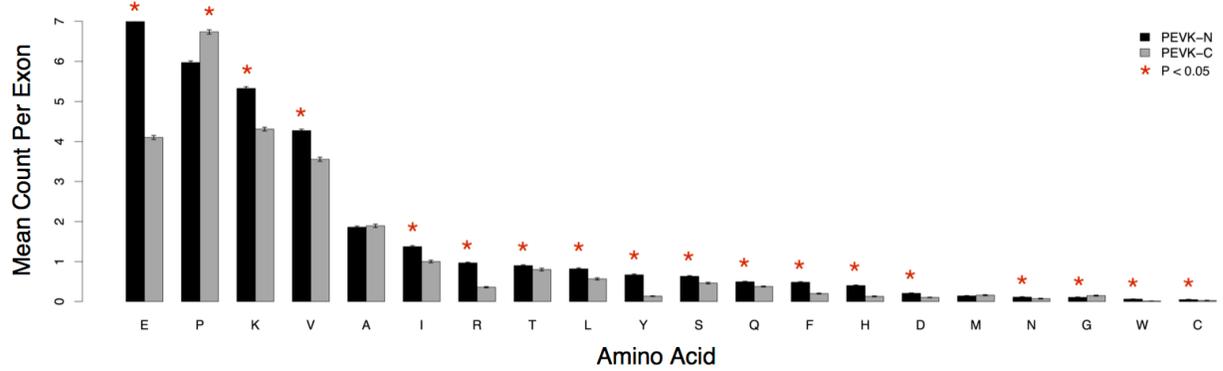








Figure S11.

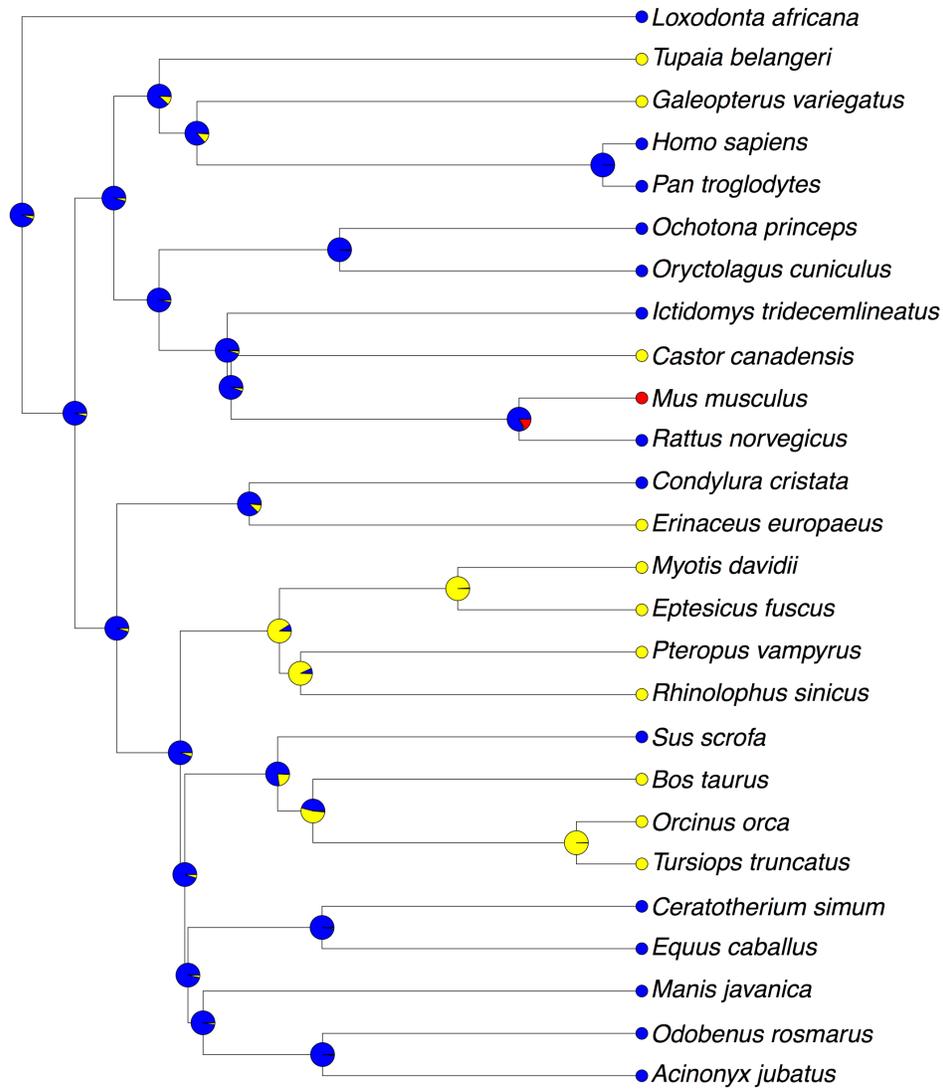


Figure S12.

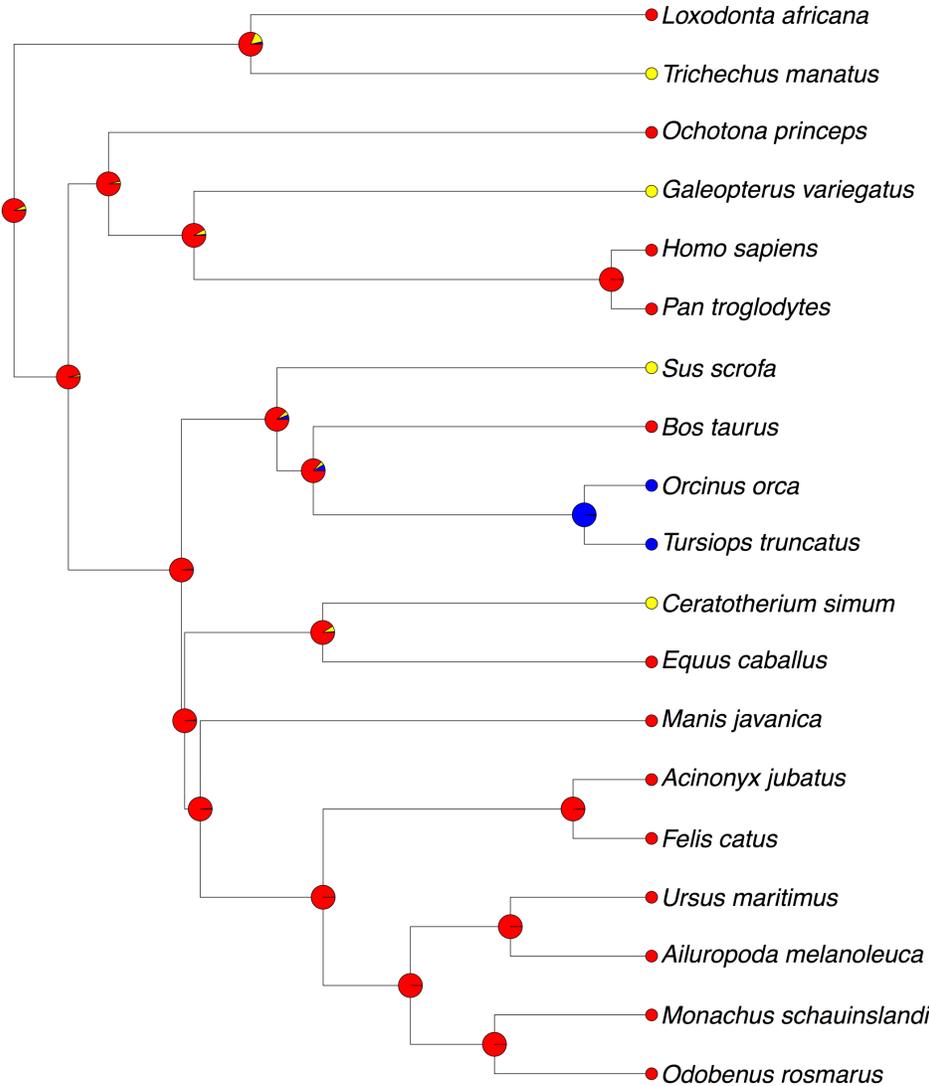


Figure S13.

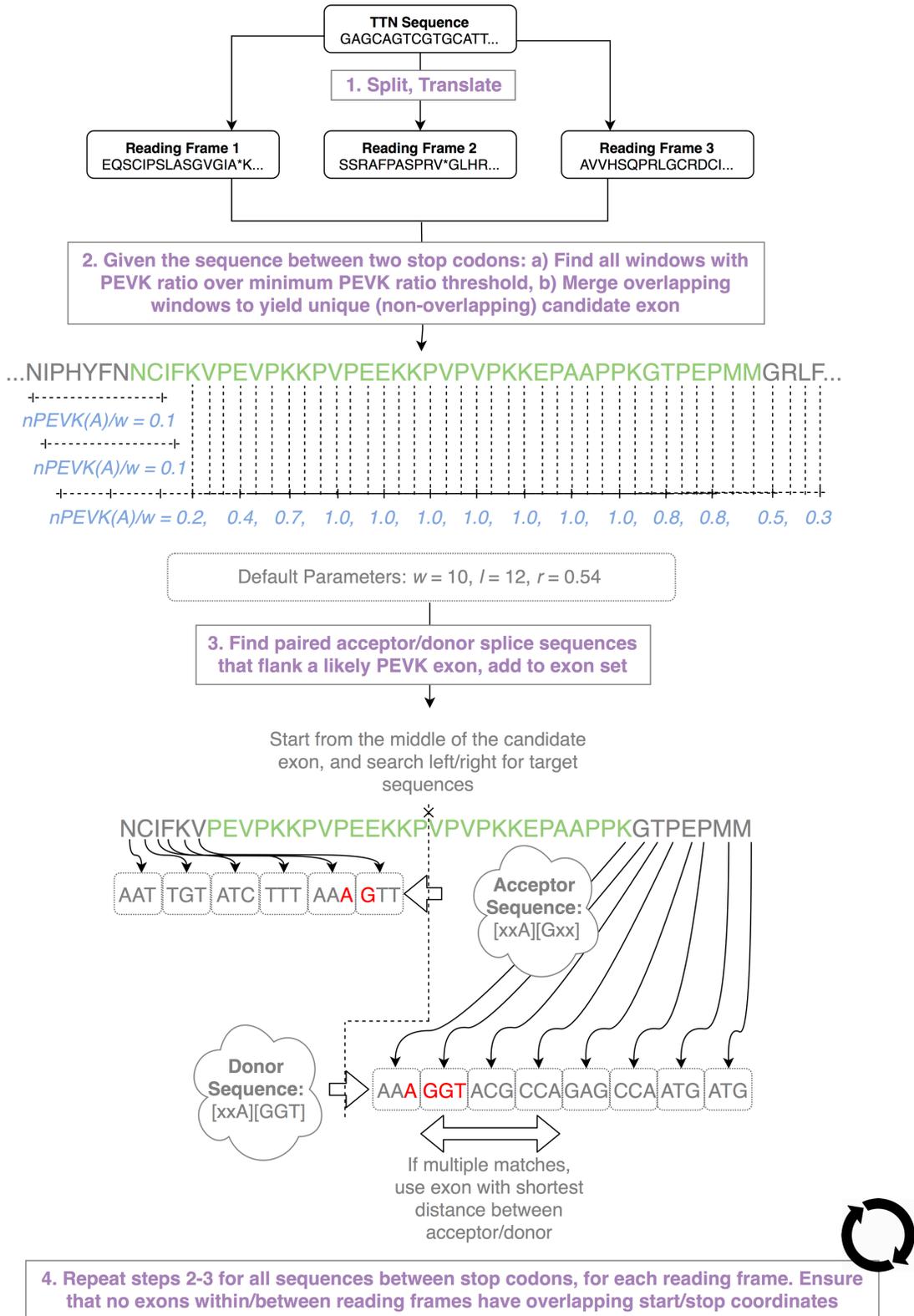


Figure S14.

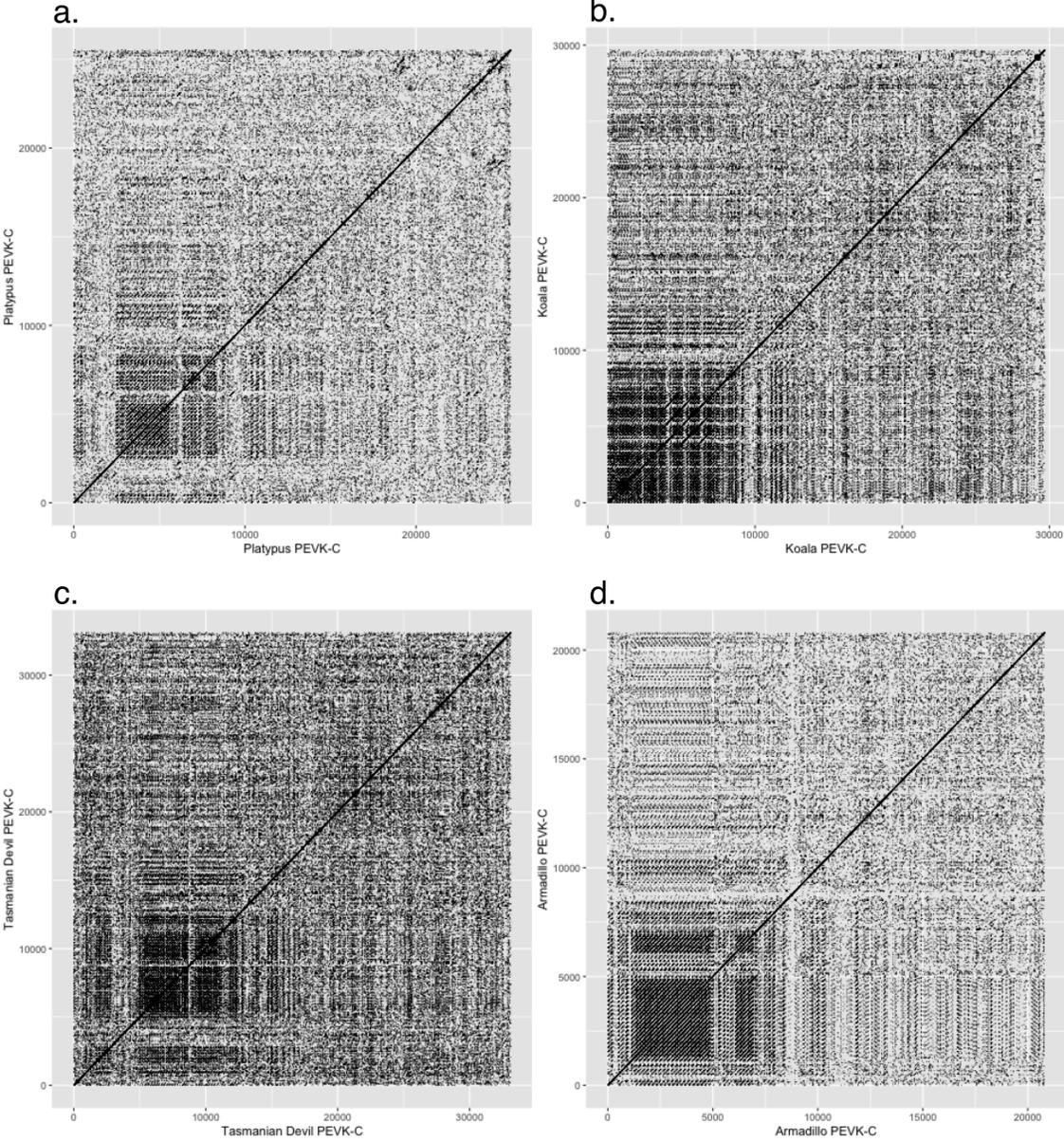


Figure S15.

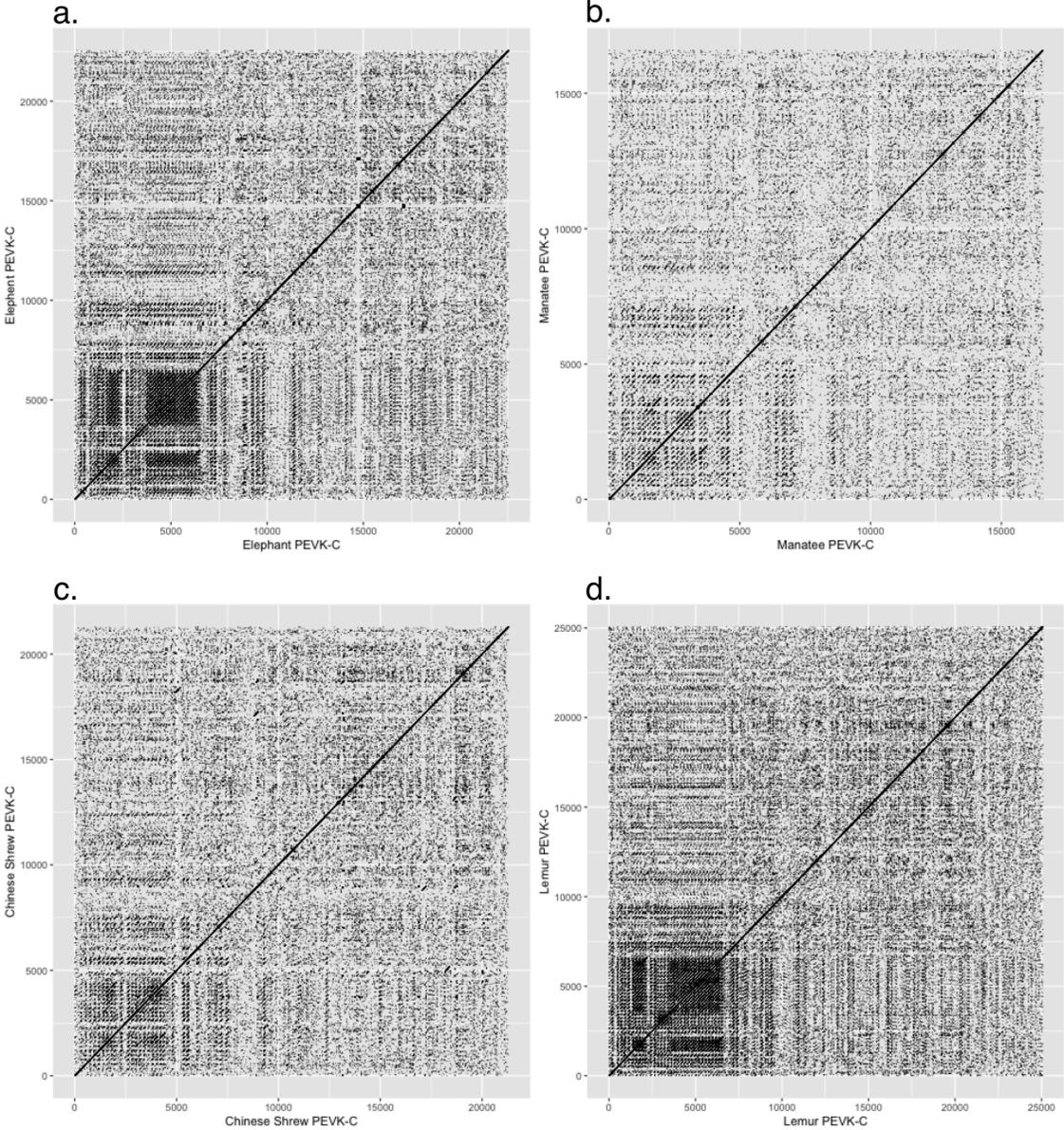


Figure S16.

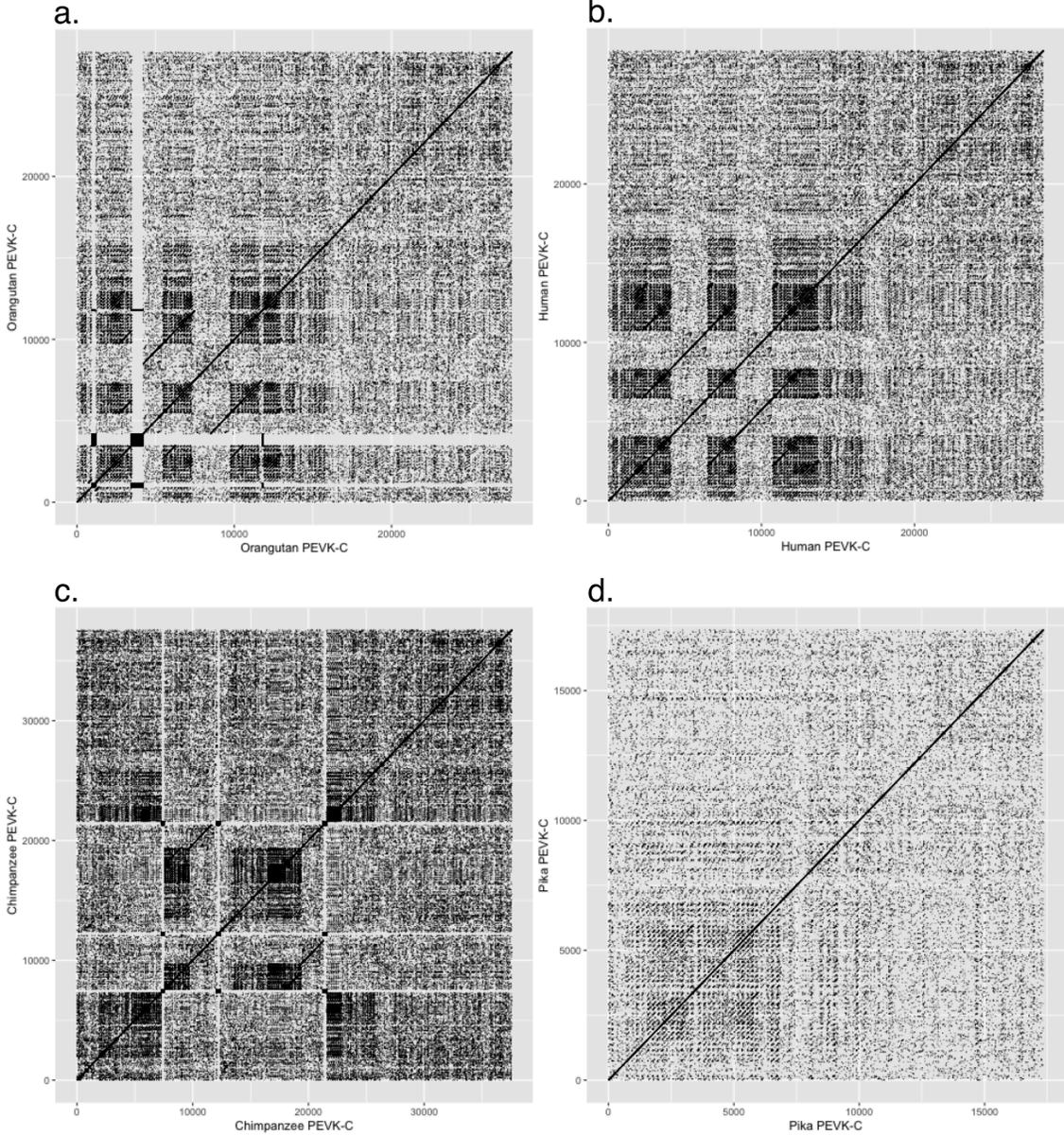


Figure S17.

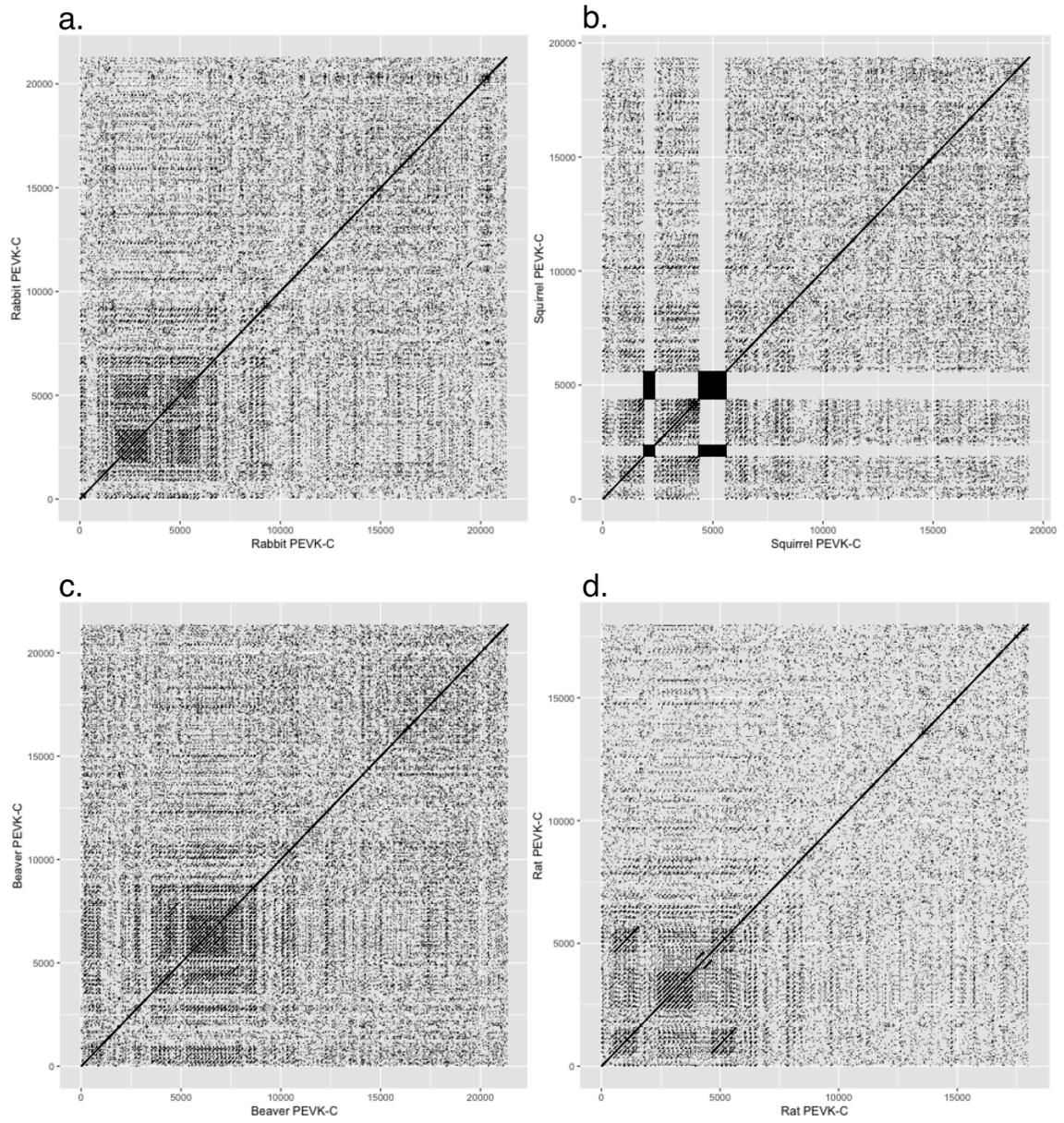
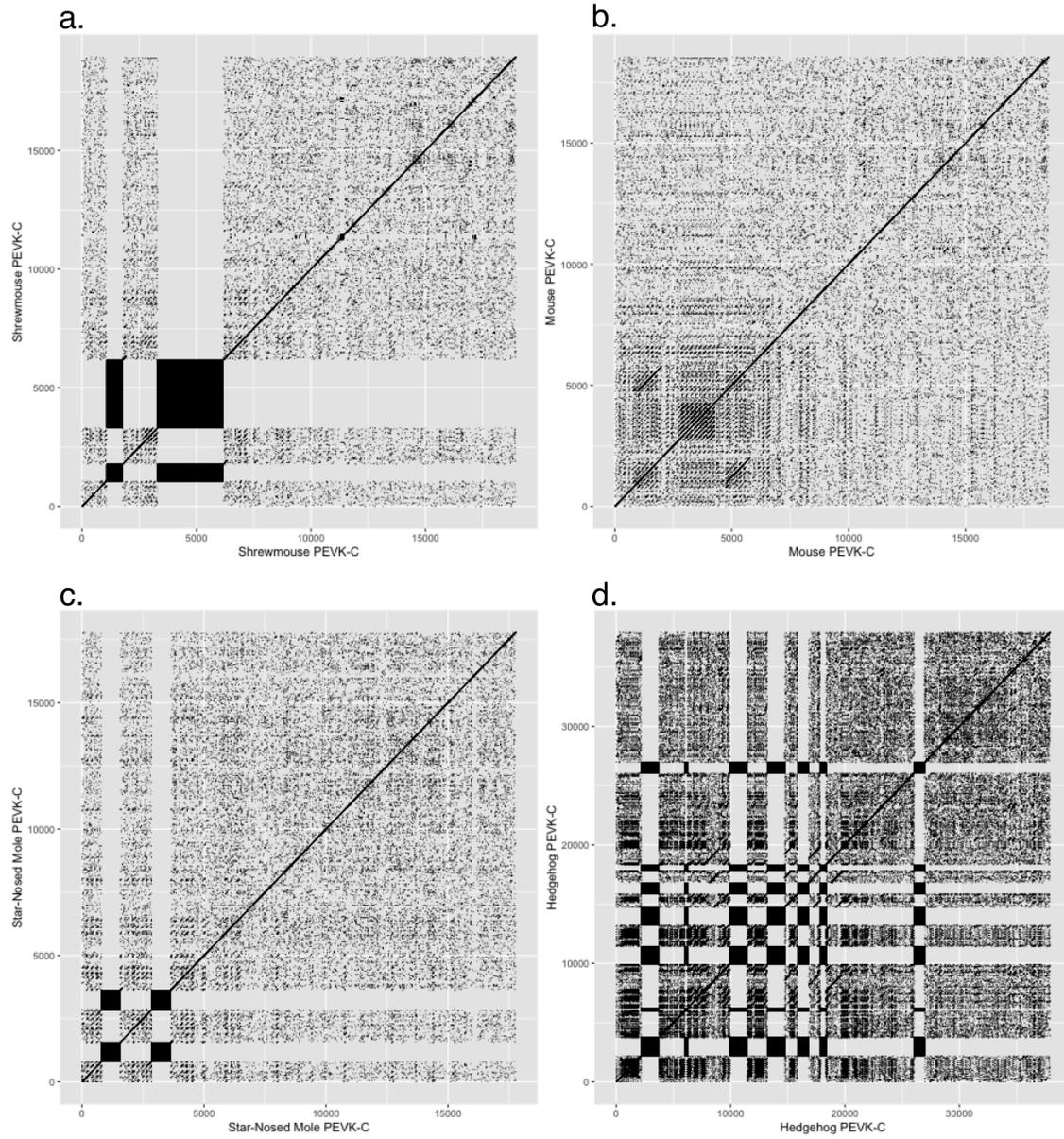


Figure S18.



**Figure S19.**

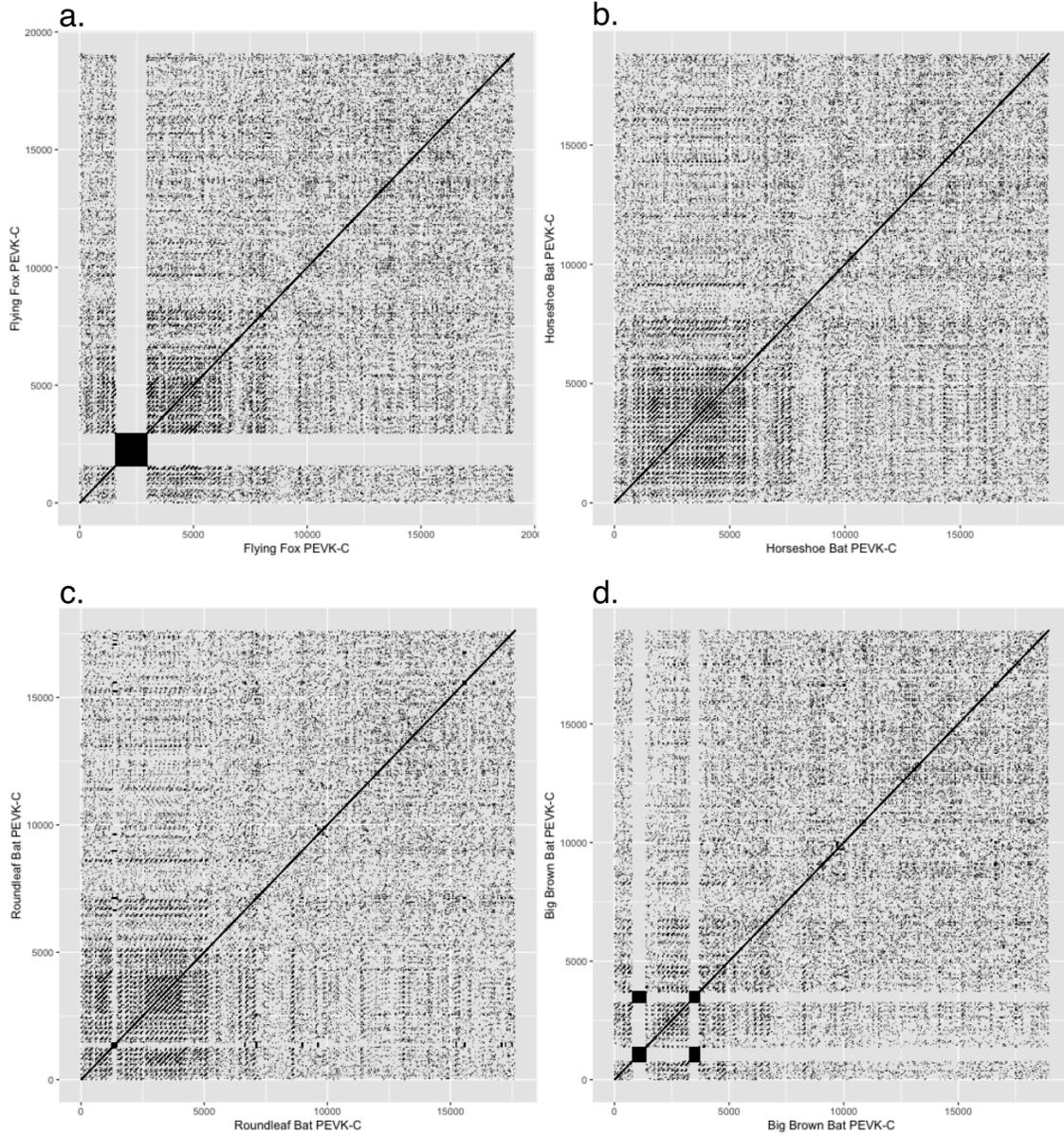


Figure S20.

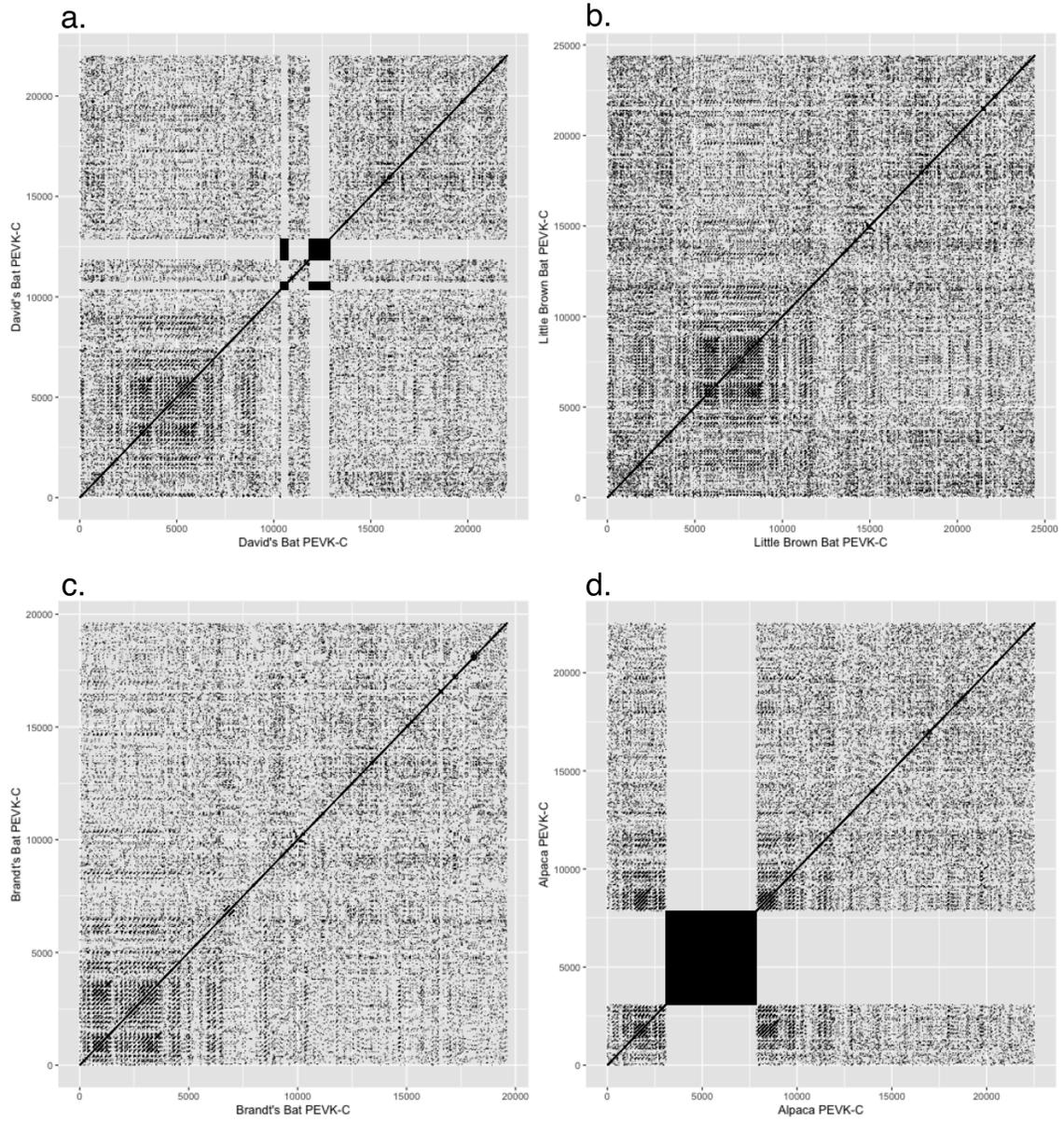


Figure S21.

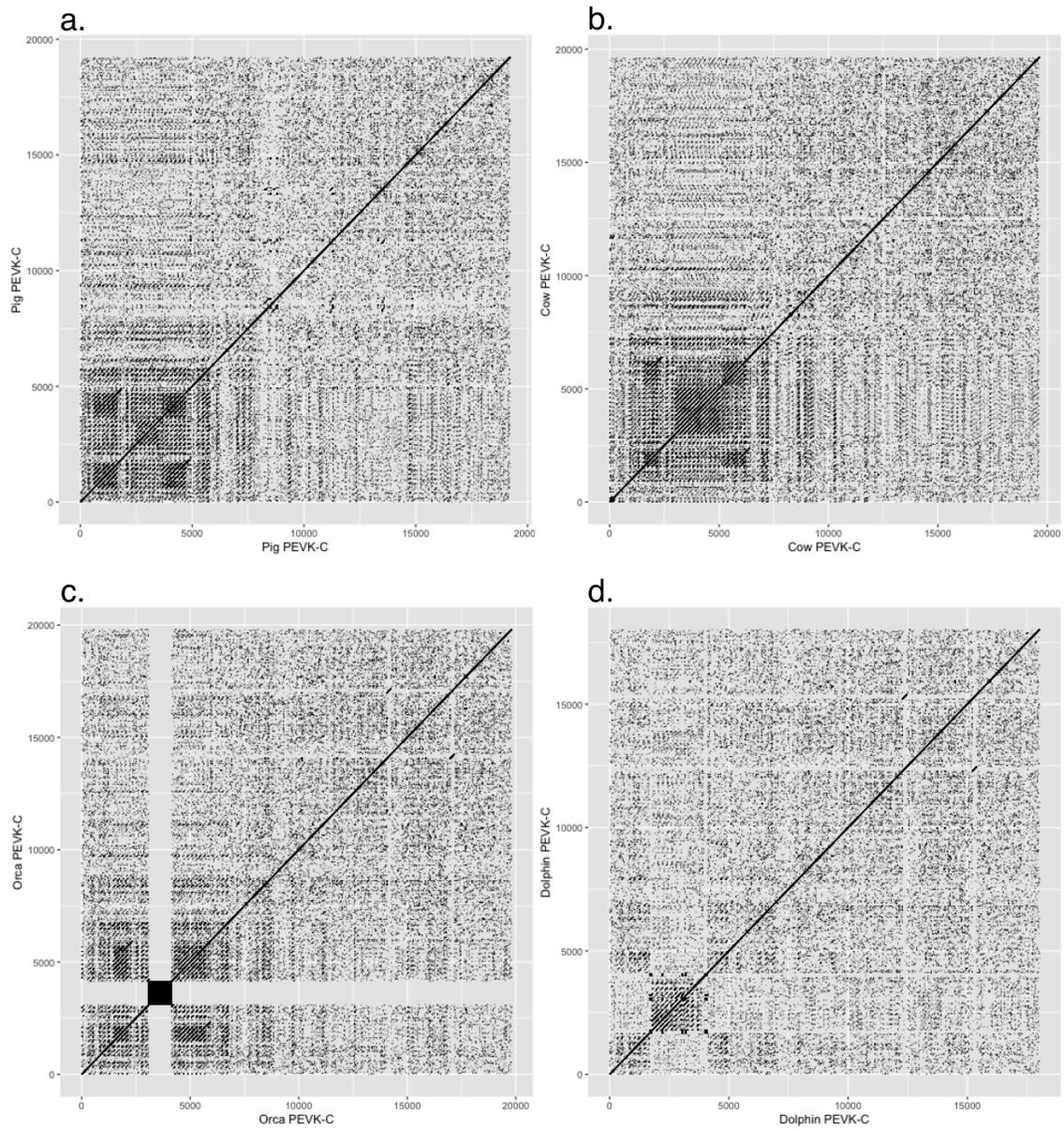


Figure S22.

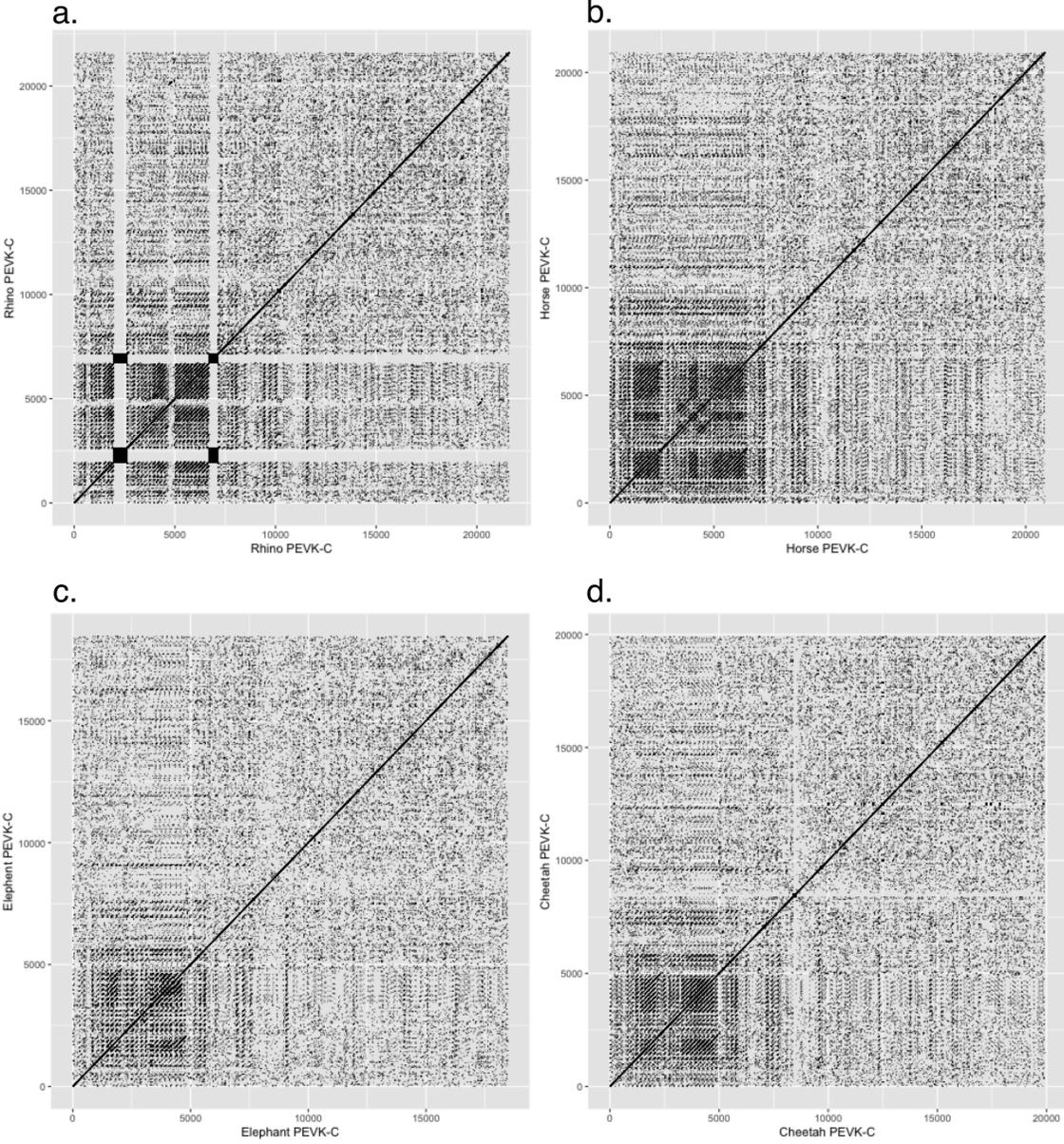


Figure S23.

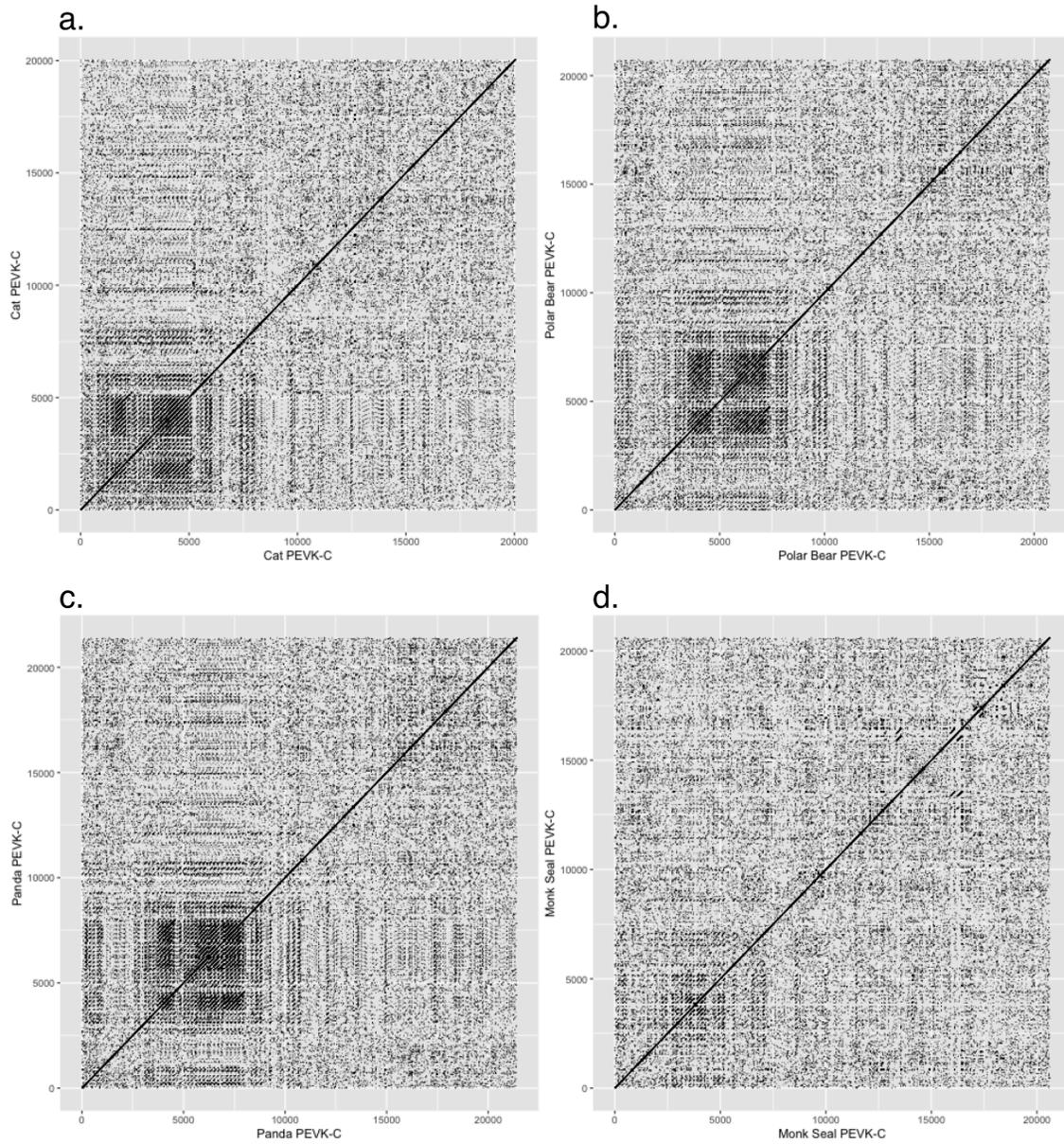


Figure S24.

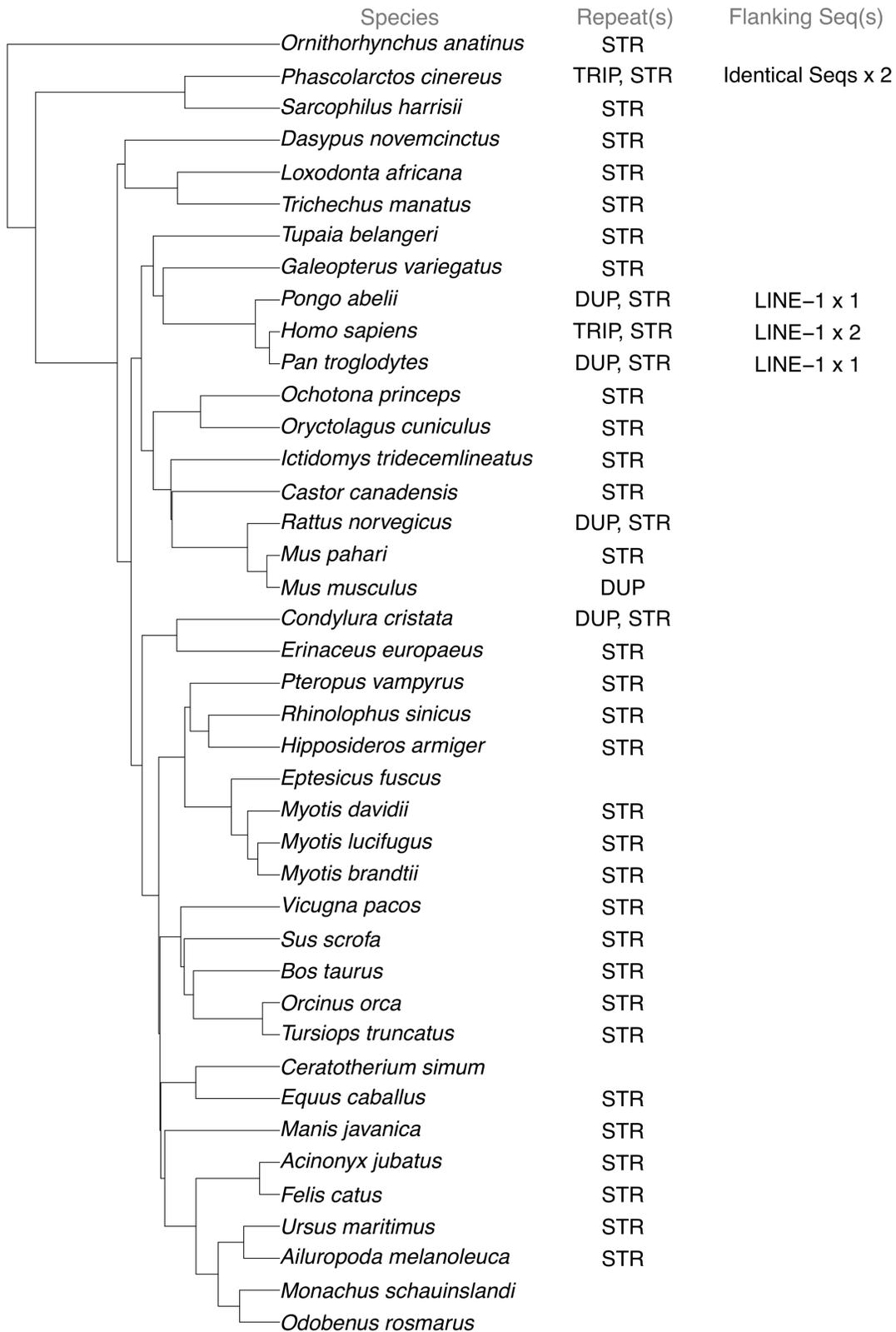
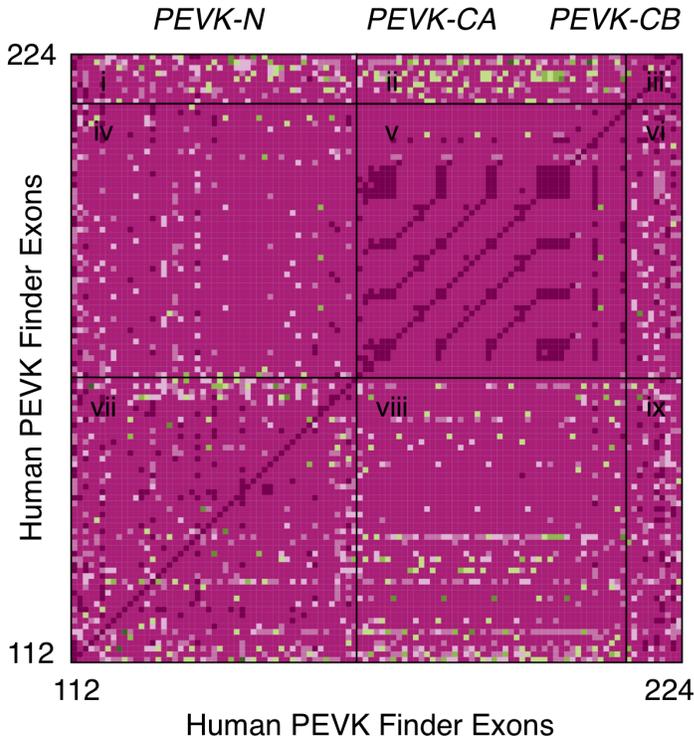
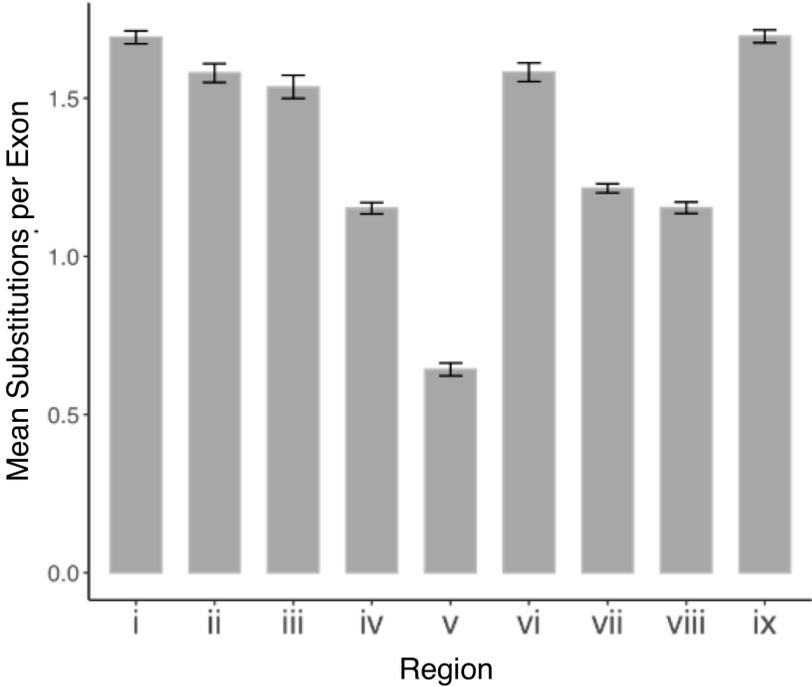


Figure S25.

a.



b.



**Supplementary Figure 1.** PEVK\_Finder tool optimization and evaluation for mouse TTN. **a)** Match scores of mouse PEVK exon sets were generated with different combinations of minimum exon length, PEVK ratio and sliding window length parameter settings. **b)** PEVK\_Finder recovered more PEVK exons than other existing gene prediction tools (GENSCAN, Augustus, FGENESH and geneid). **c)** PEVK Finder outperformed GENSCAN at recovering the exon-intron distribution of mouse TTN PEVK exons identified by cDNA. Vertical lines indicate exon boundaries and the thickness of the lines is determined by the exon coordinates. Grey circles indicate exons that were missed by either PEVK Finder or GENSCAN, and black asterisks indicate automatically annotated exons that were not annotated by cDNA. The PEVK ratio scale is given in the figure.

**Supplementary Figure 2.** Distribution of all ortholog matches in the PEVK-N and PEVK-C regions. Orange/red bars indicate a high number of ortholog matches (i.e. number of species PEVK\_Finder exon sets that contain a given ortholog), and blue bars indicate a low number of ortholog matches.

**Supplementary Figure 3.** Mean sequence divergence over all sequence pairs for each of the 35 confident orthologs (e.g. orthologs for which  $\geq 10$  species have PEVK\_Finder sets that contain a given ortholog). Estimates were made only for species with data for each ortholog. Red indicates orthologs under positive selection. Error bars indicate ranges of divergence among sequences in the alignment.

**Supplementary Figure 4.** Mean per-exon counts of the 20 standard amino acids across all 41 species, separated by PEVK-N and PEVK-C regions. Asterisks indicate significance values for paired t-tests, \*  $P < 0.05$ . Error bars indicate ranges of mean per-exon counts for each amino acid.

**Supplementary Figure 5.** Amino acid alignment and positive selection for ortholog 114. Asterisks indicate evidence of positive selection on amino acid sites with posterior probabilities of \*  $> 0.5$ ; \*\*  $> 0.75$ ; \*\*\*  $> 0.95$ .

**Supplementary Figure 6.** Amino acid-level alignment and positive site selection for ortholog 135 amino acid-level site selection. Asterisks indicate evidence of positive selection on amino acid sites with posterior probabilities of \*  $> 0.5$ ; \*\*  $> 0.75$ ; \*\*\*  $> 0.95$ .

**Supplementary Figure 7.** Amino acid-level alignment and positive site selection for ortholog 137. Asterisks indicate evidence of positive selection on amino acid sites with posterior probabilities of \*  $> 0.5$ ; \*\*  $> 0.75$ ; \*\*\*  $> 0.95$ .

**Supplementary Figure 8.** Amino acid-level alignment and positive site selection for ortholog 138. Asterisks indicate evidence of positive selection on amino acid sites with posterior probabilities of \* > 0.5; \*\* > 0.75; \*\*\* > 0.95.

**Supplementary Figure 9.** Amino acid-level alignment and positive site selection for ortholog 145. Asterisks indicate evidence of positive selection on amino acid sites with posterior probabilities of \* > 0.5; \*\* > 0.75; \*\*\* > 0.95.

**Supplementary Figure 10.** Amino acid-level alignment and positive site selection for ortholog 199. Asterisks indicate evidence of positive selection on amino acid sites with posterior probabilities of \* > 0.5; \*\* > 0.75; \*\*\* > 0.95.

**Supplementary Figure 11.** Ancestral state reconstruction of charge shifts of codon 42 in exon 114. Blue circles: hydrophobic; yellow circles: polar; red circles: positive charge.

**Supplementary Figure 12.** Ancestral state reconstruction of charge shifts of codon 6 in exon 135. Blue circles: hydrophobic; yellow circles: polar; red circles: positive charge.

**Supplementary Figure 13.** Schematic diagram of the PEVK\_Finder algorithm process for a single TTN sequence. Steps 2 and 3 show the algorithmic process for finding a single PEVK exon.

**Supplementary Figure 14.** Nucleotide self-dot plot of the PEVK-C segment for **a)** Platypus, *Ornithorhynchus anatinus*, **b)** Koala, *Phascolarctos cinereus*, **c)** Tasmanian Devil, *Sarcophilus harrisii*, and **d)** Nine-banded armadillo, *Dasyurus novemcinctus*.

**Supplementary Figure 15.** Nucleotide self-dot plot of the PEVK-C segment for **a)** African elephant, *Loxodonta africana*, **b)** West Indian manatee, *Trichechus manatus*, **c)** Chinese tree shrew, *Tupaia belangeri* and **d)** Sunda flying lemur, *Galeopterus variegatus*.

**Supplementary Figure 16.** Nucleotide self-dot plot of the PEVK-C segments for **a)** Sumatran orangutan, *Pongo abelii*, **b)** Human, *Homo sapiens*, **c)** Chimpanzee, *Pan troglodytes* and **d)** American pika, *Ochotona princeps*.

**Supplementary Figure 17.** Nucleotide self-dot plot of the PEVK-C segment for **a)** European rabbit, *Oryctolagus cuniculus*, **b)** Thirteen-lined ground squirrel, *Ictidomys tridecemlineatus*, **c)** North American beaver, *Castor canadensis* and **d)** Norway rat, *Rattus norvegicus*.

**Supplementary Figure 18.** Nucleotide self-dot plot of the PEVK-C segment for **a)** Gairdner's shrewmouse, *Mus pahari*, **b)** House mouse, *Mus musculus*, **c)** Star-nosed mole, *Condylura cristata* and **d)** European hedgehog, *Erinaceus europaeus*.

**Supplementary Figure 19.** Nucleotide self-dot plot of the PEVK-C segment for **a)** Large flying fox, *Pteropus vampyrus*, **b)** Chinese rufous horseshoe bat, *Rhinolophus sinicus*, **c)** Great roundleaf bat, *Hipposideros armiger* and **d)** Big brown bat, *Eptesicus fuscus*.

**Supplementary Figure 20.** Nucleotide self-dot plot of the PEVK-C segment for **a)** David's myotis, *Myotis davidii* **b)** Little brown bat, *Myotis lucifugus*, **c)** Brandt's bat, *Myotis brandtii* and **d)** Alpaca, *Vicugna pacos*.

**Supplementary Figure 21.** Nucleotide self-dot plot of the PEVK-C segment for **a)** Wild boar, *Sus scrofa*, **b)** Cattle, *Bos taurus*, **c)** Orca, *Orcinus orca* and **d)** Bottlenose dolphin, *Tursiops truncatus*.

**Supplementary Figure 22.** Nucleotide self-dot plot of the PEVK-C segment for **a)** White rhinoceros, *Ceratotherium simum*, **b)** Horse, *equus caballus*, **c)** Sunda pangolin, *Manis javanica* and **d)** Cheetah, *Acinonyx jubatus*.

**Supplementary Figure 23.** Nucleotide self-dot plot of the PEVK-C segment for **a)** Domestic cat, *Felis catus*, **b)** Polar bear, *Ursus maritimus*, **c)** Giant panda, *Ailuropoda melanoleuca* and **d)** Hawaiian monk seal, *(Neo)monachus schauinslandi*.

**Supplementary Figure 24.** Inferred short tandem repeats (STR), duplications (DUP), triplications (TRIP) and intronic flanking repeats in the PEVK-C regions of 41 mammal species, plotted alongside a time-calibrated phylogenetic tree. In this figure, *Neomonachus schauinslandi* and *Tupaia chinensis* are represented by their alternative names, *Monachus schauinslandi* and *Tupaia belangeri*, respectively.

**Supplementary Figure 25.** Exon comparisons of the PEVK-N, PEVK-CA and PEVK-CB regions. **a)** A heat map of substitutions among PEVK exons within a species: i) PEVK-N vs. PEVK-CB, ii) PEVK-CA vs. PEVK-CB, iii) PEVK-CB vs. PEVK-CB, iv) PEVK-N vs. PEVK-CA, v) PEVK-CA vs. PEVK-CA, vi) PEVK-CB vs. PEVK-CA, vii) PEVK-N vs. PEVK-N, viii) PEVK-CA vs. PEVK-N, and ix) PEVK-CB vs. PEVK-N. Dark pink indicates exon pairs with few substitutions, whereas light pink and green indicate exon pairs with many substitutions. **b)** Mean pairwise substitutions per exon across all 41 species for each quadrant from Supplementary Figure 25a. Bars represent mean  $\pm$  s.e.

**Table S1.** Summary of species and genome information used in this study.

**Tables S2.** NCBI CCDS and LOVD human titin exon references.

**Table S3.** Summary of PEVK\_Finder and Gnomon annotations across all mammals.