Supplementary Figures

Recurrent erosion of *COA1* exemplifies conditional gene dispensability in oxidative phosphorylation

Sagar Sharad Shinde, Sandhya Sharma, Lokdeep Teekas, Ashutosh Sharma, Nagarjun Vijay

Computational Evolutionary Genomics Lab, Department of Biological Sciences, IISER Bhopal, Bhauri, Madhya Pradesh, India

*Correspondence: nagarjun@iiserb.ac.in

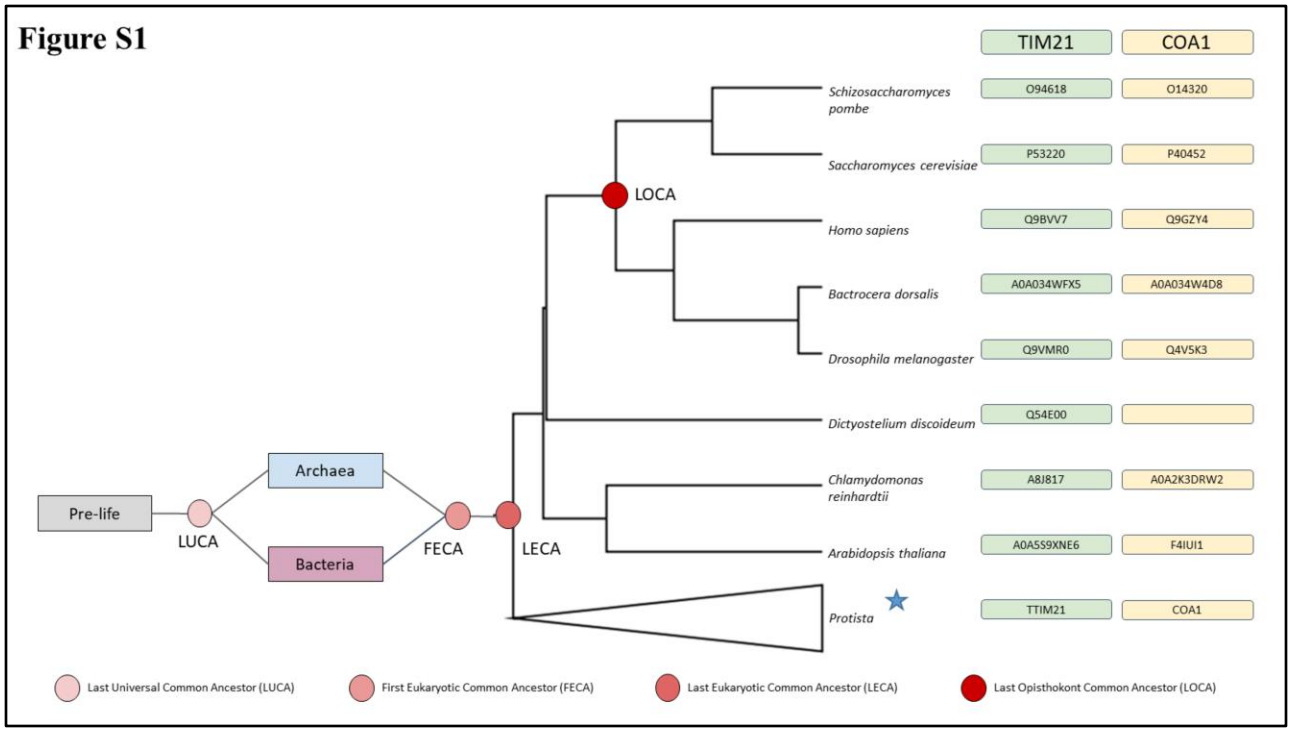


Figure S1: Presence of *TIMM21* and *COA1* gene in different clades. The blue star at Protista is representing a collapsed node of Protista. The elaborated taxanomic relationship within Protista is shown in **Supplementary Fig. S2**.

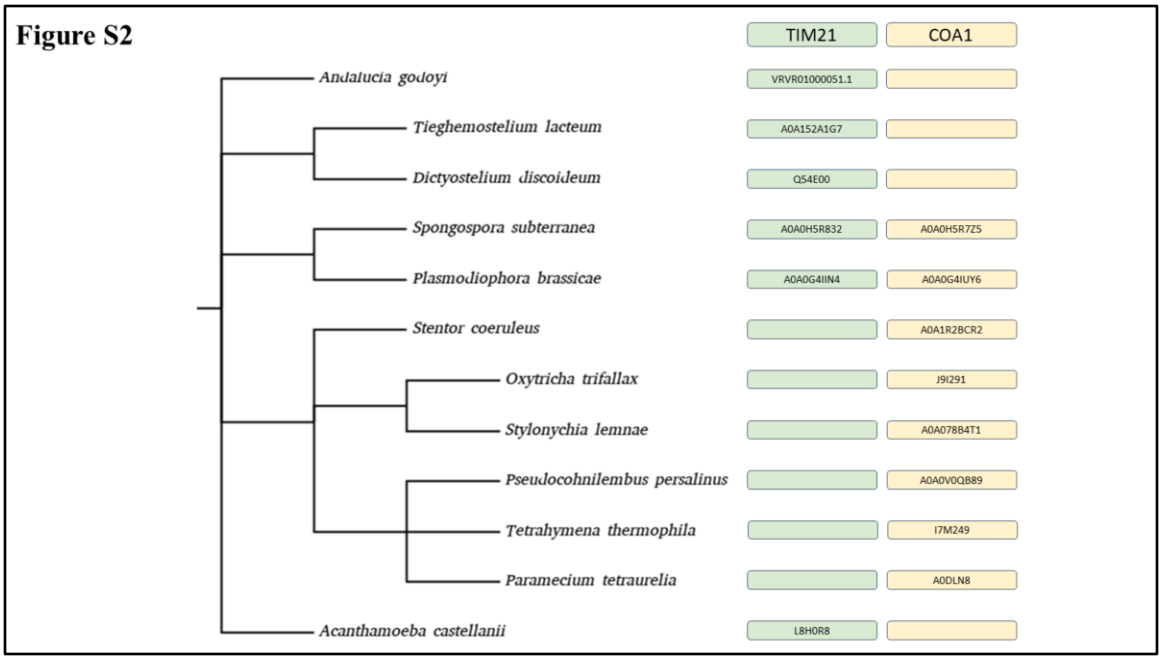


Figure S2: Taxanomic relationship of species within Protista. The tree representing taxanomic relationships is downloaded in the phylip format from NCBI taxonomy.

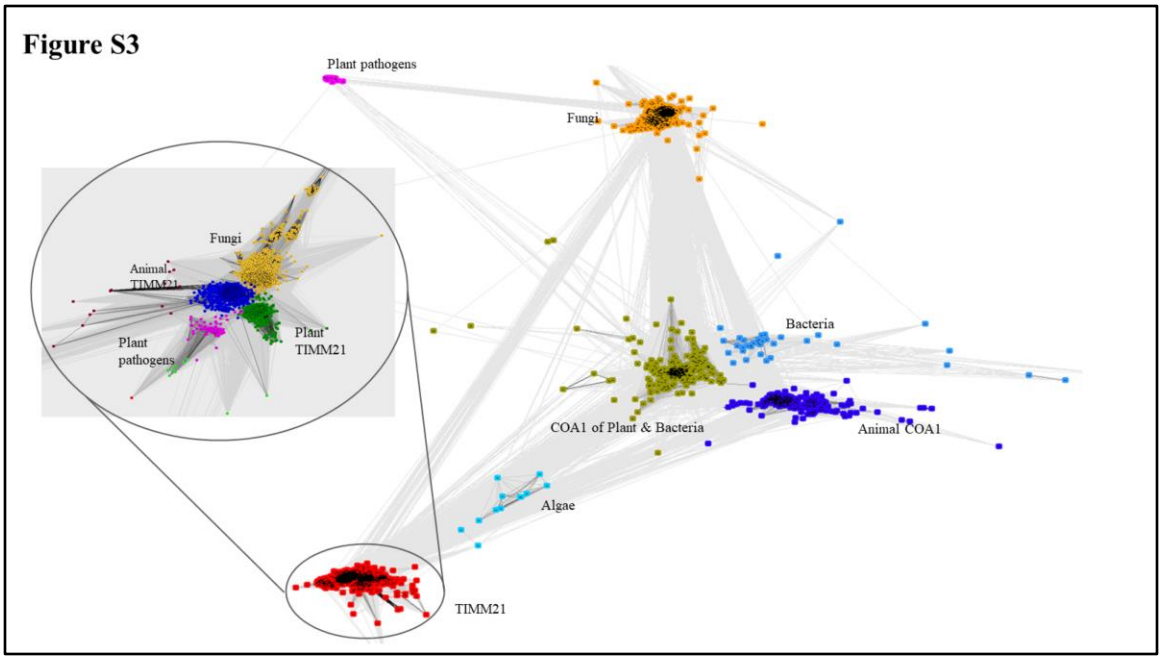


Figure S3: Cluster map of *COA1* and *TIMM21.* The amino acid sequences were downloaded from Pfam 34.0 database and used to generate the clustermap using CLANS. Different colors are used to represent distinct clusters: *COA1* is dark blue, *TIMM21* is red, homologs of *COA1* from species of fungi are in dark yellow, homologs of *COA1* from bacterial species are in the light blue cluster. *COA1* homologs in plants and bacteria are in forest green color, *COA1* from algae are in aqua blue, and homologs of *COA1* from plant pathogens are in magenta. The inset is the zoomed-in figure of *TIMM21* sequence clusters where the cluster of fungi is in yellow, animals in dark blue, plant pathogens in magenta, and the cluster of plants in dark green.

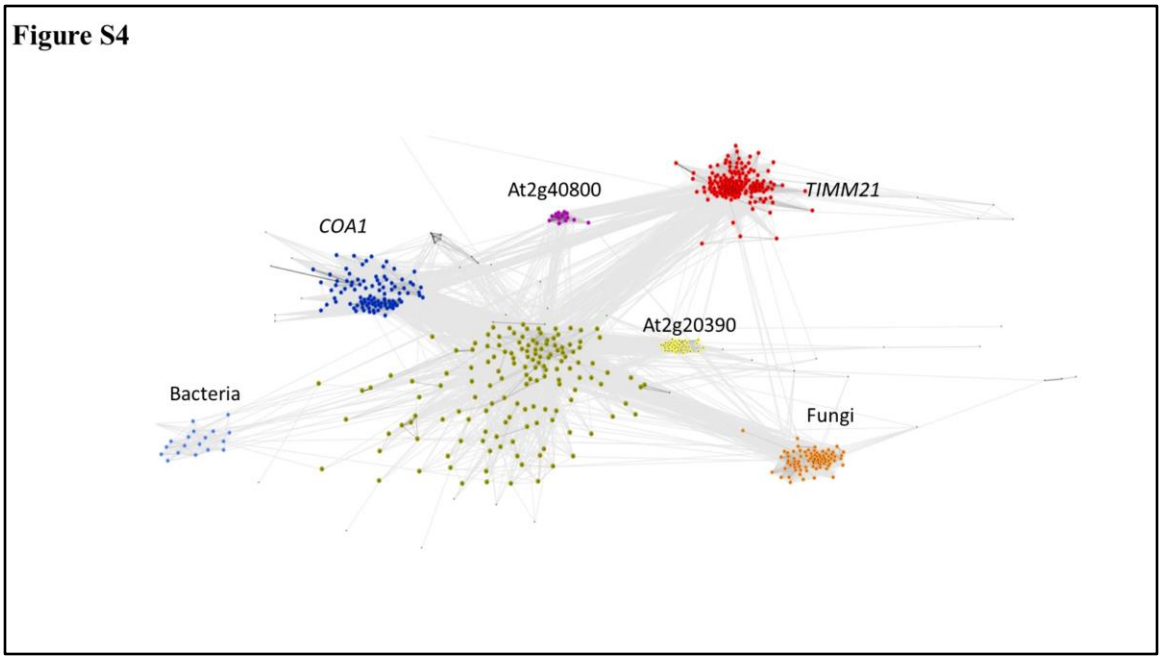


Figure S4: Cluster map of *COA1* homologs identified using profile-profile search implemented in HHblits. The cluster of *COA1*: dark blue, *TIMM21*: red, homologs of *COA1* from species of fungi: orange, homologs of *COA1* from bacterial species: light blue cluster and diffuse brown cluster, *COA1* homologs in plants, represented by *Arabidopsis thaliana* homolog At2g20390: yellow, *TIMM21*-like proteins that exist as duplicated copies in plants, represented by *Arabidopsis thaliana* homolog At2g40800: magenta.

Figure S5

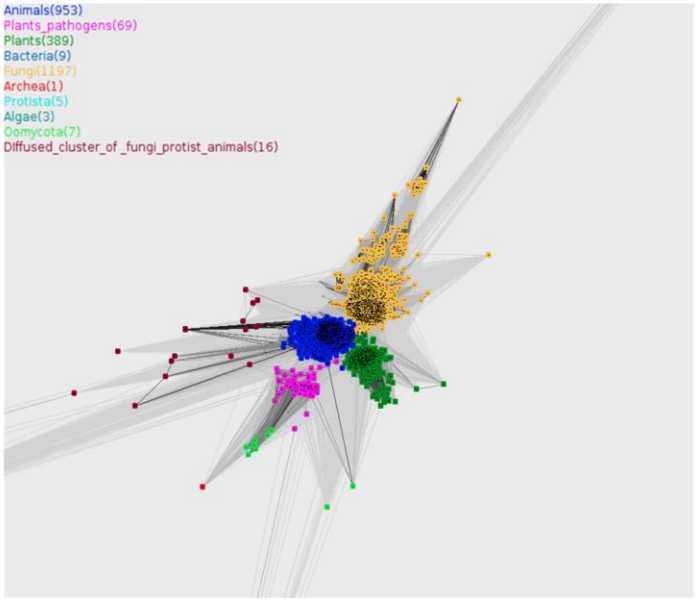


Figure S5: Cluster map of *TIMM21* amino acid sequences downloaded from Pfam 34.0 database, clusters were generated by CLANS. The clusters of fungi *TIMM21*: yellow, animal *TIMM21*: dark blue, plant pathogens cluster: magenta, plant *TIMM21* cluster: dark green, the diffused brown cluster is from fungi, protist, and animals, oomycota *TIMM21* cluster: green.

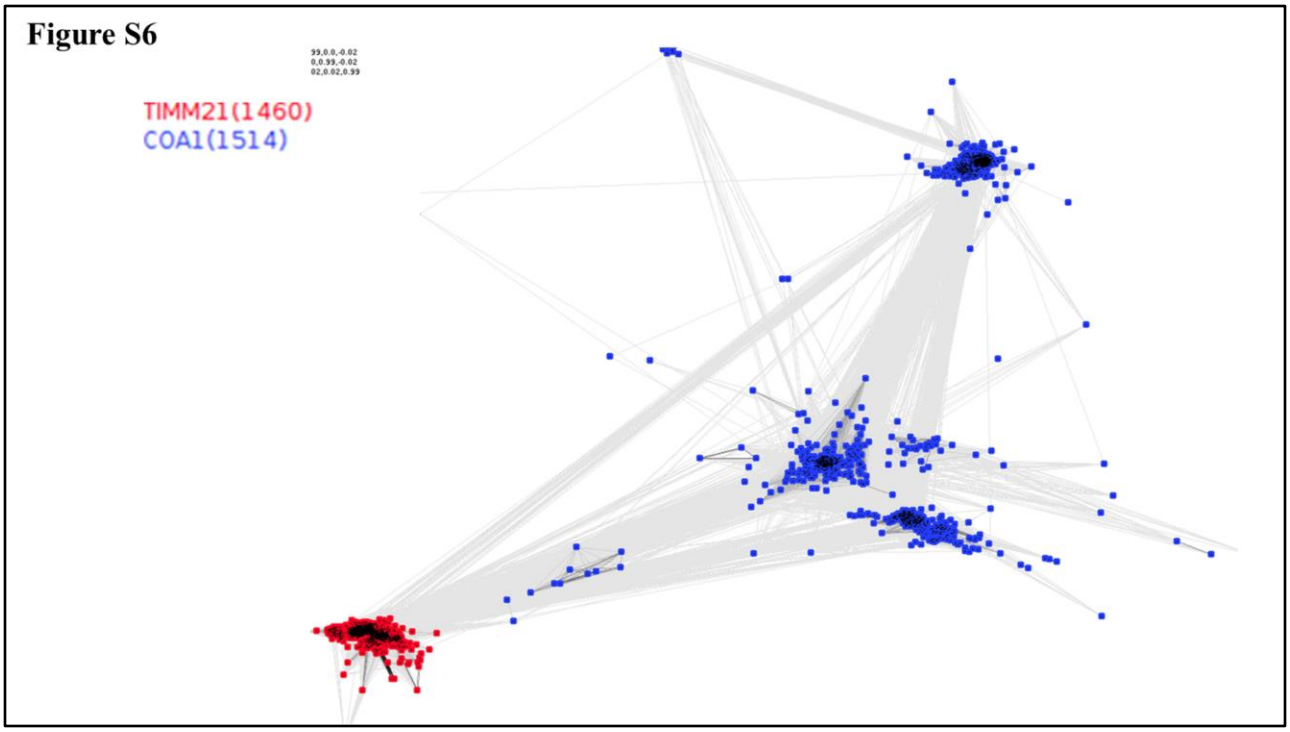
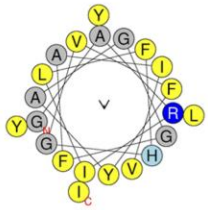


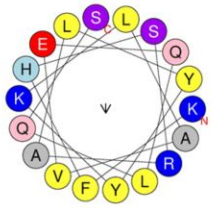
Figure S6: Cluster map of *COA1* and *TIMM21* amino acid sequences downloaded from Pfam 34.0 database, clusters were generated by CLANS. The cluster of *COA1*: dark blue, *TIMM21*: red.

Figure S7 Properties of predicted membrane helix and helix in *COA1*

COA1	Residues	Hydrophobicity	Hydrophobic moment	% polar residues
Predicted transmembrane helix (MH) in *COA1*				
MH1	16-37	0.88818	0.05041	27.3
Predicted helix (H)				
H1	39-56	0.36833	0.10717	50



Membrane helix



Helix

Figure S7: The HeliQuest predicted membrane properties of the membrane helix and other helix of human *COA1*.

Figure S8

Properties of predicted membrane helix and helix in *TIMM21*

TIM21	Residues	Hydrophobicity	Hydrophobic moment	% polar residues
Predicted transmembrane helix (MH) in *TIMM21*				
MH1	75-98	0.93458	0.07898	33.3
Predicted helix (H)				
H1	23-35	0.43231	0.06286	69.2
H2	104-118	0.23067	0.37524	60

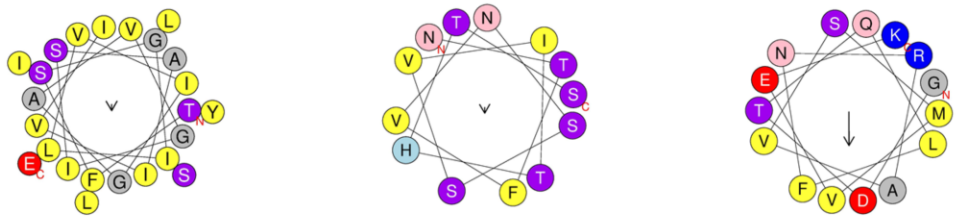


Figure S8: The HeliQuest predicted membrane properties of membrane helix and helix of yeast *TIMM21*.

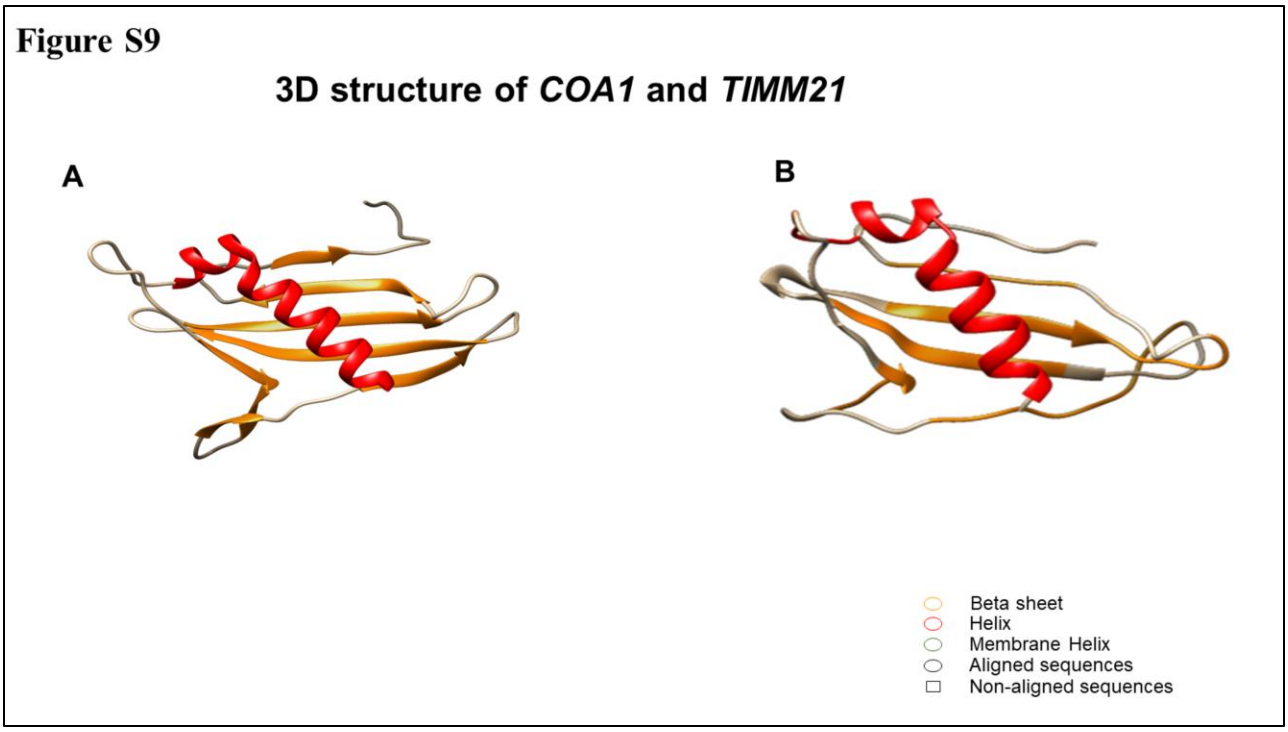


Figure S9: Structure of *COA1* and *TIMM21*. (**A**) 3D structure of *TIMM21* of *Saccharomyces cerevisiae* downloaded from the PDB database (2ciu). (**B**) The 3D structure of human *COA1* sequence homology modelled with modeller software using *TIMM21* (*Saccharomyces cerevisiae*) PDB structure.

Figure S10 HHblits output with *COA1* of DUCK

```
>UniRef100_A0A023FP12 TIM21-like protein, mitochondrial n=1 Tax=Amblyomma cajennense TaxID=34607 RepID=A0A023FP12_AMBCJ
Probab=99.79  E-value=9.6e-23  Score=155.60  Aligned_cols=121  Identities=19%  Similarity=0.336  Sum_probs=101.7
Template_Neff=3.200

Q Duck_coal         9 WQMAIFMGVFSGGG--CVVMYNLMQ-KGFA---RTQYYQQGLEQLNNNPTALEALGAPPLKVHNIRLTDG---------S   73 (139)
Q Consensus         9 ~~~a~~~~v~~~~~~-~~~~~~~v~~~~~~~~~s~~y~~Al~~~~~~~~v~~~LG~~pI~~~~~~~~~~~~~~~~~~~~   73 (139)
                        .|.|.|++|+++|+  .++++|.|+ +.|+   .+++|..|+++|.+||+|+++||+ ||++++....-+         +
T Consensus       165 TKTasYlGVIl~GiGVTG~mFYaIFrELFSSkSPNsVYS~ALekC~~DpRV~DaLGe-PIKgYGEeTrRGRR~HVSH~~Y  243 (339)
T UniRef100_A0A0  165 TKTASYLGVILAGIGVTGIMFYAVFRELFSSKSPNSVYSKALEKCINDPRVEDALGE-PIKGYGEETRRGRRRHVSHSEY  243 (339)
Confidence            57899999987665  667777776 4443   457999999999999999999996 699886543322         3


Q Duck_coal        74 NRVDTERAQIKLPVSGTKSAGYLYIDSEMDHSRQCWRLQDVTLKLRD--GQNISVYHSPV  131 (139)
Q Consensus        74 ~~~~~~~~~l~f~V~G~k~~G~v~v~a~r~~~~~~w~~~~l~l~~~~~~~~~i~L~~~~~  131 (139)
                        ..+|..|++|+|+|+|++.+|+|++++.+++ .++|+|+++.++++|  .++|.|+|++.
T Consensus       244 eKDGvkHMRMKFYIkG~R~kgTVhLEmken~~G~yeYRYLfVqlDdyp~~tIvIEDNR~  302 (339)
T UniRef100_A0A0  244 EKDGVKHMRMKFYIKGIRNKGTVHLEMKENE-SGNYEYRYLFVQLDDYPKETIVIEDNRE  302 (339)
Confidence            456889999999999999999999998885 799999999999998  78999999883
```

Figure S10: Alignment output of *COA1* of duck (*Anas platyrhynchos platyrhynchos)* and *TIMM21* of Cayenne tick (*Amblyomma cajennense)* from HHblits.

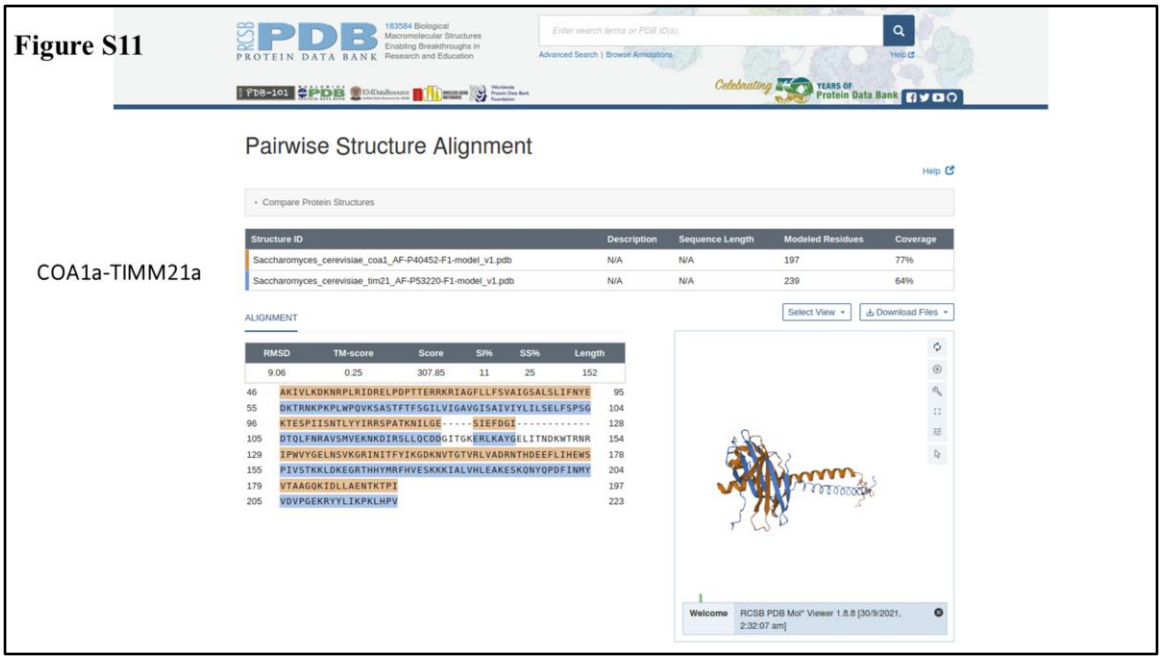


Figure S11: Pairwise comparison of predicted *COA1* (*Saccharomyces cerevisiae*) and *TIMM21* (*Saccharomyces cerevisiae*) structures downloaded from AlphaFold website.

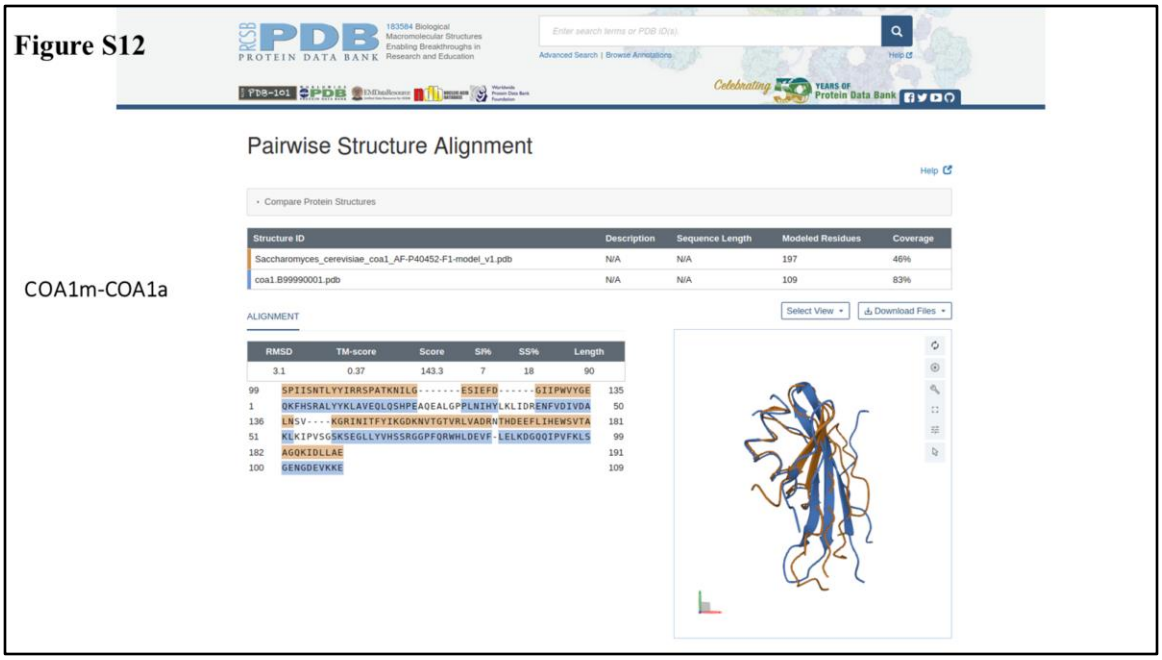


Figure S12: Pairwise alignment of human *COA1* sequence homology modelled using *TIMM21* (*Saccharomyces cerevisiae*) PDB structure and de novo predicted (*Saccharomyces cerevisiae*) *COA1* sturcture downloaded from AlphaFold website.

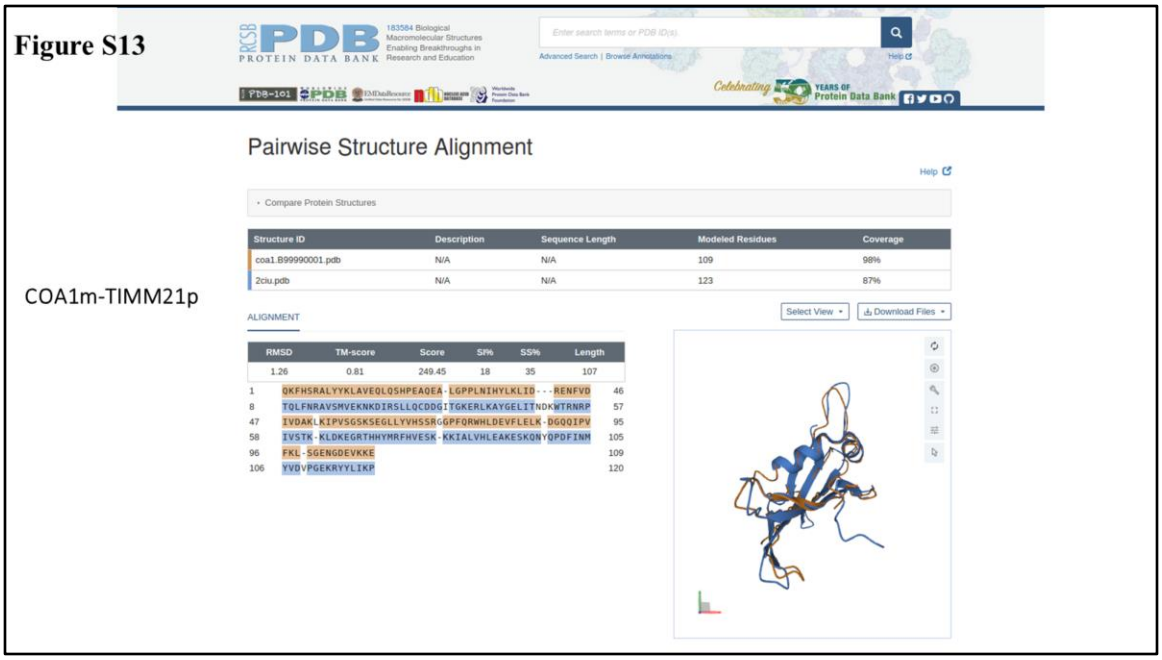


Figure S13: Pairwise alignment of human *COA1* sequence homology modelled using *TIMM21* (*Saccharomyces cerevisiae*) PDB structure and *TIMM21* structure downloaded from rcsb.org protein data bank.

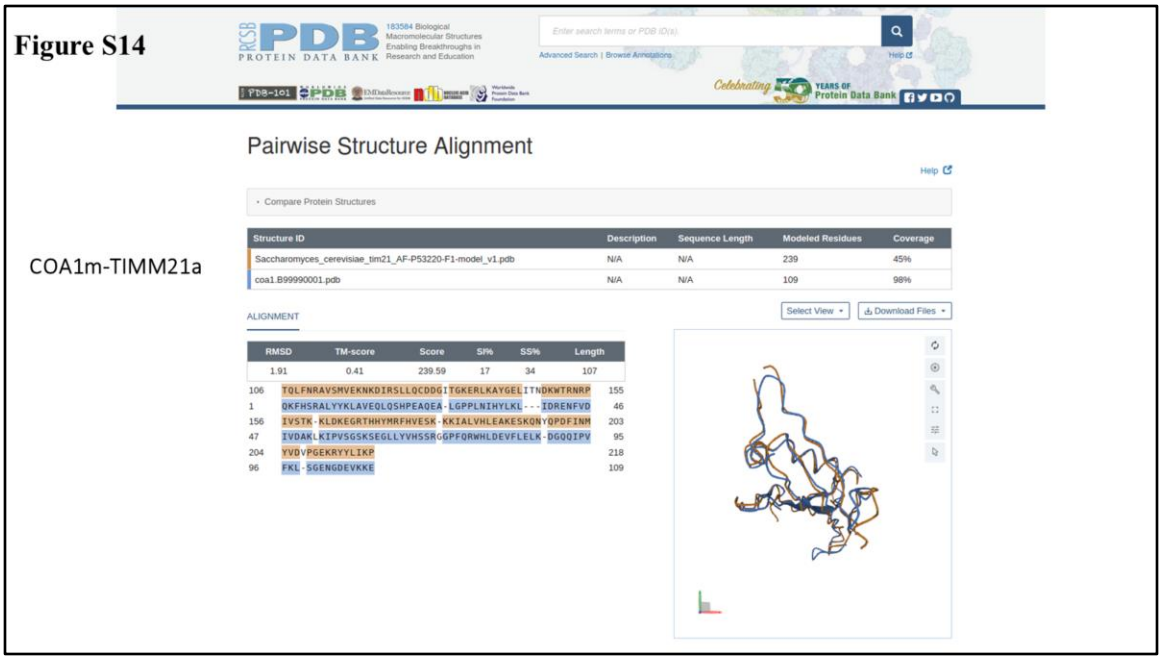


Figure S14: Pairwise alignment of human *COA1* sequence homology modelled using *TIMM21* (*Saccharomyces cerevisiae*) PDB structure and de novo predicted *TIMM21* of yeast (*Saccharomyces cerevisiae)* from AlphaFold website.

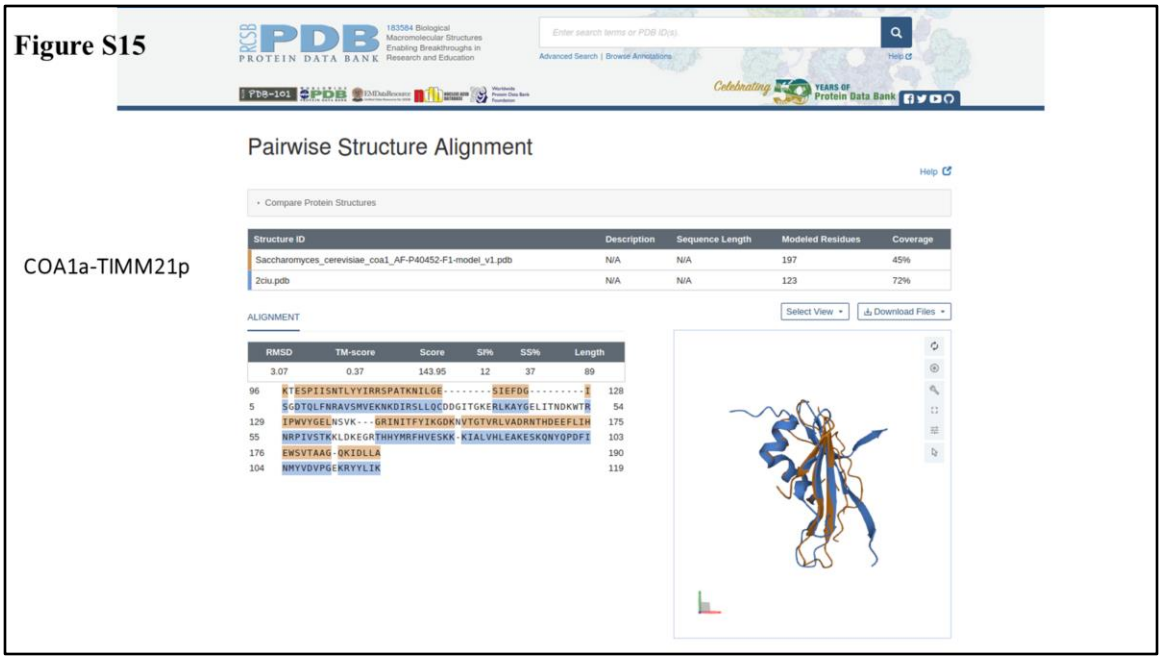


Figure S15: Pairwise structure alignmnet of de novo predicted *COA1* structure from AlphaFold website and *TIMM21* PDB structure from rcsb.org protein data bank of yeast (*Saccharomyces cerevisiae)*.

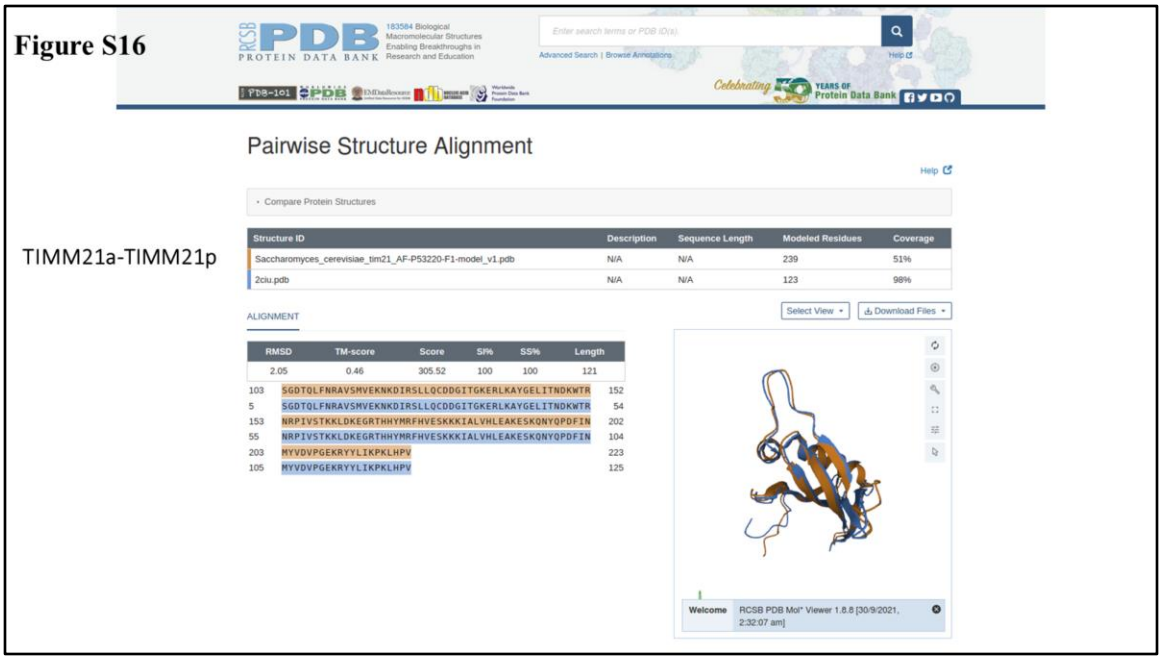
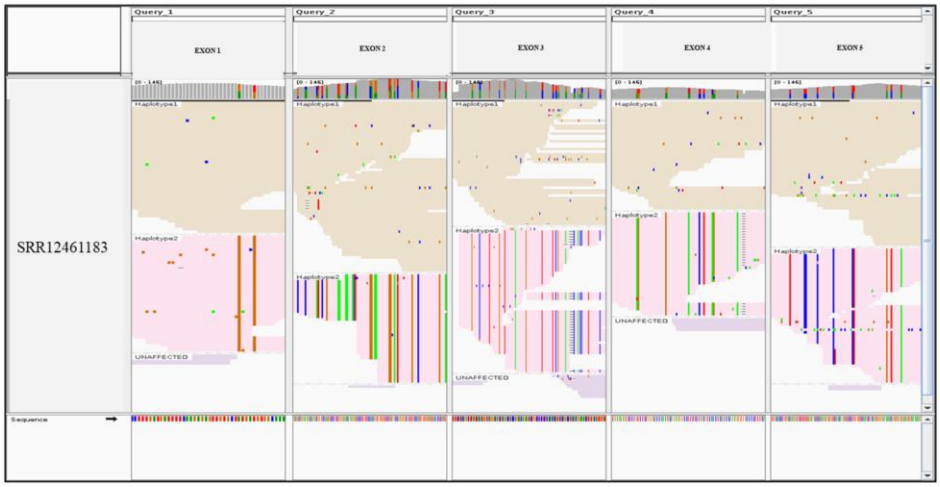


Figure S16: Pairwise structure alignment of *TIMM21* de novo predicted structure of yeast (*Saccharomyces cerevisiae)* downloaded from AlphaFold and *TIMM21* PDB structure from rcsb.org protein data bank of yeast (*Saccharomyces cerevisiae)*.

Figure S17 human Haplotype



human

Figure S17: Comparison of haplotypes of *COA1* gene inferred based on sequencing reads visualized in IGV browser. Two haplotypes of *COA1* can be seen in short-read sequencing data of human (*Homo sapiens*) and correspond to the functional and pseudogene copies.

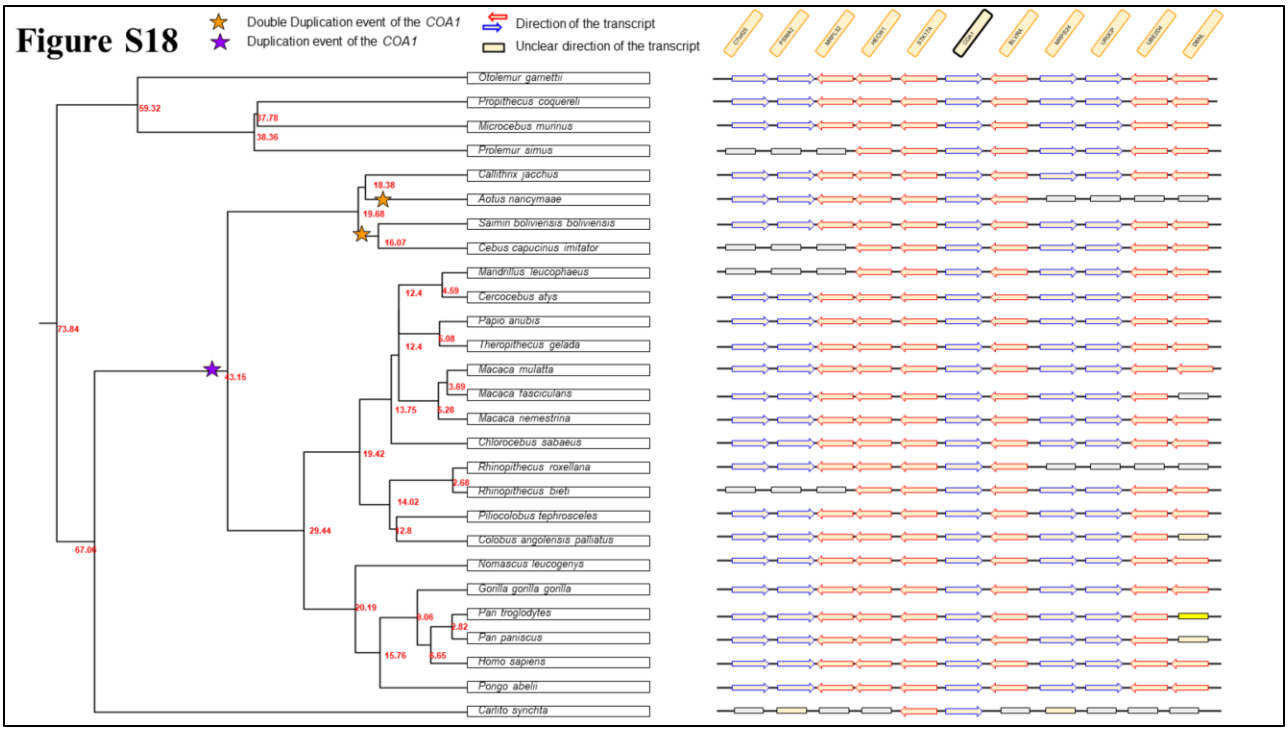


Figure S18: The duplication events of *COA1* gene in primates and their synteny with the time calibrated phylogenetic tree from TimeTree website. The first duplication occurred around 43 million year ago (MYA) and the double duplication event occurred independently at around 16 and 18 MYA as depicted. The direction of gene transcription is depicted using red and blue arrows relative to the *COA1* gene for consistency across species. Gray boxes represent the genes which are located on short scaffolds and their orientation could not be verified.

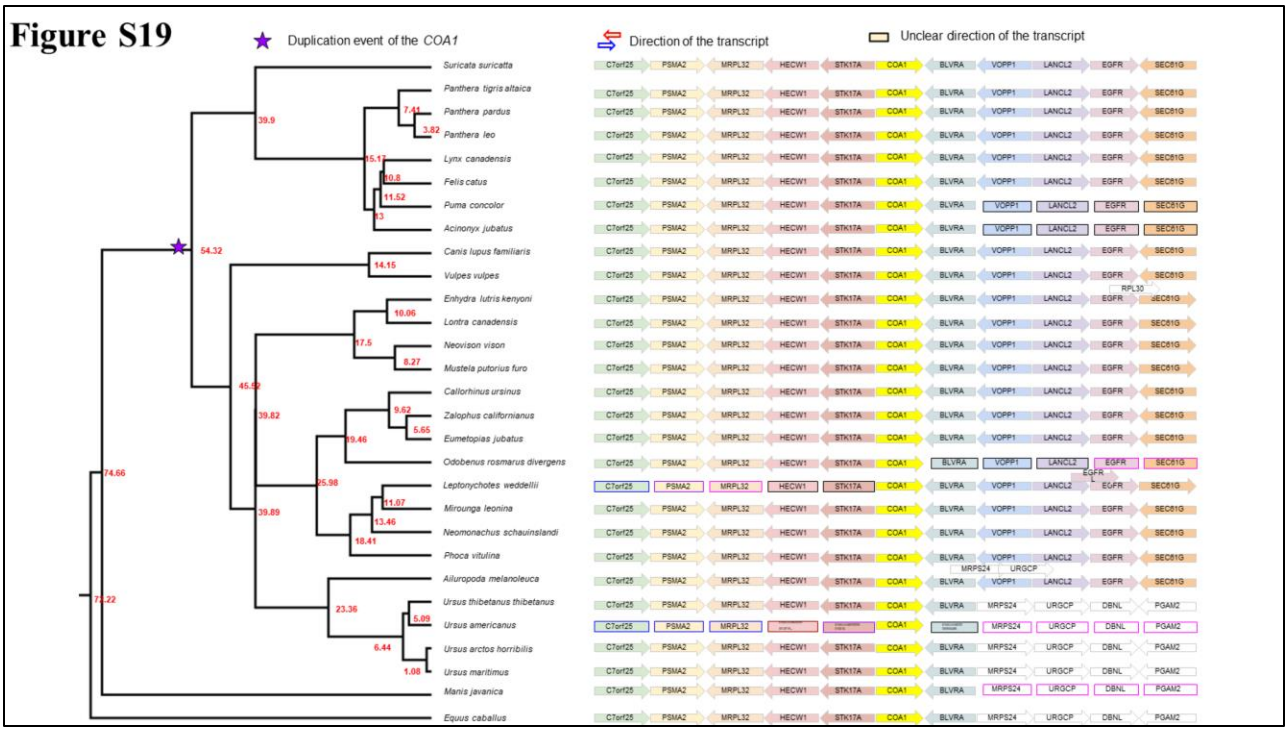


Figure S19: The independent duplication event of *COA1* gene in carnivores and their synteny with the time calibrated phylogenetic tree from TimeTree website. The duplication occurred around 54 million year ago (MYA) as depicted. The direction of gene transcription is depicted using red and blue arrows relative to the *COA1* gene for consistency across species. Gray boxes represent the genes which are located on short scaffolds and their orientation could not be verified.

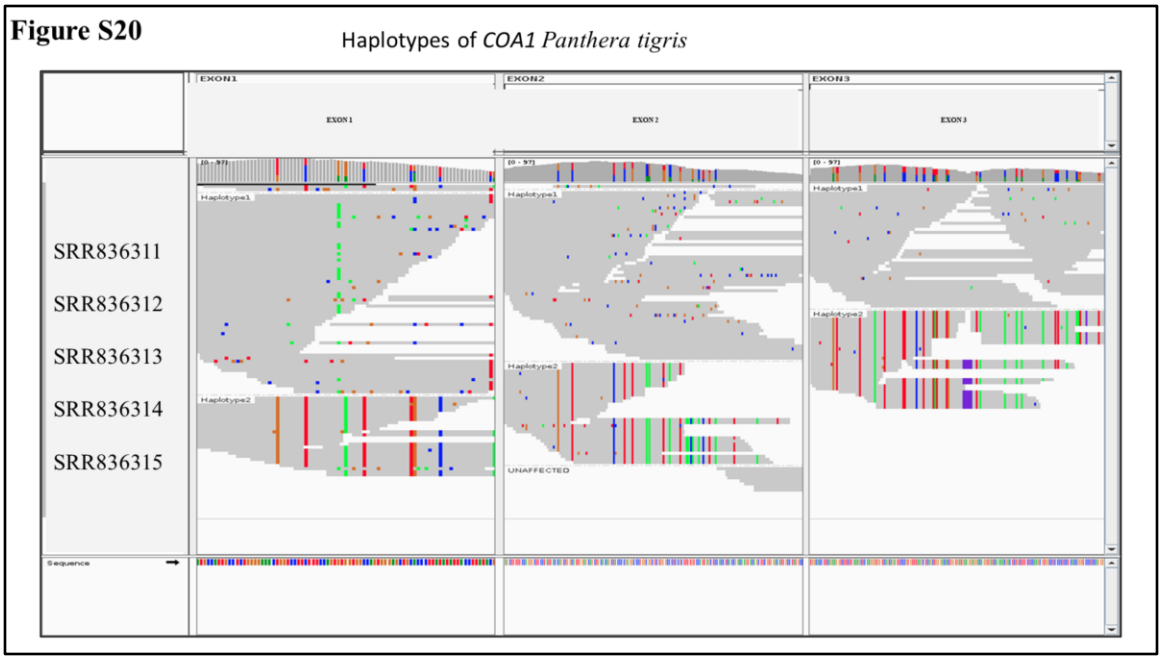


Figure S20: Two haplotypes of exon-1 to exon-4 of *COA1* in tiger (*Panthera tigris*).

Figure S21

Haplotypes of *COA1 Equus caballus*

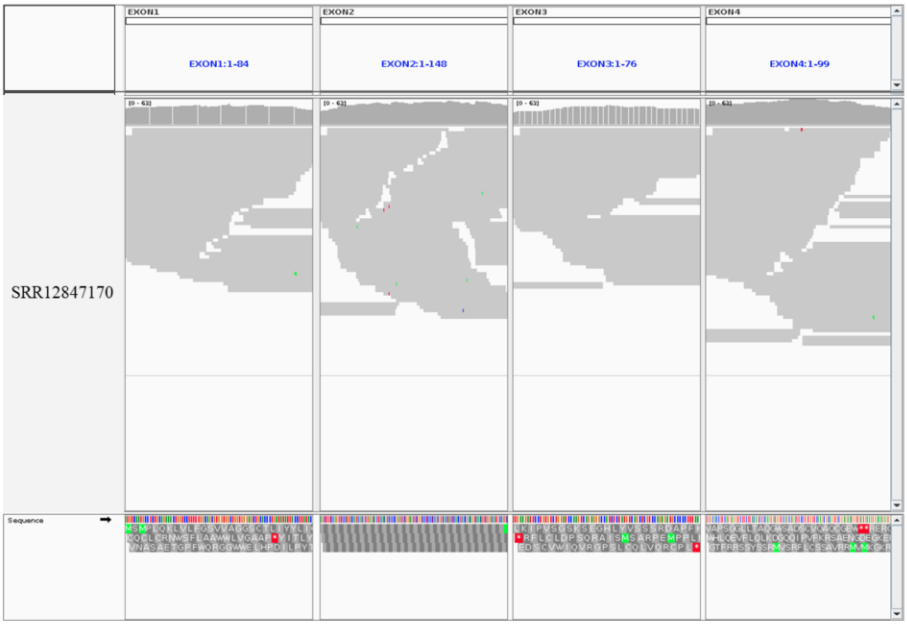


Figure S21: Single copy of *COA1* gene in horse (*Equus caballus)* :Haplotype in exon-1, exon-2, exon-3 and exon-4

Figure S22 Haplotypes of *COA1 Manis javanica*

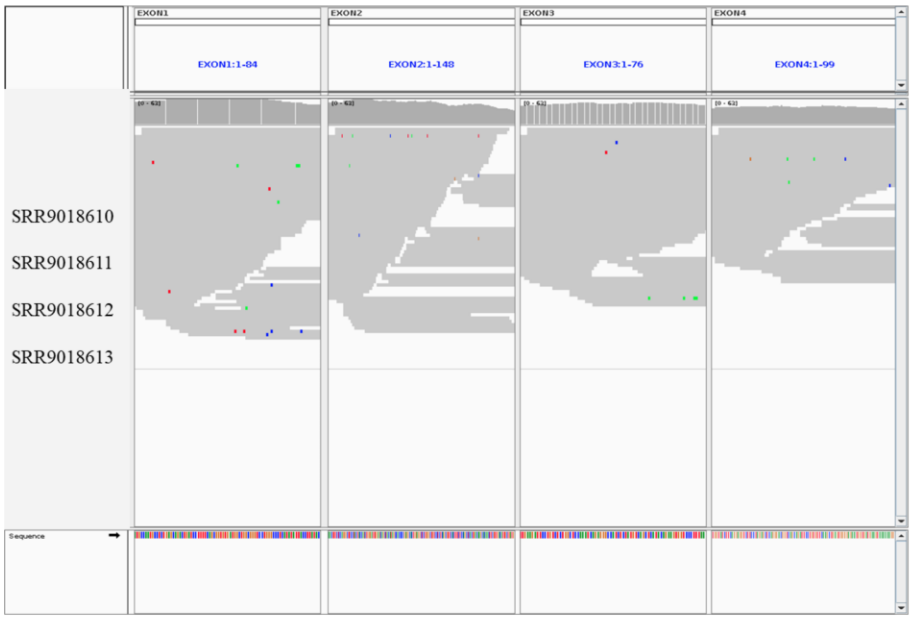


Figure S22: Single copy of *COA1* gene in pangolin (*Manis javanica)*: Haplotype in exon-1, exon-2, exon-3 and exon-4.

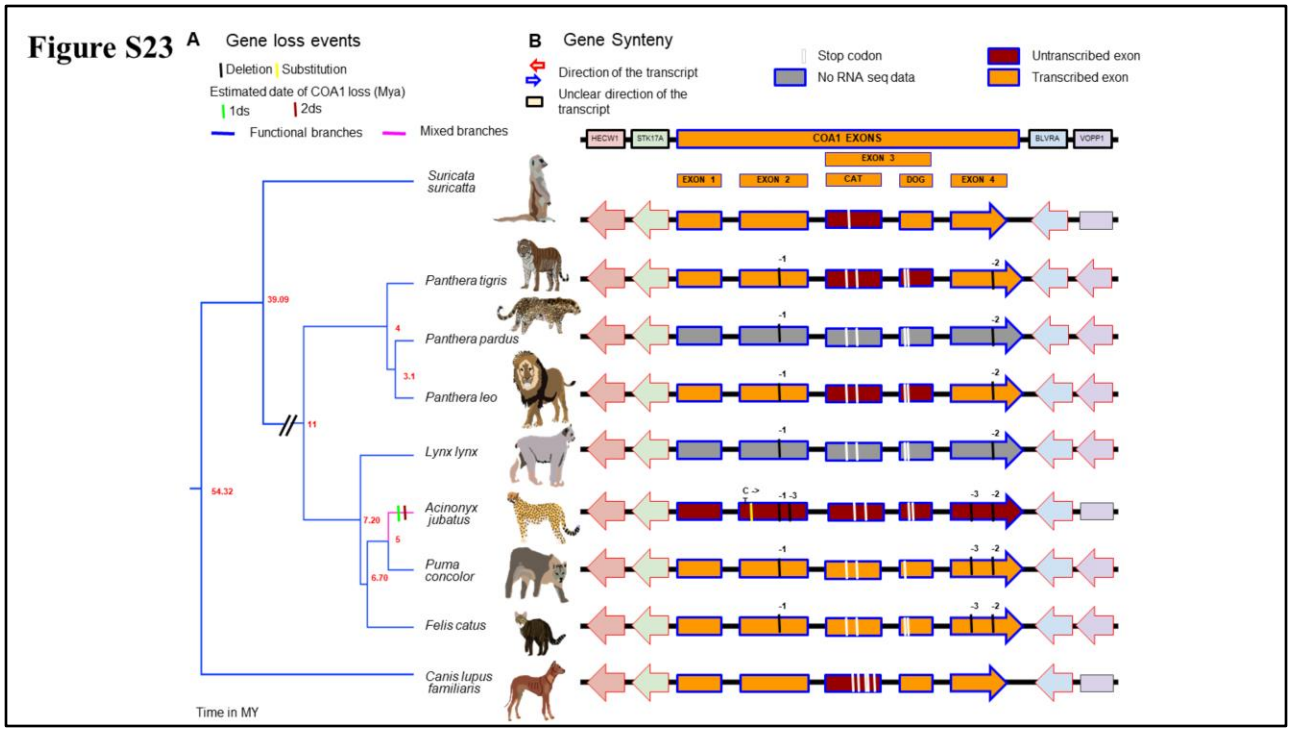


Figure S23: Loss of *COA1* gene in felids. (**A**) Gene loss event in cheetah (*Acinonyx jubatus*) besides a time-calibrated phylogenetic tree downloaded from the TimeTree website. (**B**) Gene order in the genomic region flanking the *COA1* gene and its exons in genomes. Red and blue arrows depict the direction of gene transcription relative to the *COA1* gene for consistency across species. Gray boxes represent the genes located on short scaffolds with unknown orientation.

Figure S24: Screenshot of RNA-seq reads aligned to each of the putative locations of the three *COA1* exons in cat (*Felis catus*). Gray boxes represent reads aligned to the genome and thin lines connecting these boxes represent the spliced read alignments. Datasets from spleen (SRR3218714) have been mapped to the Felis_catus.Felis_catus_9.0.dna_sm.toplevel.fa. Locations of each of the exons are depicted as bed record in the bottom row.

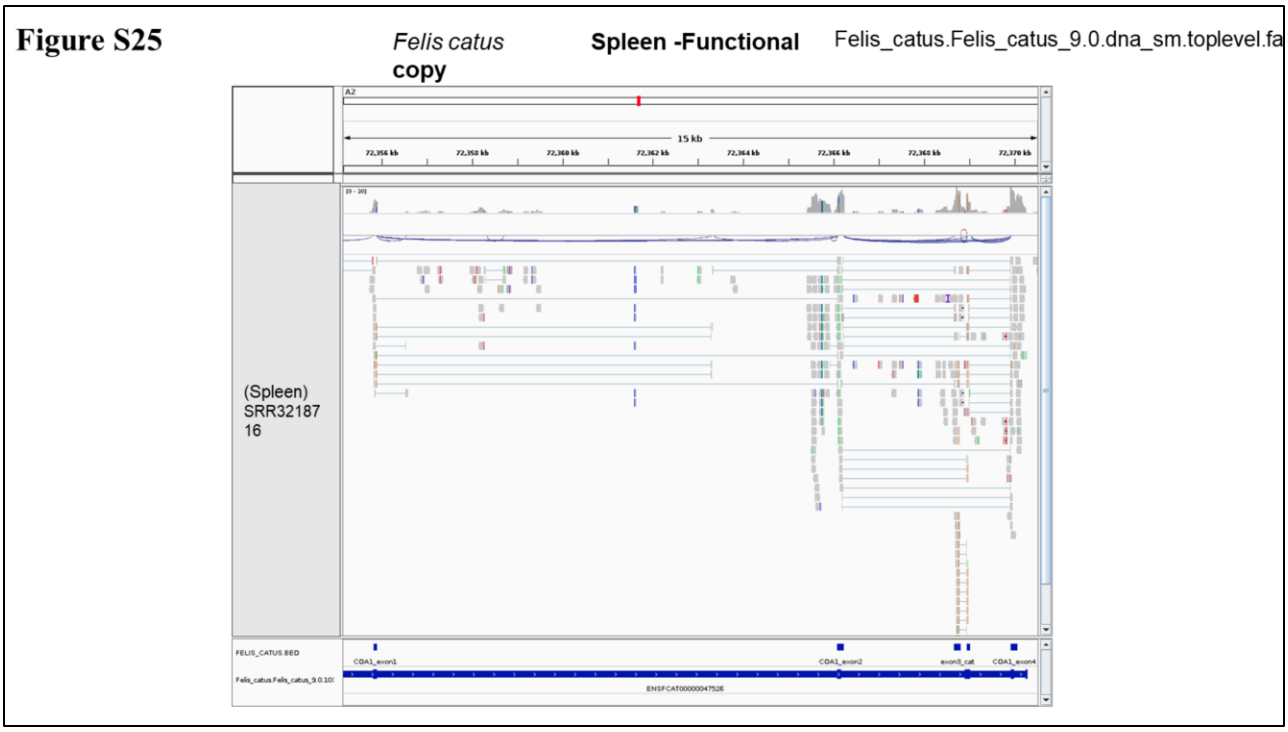


Figure S25: Screenshot of RNA-seq reads aligned to each of the putative locations of the three *COA1* exons extra 3rd exon from *Puma concolor* region in cat (*Felis catus*). Gray boxes represent reads aligned to the genome and thin lines connecting these boxes represent the spliced read alignments. Datasets from spleen (SRR3218716) have been mapped to the Felis_catus.Felis_catus_9.0.dna_sm.toplevel.fa. Locations of each of the exons are depicted as bed record in the bottom row.

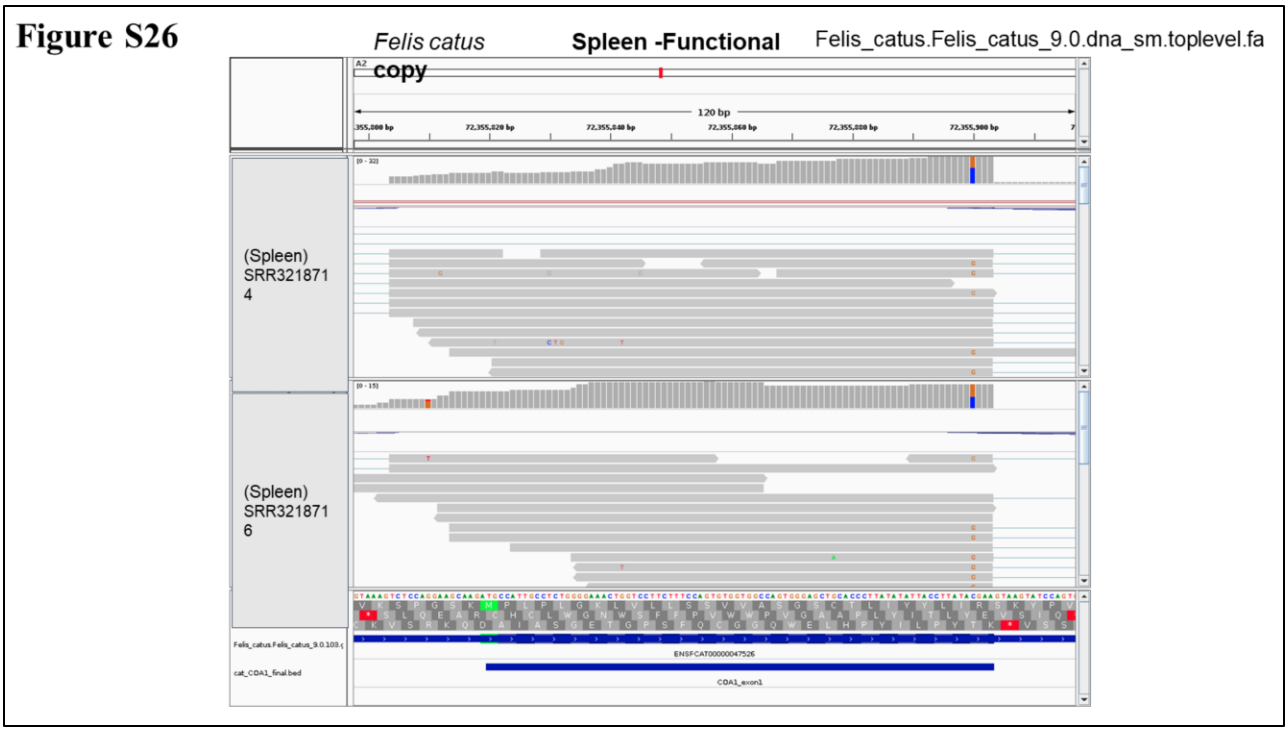


Figure S26: Zoomed view of cat *COA1* exon-1 shown in **Supplementary Fig. S26**.

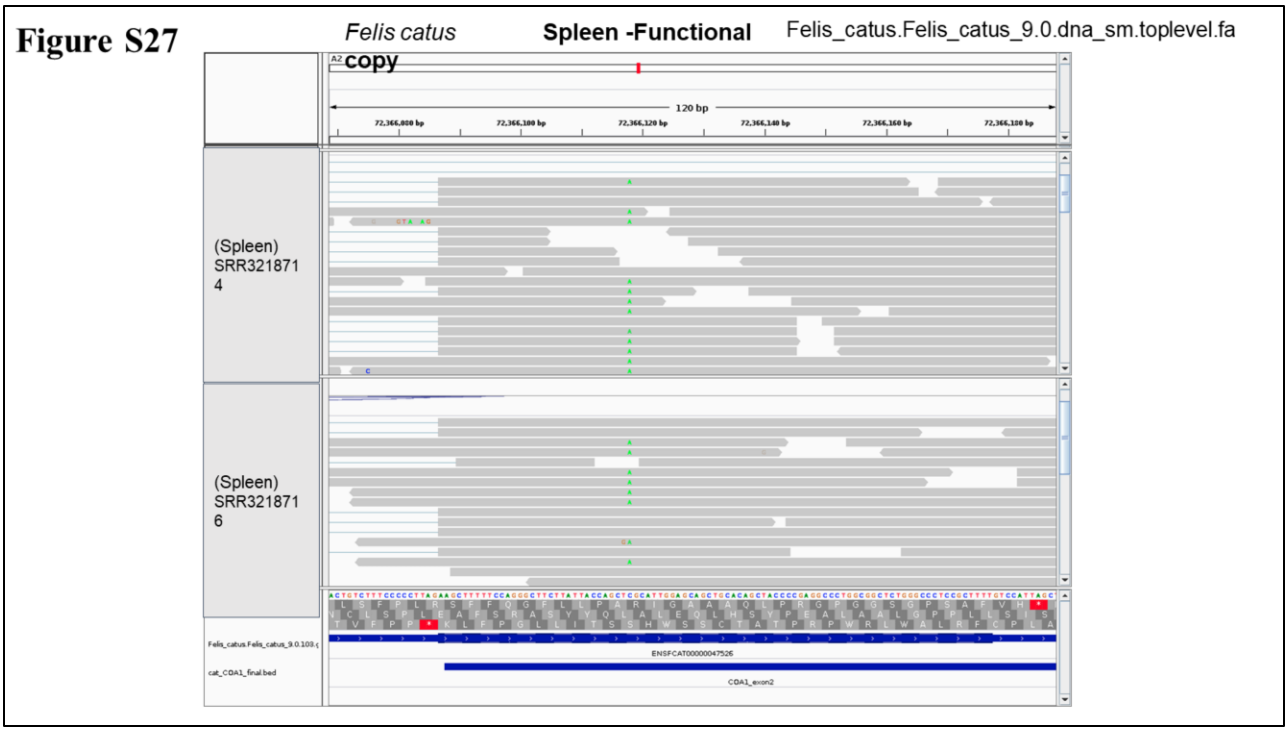


Figure S27: Zoomed view of cat *COA1* exon-2 shown in **Supplementary Fig. S26**.

.

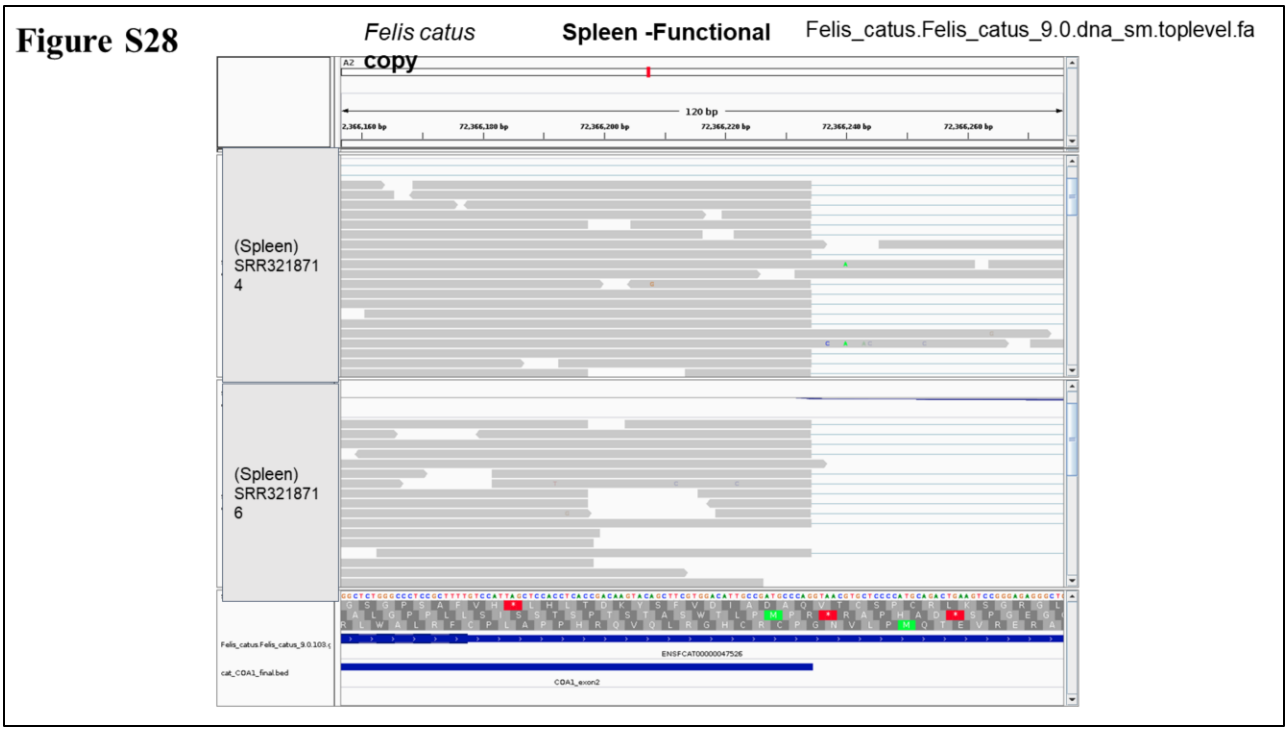


Figure S28: Zoomed view of cat *COA1* exon-2 shown in **Supplementary Fig. S26**.

.

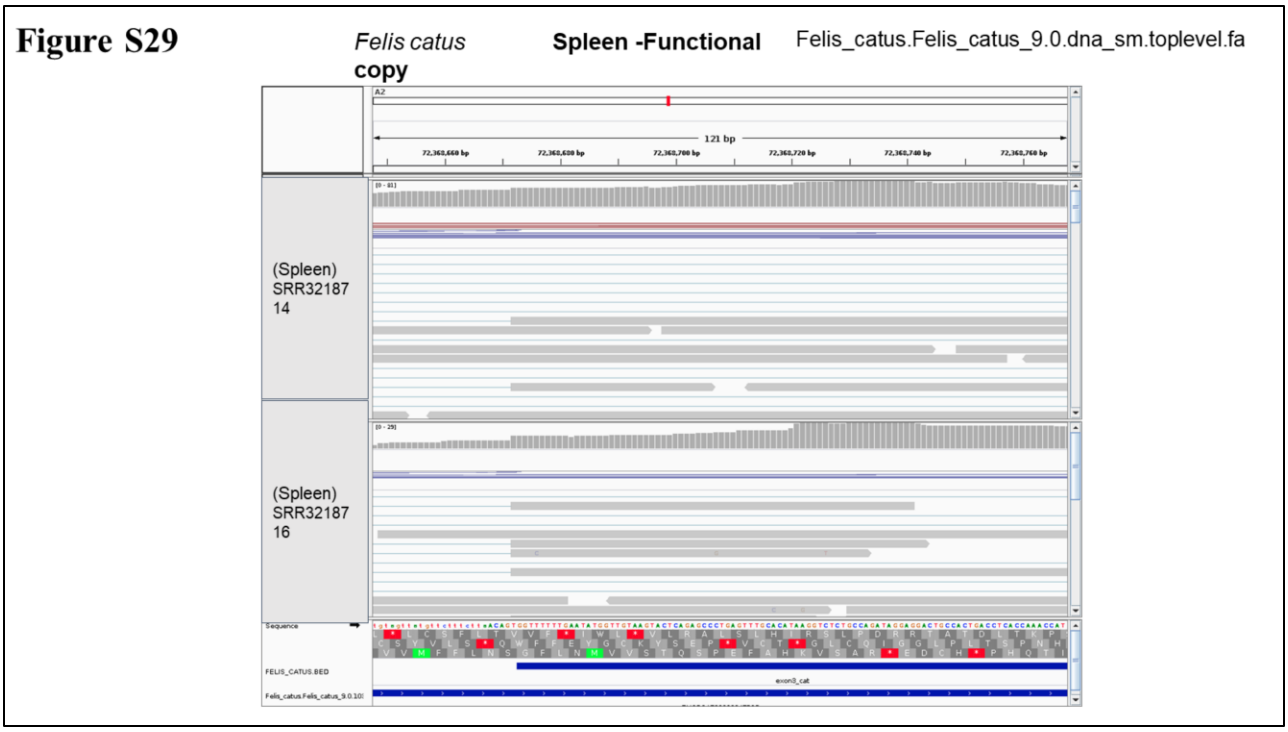


Figure S29: Zoomed view of cat *COA1* exon-3 shown in **Supplementary Fig. S26**.

.

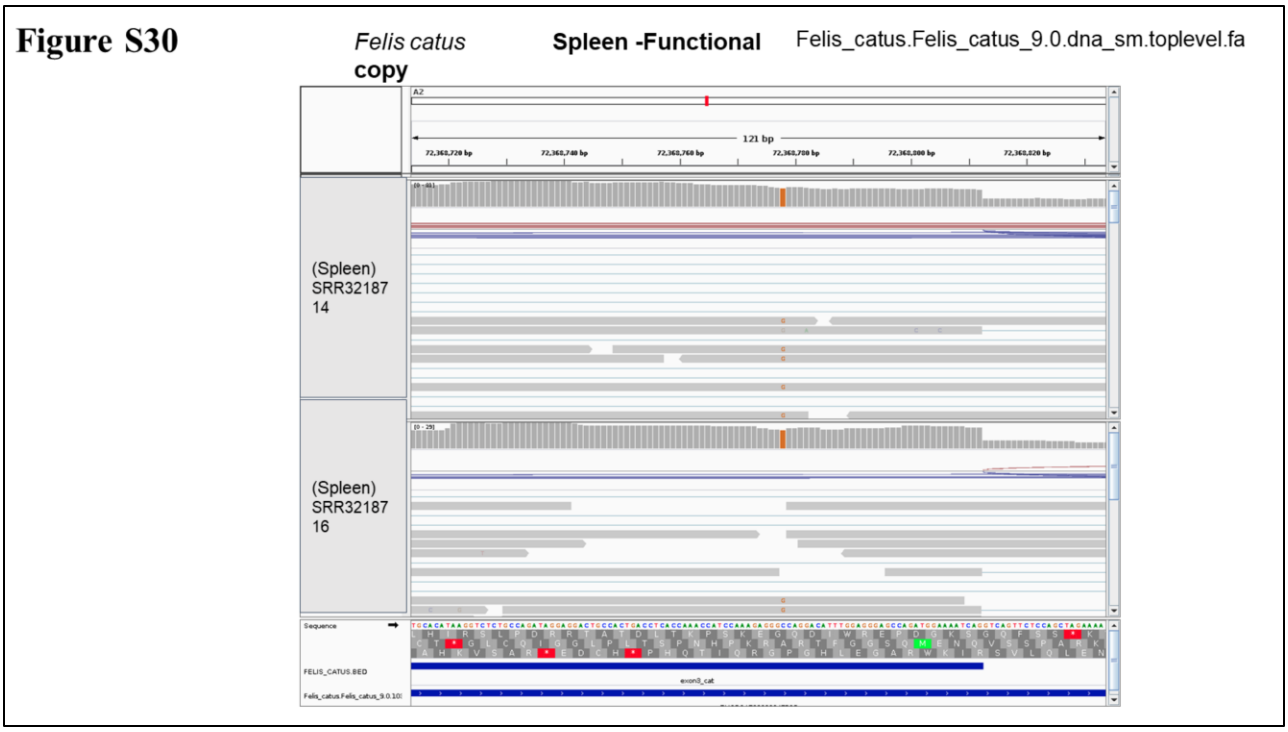


Figure S30: Zoomed view of cat *COA1* exon-3 shown in **Supplementary Fig. S26**.

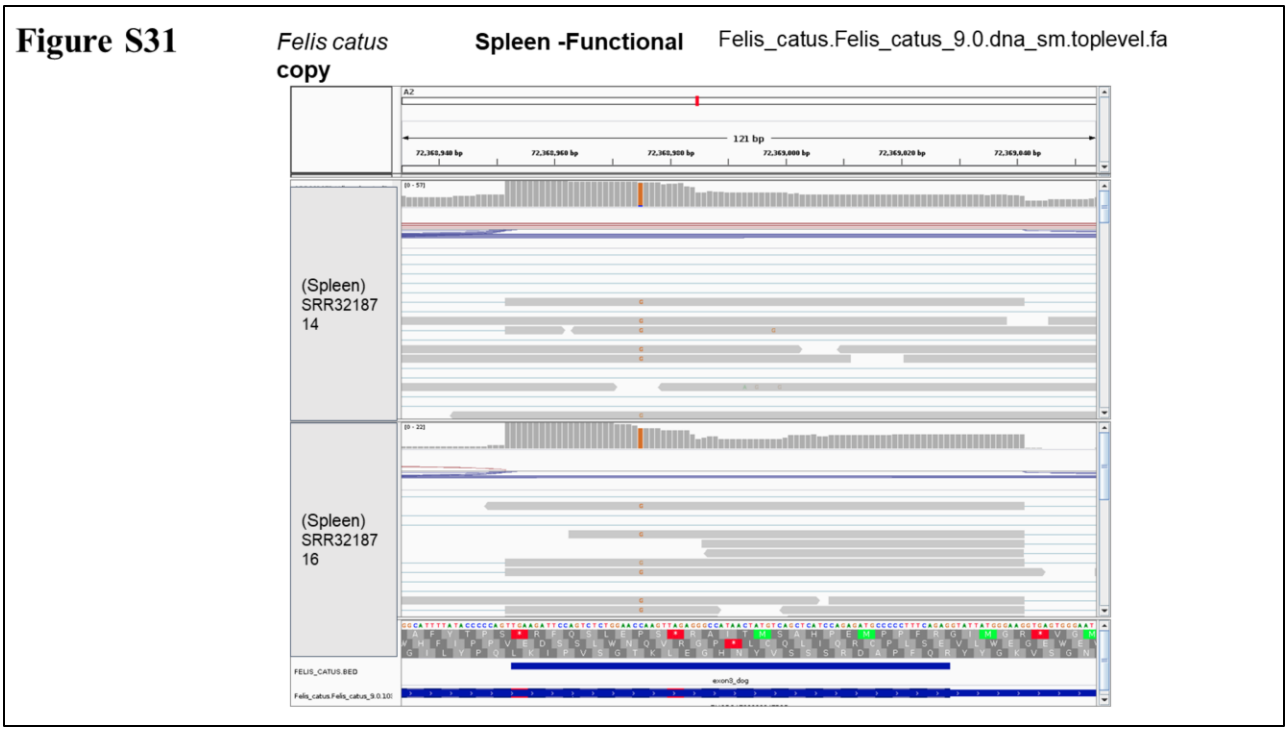


Figure S31: Zoomed view of dog *COA1* exon-3 shown in **Supplementary Fig. S26**.

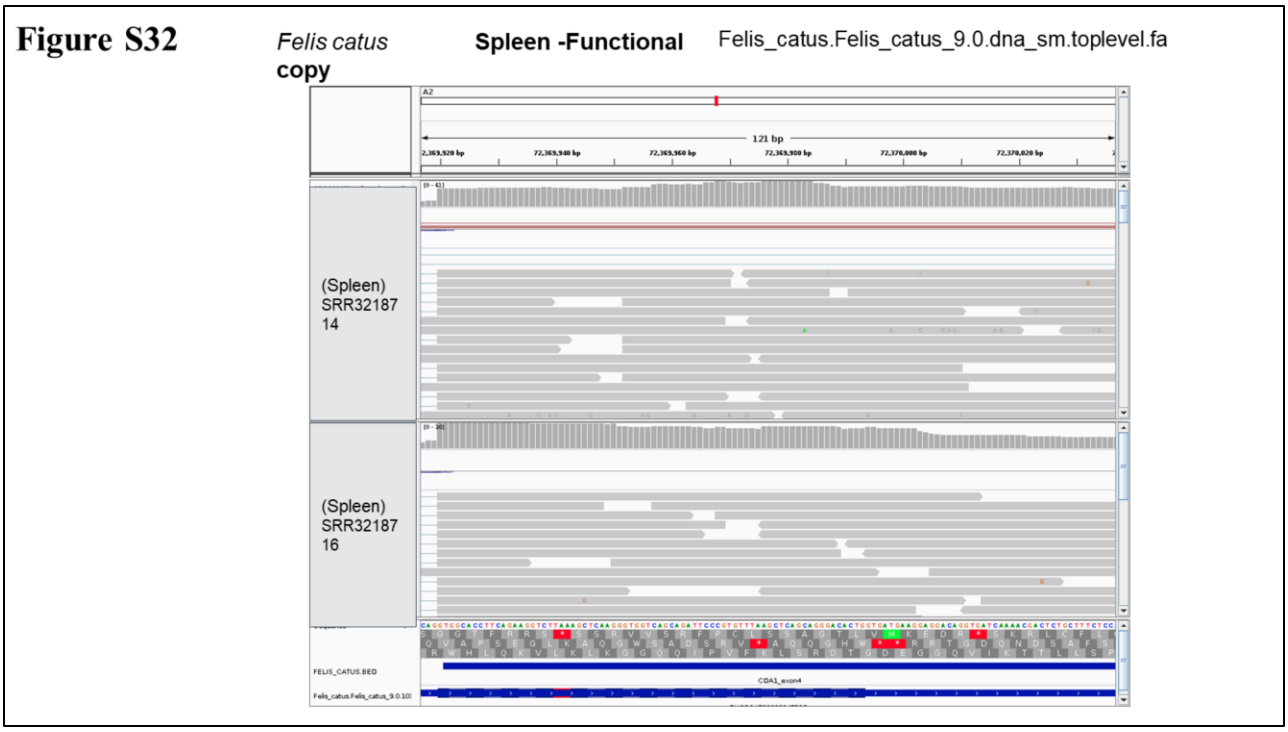


Figure S32: Zoomed view of cat *COA1* exon-4 shown in **Supplementary Fig. S26**.

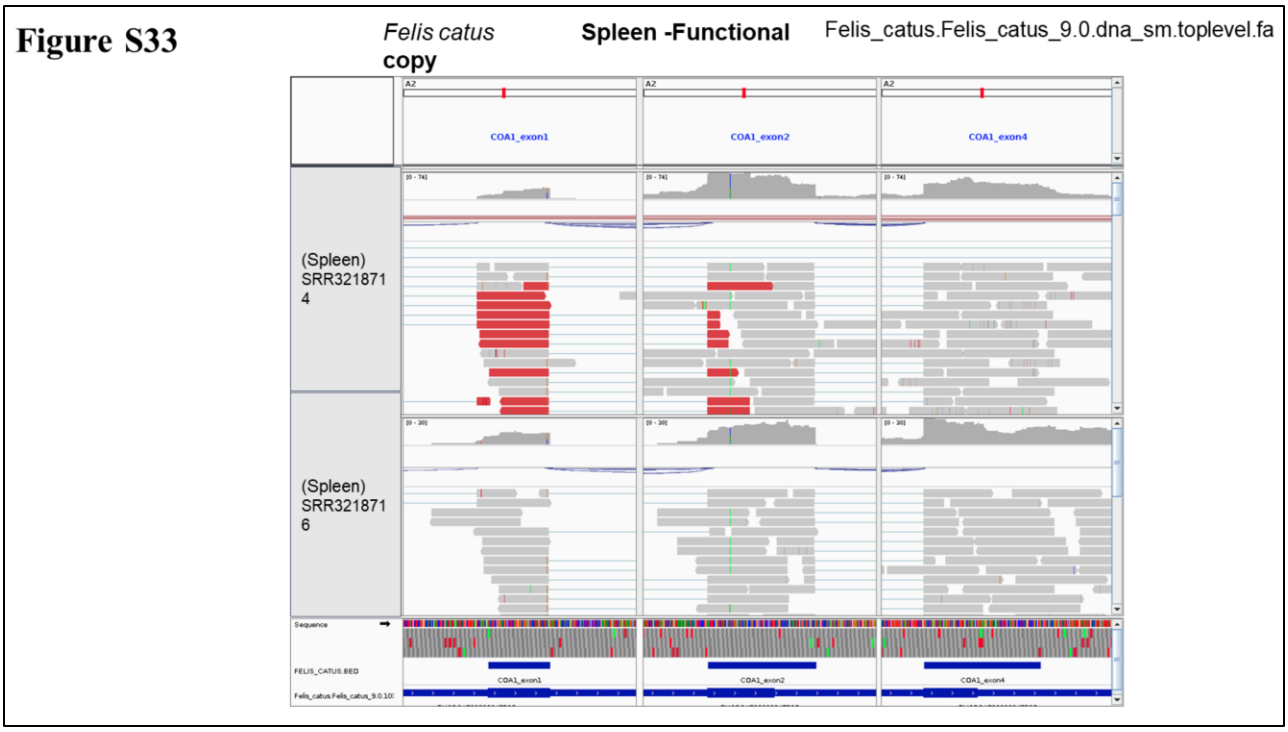


Figure S33: Zoomed view of all three cds forming exons shown in **Supplementary Fig. S26**.

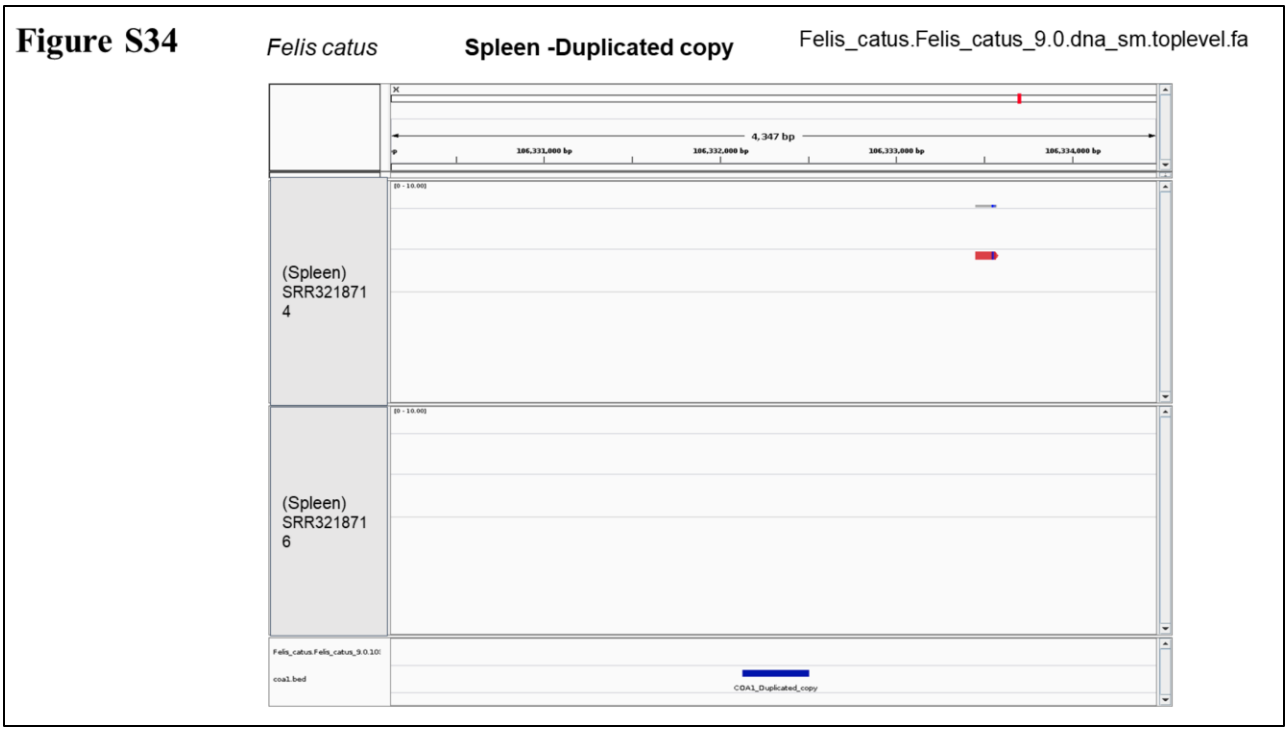


Figure S34: Screenshot of RNA-seq reads aligned to each of the putative locations of the *COA1* duplicated copy in cat (*Felis catus*). Datasets from spleen (SRR3218714 and SRR3218716) have been mapped to the Felis_catus.Felis_catus_9.0.dna_sm.toplevel.fa. Locations of of the *COA1* gene duplicated copy is depicted as bed record in the bottom row.

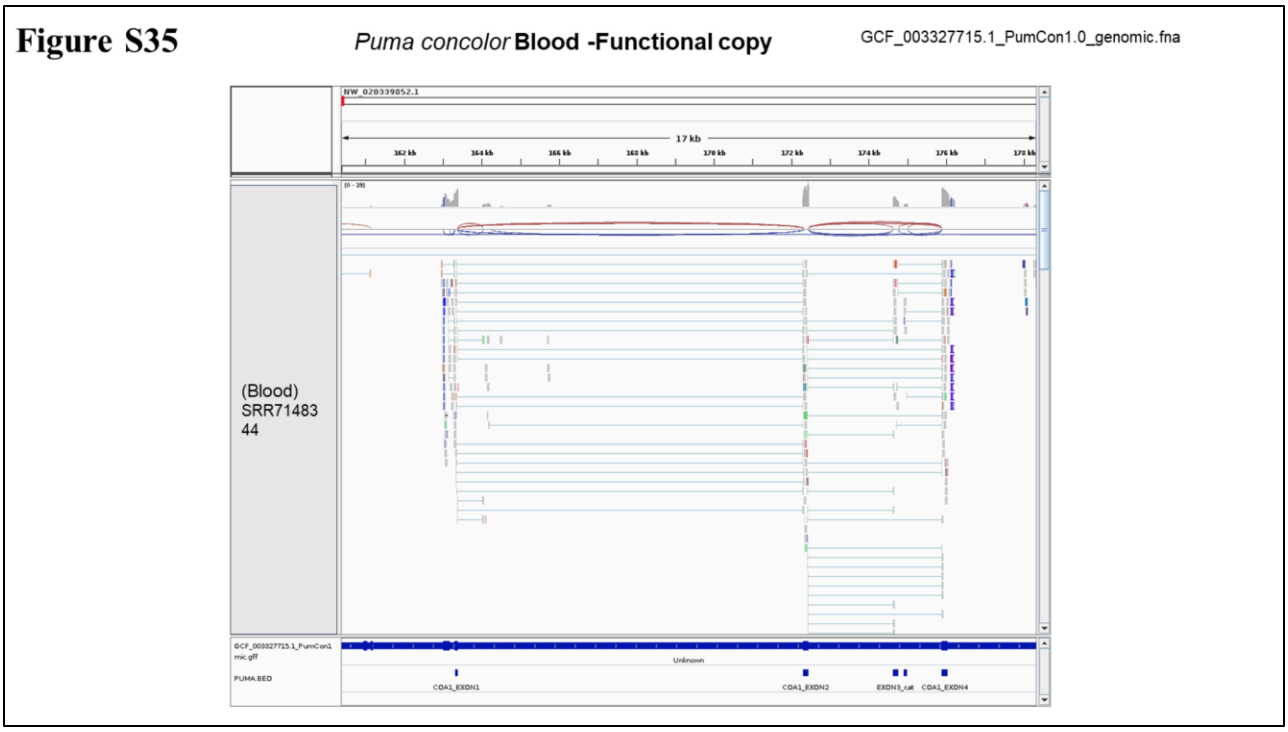


Figure S35: Screenshot of RNA-seq reads aligned to each of the putative locations of the all *COA1* exons in puma (*Puma concolor*). Gray boxes represent reads aligned to the genome and thin lines connecting these boxes represent the spliced read alignments. Datasets from blood (SRR7148344) have been mapped to the GCF_003327715.1_PumCon1.0_genomic.fna. Locations of each of the exons are depicted as bed record in the bottom row.

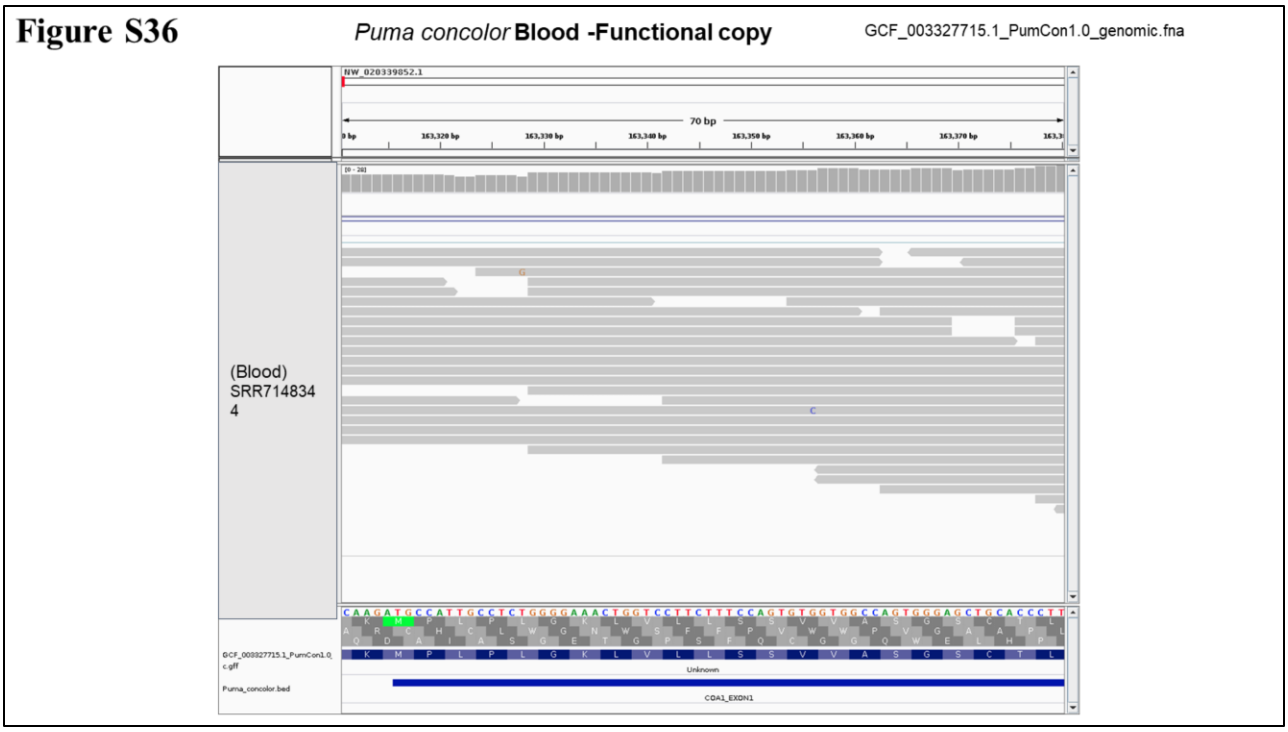


Figure S36: Zoomed view of puma *COA1* exon-1 shown in **Supplementary Fig. S35**.

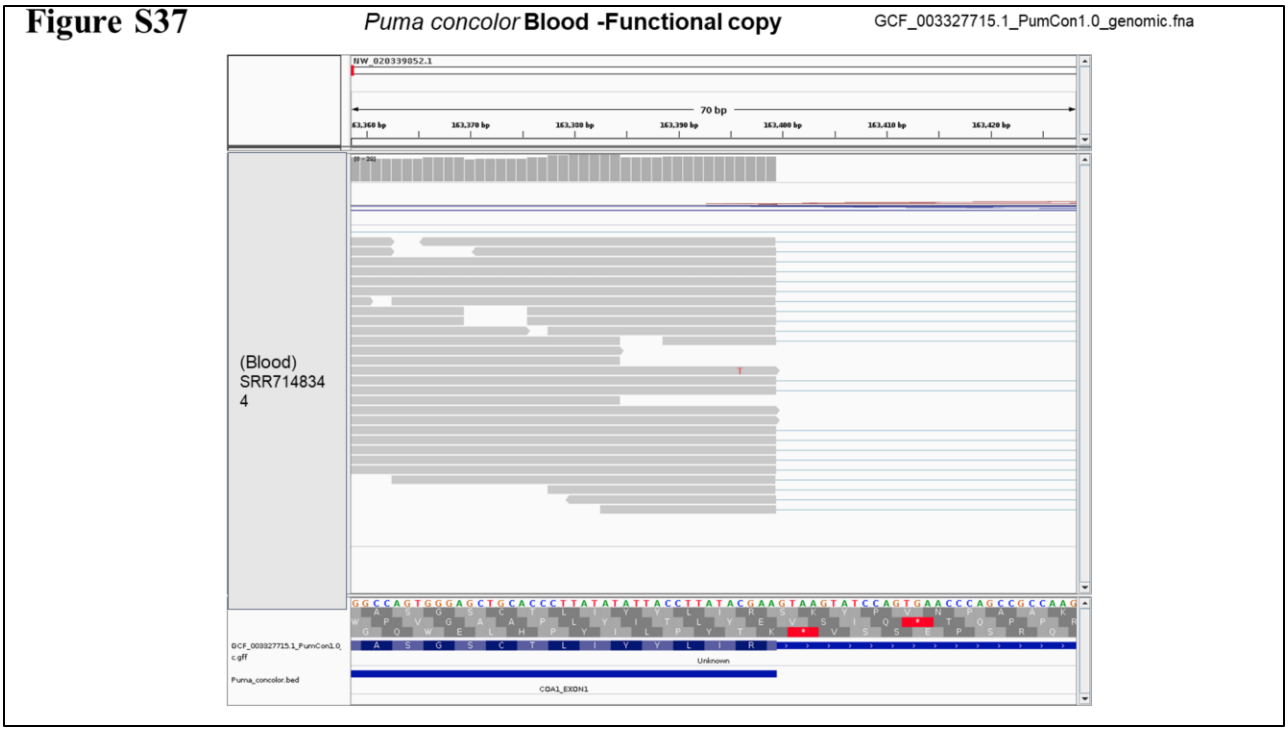


Figure S37: Zoomed view of puma *COA1* exon-1 shown in **Supplementary Fig. S35.**

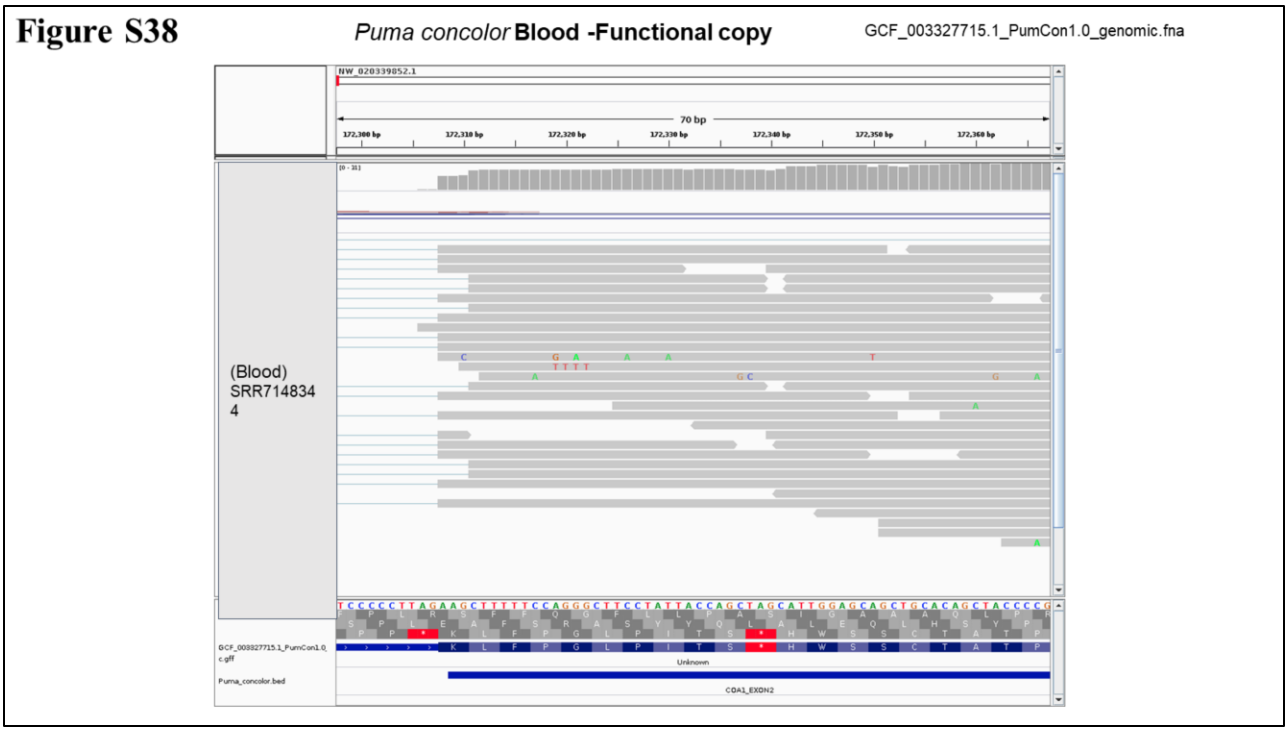


Figure S38: Zoomed view of puma *COA1* exon-2 shown in **Supplementary Fig. S35.**

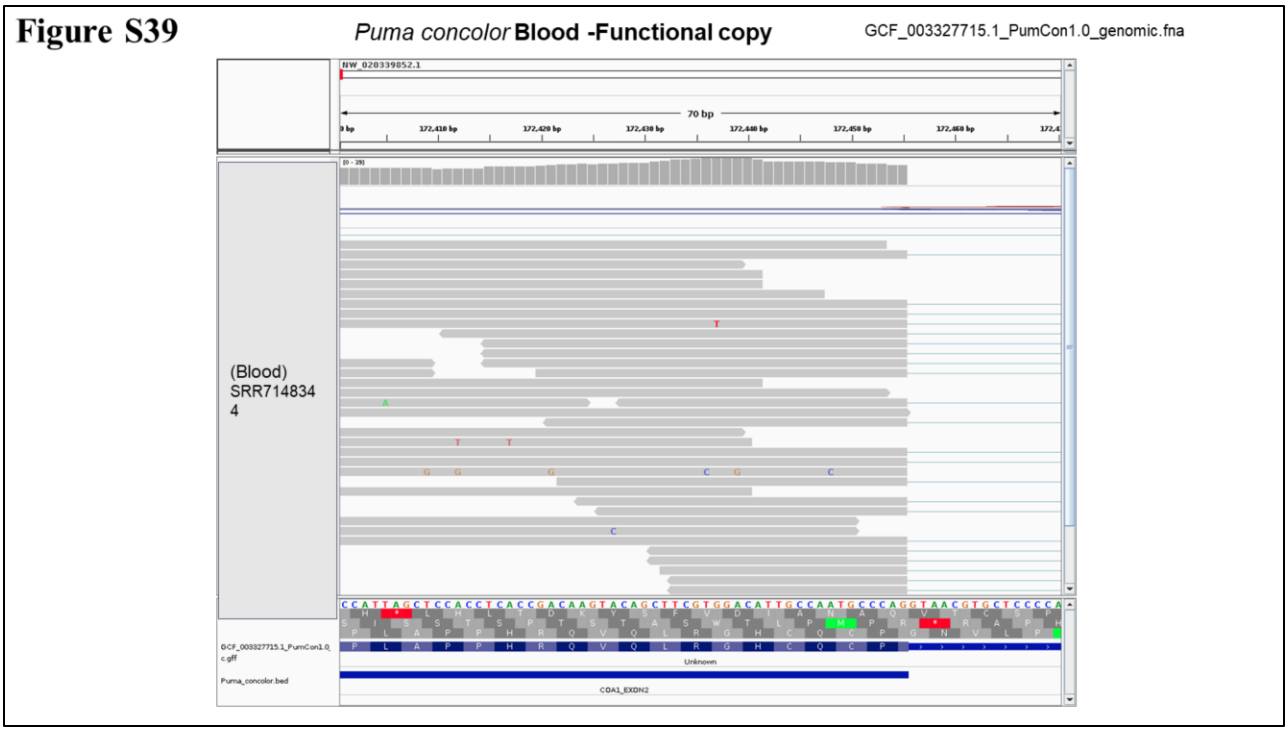


Figure S39: Zoomed view of puma *COA1* exon-2 shown in **Supplementary Fig. S35.**

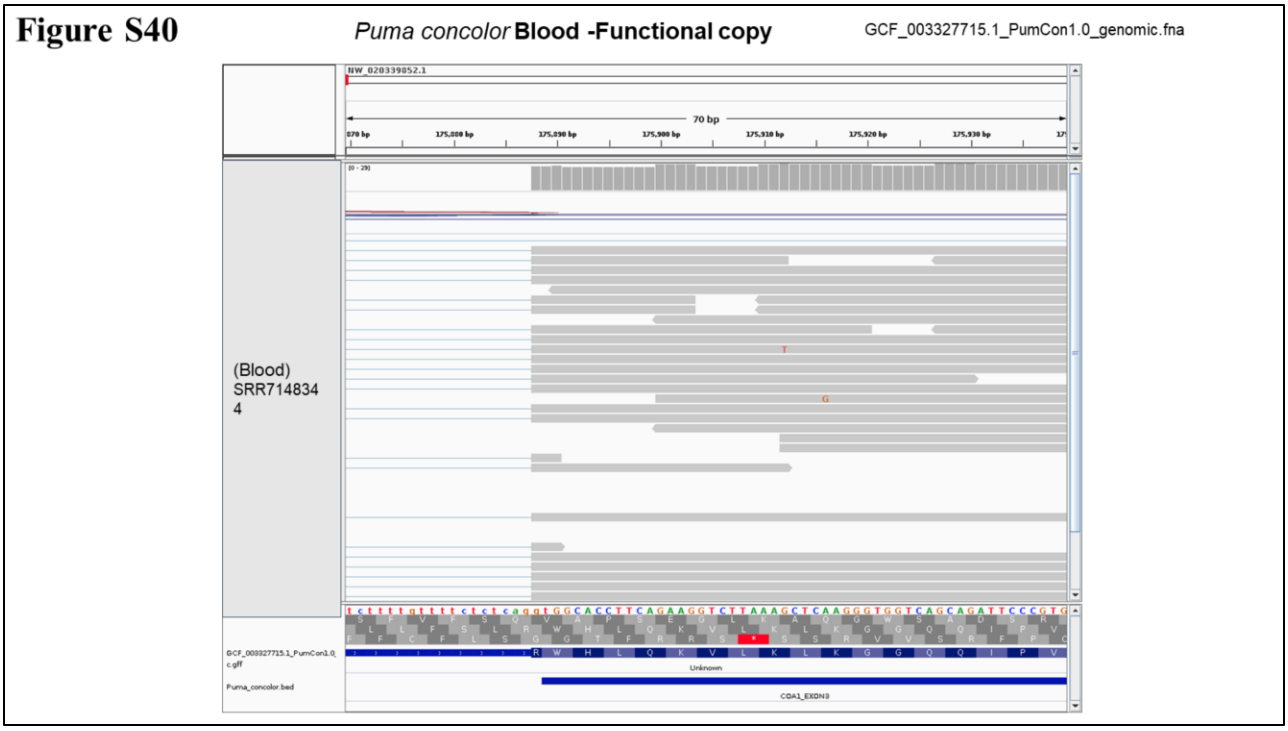


Figure S40: Zoomed view of puma *COA1* exon-3 shown in **Supplementary Fig. S35.**

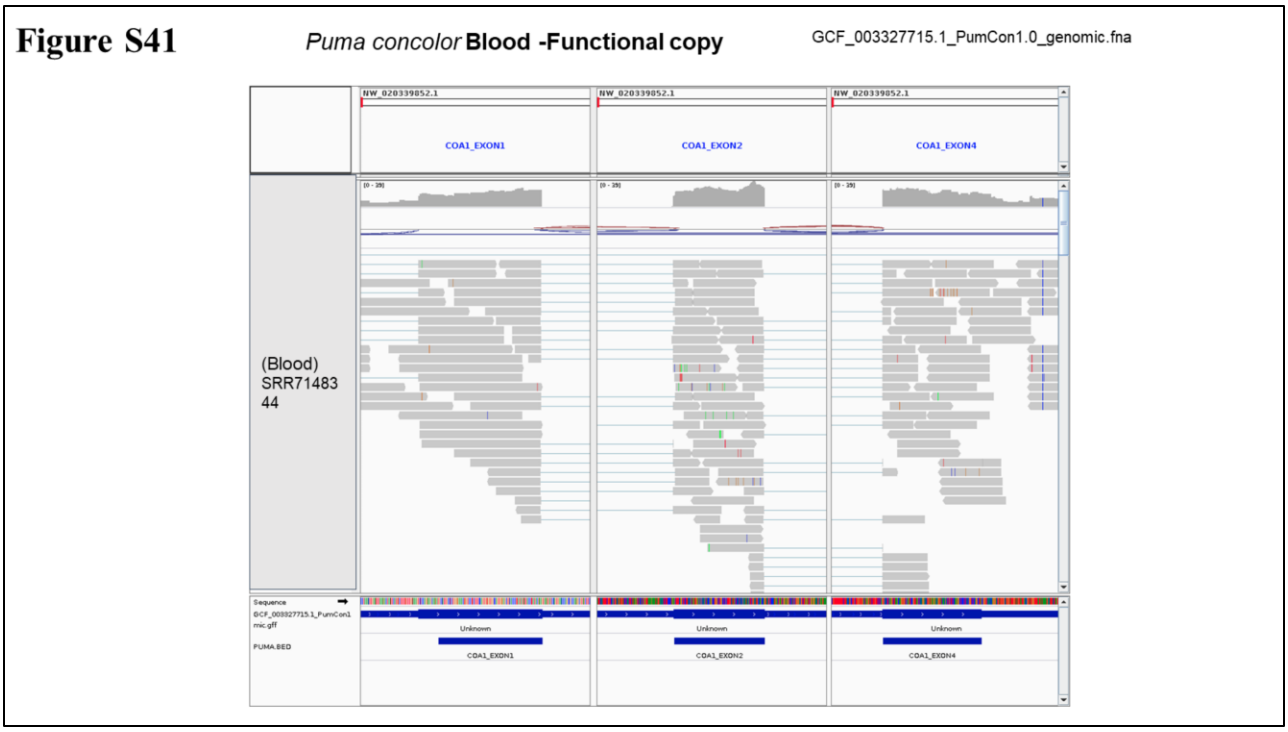


Figure S41: Zoomed view of all three cds coding exons shown in **Supplementary Fig. S35.**

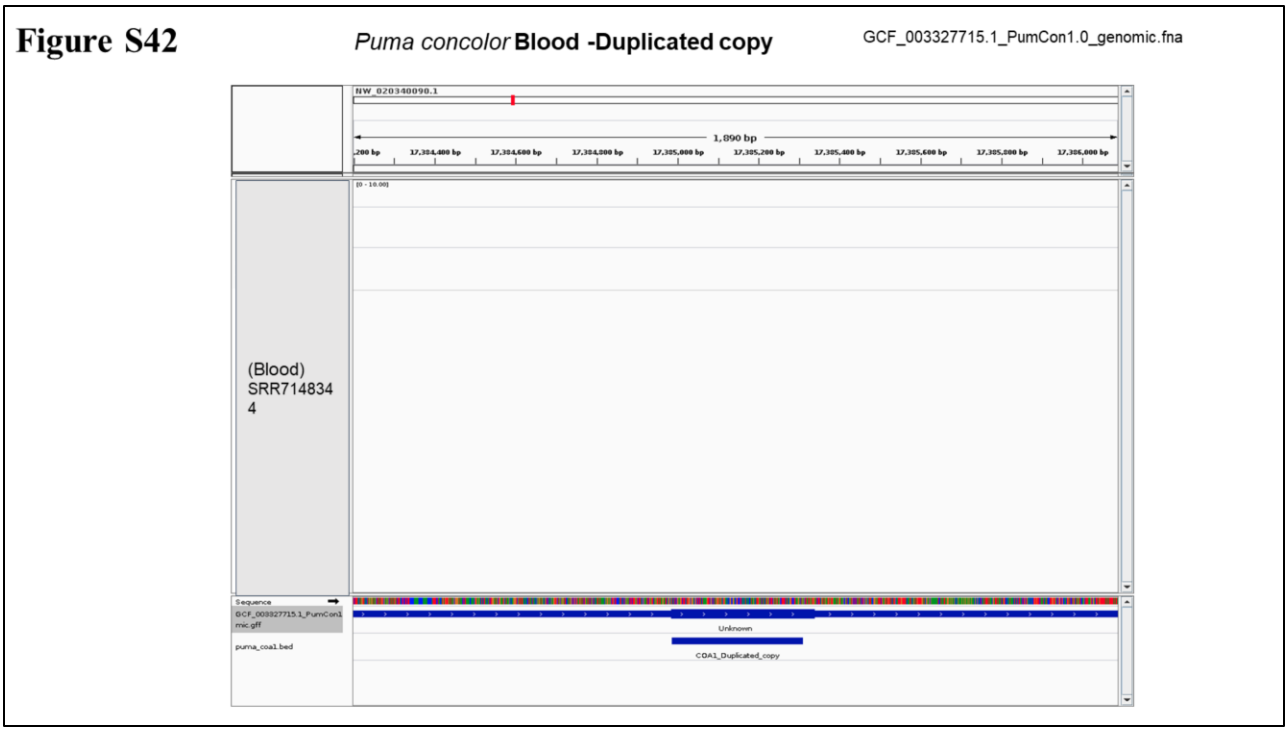


Figure S42: Screenshot of RNA-seq reads aligned to each of the putative locations of the *COA1* duplicated copy in puma (*Puma concolor*). Datasets from blood (SRR7148344) have been mapped to the GCF_003327715.1_PumCon1.0_genomic.fna. Locations of of the *COA1* gene duplicated copy is depicted as bed record in the bottom row.

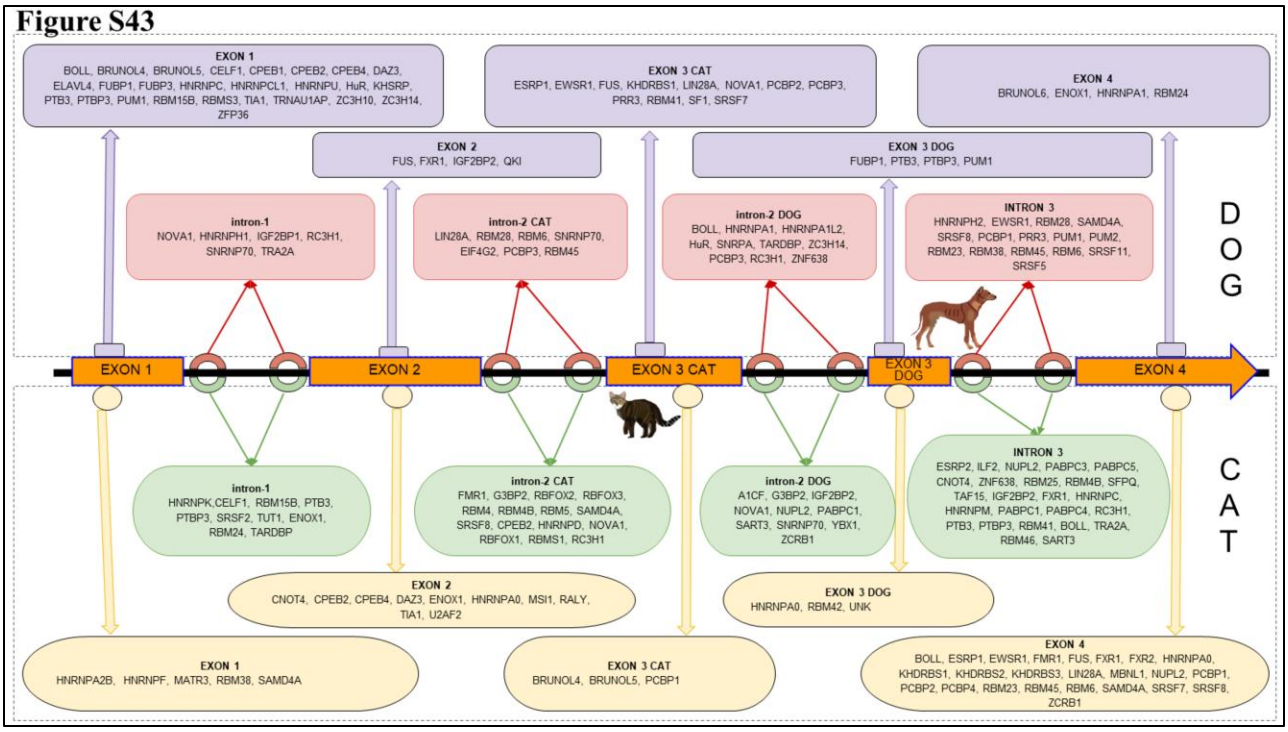


Figure S43: Different types of enhancers and silencers for different motifs of exons and introns of *COA1* gene unique in *Canis lupus familiaris* and *Felis catus* downloaded from RBPmap website (**Supplementary Table S25** and **S26**).

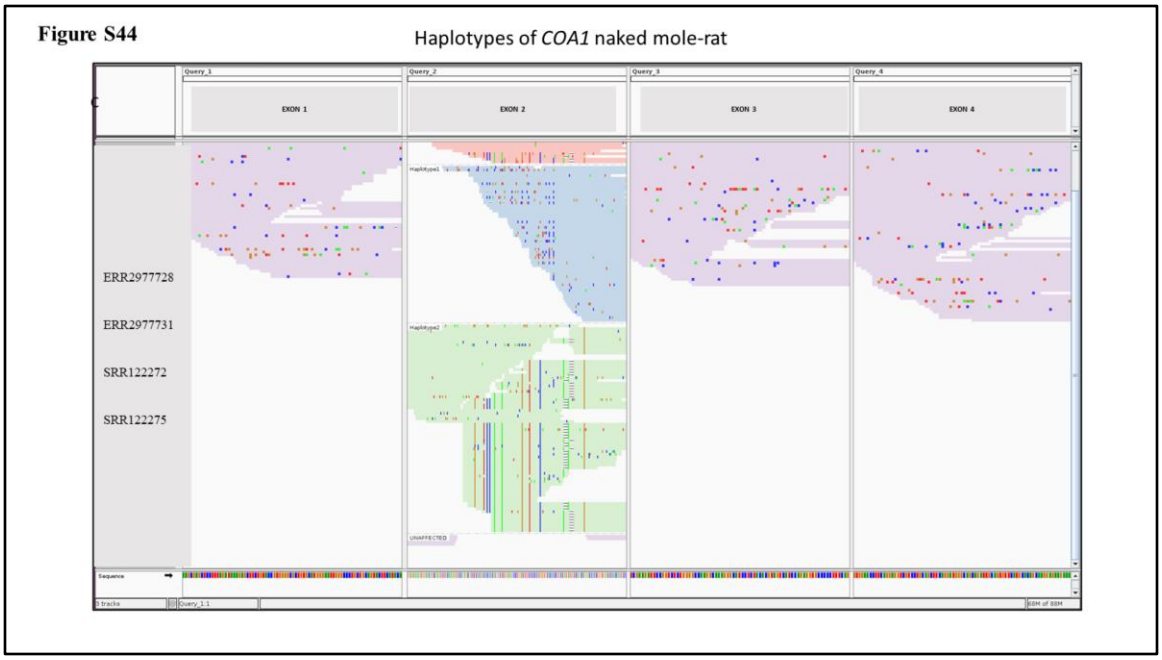


Figure S44: Two haplotypes in exon two and one haplotype of remaining exons of *COA1* in naked mole-rat (*Heterocephalus glaber*).

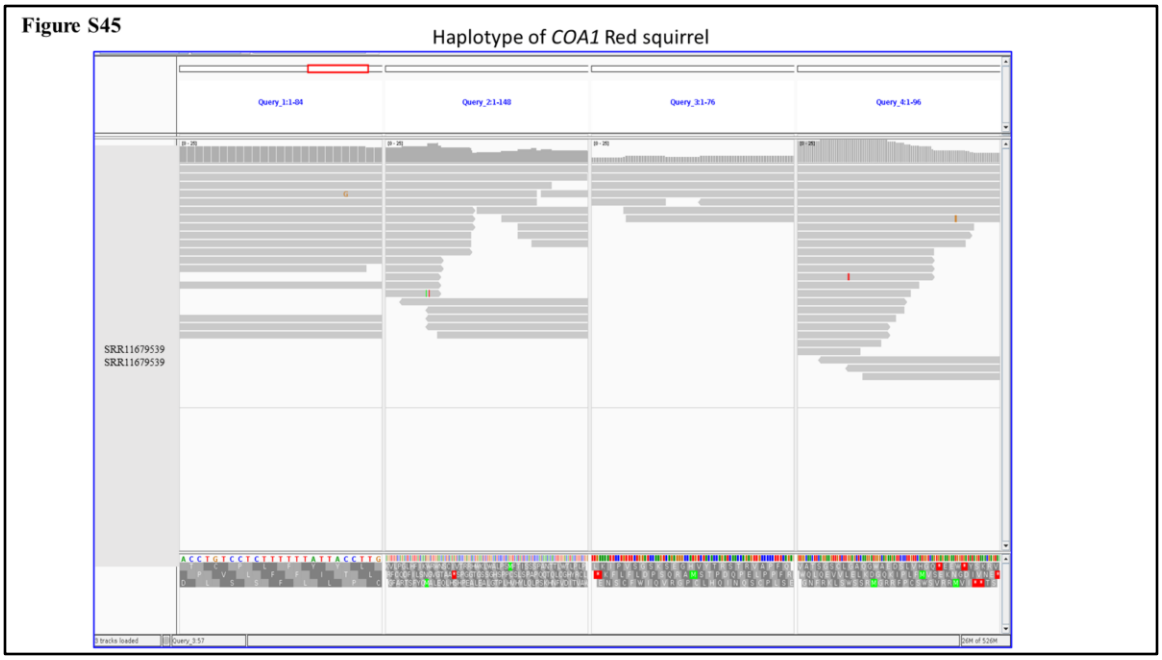


Figure S45: The single haplotype of *COA1* in red squirrel (*Sciurus vulgaris*).

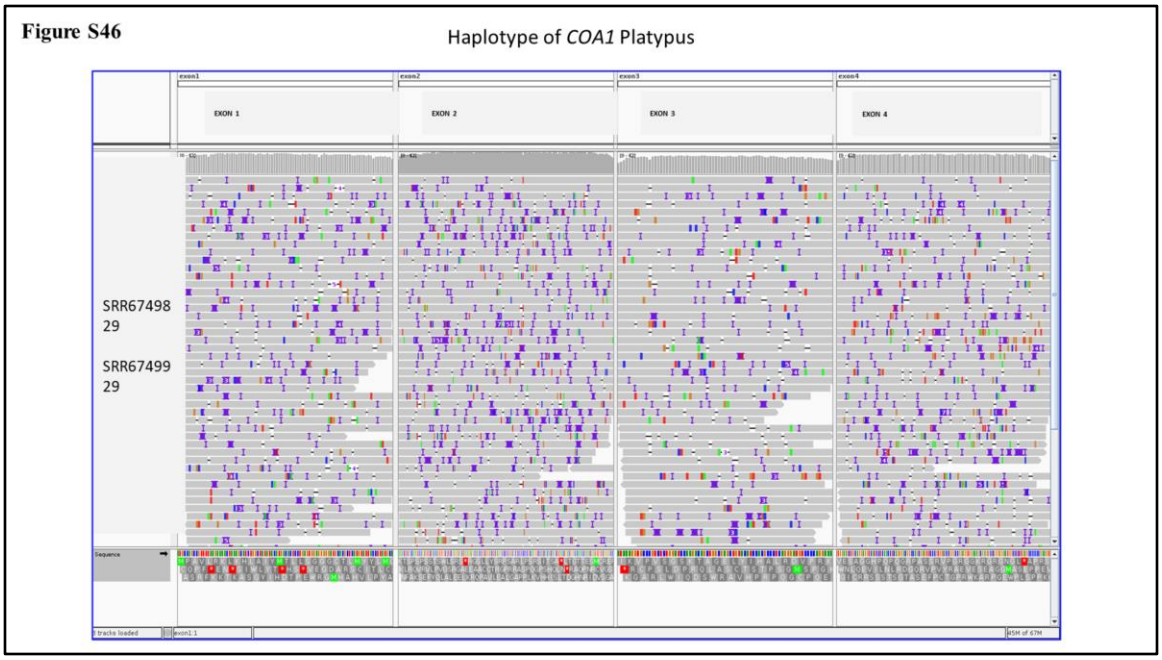


Figure S46: The single haplotype of *COA1* in the platypus (*Ornithorhynchus anatinus*).

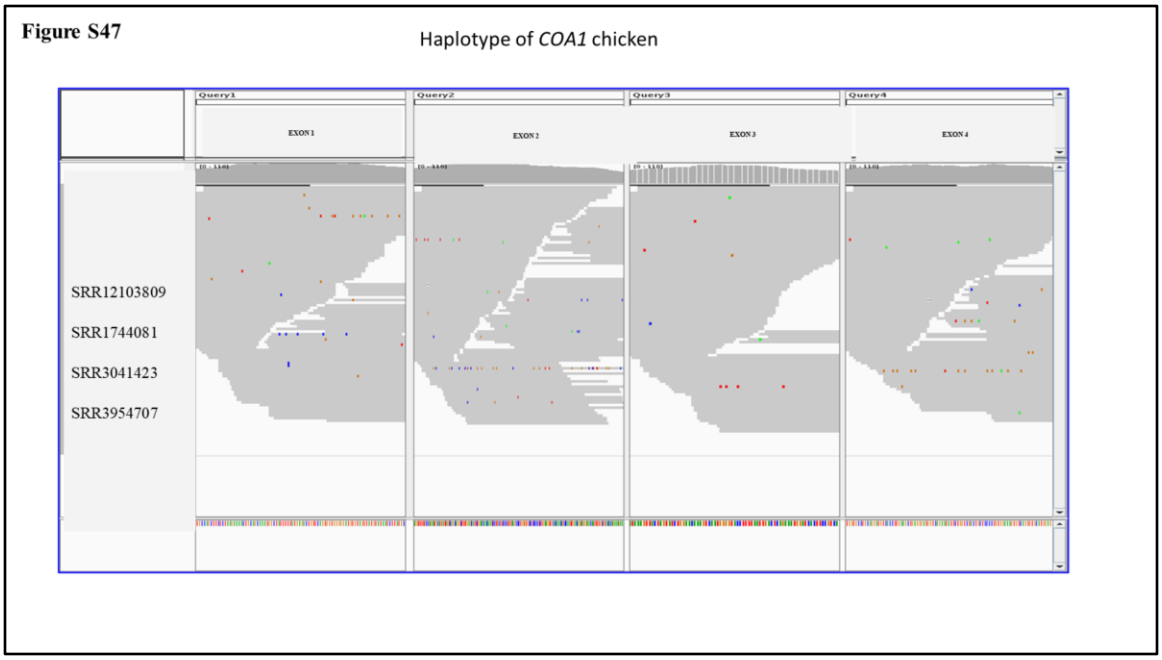


Figure S47: The single haplotype of *COA1* gene in chicken (*Gallus gallus*).

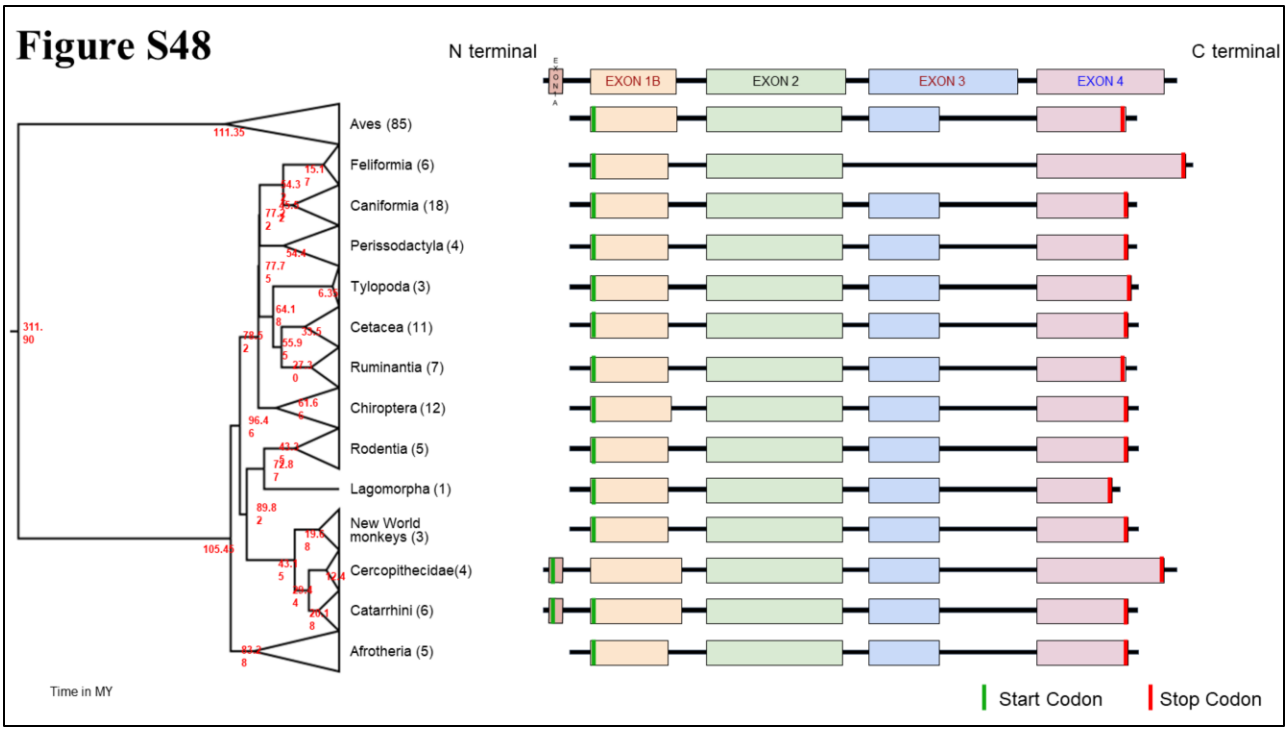


Figure S48: Overall length variation of *COA1* transcript with their start and stop codon across various groups. The numbers in brackets represents the number of species clubbed together in the time calibrated phylogenetic tree taken from TimeTree website.

Figure S49 *Acinonyx jubatus*(cheetah) *COA1* haplotype 1 figure of Illumina database

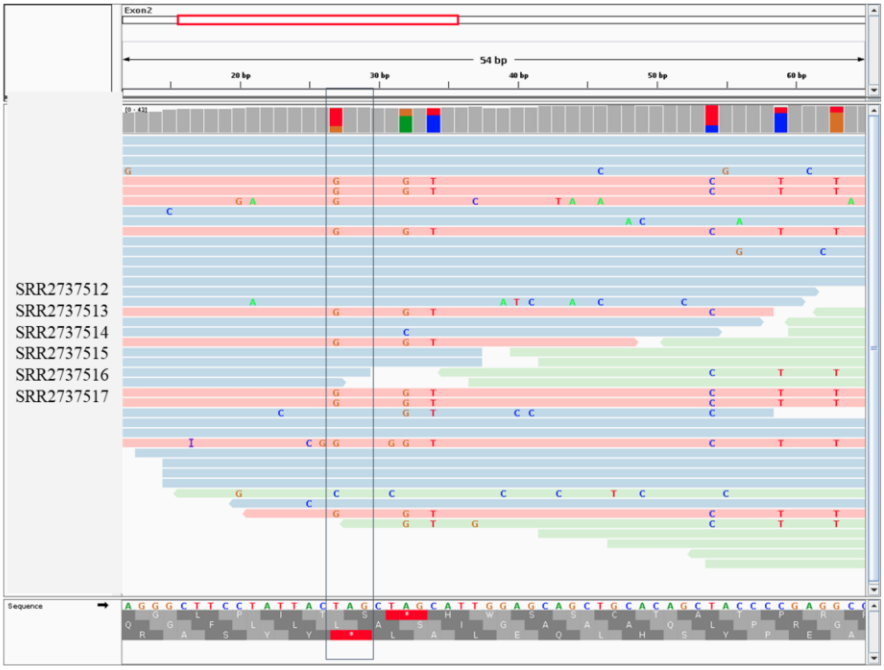


Figure S49: Zoomed view of exon-2 of *COA1*, vertical box is indicating the polymorphic site at position 27 where proportion of T is higher than G which leads a premature stop codon in haplotype 1 of *COA1* in cheetah.

Figure S50 *Acinonyx jubatus*(cheetah) COA1 Haplotype 1 figure of Illumina database

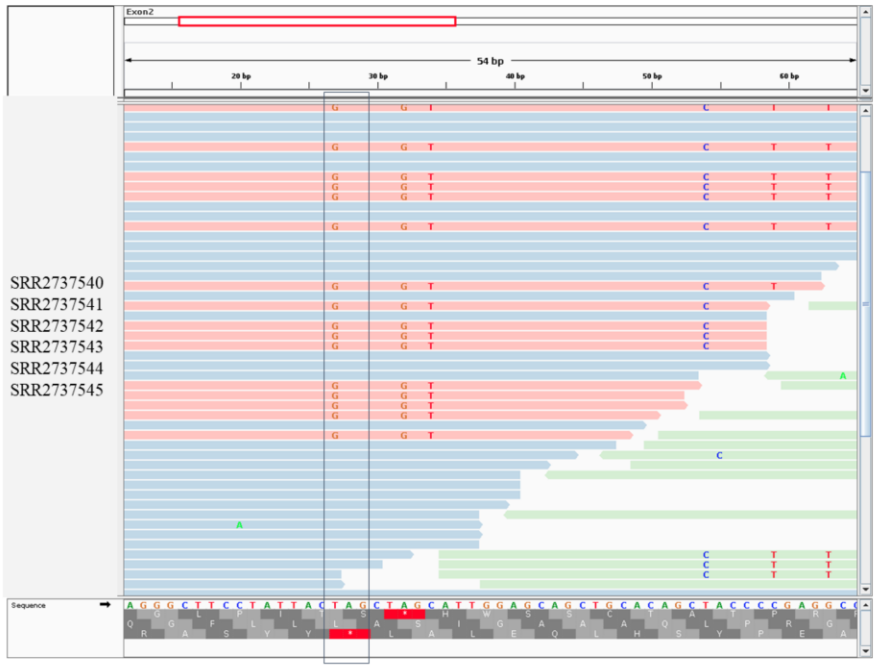


Figure S50: Zoomed view of exon-2 of *COA1*, vertical box is indicating the polymorphic site at position 27 where proportion of T is higher than G which leads a premature stop codon in cheetah.

Figure S51

Acinonyx jubatus(cheetah) *COA1* Haplotype 1 figure of Hi-C database

Figure S51: Zoomed view of exon-2 of *COA1*, vertical box is indicating the polymorphic site at position 27 where proportion of T is higher than G which leads a premature stop codon in cheetah.

Figure S52 *Acinonyx jubatus*(cheetah) *COA1* Haplotype 2 figure of Illumina database

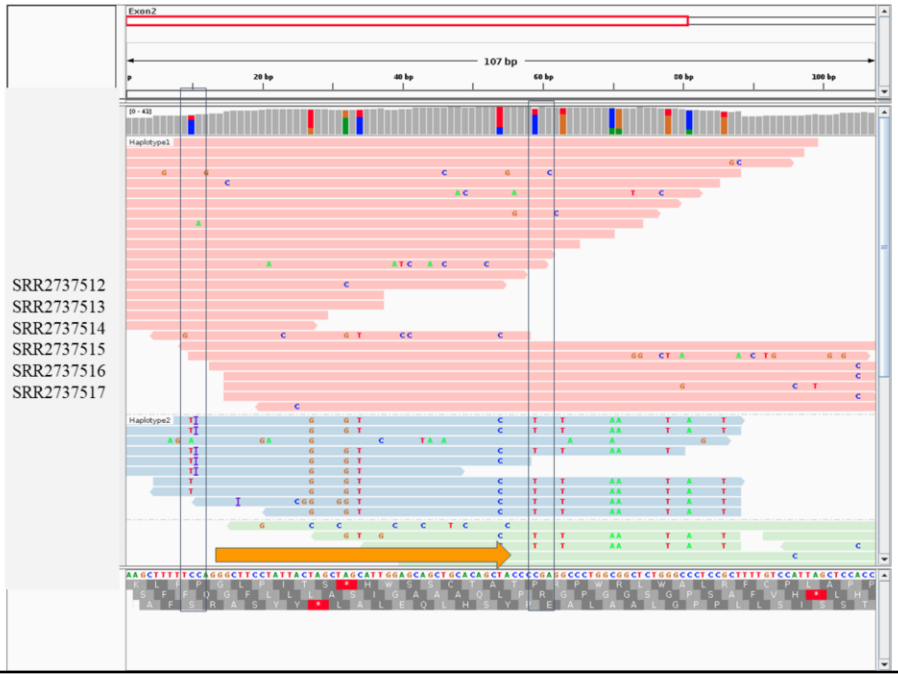


Figure S52: Zoomed view of exon-2 of *COA1*, first vertical box is indicating the insertion of T at position 10, which leads a premature stop codon at position 59 (2nd vertical box) in haplotype 2 of *COA1* in cheetah.

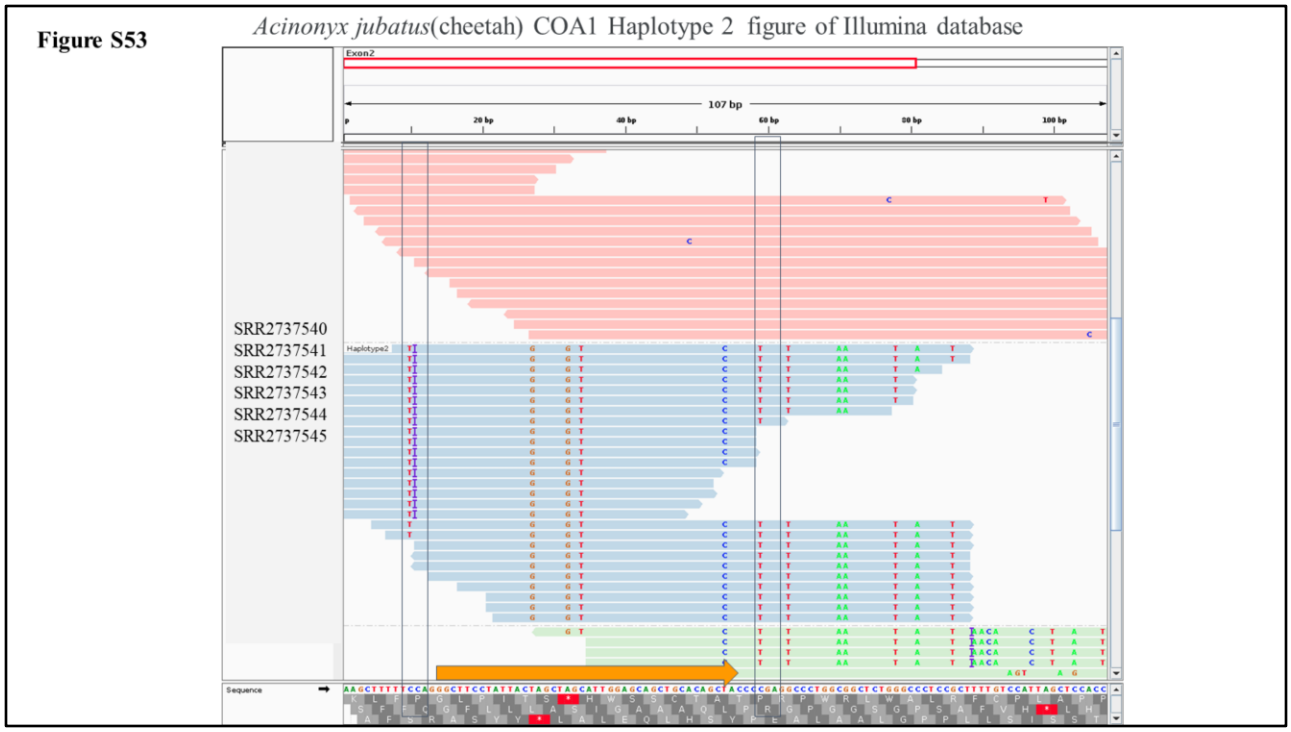


Figure S53: Zoomed view of exon-2 of *COA1*, first vertical box is indicating the insertion of T at position 10, which leads a premature stop codon at position 59 (2nd vertical box) in haplotype 2 of *COA1* in cheetah.

Figure S54 *Acinonyx jubatus*(cheetah) *COA1* Haplotype 2 figure of Hi-C database

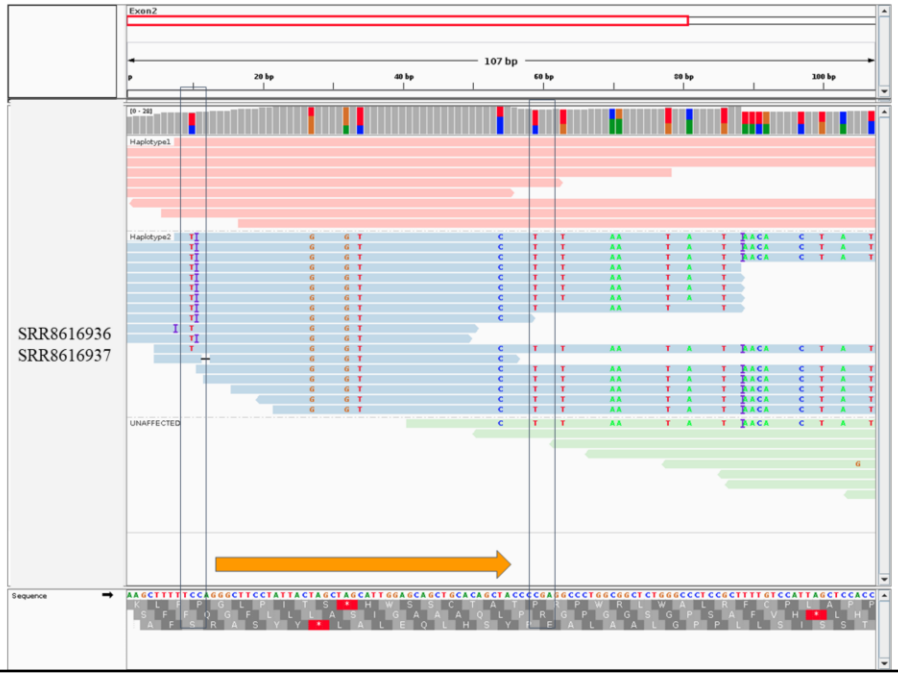


Figure S54: Zoomed view of exon-2 of *COA1*, first vertical box is indicating the insertion of T at position 10, which leads a premature stop codon at position 59 (2nd vertical box) in haplotype2 of *COA1* in cheetah.

Figure S55: Screenshot of RNA-seq reads aligned to each of the putative locations of the three *COA1* exons in cheetah *(SRR11679539*). Gray boxes represent reads aligned to the genome and thin lines connecting these boxes represent the spliced read alignments. Datasets from (SRR084769 SRR084770 SRR531831 SRR531832 SRR531833 SRR531834 SRR531835 SRR531836 SRR531837 SRR531838) have been mapped to the GCF_003709585.1_Aci_jub_2_genomic.fna. Locations of each of the exons are depicted as bed record in the bottom row.

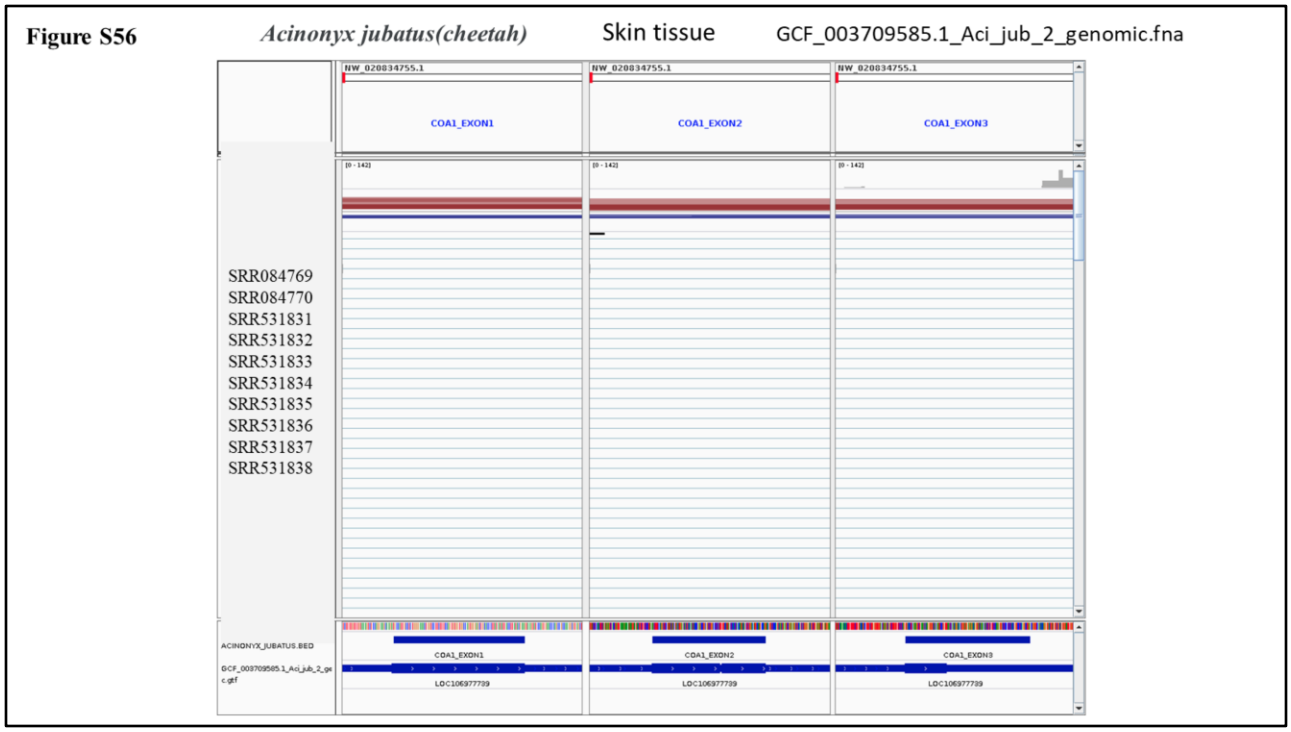


Figure S56: A zoomed in view of each of the *COA1* exons shown in **Supplementary Fig. S55**.

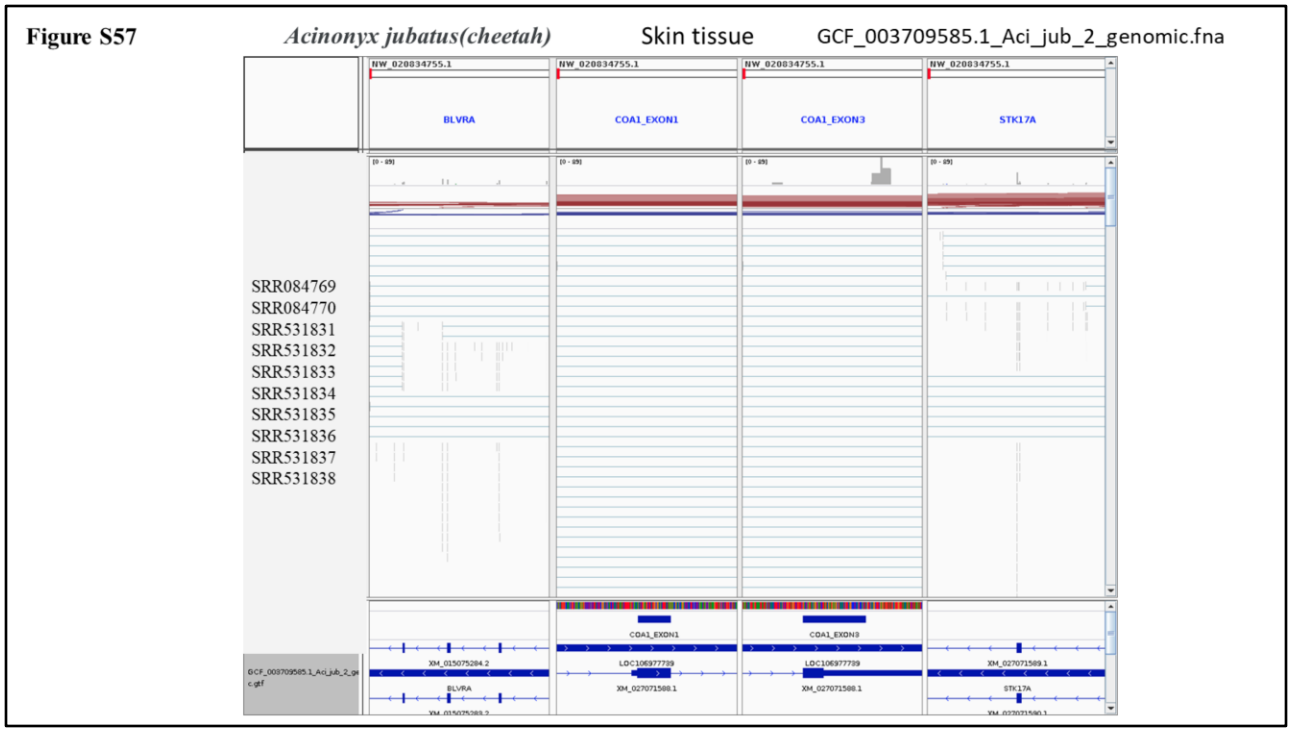


Figure S57: Screenshot of RNA-seq reads aligned to each of the putative locations of the three *COA1* exons and its nearby gene *BLVRA* and *STK17A* in cheetah (*Acinonyx jubatus*). Gray boxes represent reads aligned to the genome and thin lines connecting these boxes represent the spliced read alignments. Datasets from (SRR084769 SRR084770 SRR531831 SRR531832 SRR531833 SRR531834 SRR531835 SRR531836 SRR531837 and SRR531838) have been mapped to the GCF_003709585.1_Aci_jub_2_genomic.fna. Locations of each of the exons are depicted as bed record in the bottom row.

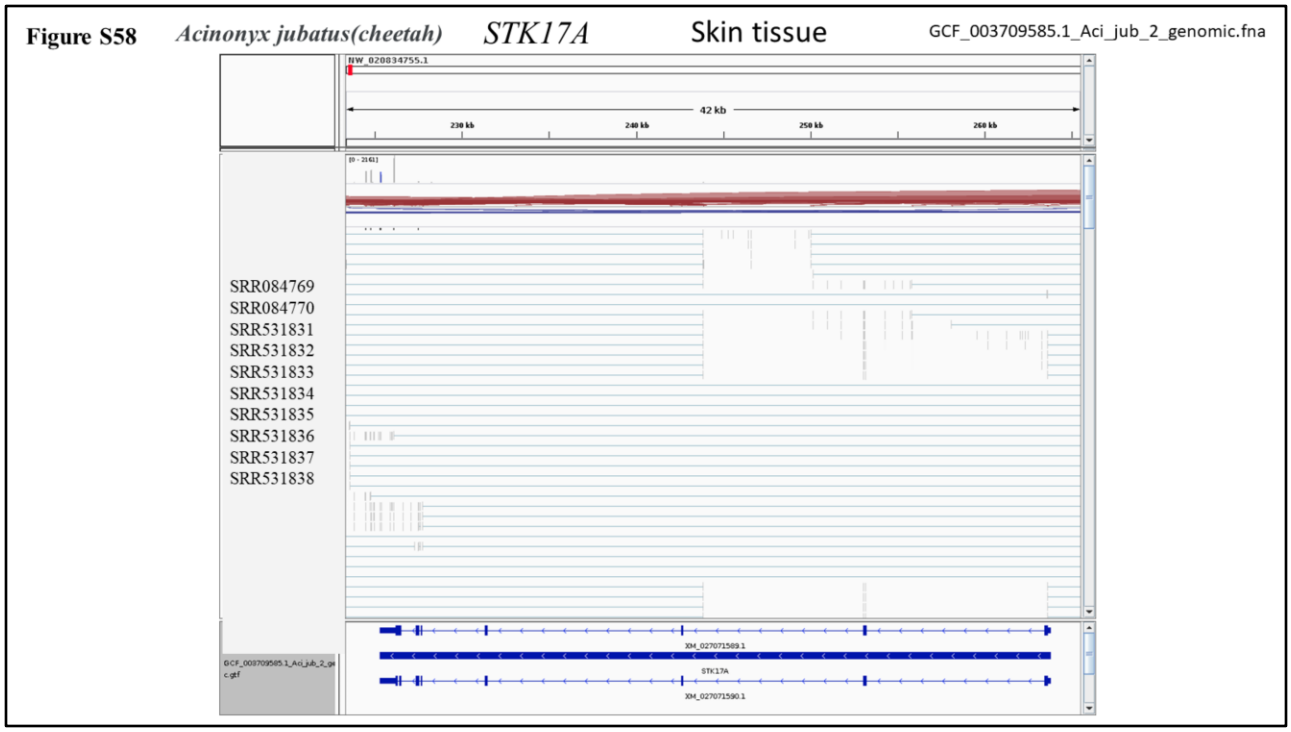


Figure S58: Screenshot of RNA-seq reads aligned to each of the putative locations of the *STK17A* in cheetah *(Acinonyx jubatus*). Gray boxes represent reads aligned to the genome and thin lines connecting these boxes represent the spliced read alignments. Datasets from (SRR084769 SRR084770 SRR531831 SRR531832 SRR531833 SRR531834 SRR531835 SRR531836 SRR531837 SRR531838) have been mapped to the GCF_003709585.1_Aci_jub_2_genomic.fna. Locations of each of the exons are depicted as bed record in the bottom row.

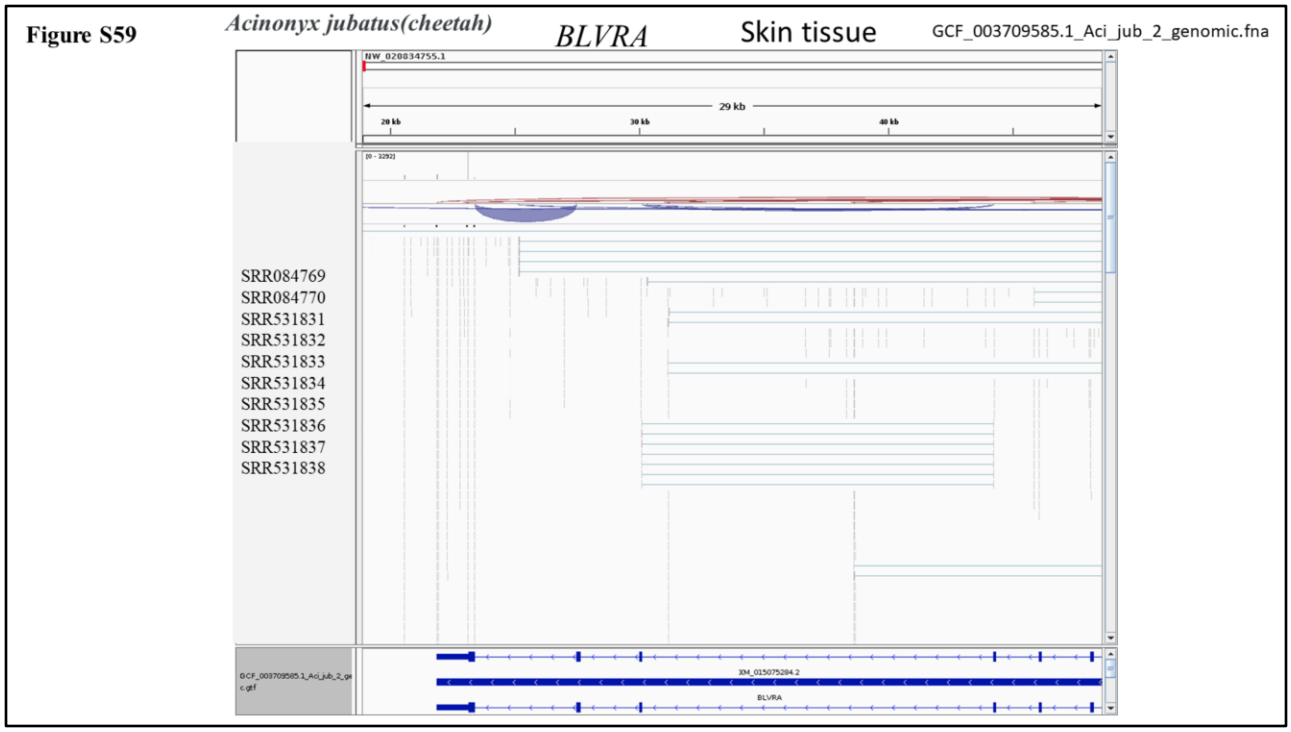


Figure S59: Screenshot of RNA-seq reads aligned to each of the putative locations of the *BLVRA* in cheetah *(Acinonyx jubatus*). Gray boxes represent reads aligned to the genome and thin lines connecting these boxes represent the spliced read alignments. Datasets from (SRR084769 SRR084770 SRR531831 SRR531832 SRR531833 SRR531834 SRR531835 SRR531836 SRR531837 SRR531838) have been mapped to the GCF_003709585.1_Aci_jub_2_genomic.fna. Locations of each of the exons are depicted as bed record in the bottom row.

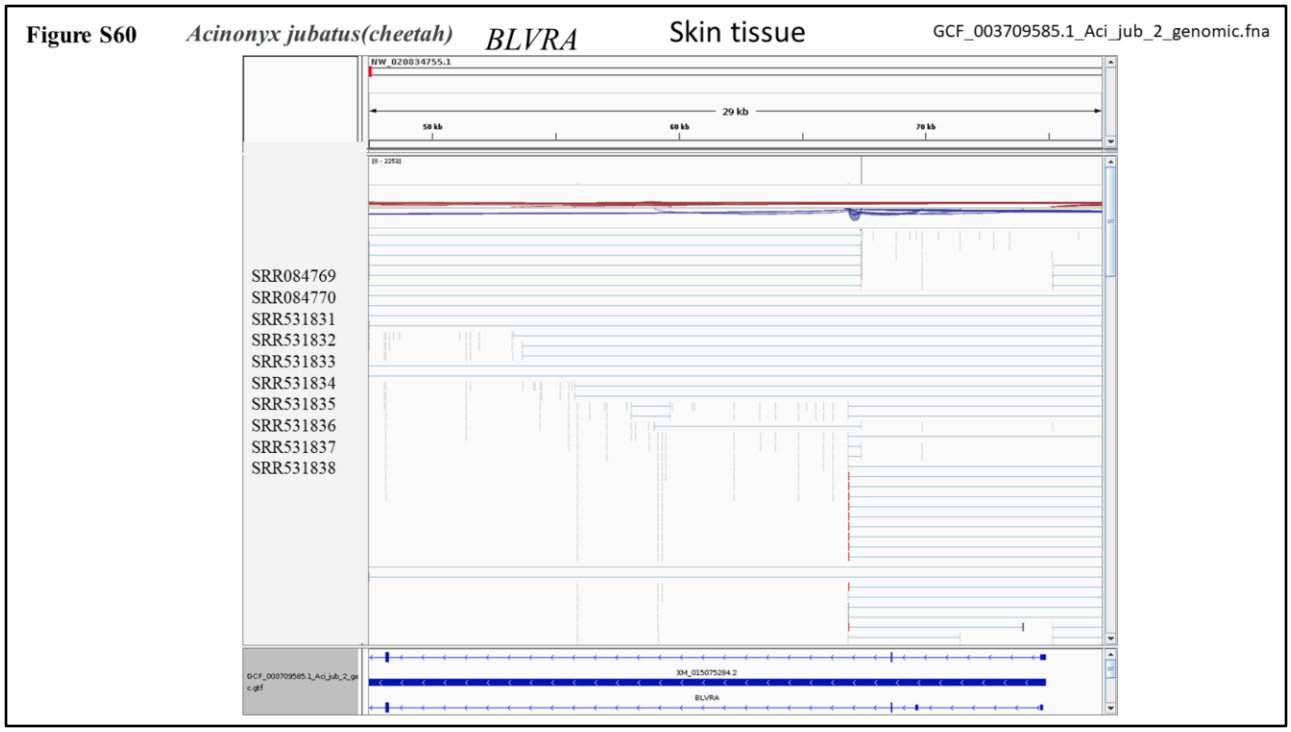


Figure S60: Screenshot of RNA-seq reads aligned to each of the putative locations of the *BLVRA* in cheetah *(Acinonyx jubatus*). Gray boxes represent reads aligned to the genome and thin lines connecting these boxes represent the spliced read alignments. Datasets from (SRR084769 SRR084770 SRR531831 SRR531832 SRR531833 SRR531834 SRR531835 SRR531836 SRR531837 SRR531838) have been mapped to the GCF_003709585.1_Aci_jub_2_genomic.fna. Locations of each of the exons are depicted as bed record in the bottom row.

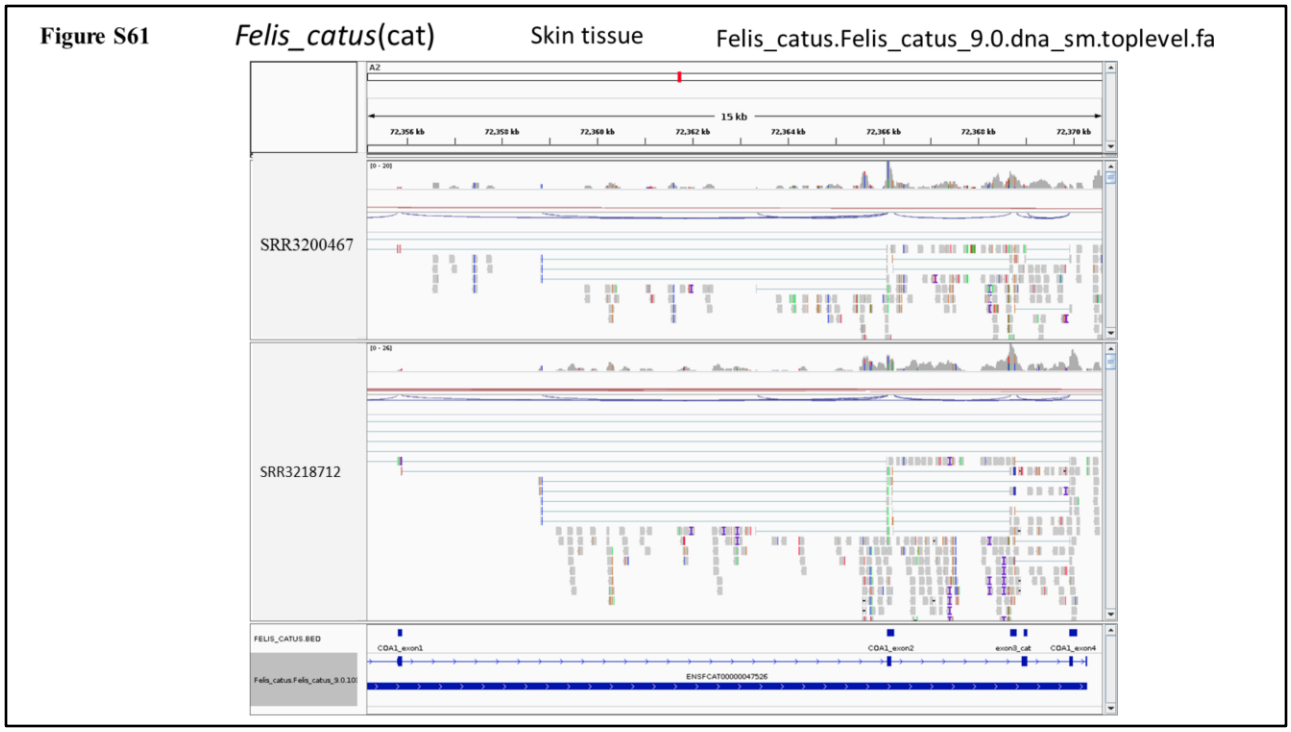


Figure S61: Screenshot of RNA-seq reads aligned to each of the putative locations of the three *COA1* exons in cat (*Felis catus*). Gray boxes represent reads aligned to the genome and thin lines connecting these boxes represent the spliced read alignments. Datasets from skin (SRR3200467 and SRR3218712) have been mapped to the Felis_catus.Felis_catus_9.0.dna_sm.toplevel.fa. Locations of each of the exons are depicted as bed record in the bottom row.

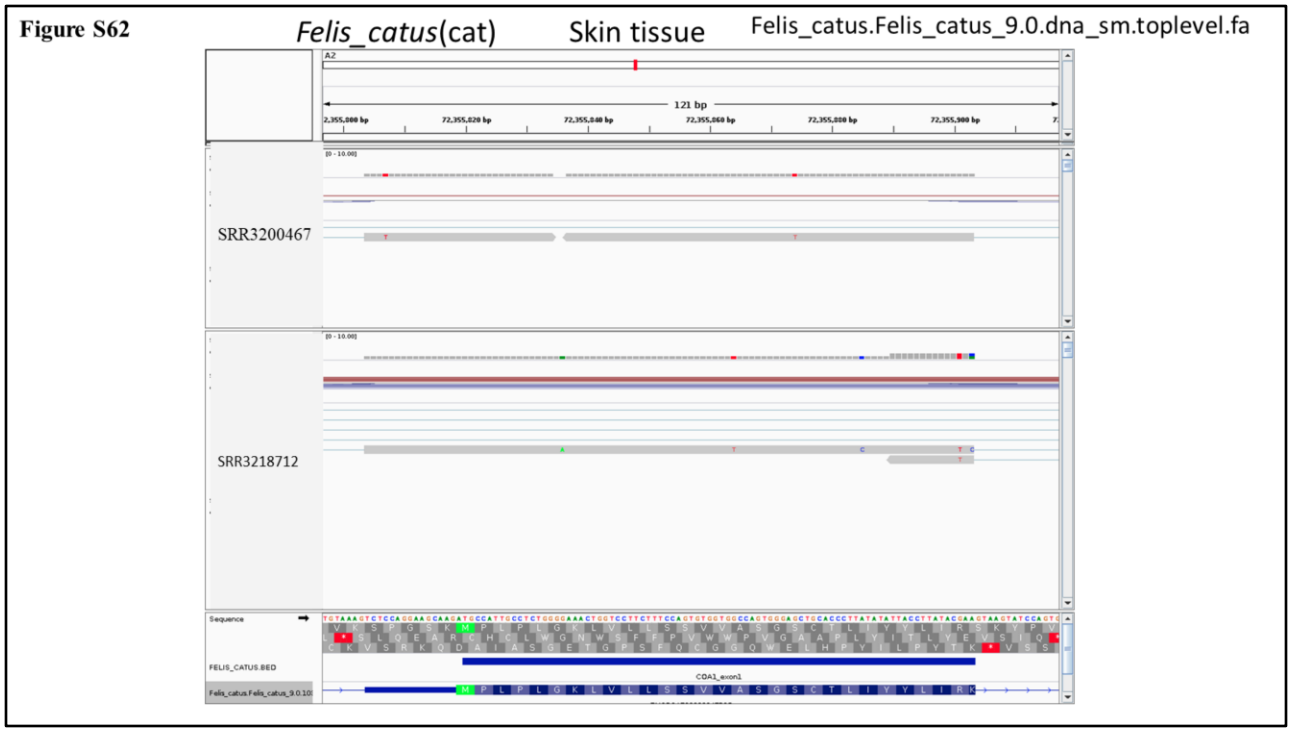


Figure S62: A zoomed in view of exon1 of the *COA1* shown in **Supplementary Fig. S61**.

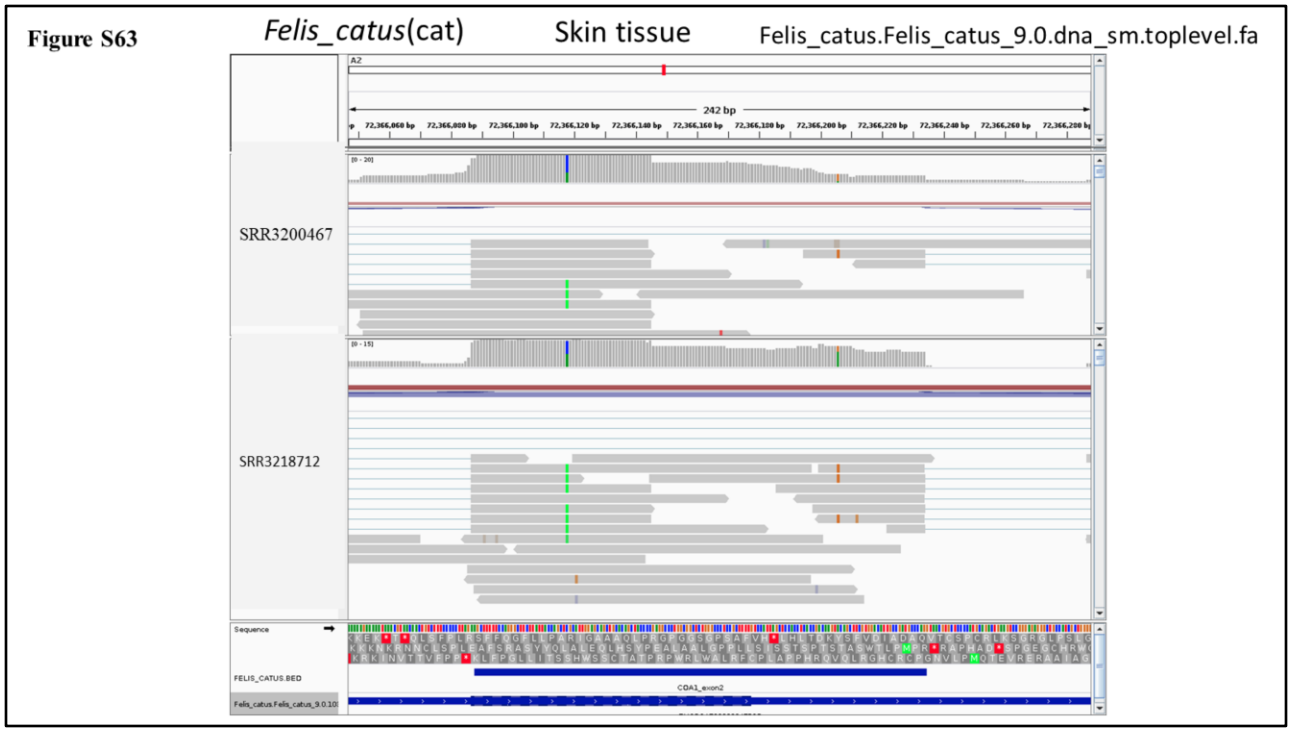


Figure S63: A zoomed in view of exon-2 of the *COA1* shown in **Supplementary Fig. S61**.

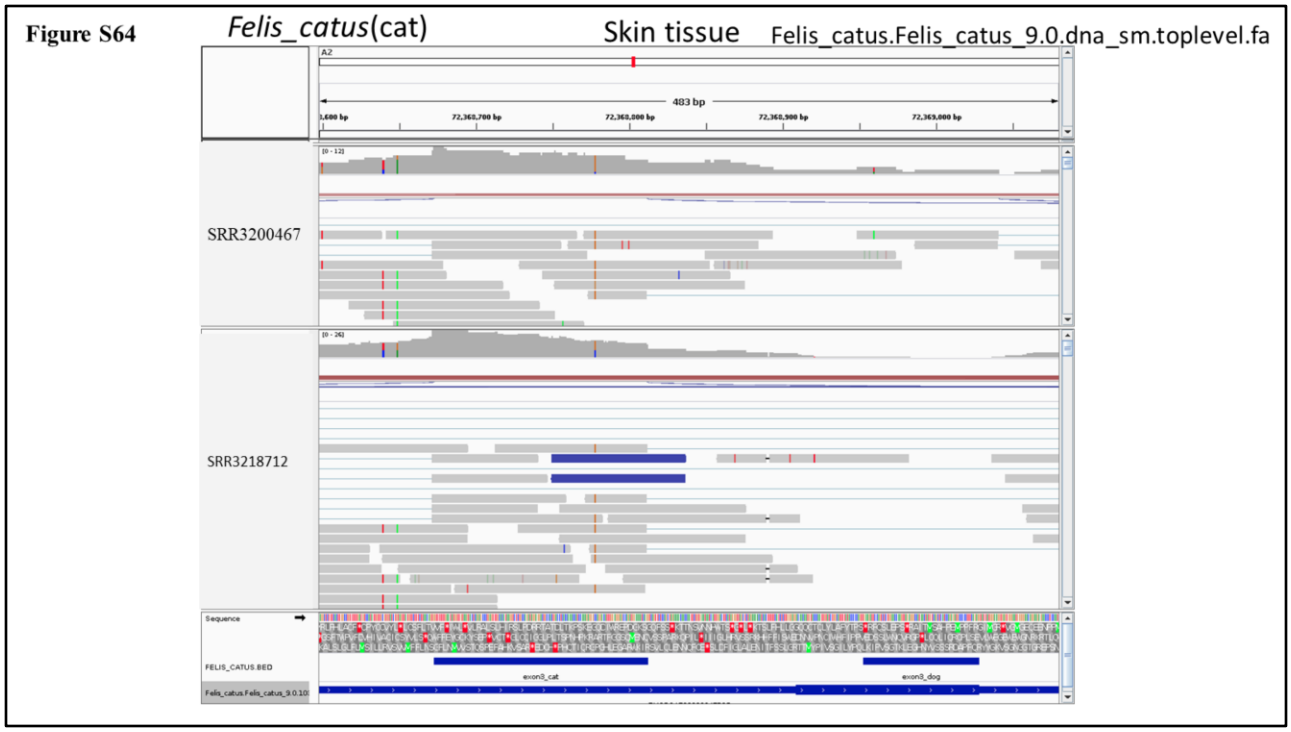


Figure S64: A zoomed in view of exon-3 of the *COA1* shown in **Supplementary Fig. S61**.

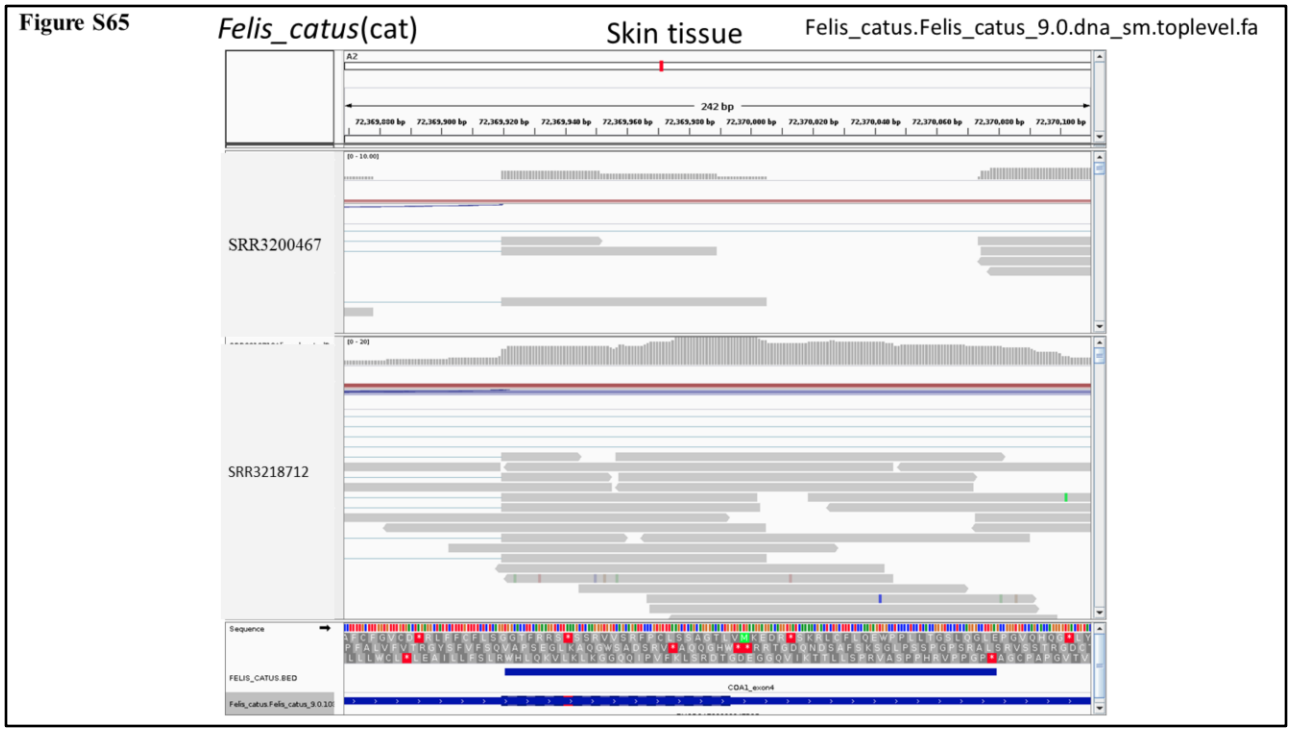


Figure S65: A zoomed in view of exon-4 of the *COA1* shown in **Supplementary Fig. S61**.

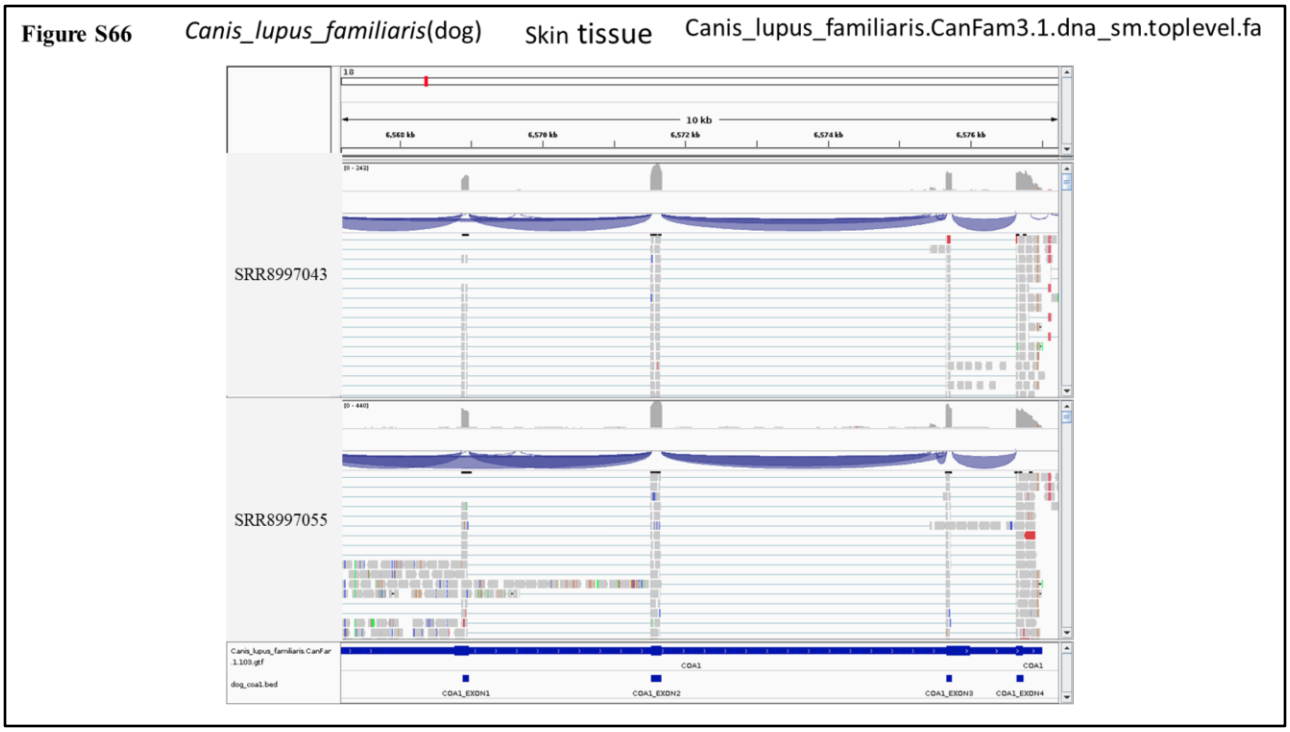


Figure S66: Screenshot of RNA-seq reads aligned to each of the putative locations of the four *COA1* exons in dog (*Canis lupus familiaris*). Gray boxes represent reads aligned to the genome and thin lines connecting these boxes represent the spliced read alignments. Datasets from skin muscle (SRR8997043 and SRR8997055) have been mapped to the Canis_lupus_familiaris.CanFam3.1.dna_sm.toplevel.fa. Locations of each of the exons are depicted as bed record in the bottom row.

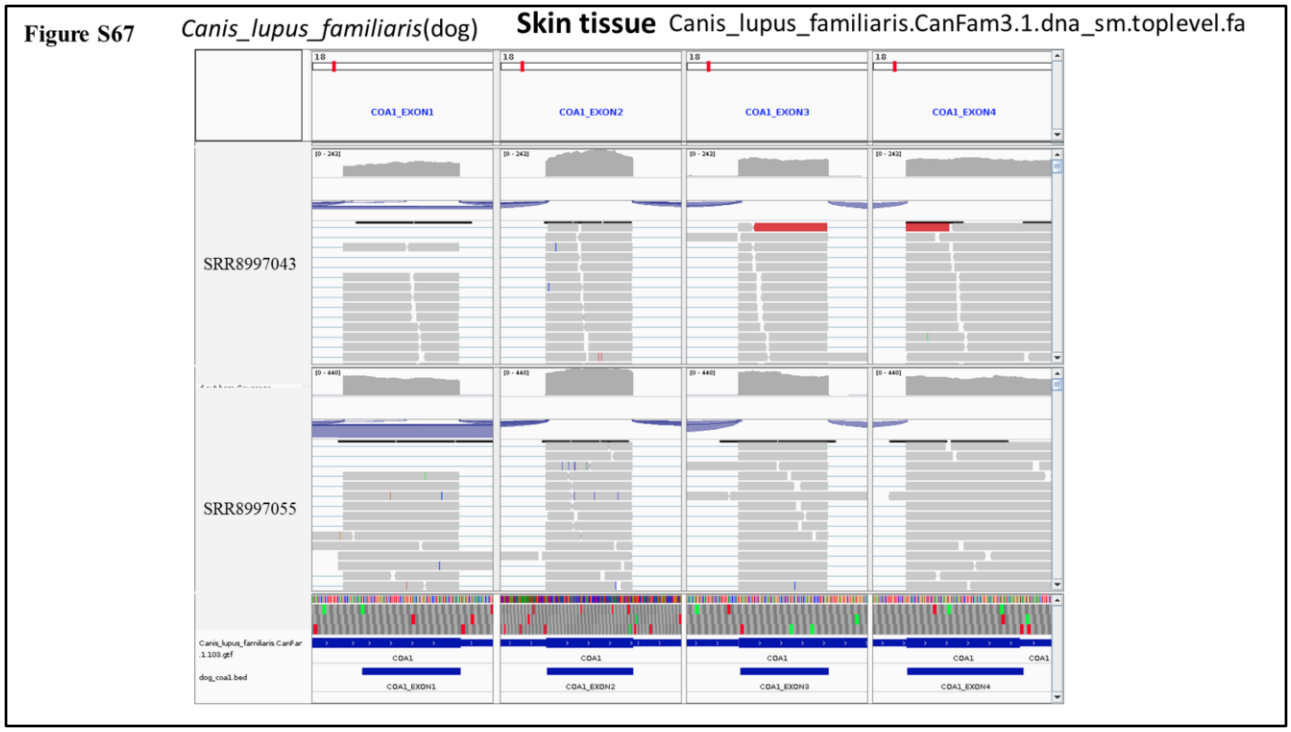


Figure S67: A zoomed in view of each of the *COA1* shown in **Supplementary Fig. S66**.

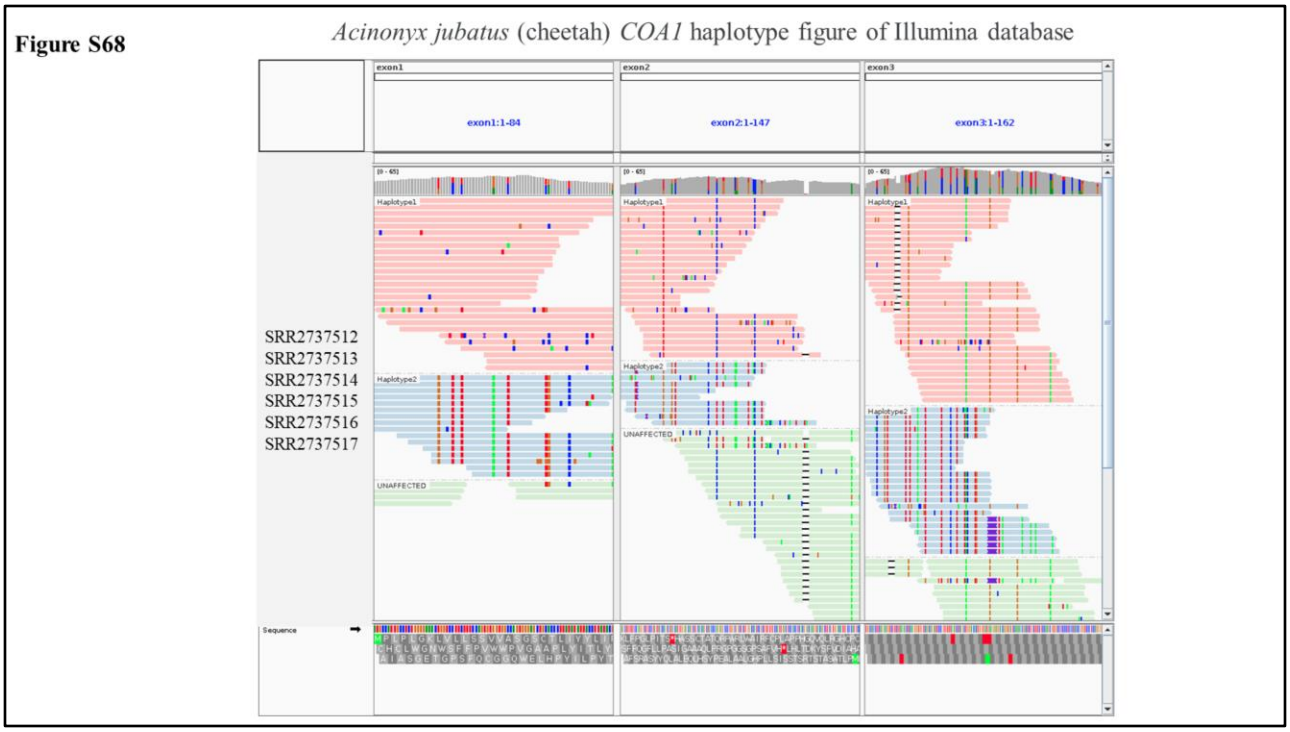


Figure S68: Comparison of haplotypes of *COA1* gene in cheetah :Two haplotypes in exon 1, 2 and single haplotype in exon-3.

Figure S69 *Acinonyx jubatus*(cheetah) *COA1* haplotype figure of Illumina database

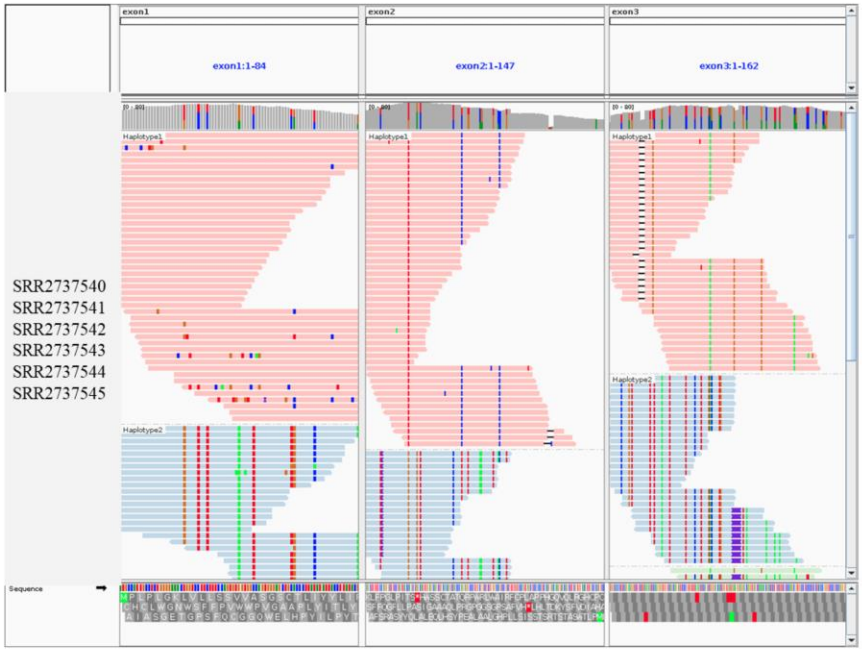


Figure S69: Comparison of haplotypes of *COA1* gene in cheetah :Two haplotypes in exon 1, 2 and single haplotype in exon-3.

Figure S70

Acinonyx jubatus(cheetah) *COA1* haplotype figure of Hi-C database

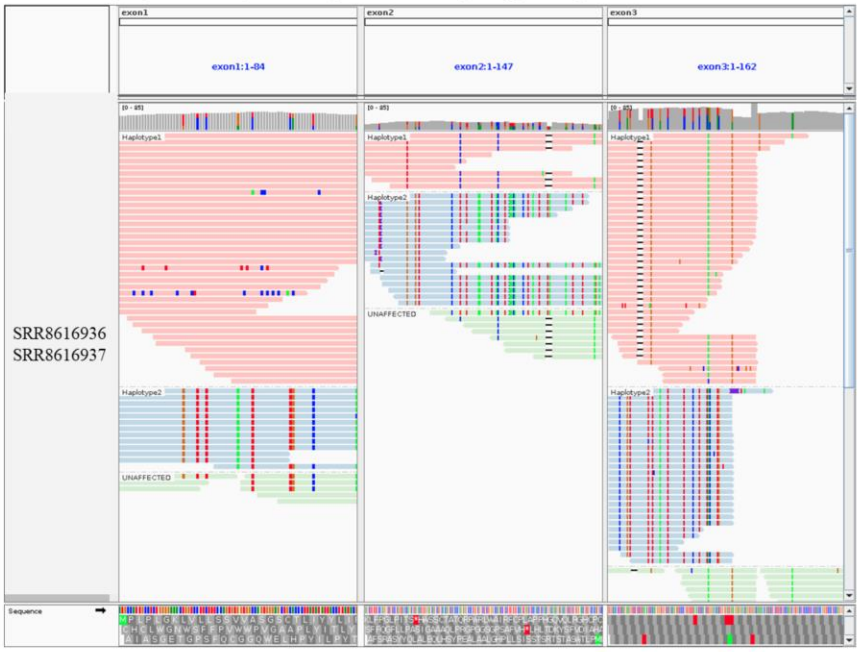


Figure S70: Comparison of haplotypes of *COA1* gene in cheetah :Two haplotypes in exon 1, 2 and single haplotype in exon-3.

Figure S71

Acinonyx jubatus(cheetah) *COA1* haplotype 2 figure of Illumina database

Figure S71: Zoomed view of exon-3 of *COA1* shown in **Supplementary Fig. S68**, vertical box is indicating the position 265 where T is present in haplotype 2 for those reads which don't have insertion of T at position 10 and leads to a premature stop codon in cheetah.

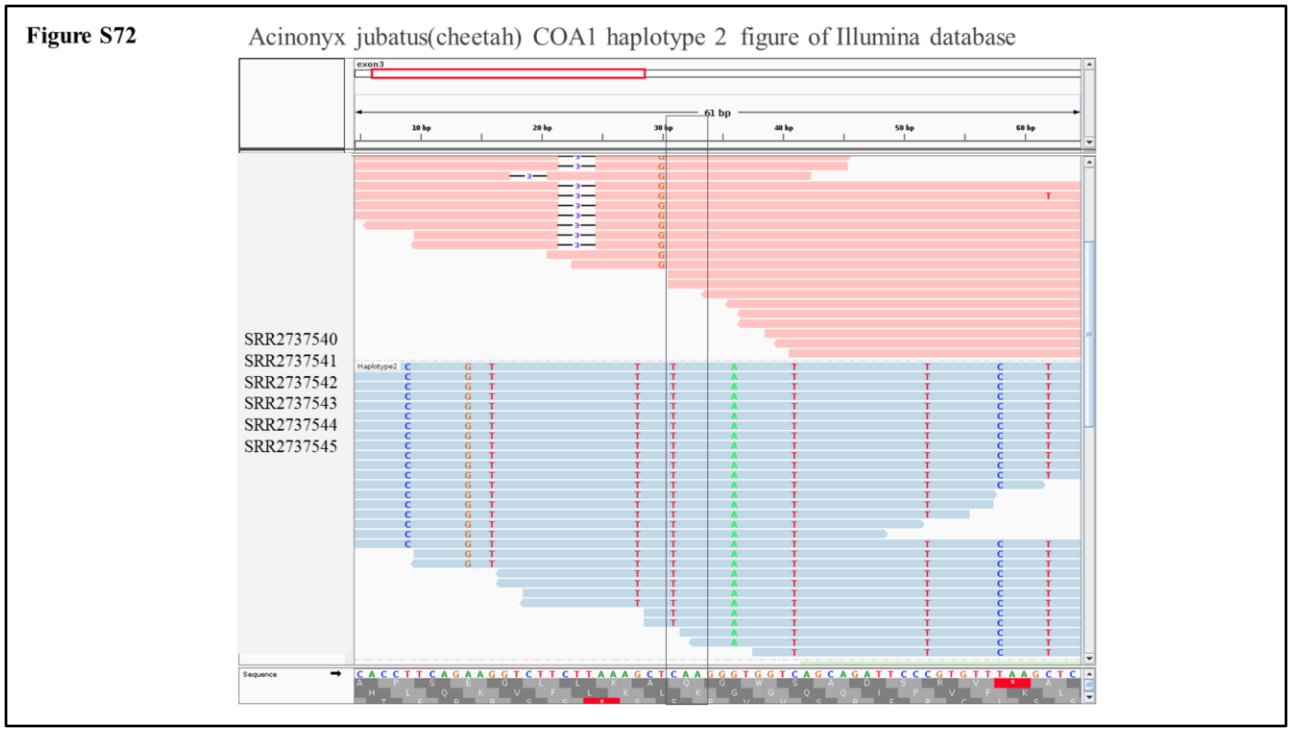


Figure S72: Zoomed view of exon-3 of *COA1* shown in **Supplementary Fig. S69**, vertical box is indicating the position 265 where T is present in haplotype 2 for those reads which don't have insertion of T at position 11 and leads to a premature stop codon in cheetah.

Figure S73 *Acinonyx jubatus*(cheetah) *COA1* haplotype 2 figure of Hi-C database

Figure S73: Zoomed view of exon-3 of *COA1* shown in **Supplementary Fig. S70**, vertical box is indicating the position 265 where T is present in haplotype 2 for those reads which don't have insertion of T at position 10 and leads to a premature stop codon in cheetah.

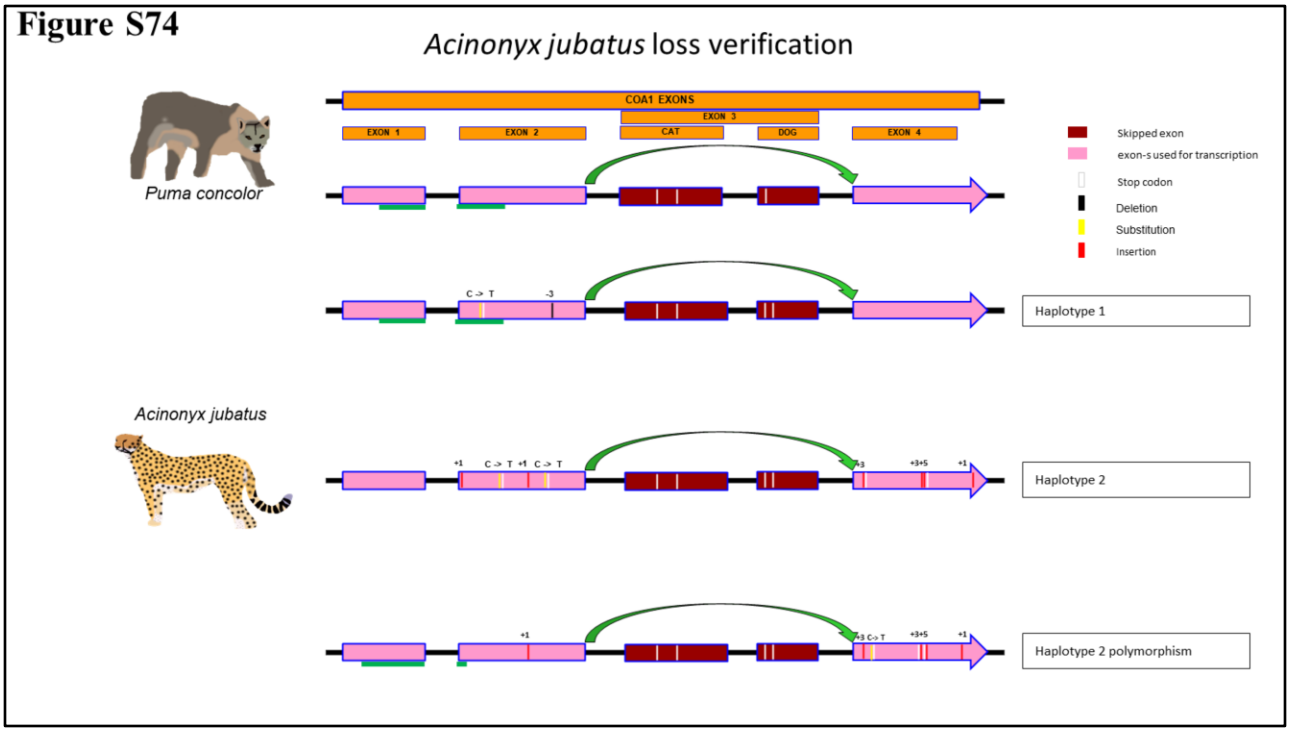


Figure S74: Verification of gene loss in two haplotypes with haplotype two polymorphism showed in **Supplementary Fig. S74**. Different events are mentioned in figure legends with different colors. Horizontal green bars represent the domain region. Puma concolor is used as reference species intact *COA1*.

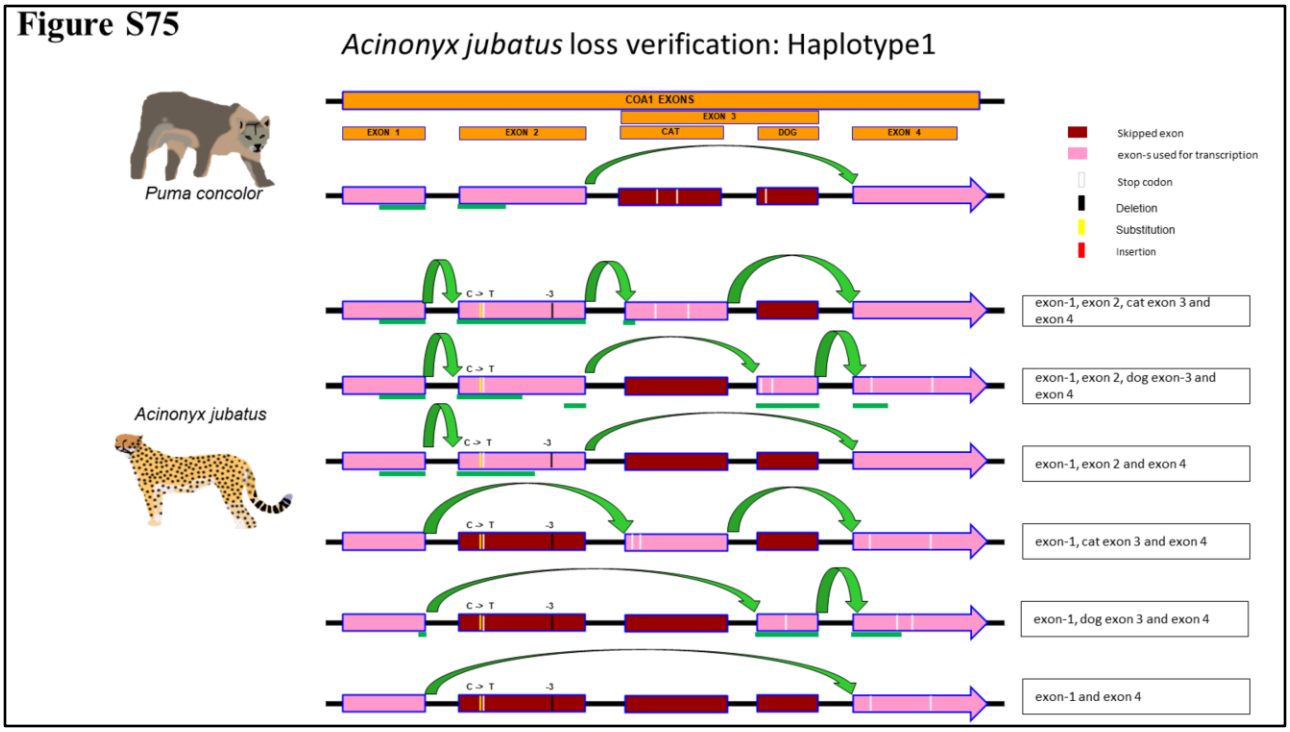


Figure S75: Different combinations of the exon with domain region in haplotype 1, green arrows represent exon splicing. More detail of colors mentioned in figure legends. *Puma concolor* is used as reference species intact *COA1*.

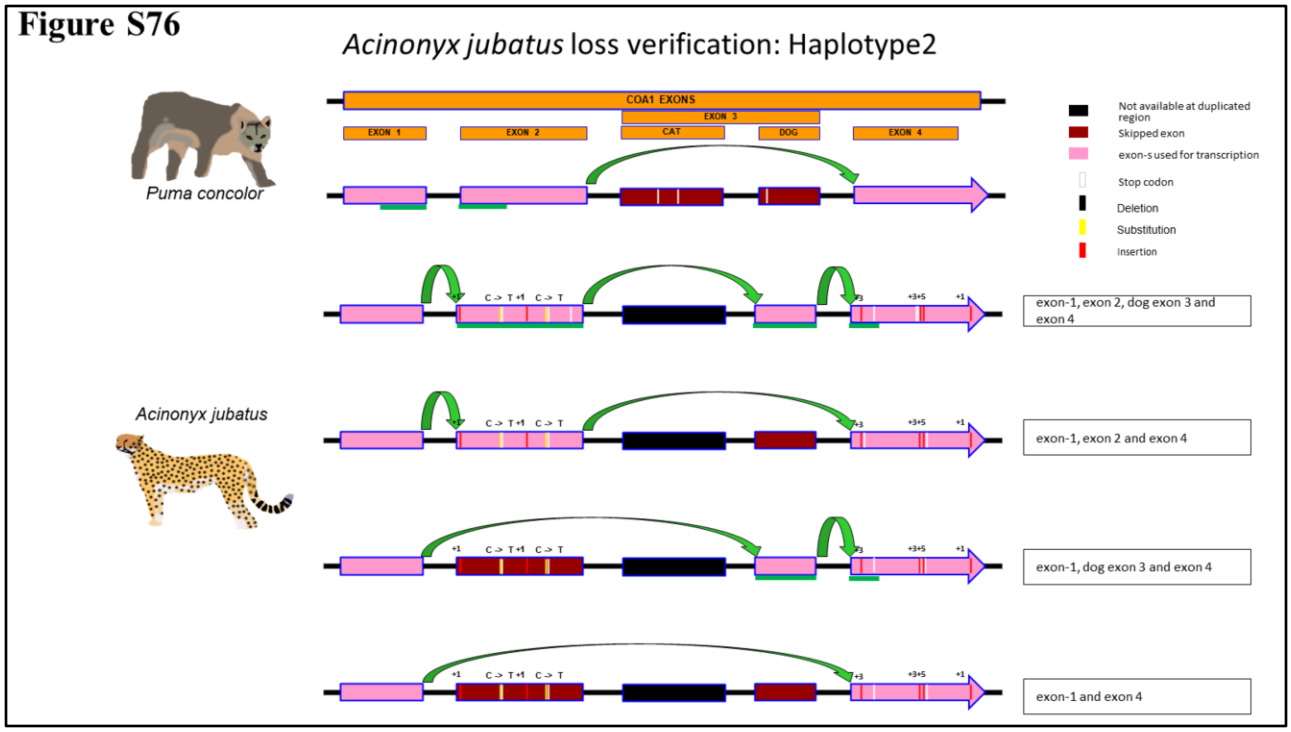


Figure S76: Different combinations of the exon with domain region in haplotype 2, green arrows represent exon splicing. More detail of colors mentioned in figure legends. *Puma concolor* is used as reference species intact *COA1*.

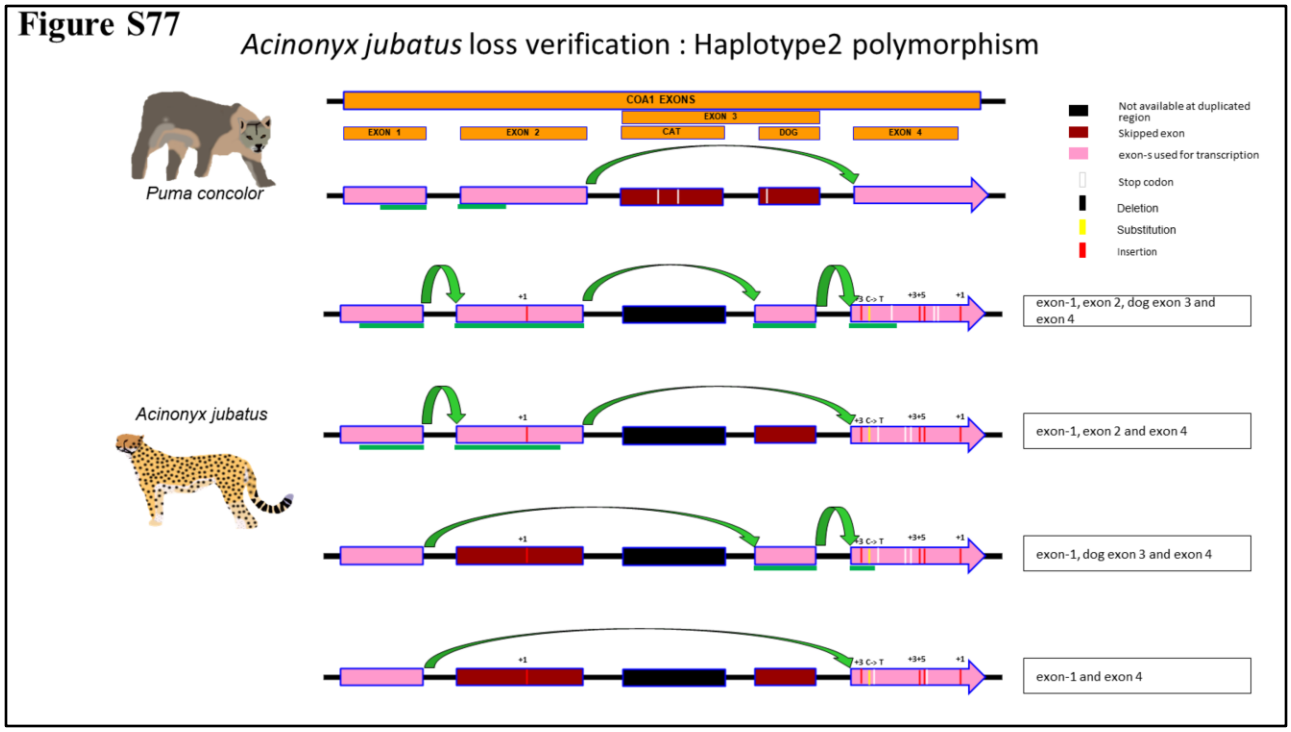


Figure S77: Different combinations of the exon with domain region in haplotype 2 polymorphism, green arrows represent exon splicing. More detail of colors mentioned in figure legends. *Puma concolor* is used as reference species intact *COA1*.

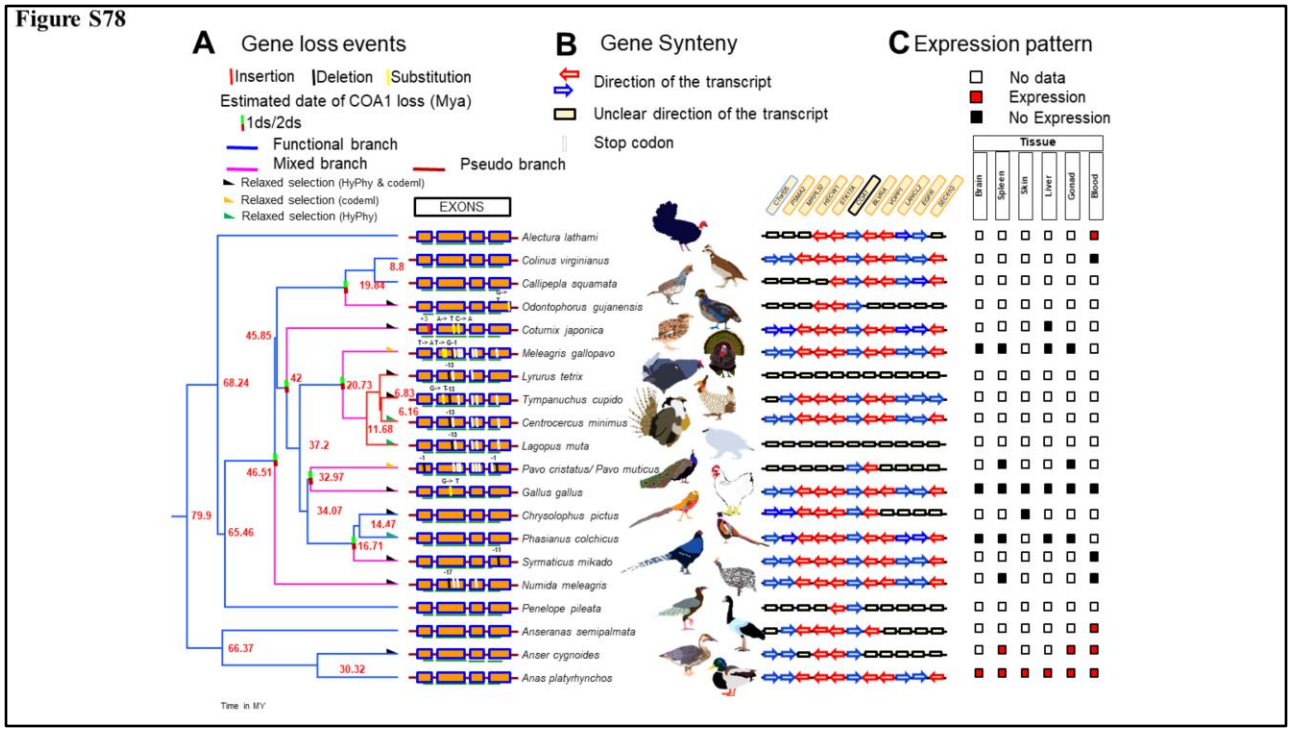


Figure S78: Recurrent loss of *COA1* gene in Galliform species. (**A**) Gene loss events in ten Galliform species besides a time-calibrated phylogenetic tree downloaded from the TimeTree website. Blue branches in the tree represent functional branches, and the pink colored branches represent mixed (functional + pseudogenic) branches and red colored branches represent pseudo branch. The method proposed by Meredith et al., 2009 was used to estimate the time of gene loss using two different substitution rates (1ds and 2ds). Short colored bars depict the locations of the gene disrupting mutations on the four exons of *COA1*. (**B**) Gene order in the

genomic region flanking the *COA1* gene in bird genomes. Red and blue arrows depict the direction of gene transcription relative to the *COA1* gene for consistency across species. Gray boxes represent the genes located on short scaffolds with unknown orientation. (**C**) The gene expression pattern of the *COA1* gene in six tissues (brain, spleen, skin, liver, gonad, and blood) was assessed by screening RNA-seq datasets. The red-colored blocks depict the robust expression of the *COA1* gene, the black-colored blocks depict a lack of *COA1* gene expression in that particular tissue, and the white-colored blocks represent a lack of data for that tissue.

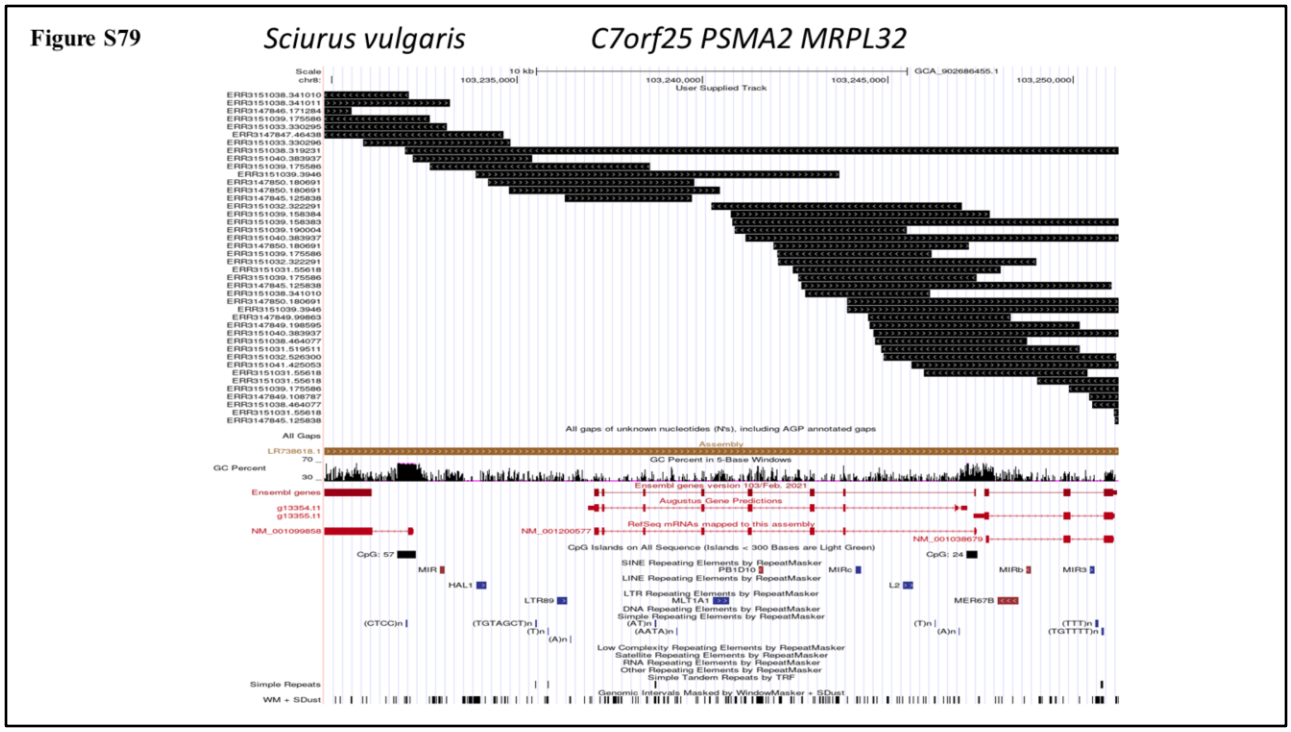


Figure S79: Screenshot of UCSC genome browser showing the alignment of > 3 kb long reads generated using PacBio technology to the Eurasian red squirrel (*Sciurus vulgaris*) **GCA_902686455.1** at the location of the *C7orf25*, *PSMA2* and *MRPL32*. Each black rectangular box represents a read that aligns to that particular position. A series of overlapping reads that span the entire region up to adjoining genes are shown. The accession numbers of each of the reads are provided beside the alignments.

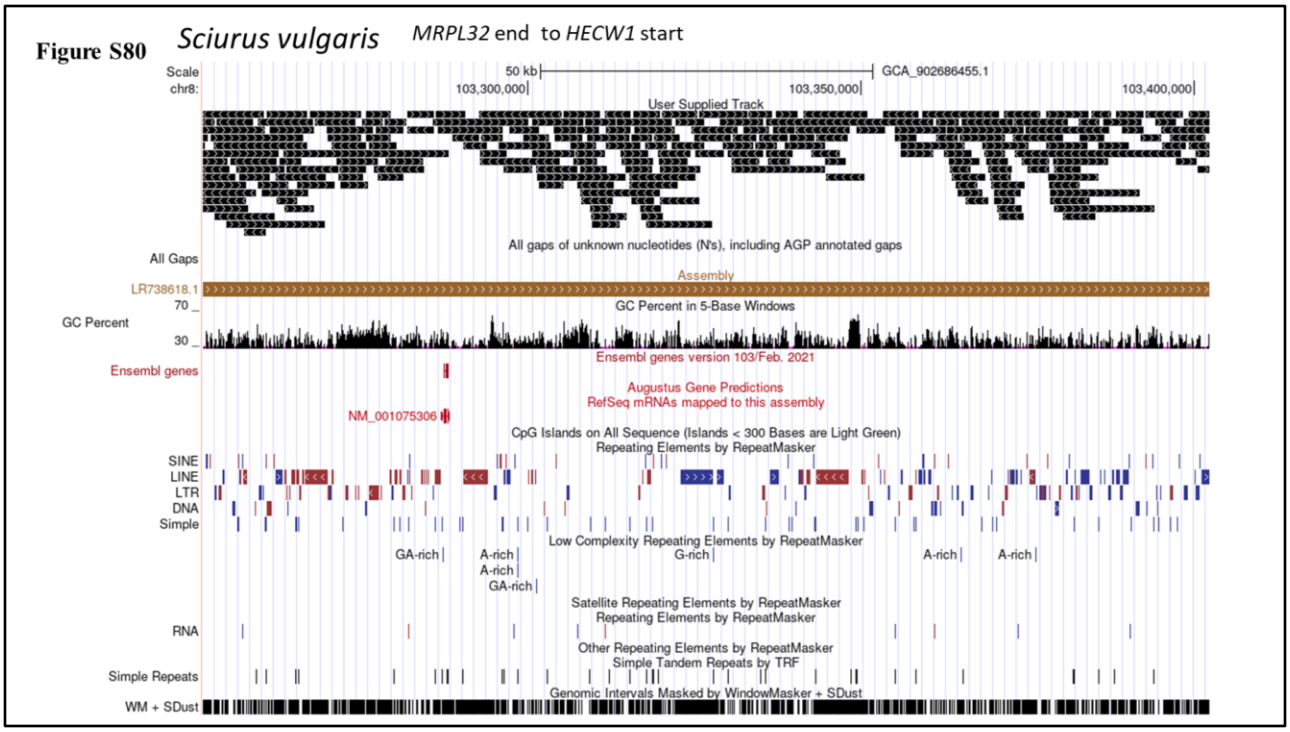


Figure S80: Screenshot of UCSC genome browser showing the alignment of > 3 kb long reads generated using PacBio technology to the Eurasian red squirrel (*Sciurus vulgaris*) **GCA_902686455.1** at the location of the *MRPL32* end to *HECW1* start. Each black rectangular box represents a read that aligns to that particular position. A series of overlapping reads that span the entire region up to adjoining genes are shown.

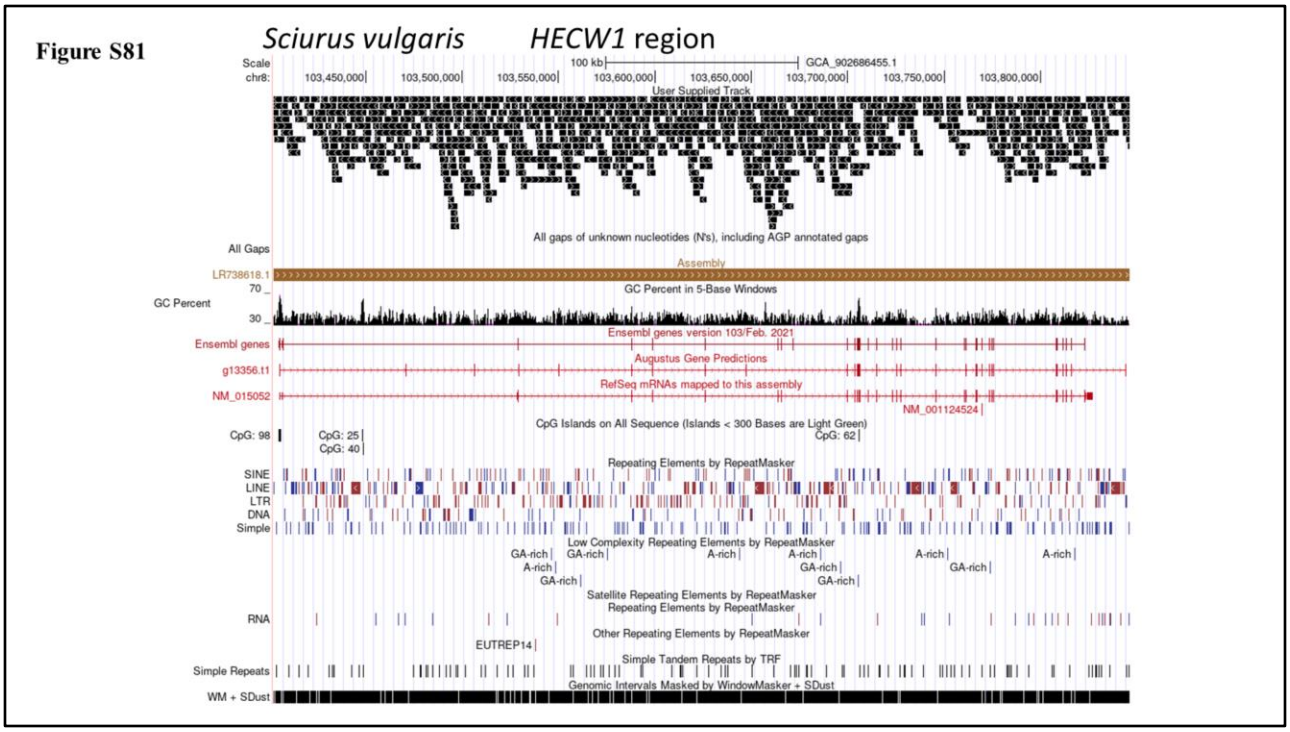


Figure S81: Screenshot of UCSC genome browser showing the alignment of > 3 kb long reads generated using PacBio technology to the Eurasian red squirrel (*Sciurus vulgaris*) **GCA_902686455.1** at the location of the *HECW1*. Each black rectangular box represents a read that aligns to that particular position. A series of overlapping reads that span the entire region up to adjoining genes are shown.

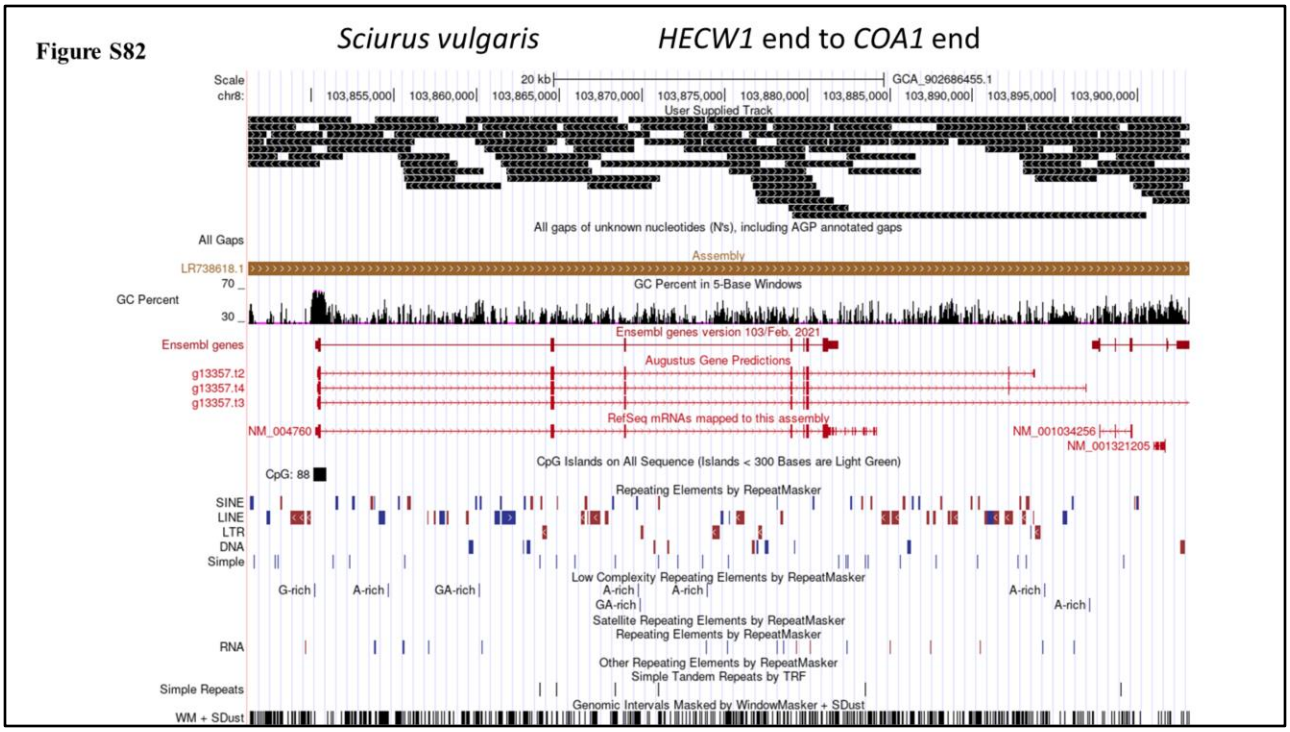


Figure S82: Screenshot of UCSC genome browser showing the alignment of > 3 kb long reads generated using PacBio technology to the Eurasian red squirrel (*Sciurus vulgaris*) **GCA_902686455.1** at the location of the *HECW1* end to *COA1* start. Each black rectangular box represents a read that aligns to that particular position. A series of overlapping reads that span the entire region up to adjoining genes are shown.

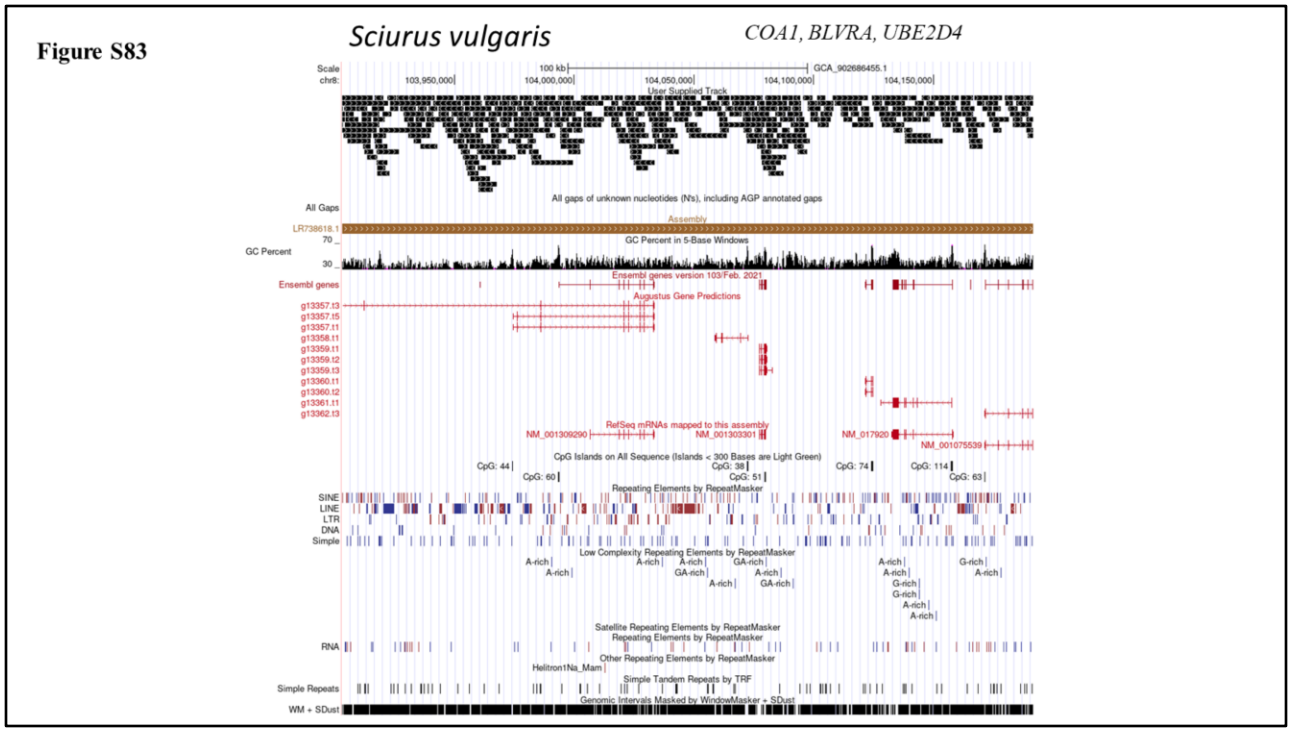


Figure S83: Screenshot of UCSC genome browser showing the alignment of > 3 kb long reads generated using PacBio technology to the Eurasian red squirrel (*Sciurus vulgaris*) **GCA_902686455.1** at the location of the *COA1* end to *UBE2D4*. Each black rectangular box represents a read that aligns to that particular position. A series of overlapping reads that span the entire region up to adjoining genes are shown.

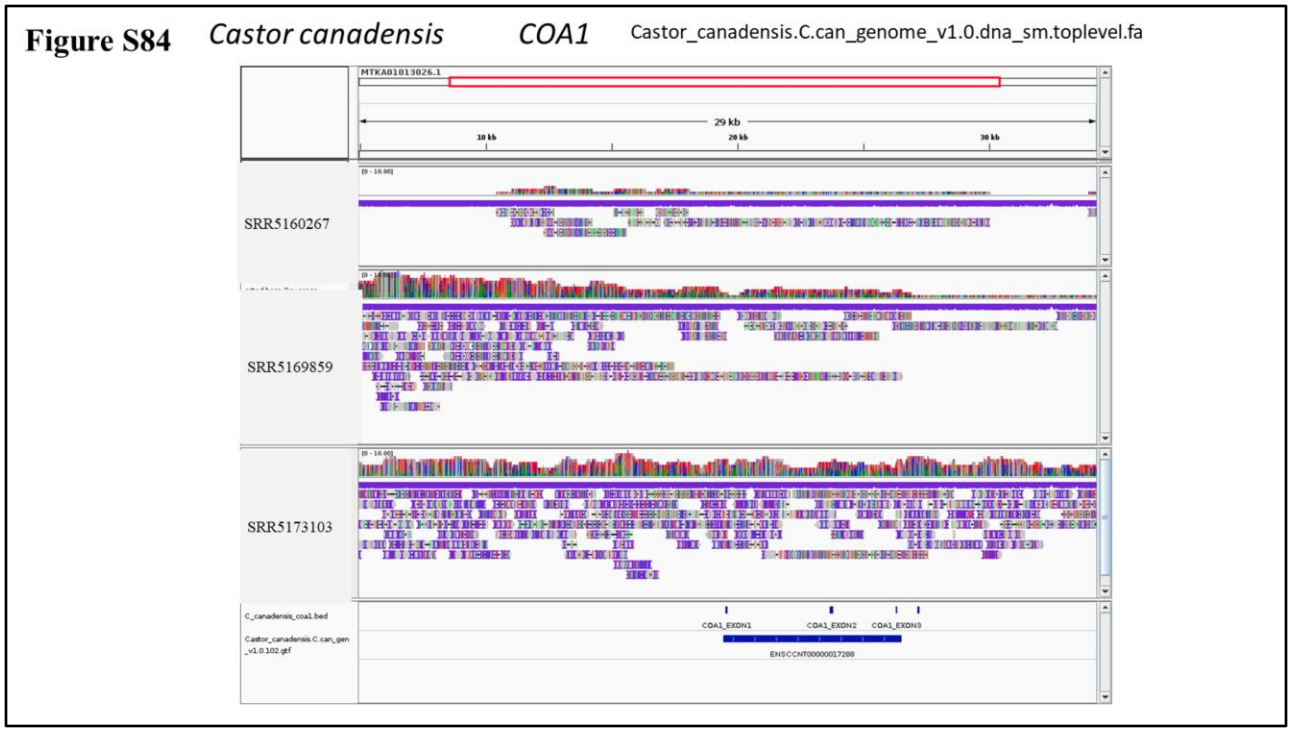


Figure S84: Screenshot of PacBio reads aligned to each of the putative locations of the *COA1* in American beaver (*Castor canadensis*). Gray boxes represent reads aligned to the genome. Raw read dataset which is mentioned above have been mapped to the Castor_canadensis.C.can_genome_v1.0.dna_sm.toplevel.fa.

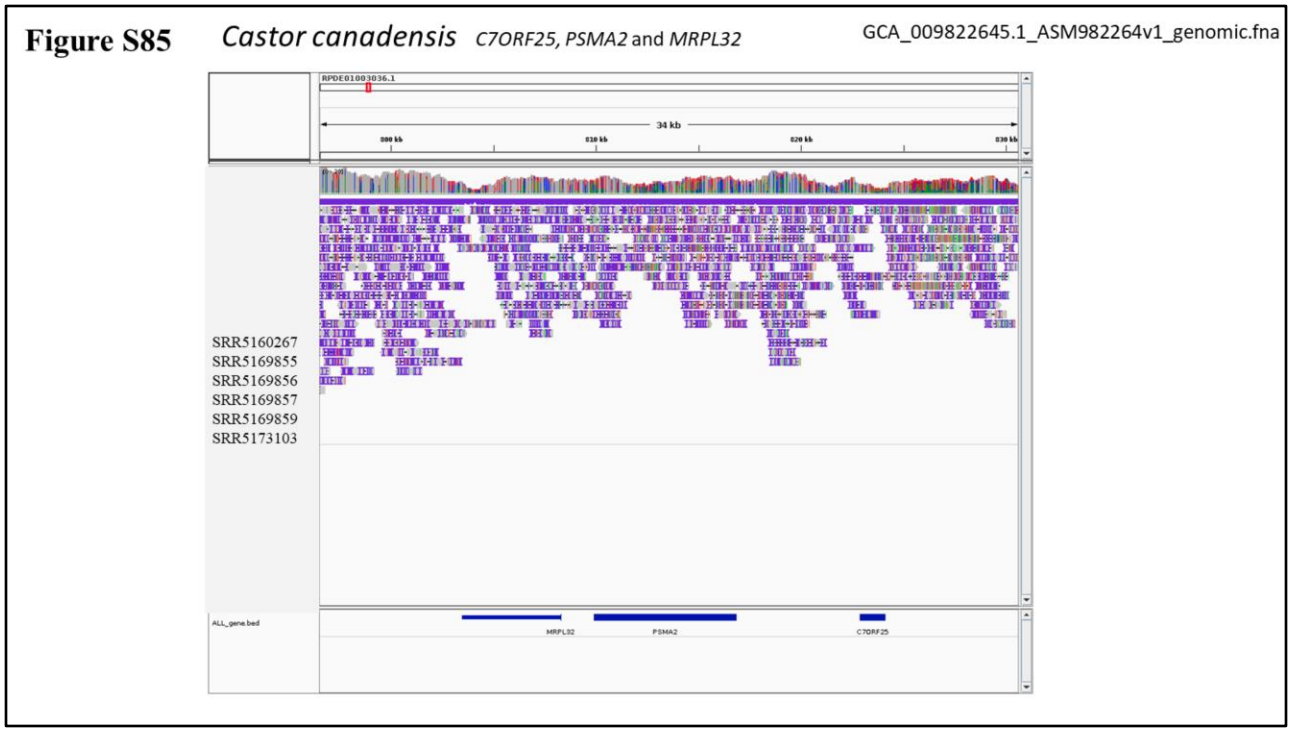


Figure S85: Screenshot of PacBio reads aligned to each of the putative locations of the *C7orf25*, *PSMA2* and *MRPL32* in American beaver (*Castor canadensis*). Gray boxes represent reads aligned to the genome. Raw read dataset which is mentioned above have been mapped to the GCA_009822645.1_ASM982264v1_genomic.fna.

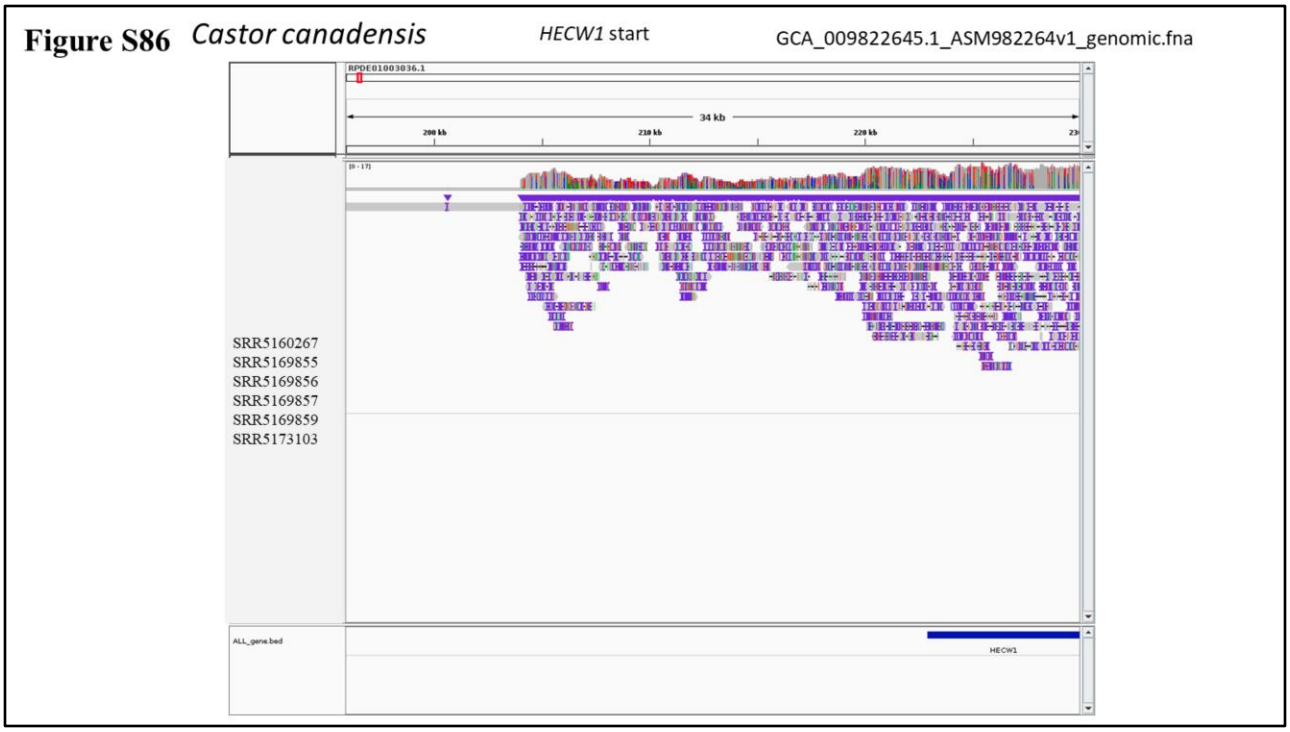


Figure S86: Screenshot of PacBio reads aligned to each of the putative locations of the *HECWI* start in American beaver (*Castor canadensis*). Gray boxes represent reads aligned to the genome. Raw read dataset which is mentioned above have been mapped to the GCA_009822645.1_ASM982264v1_genomic.fna.

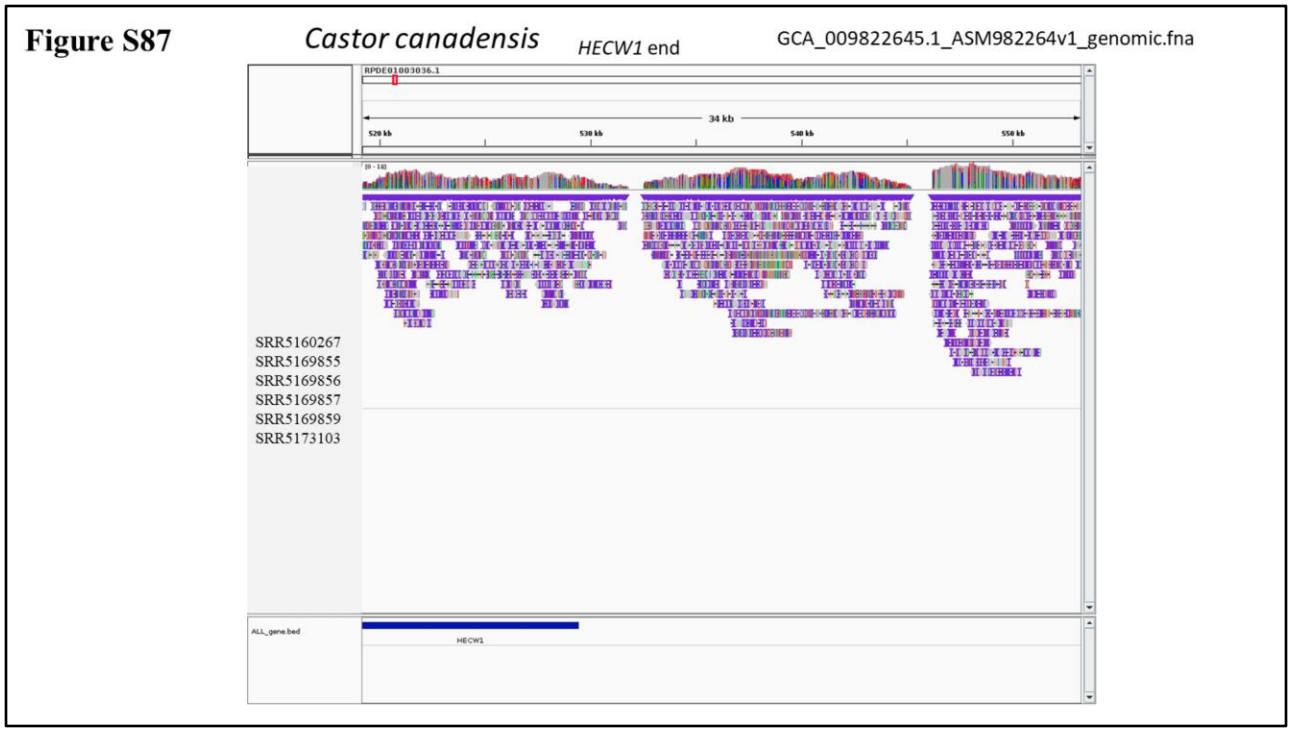


Figure S87: Screenshot of PacBio reads aligned to each of the putative locations of the *HECW1* end in American beaver (*Castor canadensis*). Gray boxes represent reads aligned to the genome. Raw read dataset which is mentioned above have been mapped to the GCA_009822645.1_ASM982264v1_genomic.fna.

Figure S88: Screenshot of PacBio reads aligned to each of the putative locations of the *STK17A* start in American beaver (*Castor canadensis*). Gray boxes represent reads aligned to the genome. Raw read dataset which is mentioned above have been mapped to the GCA_009822645.1_ASM982264v1_genomic.fna.

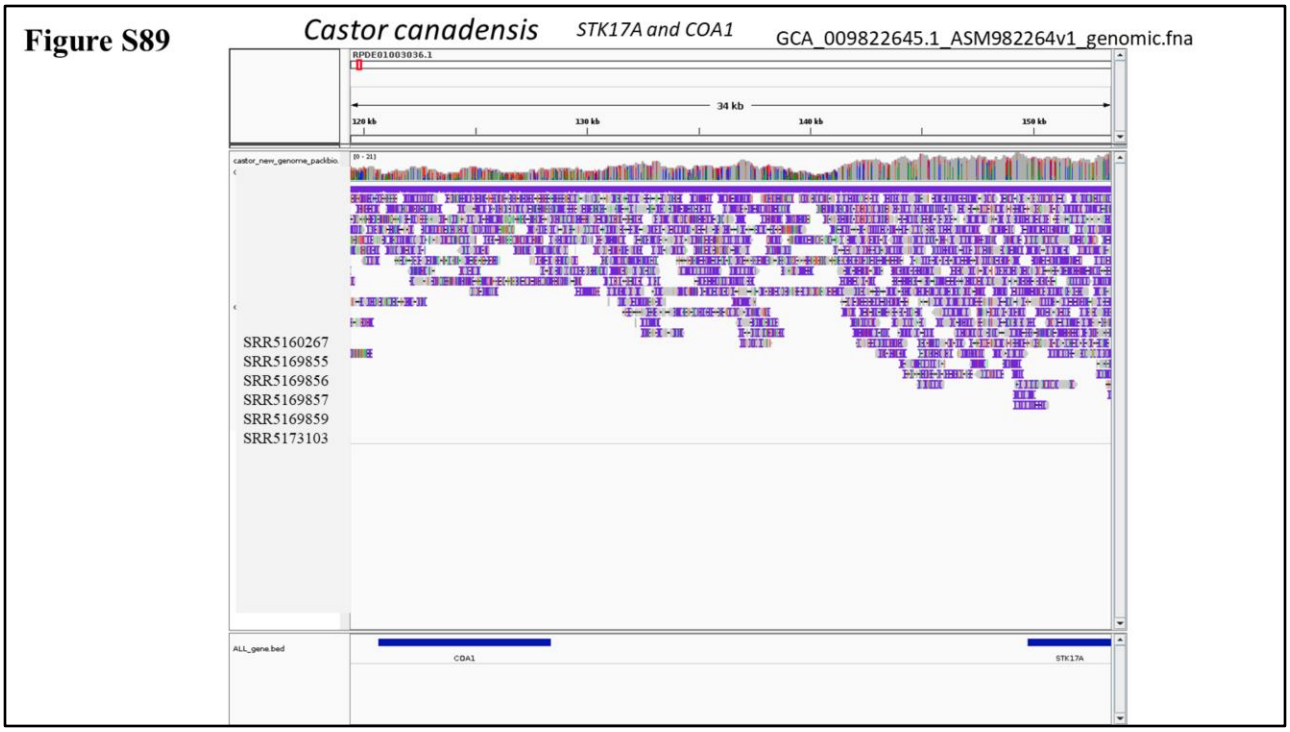


Figure S89: Screenshot of PacBio reads aligned to each of the putative locations of the *STK17A* and *COA1* in American beaver (*Castor canadensis*). Gray boxes represent reads aligned to the genome. Raw read dataset which is mentioned above have been mapped to the GCA_009822645.1_ASM982264v1_genomic.fna.

Figure S90: Screenshot of PacBio reads aligned to each of the putative locations of the *BLVRA* in American beaver (*Castor canadensis*). Gray boxes represent reads aligned to the genome. Raw read dataset which is mentioned above have been mapped to the GCA_009822645.1_ASM982264v1_genomic.fna.

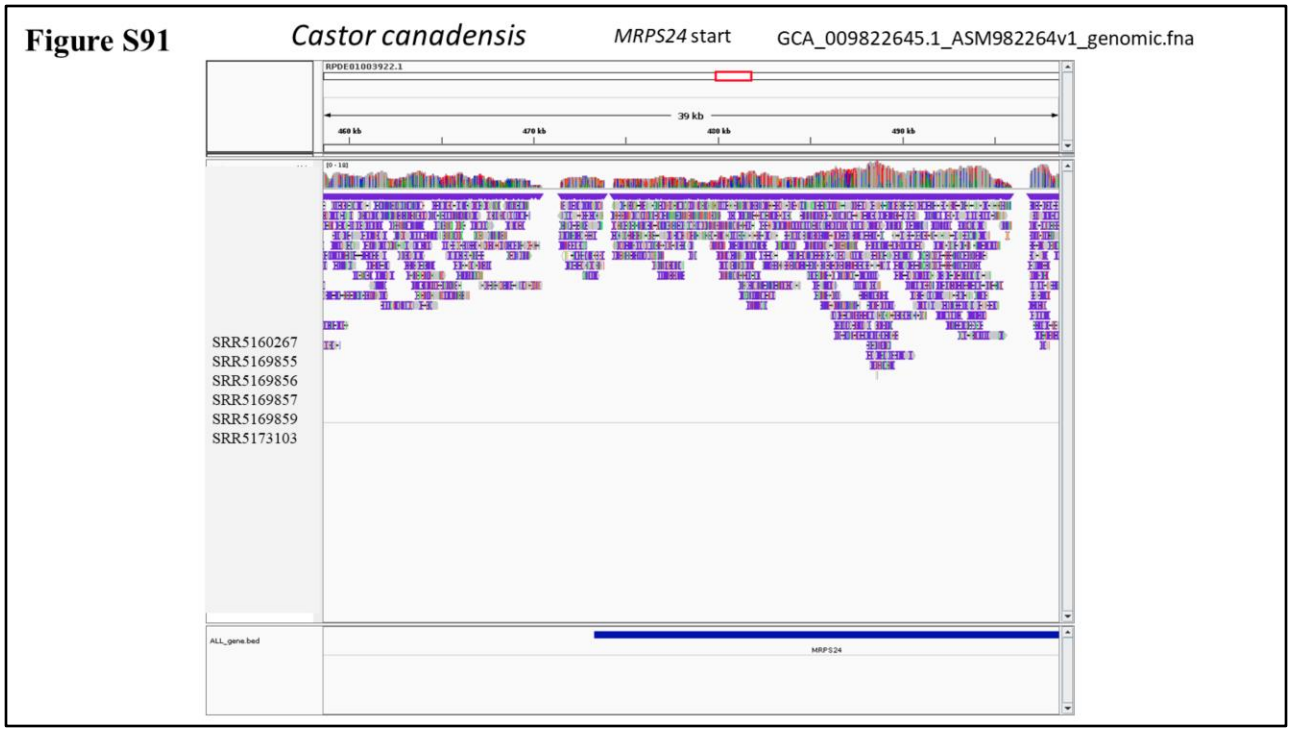


Figure S91: Screenshot of PacBio reads aligned to each of the putative locations of the *MRPS24* start in American beaver (*Castor canadensis*). Gray boxes represent reads aligned to the genome. Raw read dataset which is mentioned above have been mapped to the GCA_009822645.1_ASM982264v1_genomic.fna.

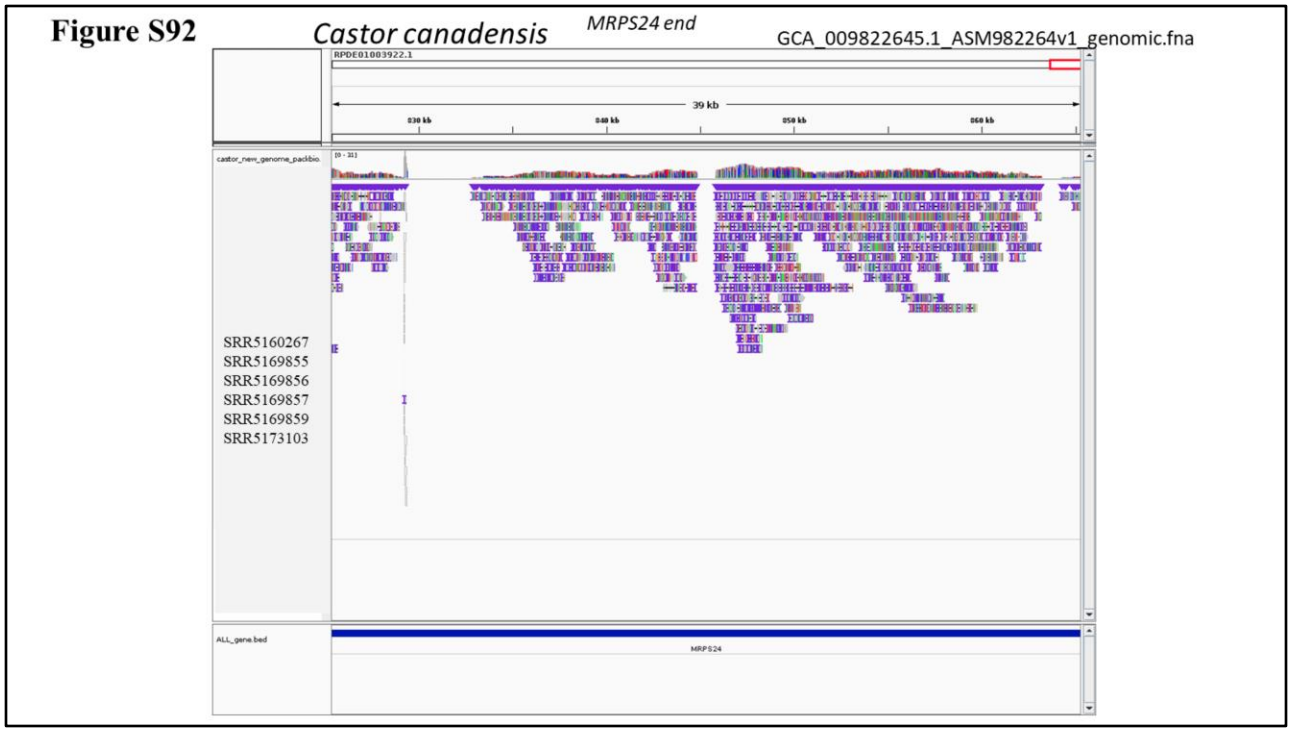


Figure S92: Screenshot of PacBio reads aligned to each of the putative locations of the *MRPS24* in American beaver (*Castor canadensis*). Gray boxes represent reads aligned to the genome. Raw read dataset which is mentioned above have been mapped to the GCA_009822645.1_ASM982264v1_genomic.fna.

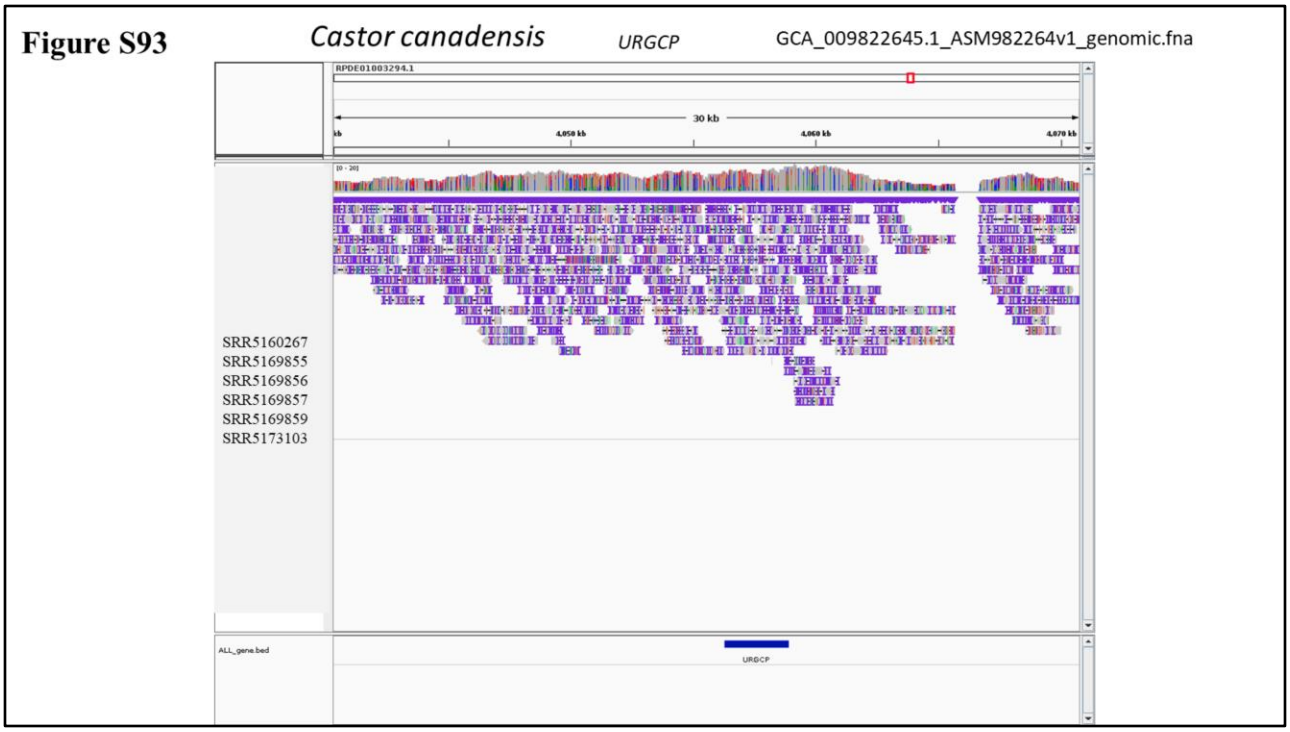


Figure S93: Screenshot of PacBio reads aligned to each of the putative locations of the *URGCP* in American beaver (*Castor canadensis*). Gray boxes represent reads aligned to the genome. Raw read dataset which is mentioned above have been mapped to the GCA_009822645.1_ASM982264v1_genomic.fna.

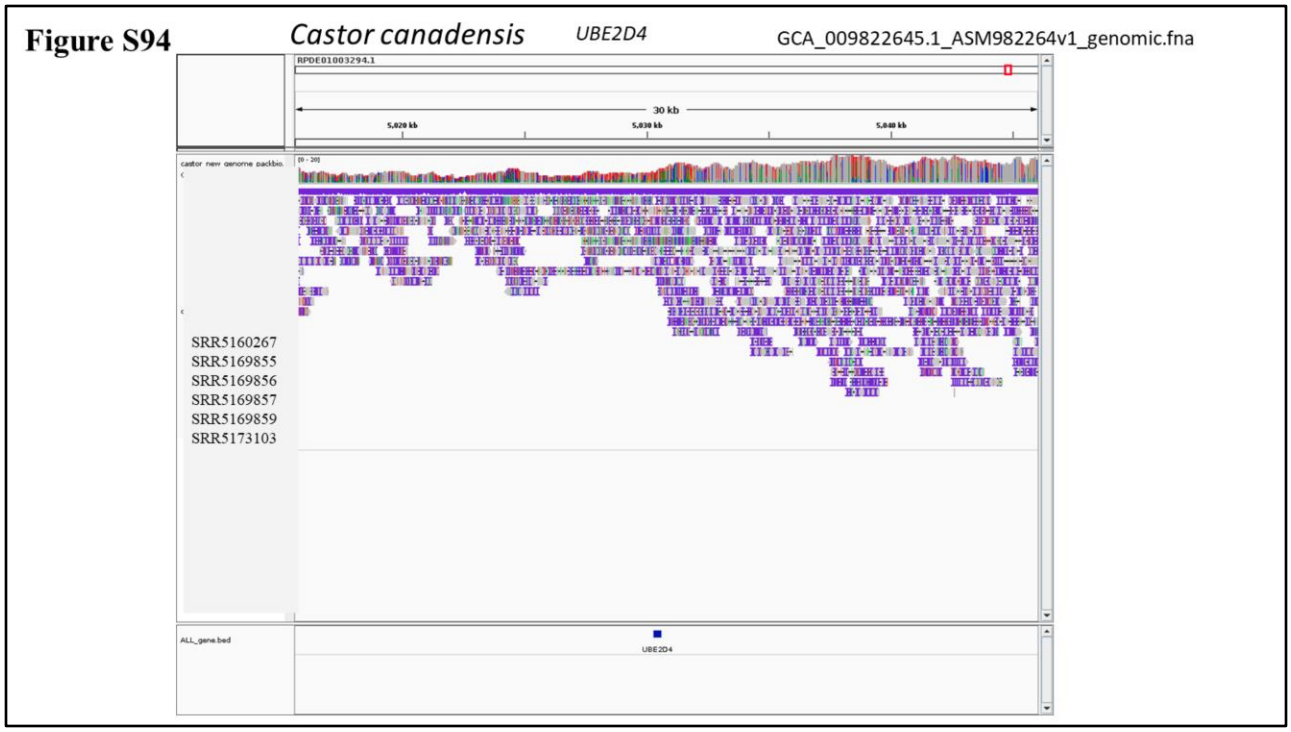


Figure S94: Screenshot of PacBio reads aligned to each of the putative locations of the *UBE2D4* in American beaver (*Castor canadensis*). Gray boxes represent reads aligned to the genome. Raw read dataset which is mentioned above have been mapped to the GCA_009822645.1_ASM982264v1_genomic.fna.

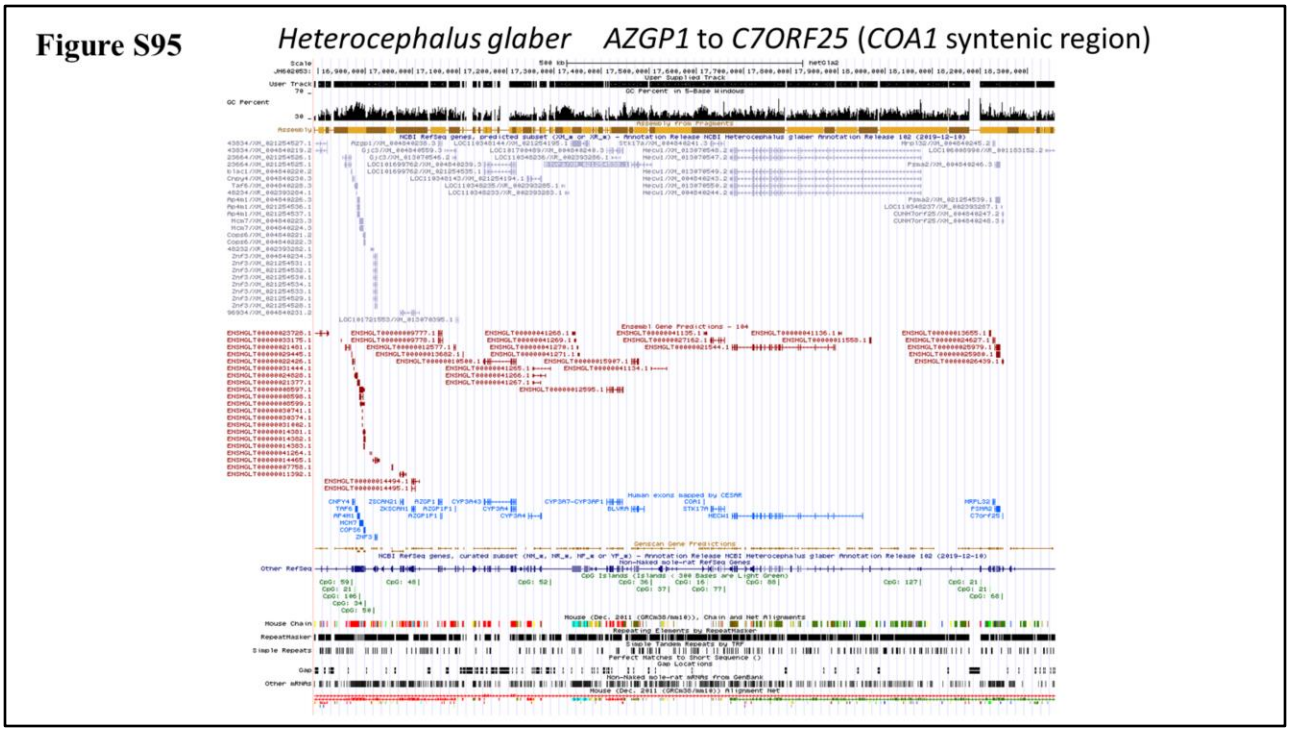


Figure S95: Screenshot of UCSC genome browser showing the alignment of > 3 kb long reads generated using PacBio technology to the naked mole-rat (*Heterocephalus glaber*) **GCA_000247695.1** at the location of the *AZGP1* to *C7orf25* (*COA1* syntenic region). Each black rectangular box represents a read that aligns to that particular position. A series of overlapping reads that span the entire region up to adjoining genes are shown.

Figure S96: Screenshot of PacBio reads aligned to each of the putative locations of the *C7orf25*, *PSMA2* and *MRPL32* in naked mole-rat (*Heterocephalus glaber*). Gray boxes represent reads aligned to the genome. Raw read dataset which is mentioned above have been mapped to the GCA_014060925.1_Heter_glaber.v1.7_hic_pac_genomic.fna.

Figure S97: Screenshot of PacBio reads aligned to each of the putative locations of the *HECW1* start in naked mole-rat (*Heterocephalus glaber*). Gray boxes represent reads aligned to the genome. Raw read dataset which is mentioned above have been mapped to the GCA_014060925.1_Heter_glaber.v1.7_hic_pac_genomic.fna.

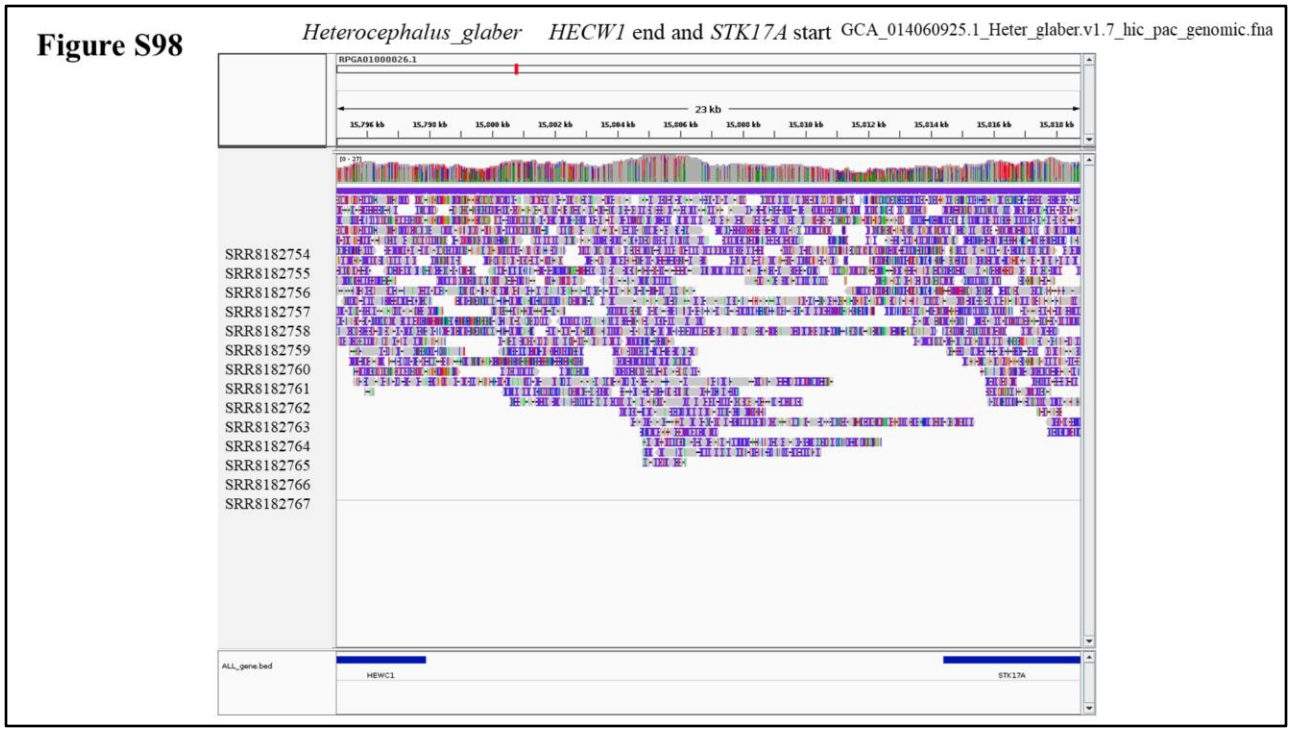


Figure S98: Screenshot of PacBio reads aligned to each of the putative locations of the *HECW1* and *STK17A* in naked mole-rat (*Heterocephalus glaber*). Gray boxes represent reads aligned to the genome. Raw read dataset which is mentioned above have been mapped to the GCA_014060925.1_Heter_glaber.v1.7_hic_pac_genomic.fna.

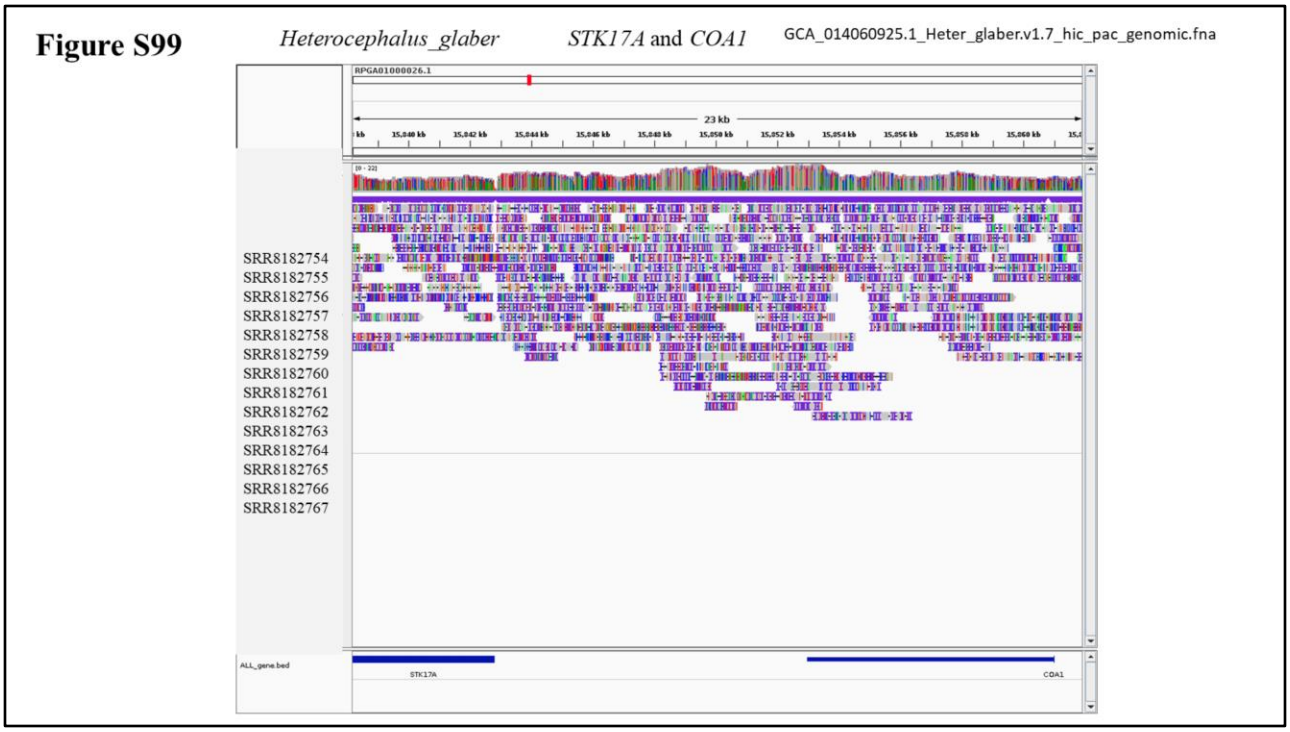


Figure S99: Screenshot of PacBio reads aligned to each of the putative locations of the *STK17A* and *COA1* in naked mole-rat (*Heterocephalus glaber*). Gray boxes represent reads aligned to the genome. Raw read dataset which is mentioned above have been mapped to the GCA_014060925.1_Heter_glaber.v1.7_hic_pac_genomic.fna.

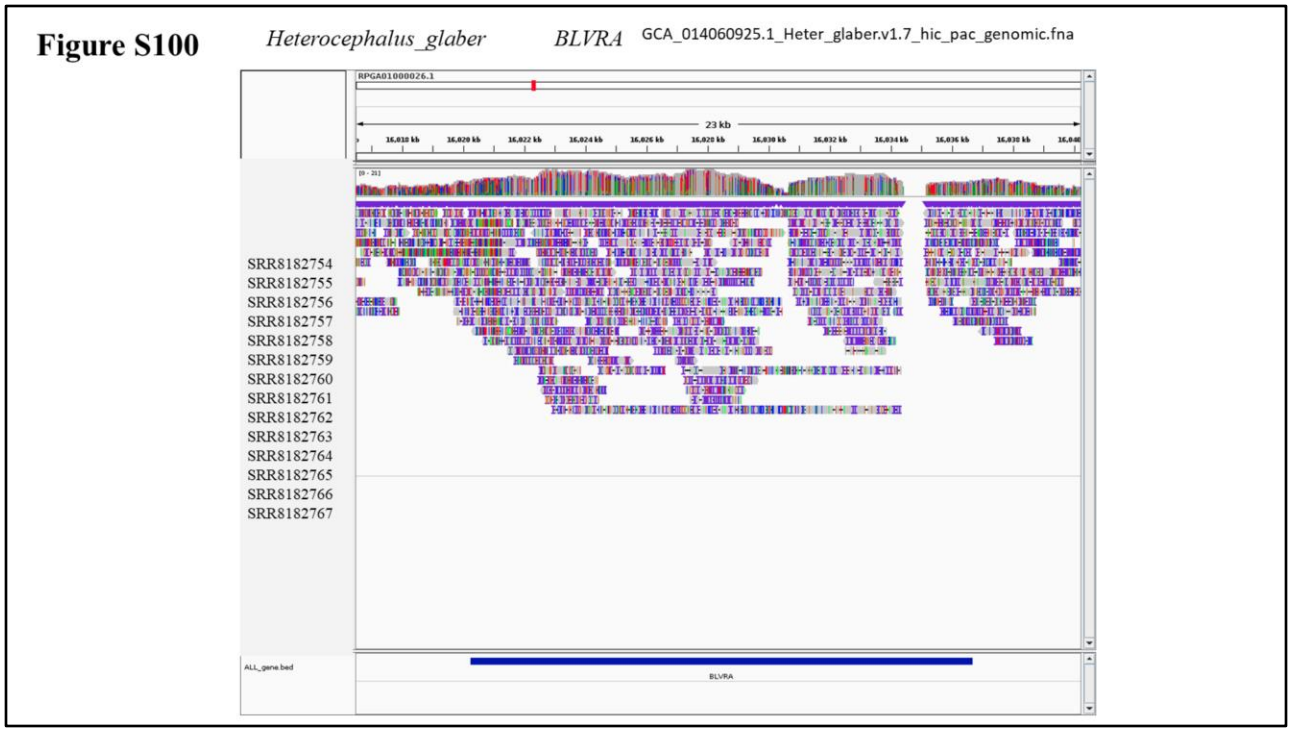


Figure S100: Screenshot of PacBio reads aligned to each of the putative locations of the *BLVRA* in naked mole-rat (*Heterocephalus glaber*). Gray boxes represent reads aligned to the genome. Raw read dataset which is mentioned above have been mapped to the GCA_014060925.1_Heter_glaber.v1.7_hic_pac_genomic.fna.

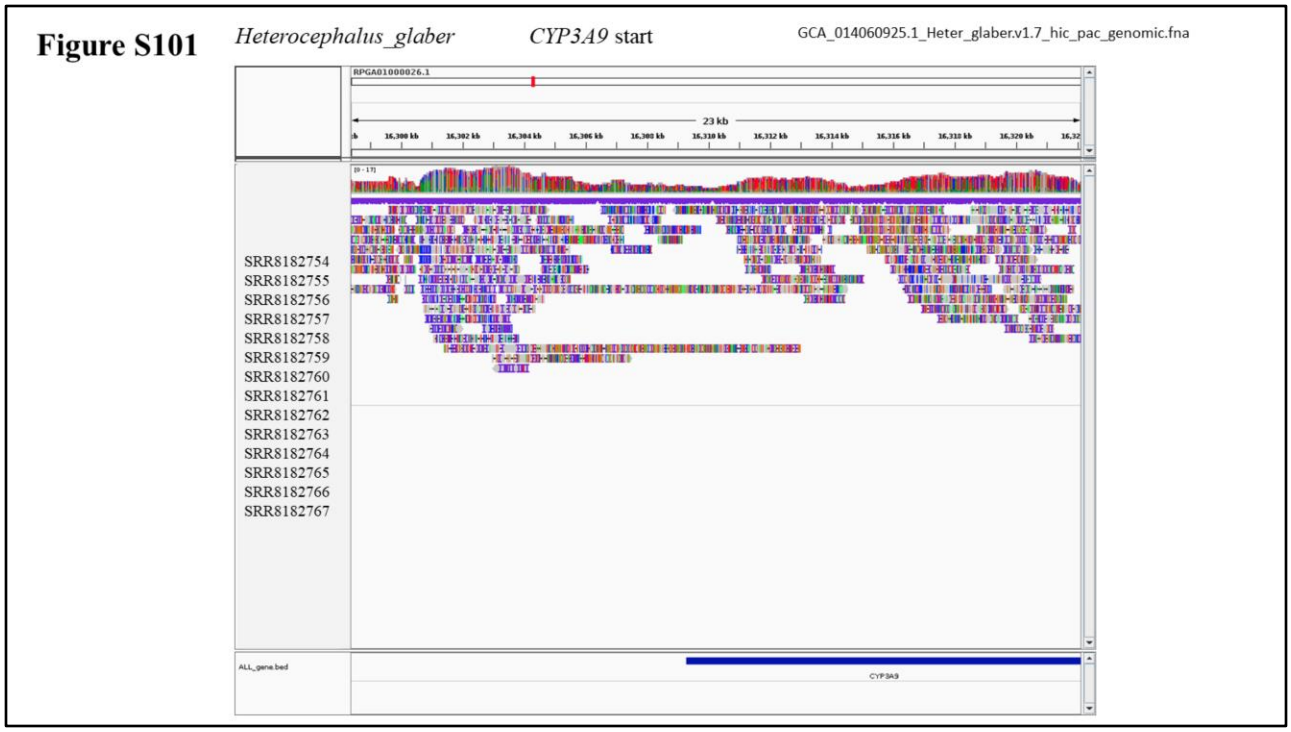


Figure S101: Screenshot of PacBio reads aligned to each of the putative locations of the *CYP2A9* start in naked mole-rat (*Heterocephalus glaber*). Gray boxes represent reads aligned to the genome. Raw read dataset which is mentioned above have been mapped to the GCA_014060925.1_Heter_glaber.v1.7_hic_pac_genomic.fna.

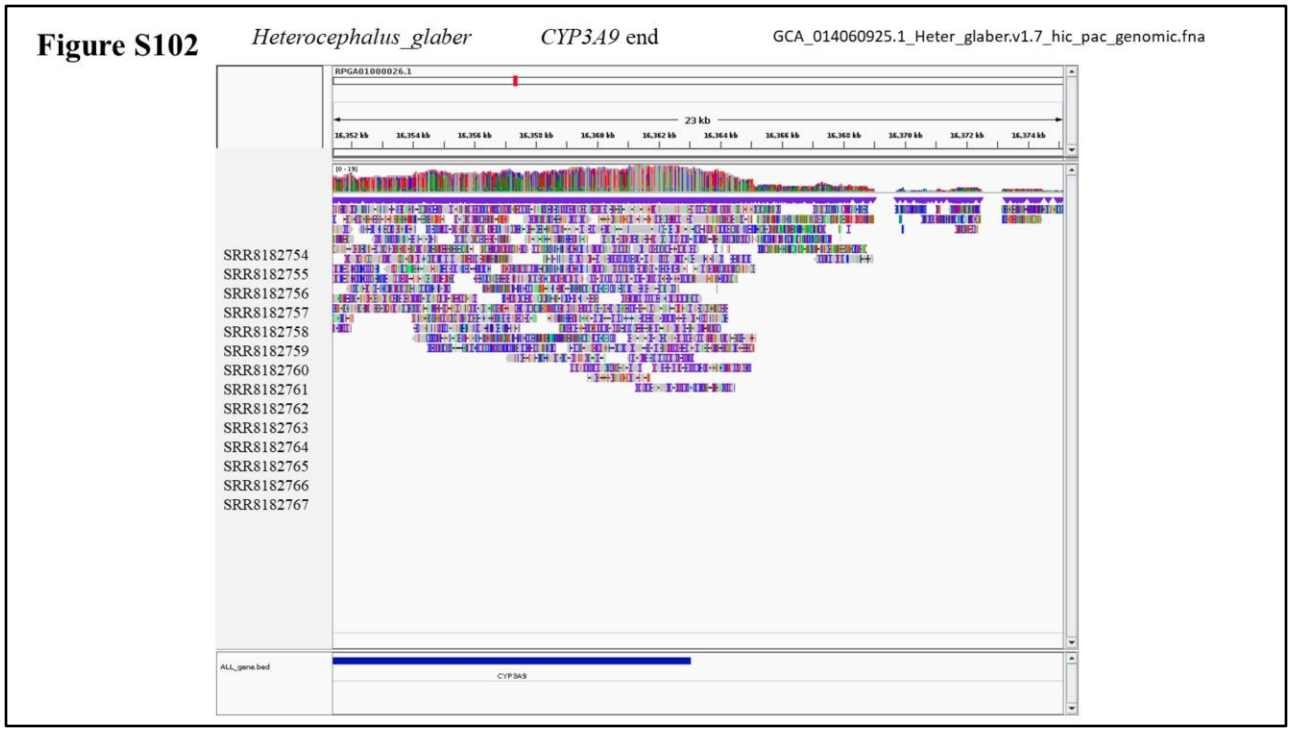


Figure S102: Screenshot of PacBio reads aligned to each of the putative locations of the *CYP3A9* end in naked mole-rat (*Heterocephalus glaber*). Gray boxes represent reads aligned to the genome. Raw read dataset which is mentioned above have been mapped to the GCA_014060925.1_Heter_glaber.v1.7_hic_pac_genomic.fna.

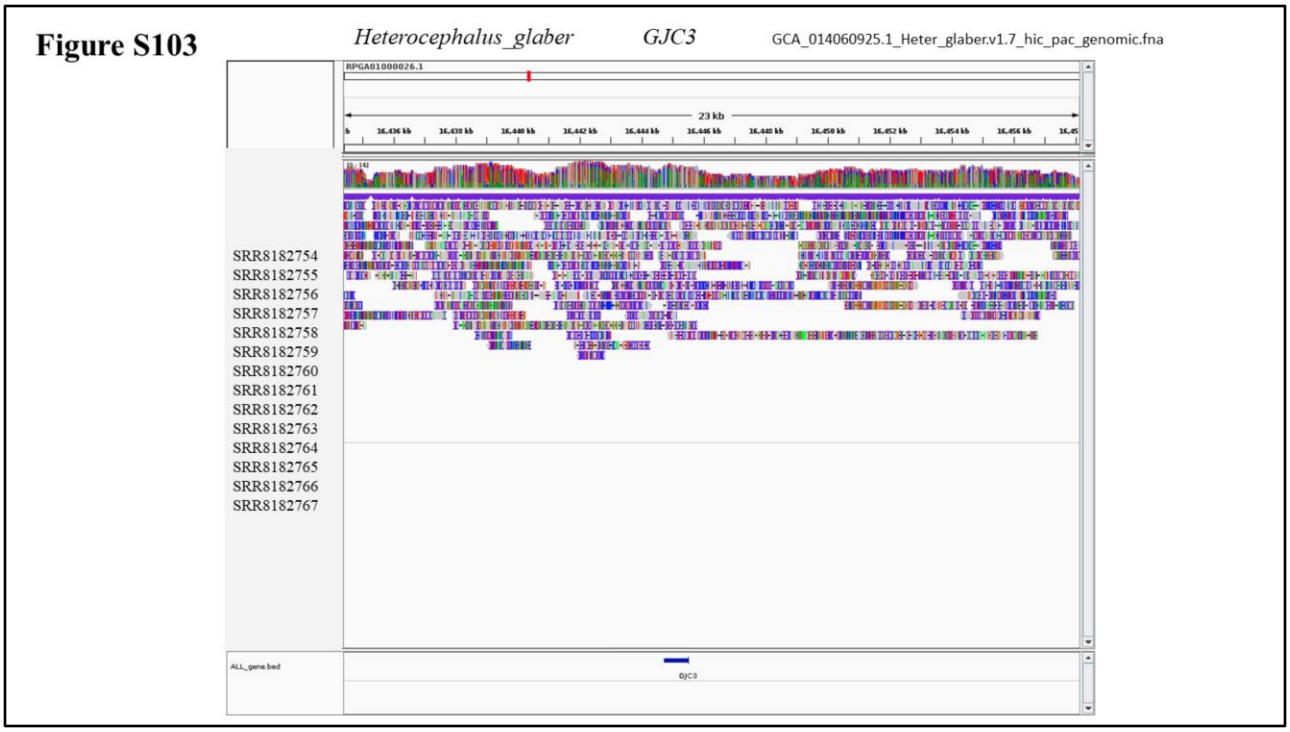


Figure S103: Screenshot of PacBio reads aligned to each of the putative locations of the *GJC3* in naked mole-rat (*Heterocephalus glaber*). Gray boxes represent reads aligned to the genome. Raw read dataset which is mentioned above have been mapped to the GCA_014060925.1_Heter_glaber.v1.7_hic_pac_genomic.fna.

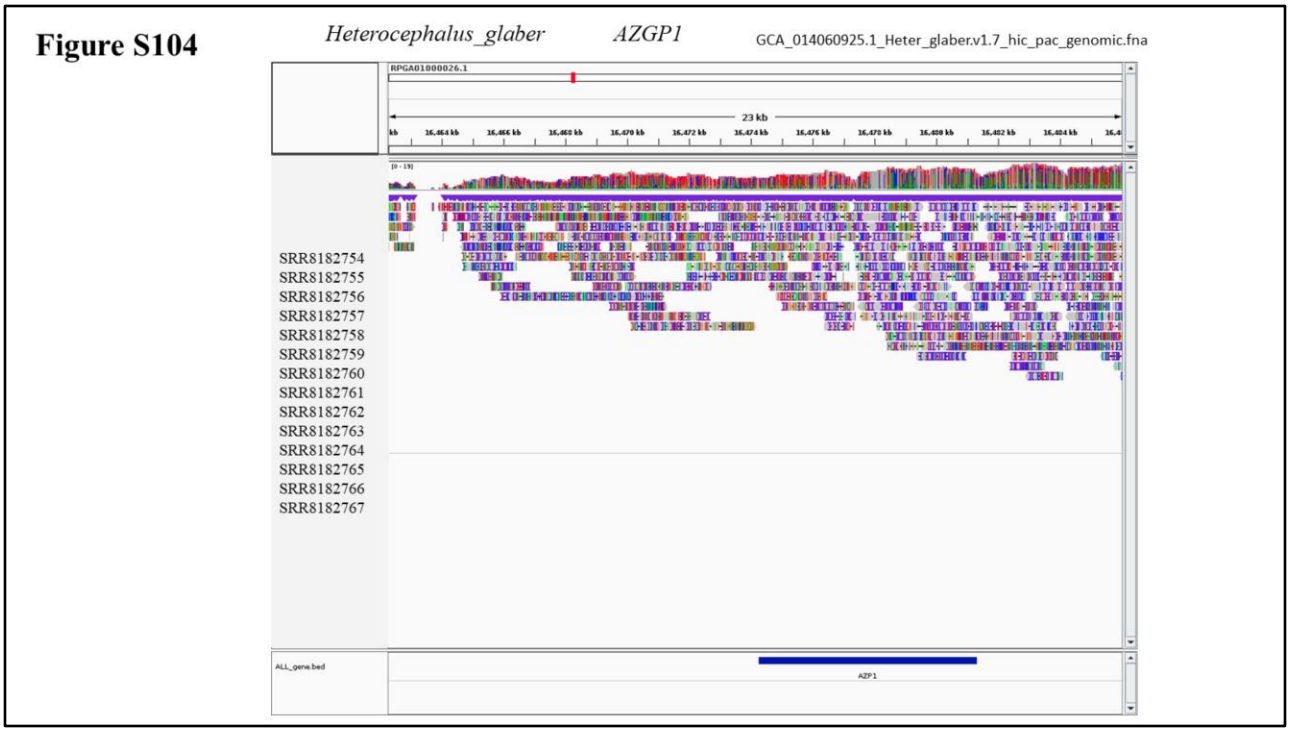


Figure S104: Screenshot of PacBio reads aligned to each of the putative locations of the *AZGP1* in naked mole-rat (*Heterocephalus glaber*). Gray boxes represent reads aligned to the genome. Raw read dataset which is mentioned above have been mapped to the GCA_014060925.1_Heter_glaber.v1.7_hic_pac_genomic.fna.

Figure S105 naked mole-rat embryonic fibroblast cells Hi-C (SRR8204318) *COA1* region

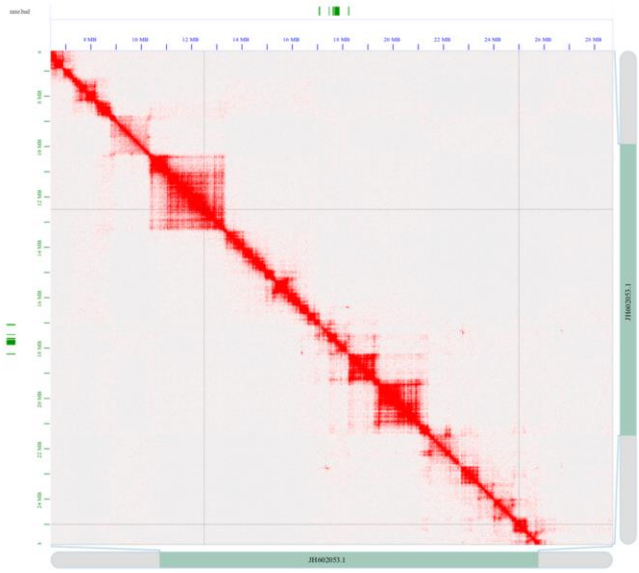


Figure S105: Hi-C figure generated using Juicebox of naked mole-rat embryonic fibroblast cells data (SRR8204318) where red dots show the genomic loci interaction between adjacent regions and gap shows there is no interaction between the nearby region. We loaded the bed file on the top and left sides of the Hi-C plot, where each line of the green color indicates the gene of interest (*COA1)* and adjacent syntenic regions. chromosome or scaffold is mentioned on the bottom and right sides of the plot.

Figure S106 naked mole-rat embryonic fibroblast cells Hi-C (SRR8204318) *COA1* duplicate region

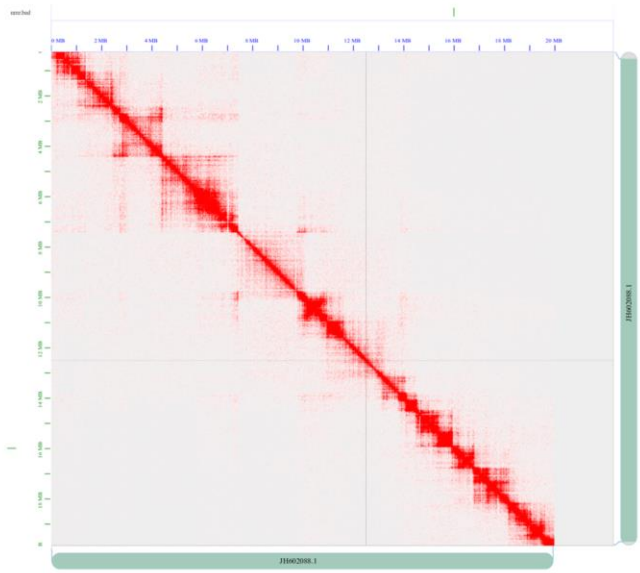


Figure S106: Hi-C figure generated using Juicebox of naked mole-rat embryonic fibroblast cells data (SRR8204318) where red dots show the genomic loci interaction between adjacent regions and gap shows there is no interaction between the nearby region. We loaded the bed file on the top and left sides of the Hi-C plot, where the green color indicates the gene of interest (*COA1* duplicate*)* and adjacent syntenic regions. chromosome or scaffold is mentioned on the bottom and right sides of the plot.

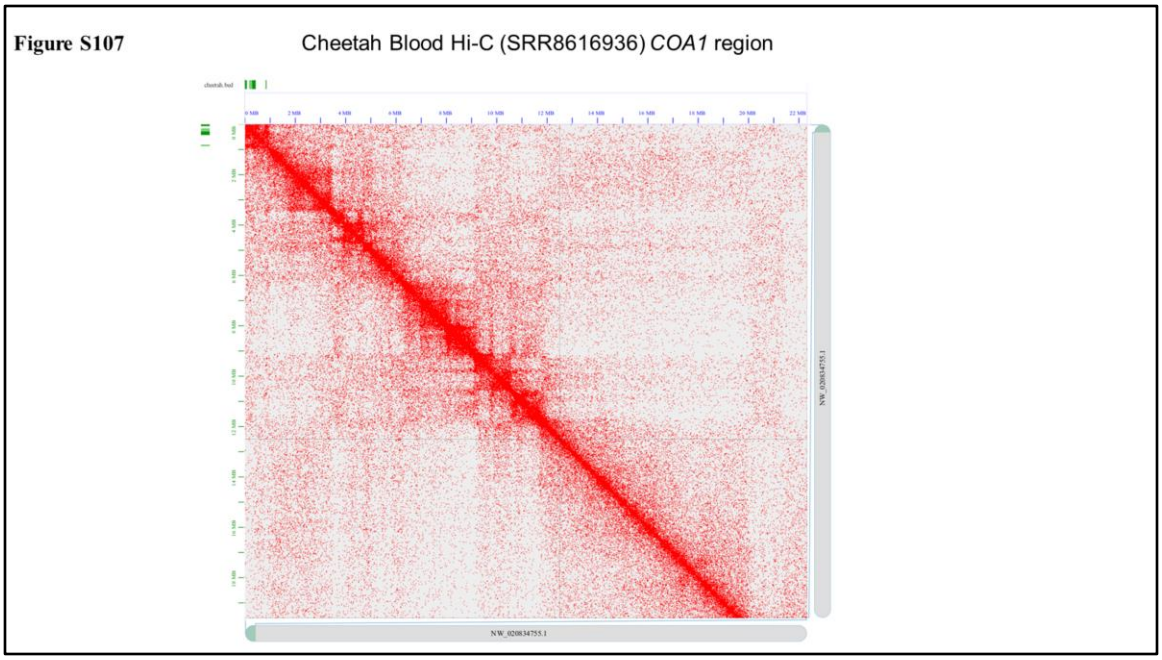


Figure S107: Hi-C figure generated using Juicebox of cheetah blood data (SRR8616936) where red dots show the genomic loci interaction between adjacent regions and gap shows there is no interaction between the nearby region. We loaded the bed file on the top and left sides of the Hi-C plot, where each line of the green color indicates the gene of interest (*COA1)* and adjacent syntenic regions. chromosome or scaffold is mentioned on the bottom and right sides of the plot.

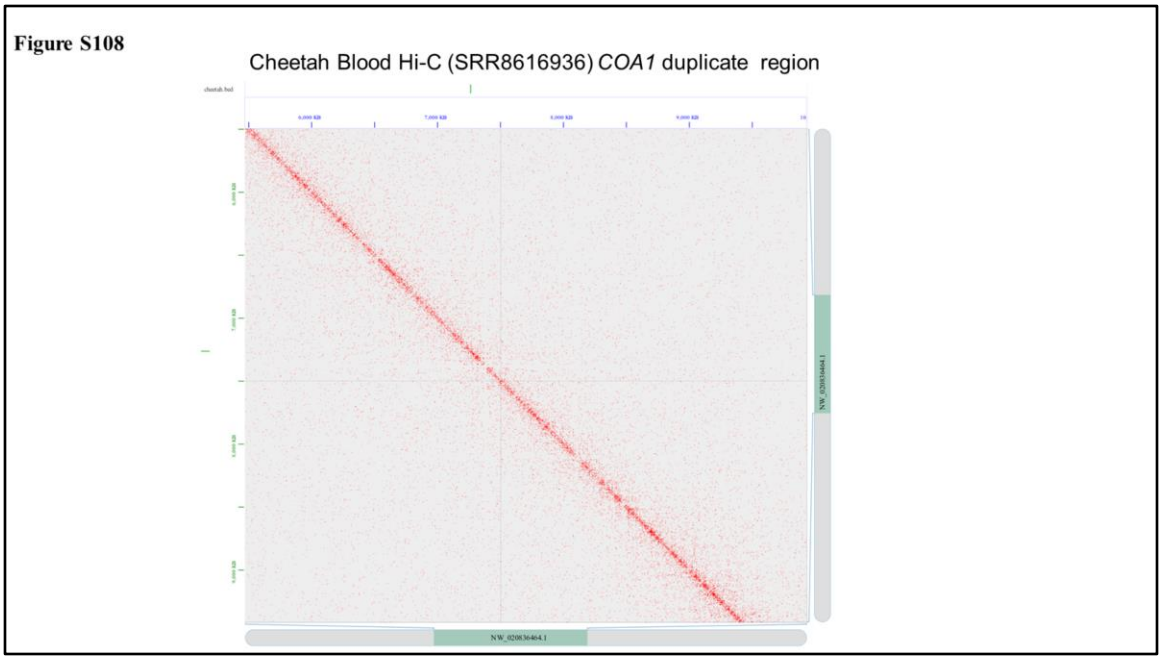


Figure S108: Hi-C figure generated using Juicebox of Cheetah blood data (SRR8616936) where red dots show the genomic loci interaction between adjacent regions and gap shows there is no interaction between the nearby region. We loaded the bed file on the top and left sides of the Hi-C plot, where the green color indicates the gene of interest (*COA1* duplicate*)* and adjacent syntenic regions. chromosome or scaffold is mentioned on the bottom and right sides of the plot.

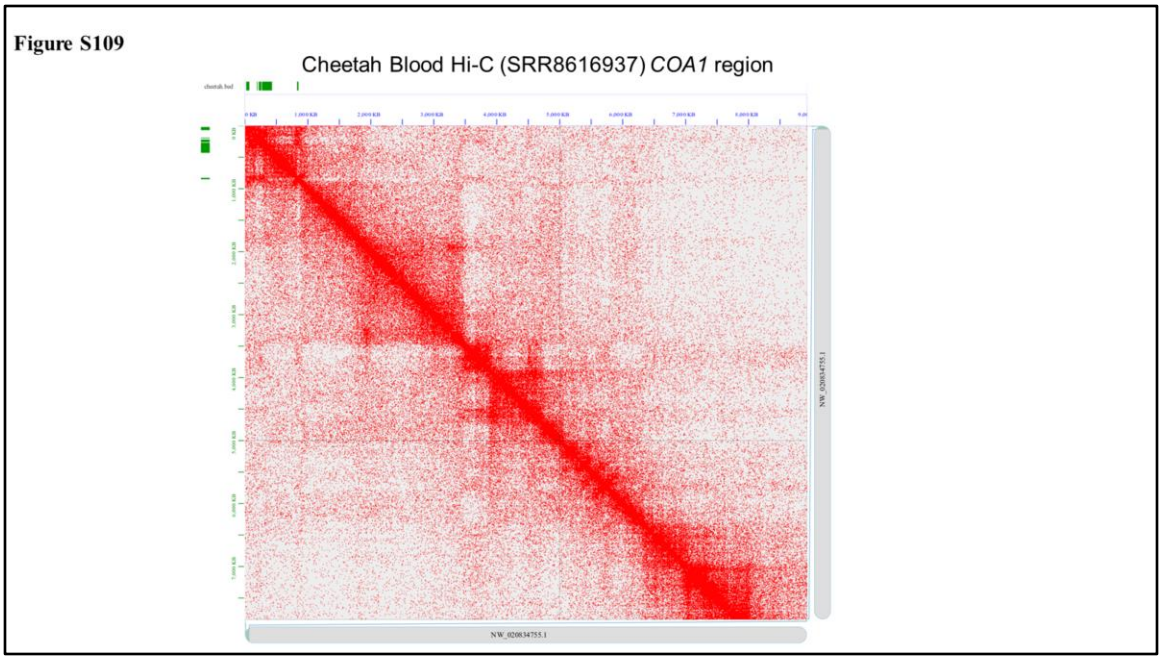


Figure S109: Hi-C figure generated using Juicebox of Cheetah blood data (SRR8616937) where red dots show the genomic loci interaction between adjacent regions and gap shows there is no interaction between the nearby region. We loaded the bed file on the top and left sides of the Hi-C plot, where each line of the green color indicates the gene of interest (*COA1)* and adjacent syntenic regions. chromosome or scaffold is mentioned on the bottom and right sides of the plot.

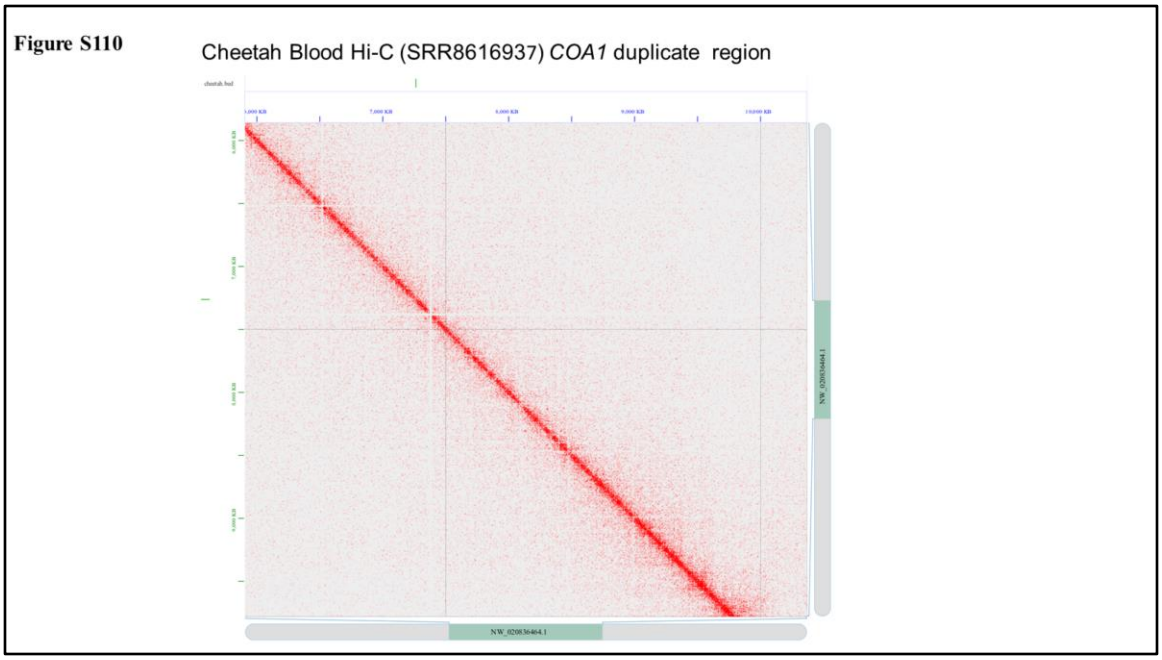


Figure S110: Hi-C figure generated using Juicebox of Cheetah blood data (SRR8616937) where red dots show the genomic loci interaction between adjacent regions and gap shows there is no interaction between the nearby region. We loaded the bed file on the top and left sides of the Hi-C plot, where the green color indicates the gene of interest (*COA1* duplicate*)* and adjacent syntenic regions. chromosome or scaffold is mentioned on the bottom and right sides of the plot.

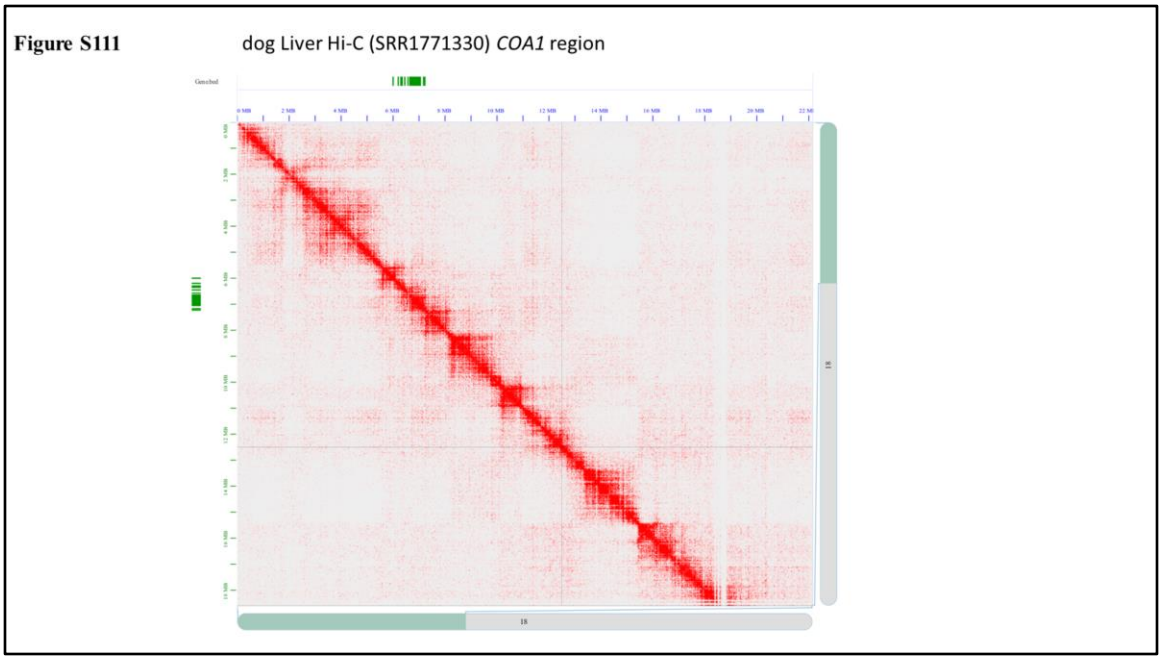


Figure S111: Hi-C figure generated using Juicebox of dog liver data (SRR1771330) where red dots show the genomic loci interaction between adjacent regions and gap shows there is no interaction between the nearby region. We loaded the bed file on the top and left sides of the Hi-C plot, where each line of the green color indicates the gene of interest (*COA1)* and adjacent syntenic regions. chromosome or scaffold is mentioned on the bottom and right sides of the plot.

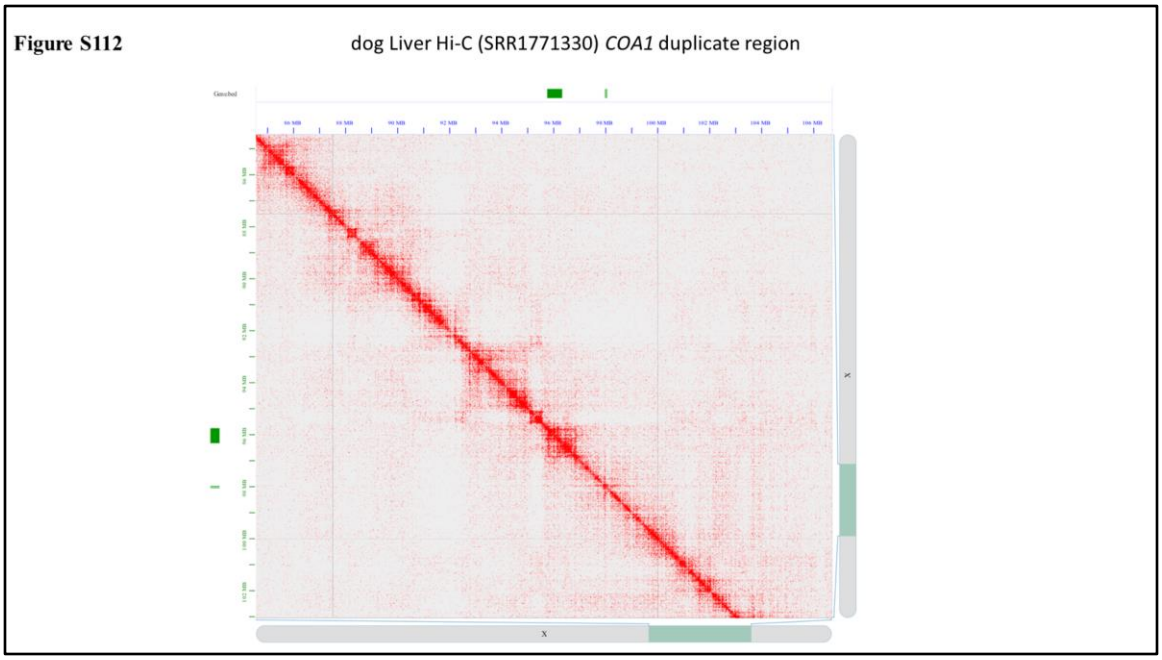


Figure S112: Hi-C figure generated using Juicebox of dog liver data (SRR1771330) where red dots show the genomic loci interaction between adjacent regions and gap shows there is no interaction between the nearby region. We loaded the bed file on the top and left sides of the Hi-C plot, where the green color indicates the gene of interest (*COA1* duplicate*)* and adjacent syntenic regions. chromosome or scaffold is mentioned on the bottom and right sides of the plot.

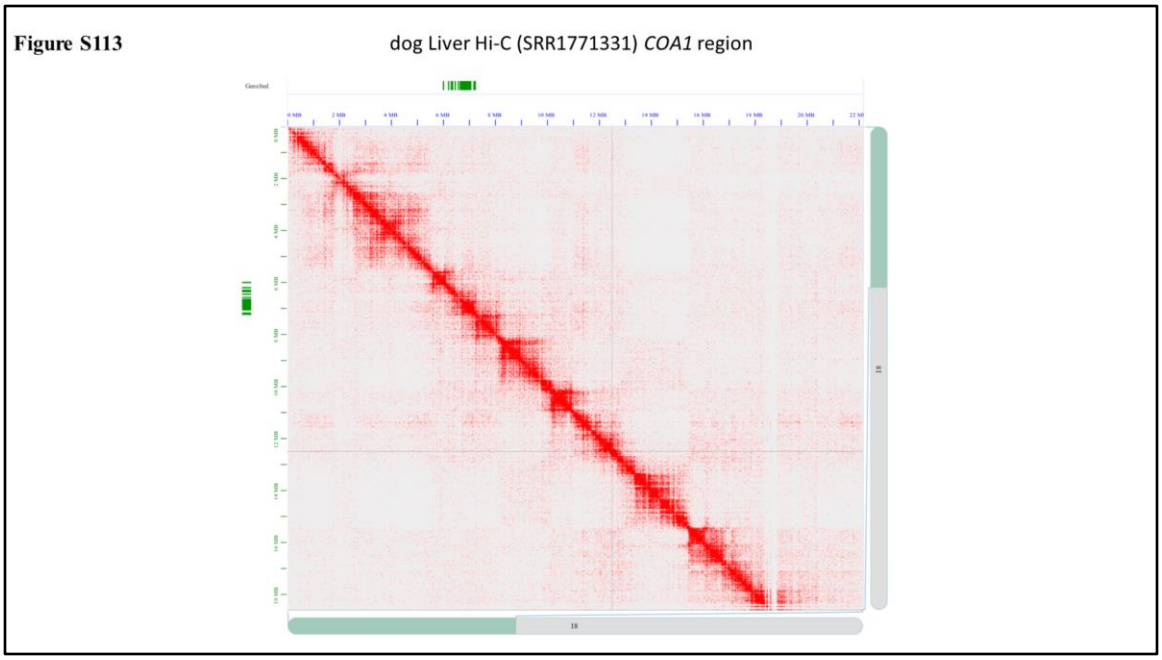


Figure S113: Hi-C figure generated using Juicebox of dog liver data (SRR1771331) where red dots show the genomic loci interaction between adjacent regions and gap shows there is no interaction between the nearby region. We loaded the bed file on the top and left sides of the Hi-C plot, where each line of the green color indicates the gene of interest (*COA1)* and adjacent syntenic regions. chromosome or scaffold is mentioned on the bottom and right sides of the plot.

Figure S114 dog Liver Hi-C (SRR1771331) *COA1* duplicate region

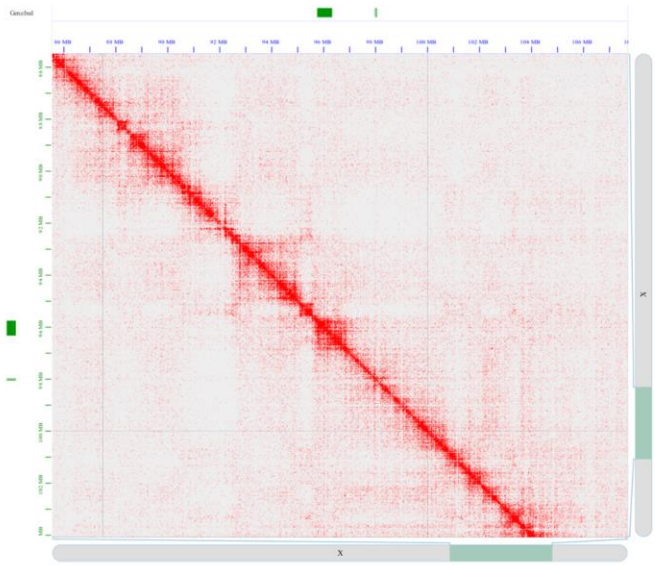


Figure S114: Hi-C figure generated using Juicebox of dog liver data (SRR1771331) where red dots show the genomic loci interaction between adjacent regions and gap shows there is no interaction between the nearby region. We loaded the bed file on the top and left sides of the Hi-C plot, where the green color indicates the gene of interest (*COA1* duplicate*)* and adjacent syntenic regions. chromosome or scaffold is mentioned on the bottom and right sides of the plot.

Figure S115 dog Liver Hi-C (SRR1771332) *COA1* region

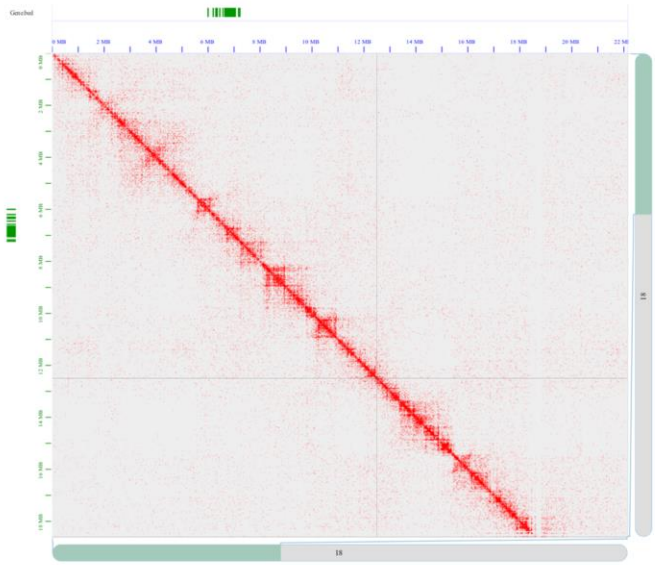


Figure 115: Hi-C figure generated using Juicebox of dog liver data (SRR1771332) where red dots show the genomic loci interaction between adjacent regions and gap shows there is no interaction between the nearby region. We loaded the bed file on the top and left sides of the Hi-C plot, where each line of the green color indicates the gene of interest (*COA1)* and adjacent syntenic regions. chromosome or scaffold is mentioned on the bottom and right sides of the plot.

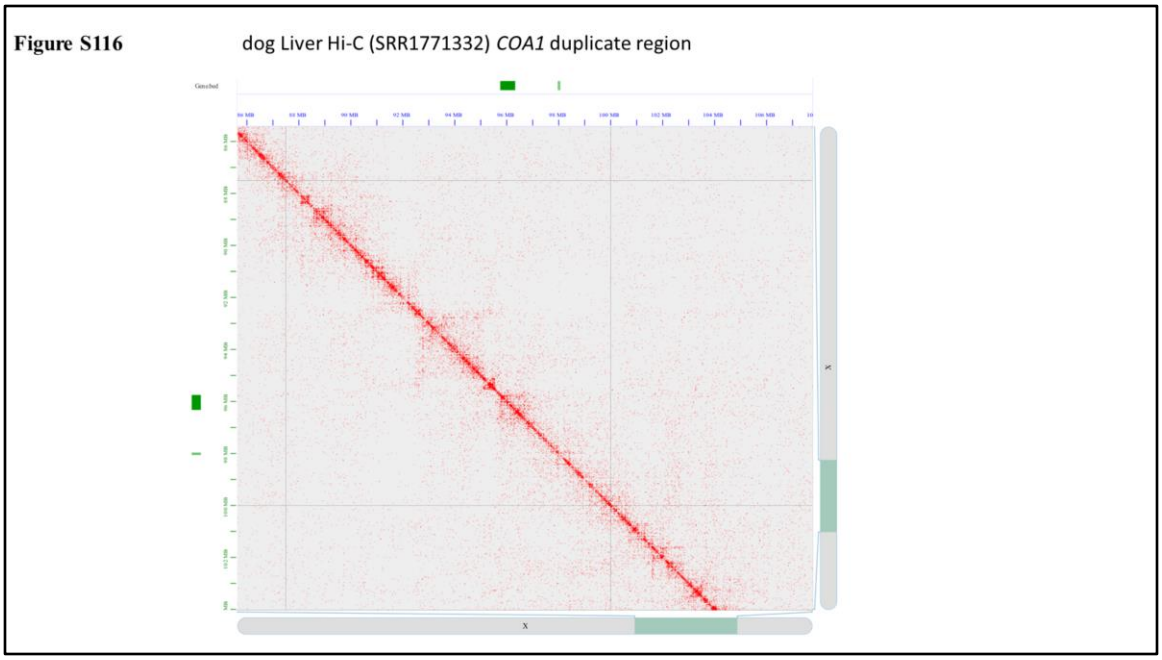


Figure S116: Hi-C figure generated using Juicebox of dog liver data (SRR1771332) where red dots show the genomic loci interaction between adjacent regions and gap shows there is no interaction between the nearby region. We loaded the bed file on the top and left sides of the Hi-C plot, where the green color indicates the gene of interest (*COA1* duplicate*)* and adjacent syntenic regions. chromosome or scaffold is mentioned on the bottom and right sides of the plot.

Figure S117 dog Liver Hi-C (SRR1771333) *COA1* region

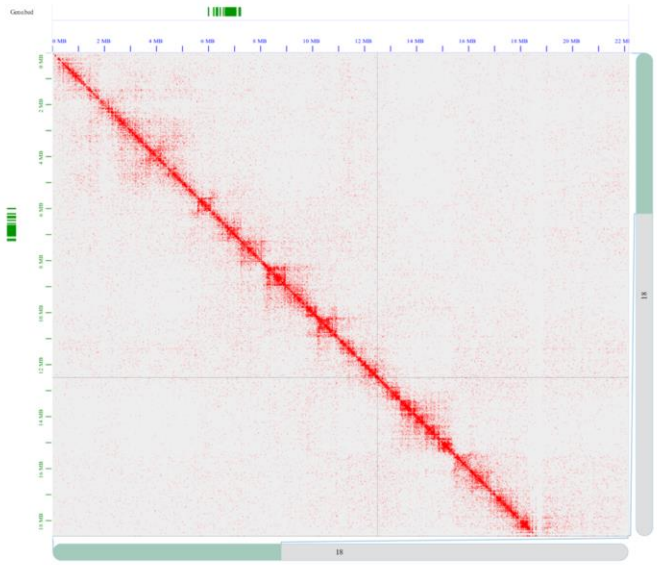


Figure 117: Hi-C figure generated using Juicebox of dog liver data (SRR1771333) where red dots show the genomic loci interaction between adjacent regions and gap shows there is no interaction between the nearby region. We loaded the bed file on the top and left sides of the Hi-C plot, where each line of the green color indicates the gene of interest (*COA1)* and adjacent syntenic regions. chromosome or scaffold is mentioned on the bottom and right sides of the plot.

Figure S118 dog Liver Hi-C (SRR1771333) *COA1* duplicate region

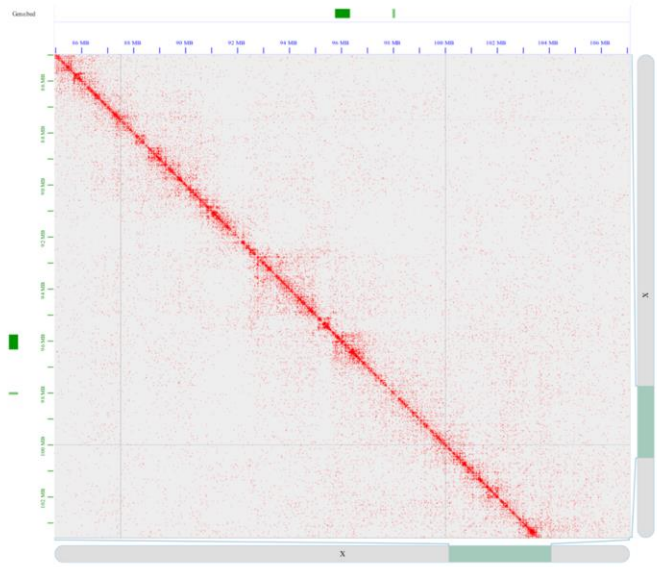


Figure S118: Hi-C figure generated using Juicebox of dog liver data (SRR1771333) where red dots show the genomic loci interaction between adjacent regions and gap shows there is no interaction between the nearby region. We loaded the bed file on the top and left sides of the Hi-C plot, where the green color indicates the gene of interest (*COA1* duplicate*)* and adjacent syntenic regions. chromosome or scaffold is mentioned on the bottom and right sides of the plot.

Figure S119 dog Liver Hi-C (SRR1771334) *COA1* region

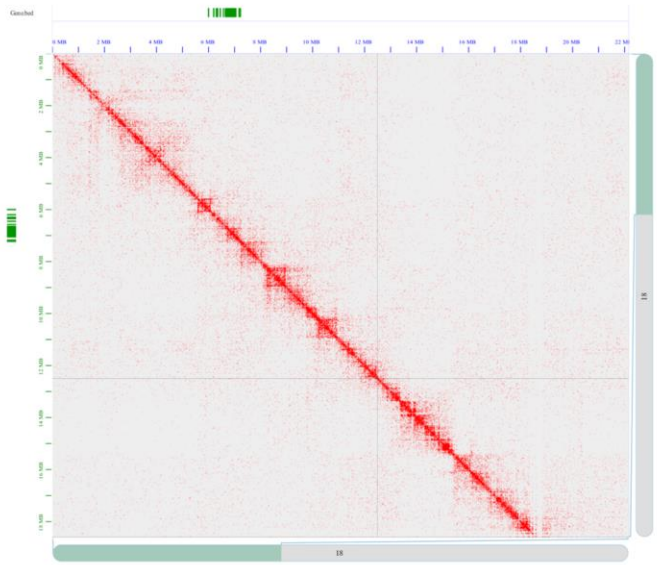


Figure S119: Hi-C figure generated using Juicebox of dog liver data (SRR1771334) where red dots show the genomic loci interaction between adjacent regions and gap shows there is no interaction between the nearby region. We loaded the bed file on the top and left sides of the Hi-C plot, where each line of the green color indicates the gene of interest (*COA1)* and adjacent syntenic regions. chromosome or scaffold is mentioned on the bottom and right sides of the plot.

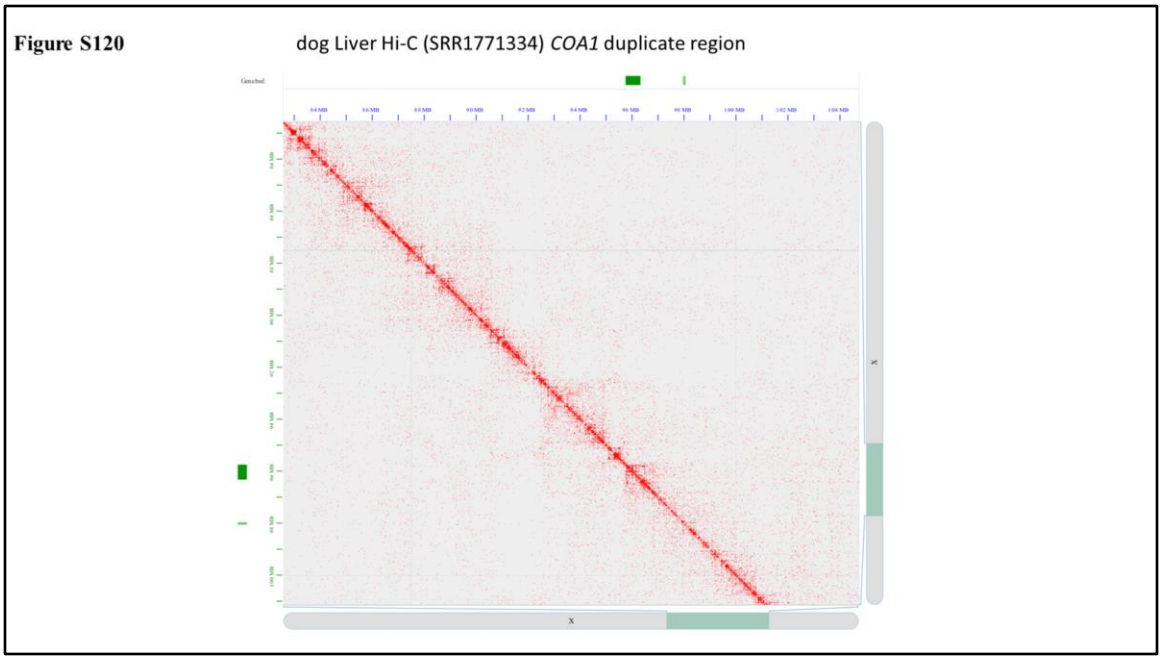


Figure S120: Hi-C figure generated using Juicebox of dog liver data (SRR1771334) where red dots show the genomic loci interaction between adjacent regions and gap shows there is no interaction between the nearby region. We loaded the bed file on the top and left sides of the Hi-C plot, where the green color indicates the gene of interest (*COA1* duplicate*)* and adjacent syntenic regions. chromosome or scaffold is mentioned on the bottom and right sides of the plot.

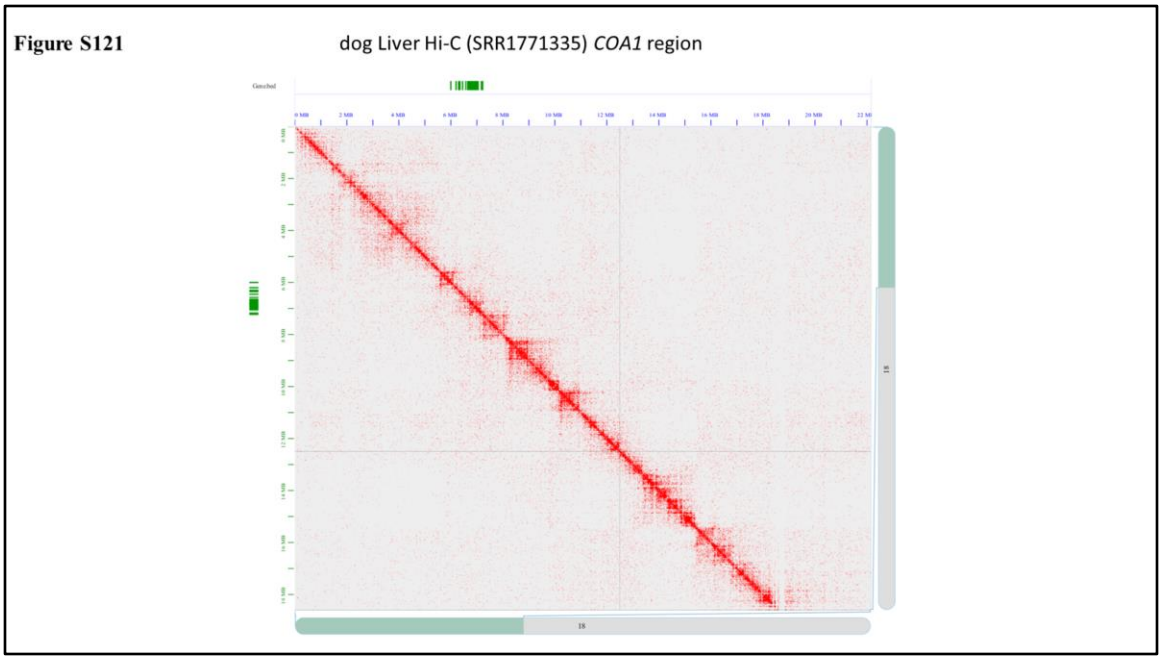


Figure S121: Hi-C figure generated using Juicebox of dog liver data (SRR1771335) where red dots show the genomic loci interaction between adjacent regions and gap shows there is no interaction between the nearby region. We loaded the bed file on the top and left sides of the Hi-C plot, where each line of the green color indicates the gene of interest (*COA1)* and adjacent syntenic regions. chromosome or scaffold is mentioned on the bottom and right sides of the plot.

Figure S122 dog Liver Hi-C (SRR1771335) *COA1* duplicate region

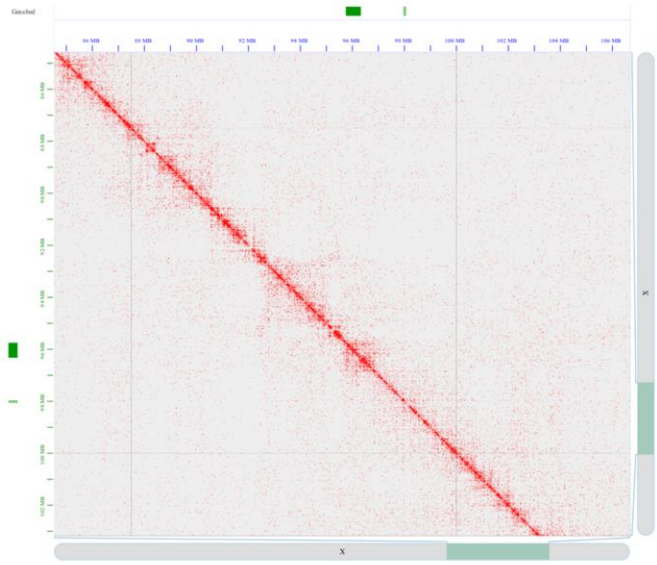


Figure S122: Hi-C figure generated using Juicebox of dog liver data (SRR1771335) where red dots show the genomic loci interaction between adjacent regions and gap shows there is no interaction between the nearby region. We loaded the bed file on the top and left sides of the Hi-C plot, where the green color indicates the gene of interest (*COA1* duplicate*)* and adjacent syntenic regions. chromosome or scaffold is mentioned on the bottom and right sides of the plot.

Figure S123 dog Liver Hi-C (SRR1771336) *COA1* region

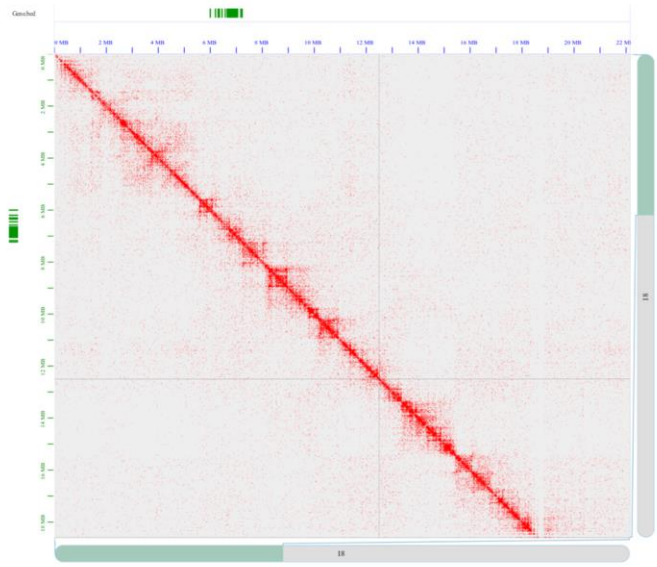


Figure S123: Hi-C figure generated using Juicebox of dog liver data (SRR1771336) where red dots show the genomic loci interaction between adjacent regions and gap shows there is no interaction between the nearby region. We loaded the bed file on the top and left sides of the Hi-C plot, where each line of the green color indicates the gene of interest (*COA1)* and adjacent syntenic regions. chromosome or scaffold is mentioned on the bottom and right sides of the plot.

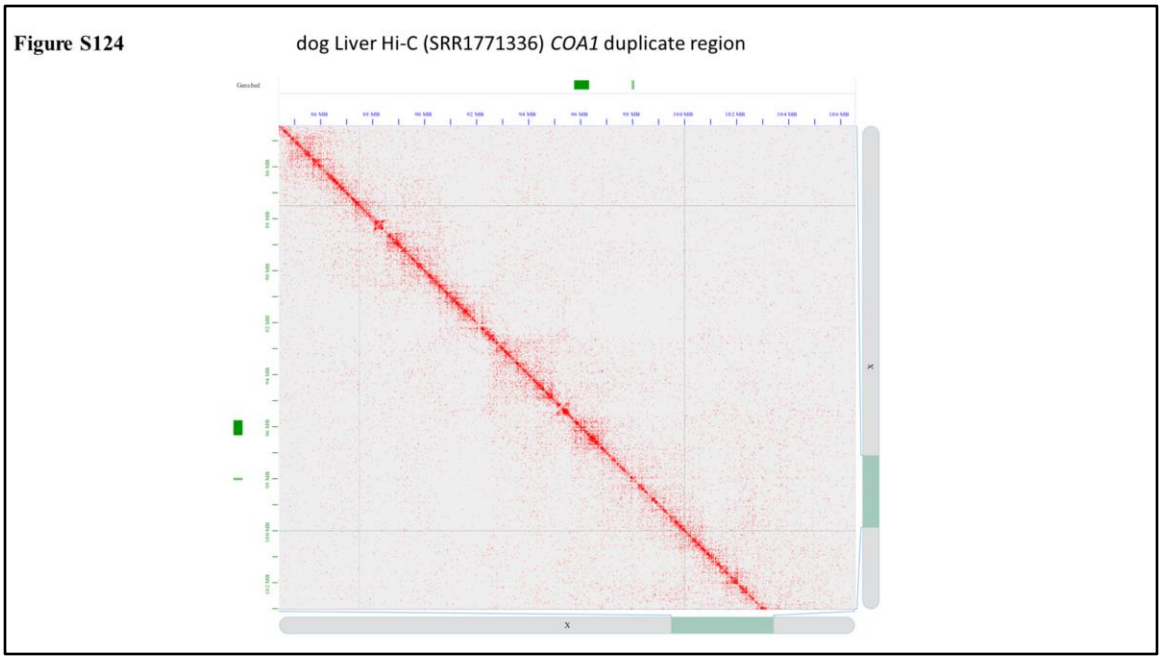


Figure S124: Hi-C figure generated using Juicebox of dog liver data (SRR1771336) where red dots show the genomic loci interaction between adjacent regions and gap shows there is no interaction between the nearby region. We loaded the bed file on the top and left sides of the Hi-C plot, where the green color indicates the gene of interest (*COA1* duplicate*)* and adjacent syntenic regions. chromosome or scaffold is mentioned on the bottom and right sides of the plot.

Figure S125 dog Liver Hi-C (SRR1771337) *COA1* region

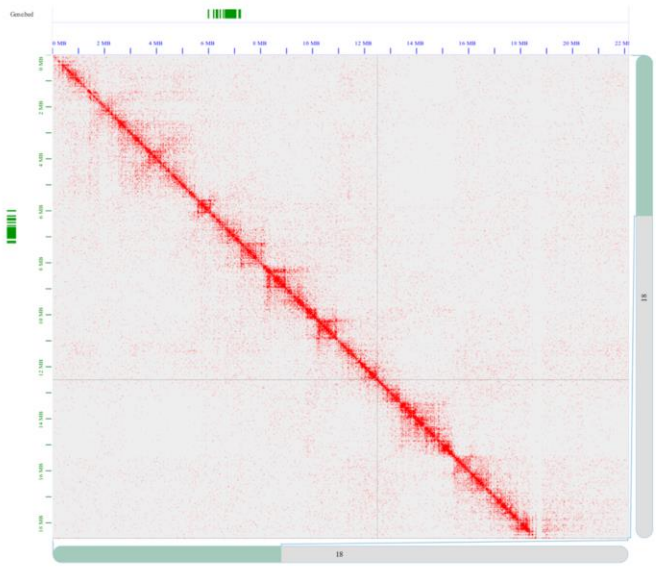


Figure S125: Hi-C figure generated using Juicebox of dog liver data (SRR1771337) where red dots show the genomic loci interaction between adjacent regions and gap shows there is no interaction between the nearby region. We loaded the bed file on the top and left sides of the Hi-C plot, where each line of the green color indicates the gene of interest (*COA1)* and adjacent syntenic regions. chromosome or scaffold is mentioned on the bottom and right sides of the plot.

Figure S126 dog Liver Hi-C (SRR1771337*) COA1* duplicate region

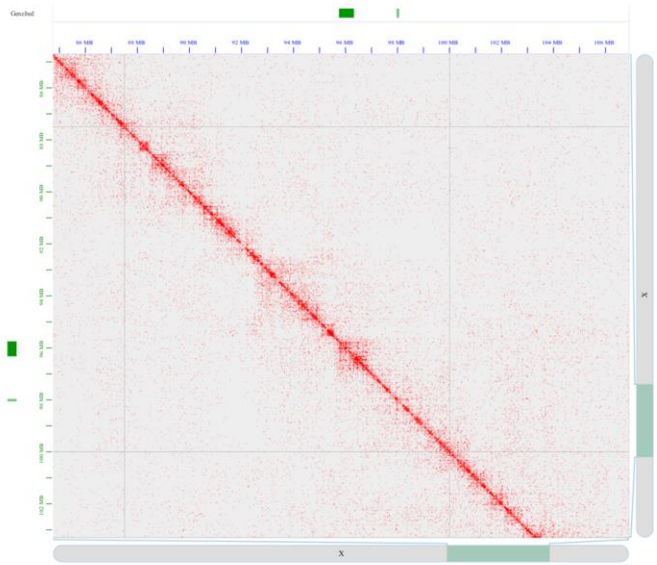


Figure S126: Hi-C figure generated using Juicebox of dog liver data (SRR1771337) where red dots show the genomic loci interaction between adjacent regions and gap shows there is no interaction between the nearby region. We loaded the bed file on the top and left sides of the Hi-C plot, where the green color indicates the gene of interest (*COA1* duplicate*)* and adjacent syntenic regions. chromosome or scaffold is mentioned on the bottom and right sides of the plot.

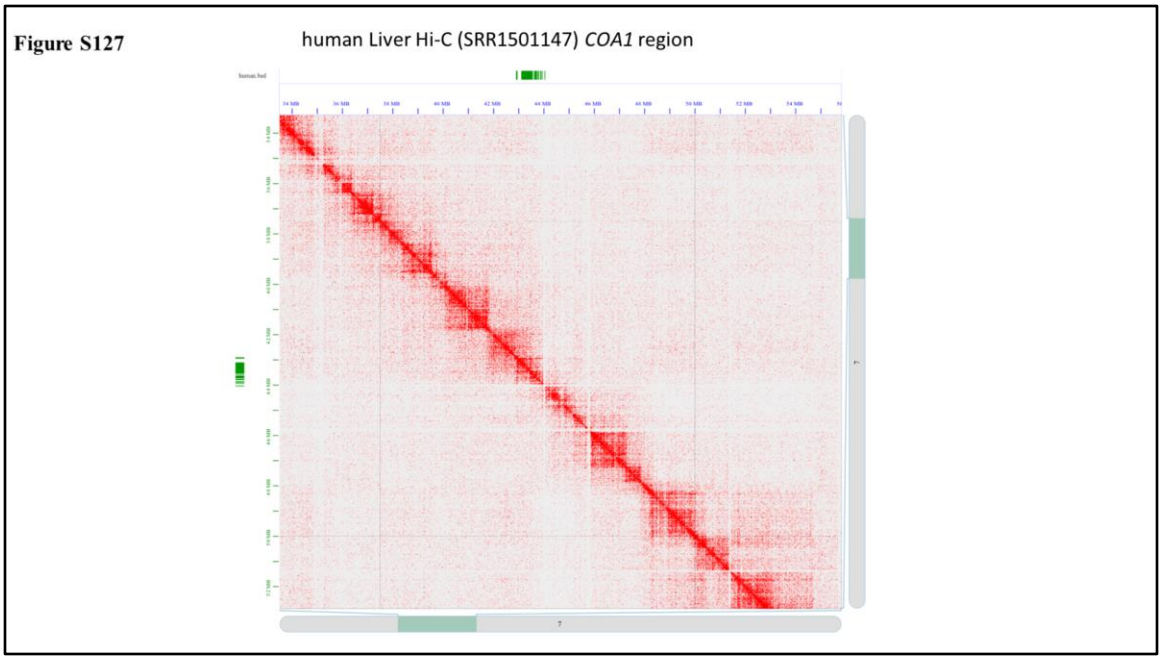


Figure S127: Hi-C figure generated using Juicebox of human liver data (SRR1501147) where red dots show the genomic loci interaction between adjacent regions and gap shows there is no interaction between the nearby region. We loaded the bed file on the top and left sides of the Hi-C plot, where the each line of green color indicates the gene of interest (*COA1*) and adjacent syntenic regions. chromosome or scaffold is mentioned on the bottom and right sides of the plot.

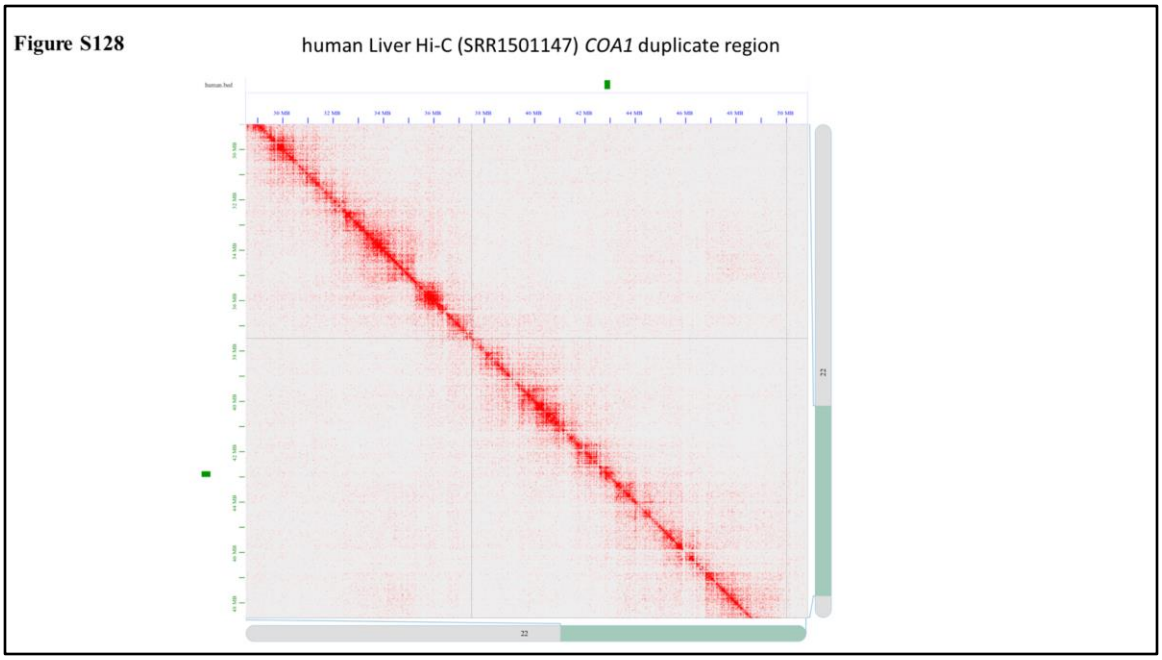


Figure S128: Hi-C figure generated using Juicebox of human liver data (SRR1501147) where red dots show the genomic loci interaction between adjacent regions and gap shows there is no interaction between the nearby region. We loaded the bed file on the top and left sides of the Hi-C plot, where the green color indicates the gene of interest (*COA1* duplicate) and adjacent syntenic regions. chromosome or scaffold is mentioned on the bottom and right sides of the plot.

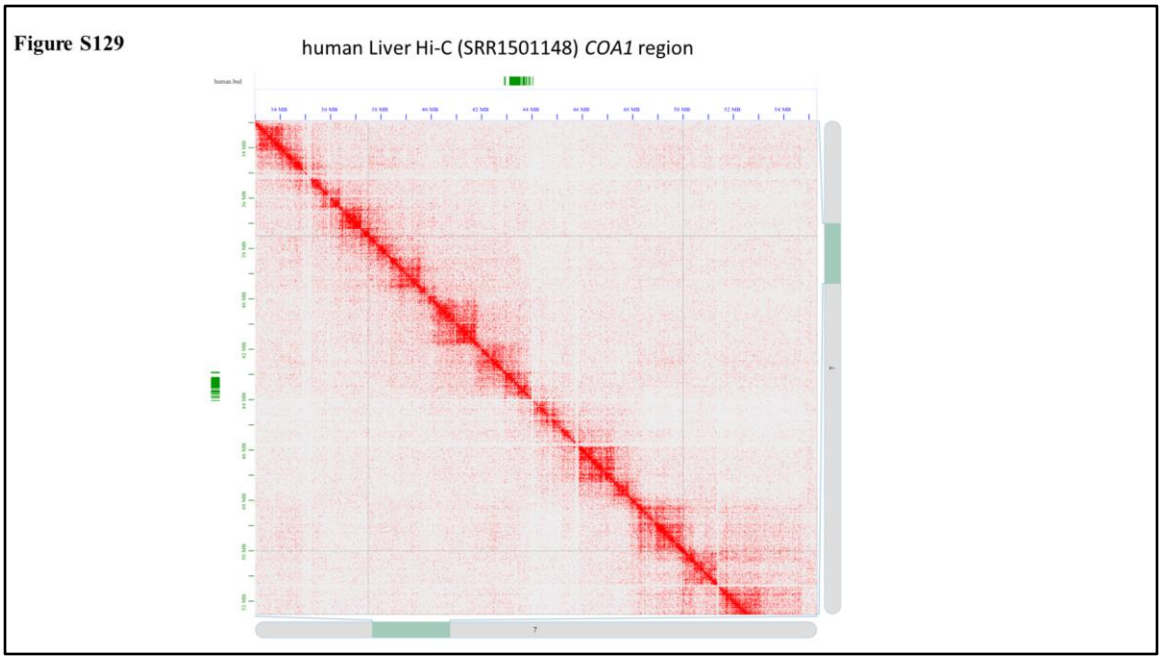


Figure S129: Hi-C figure generated using Juicebox of human liver data (SRR1501148) where red dots show the genomic loci interaction between adjacent regions and gap shows there is no interaction between the nearby region. We loaded the bed file on the top and left sides of the Hi-C plot, where the each line of green color indicates the gene of interest (*COA1*) and adjacent syntenic regions. chromosome or scaffold is mentioned on the bottom and right sides of the plot.

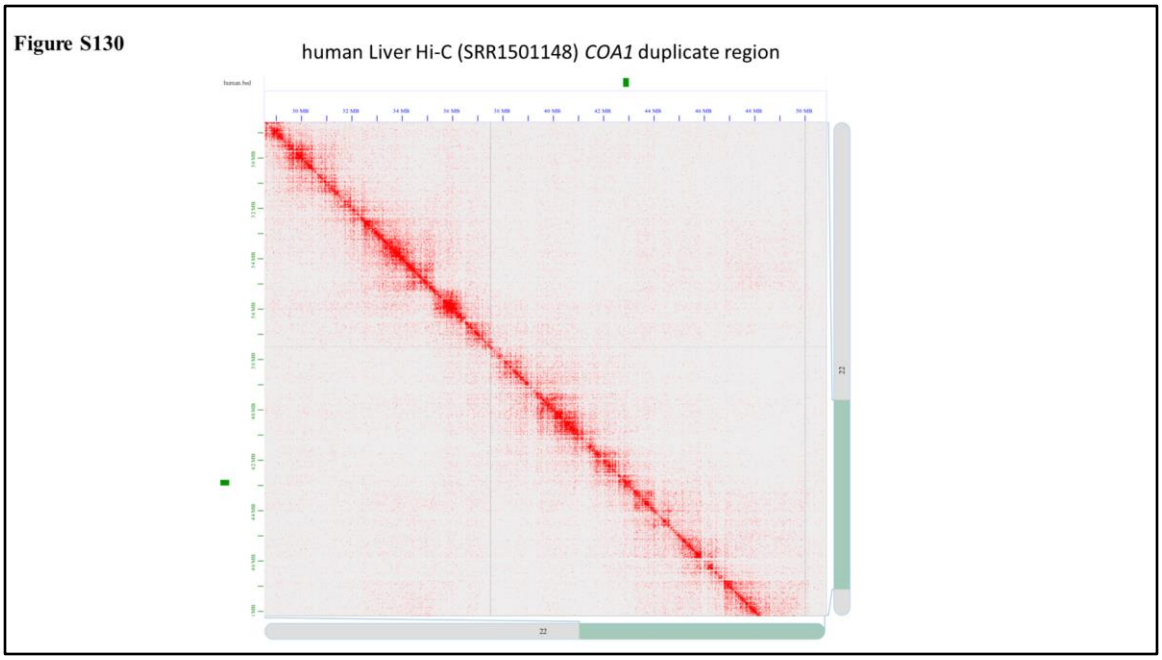


Figure S130: Hi-C figure generated using Juicebox of human liver data (SRR1501148) where red dots show the genomic loci interaction between adjacent regions and gap shows there is no interaction between the nearby region. We loaded the bed file on the top and left sides of the Hi-C plot, where the green color indicates the gene of interest (*COA1* duplicate) and adjacent syntenic regions. chromosome or scaffold is mentioned on the bottom and right sides of the plot.

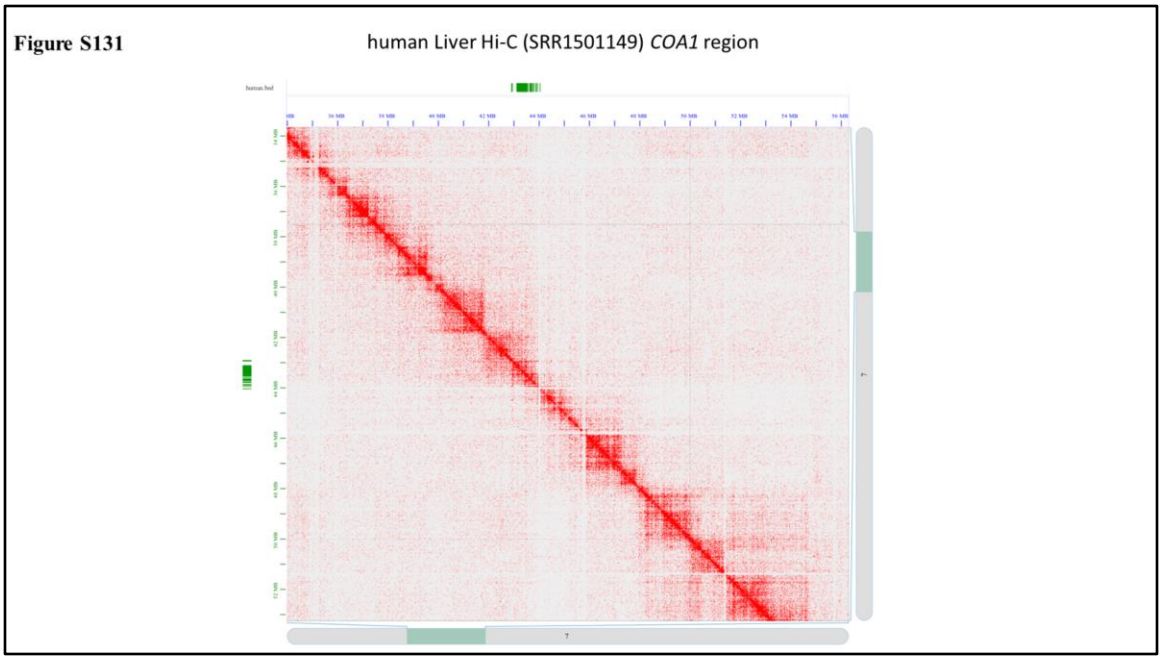


Figure S131: Hi-C figure generated using Juicebox of human liver data (SRR1501149) where red dots show the genomic loci interaction between adjacent regions and gap shows there is no interaction between the nearby region. We loaded the bed file on the top and left sides of the Hi-C plot, where the each line of green color indicates the gene of interest (*COA1*) and adjacent syntenic regions. chromosome or scaffold is mentioned on the bottom and right sides of the plot.

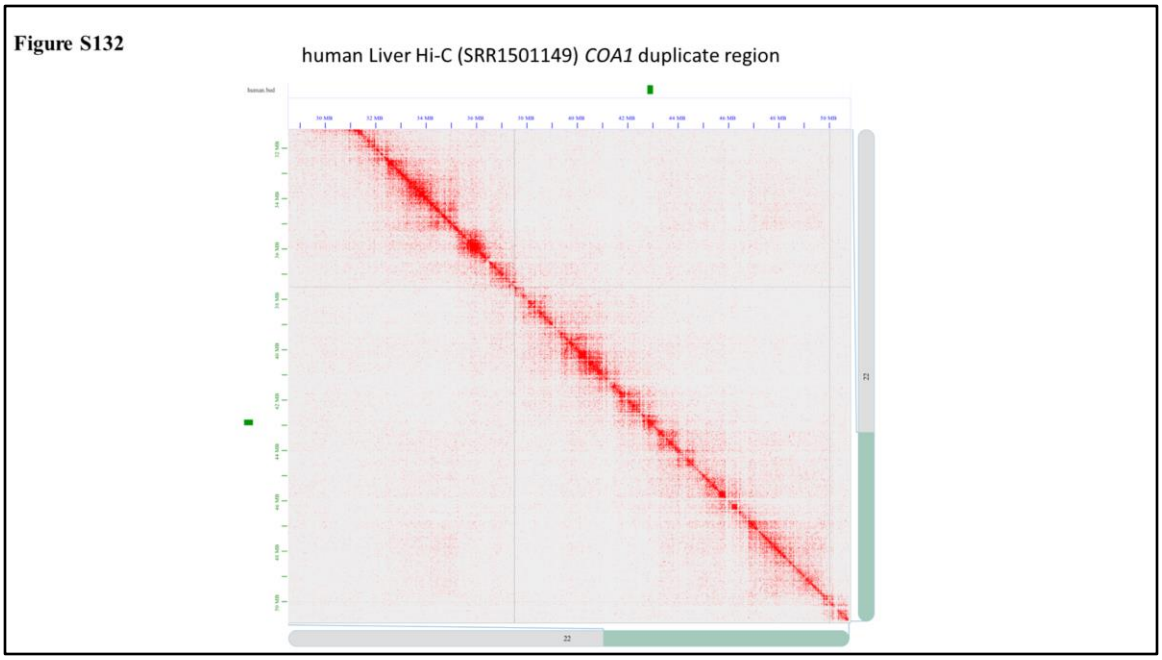


Figure S132: Hi-C figure generated using Juicebox of human liver data (SRR1501149) where red dots show the genomic loci interaction between adjacent regions and gap shows there is no interaction between the nearby region. We loaded the bed file on the top and left sides of the Hi-C plot, where the green color indicates the gene of interest (*COA1* duplicate) and adjacent syntenic regions. chromosome or scaffold is mentioned on the bottom and right sides of the plot.

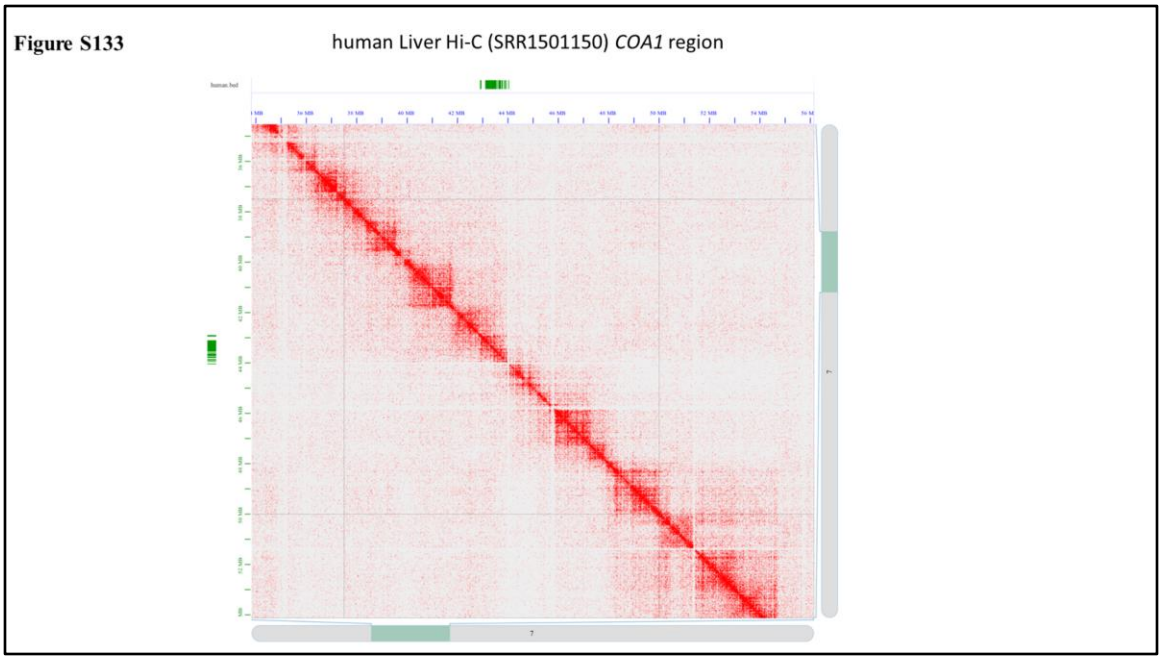


Figure S133: Hi-C figure generated using Juicebox of human liver data (SRR1501150) where red dots show the genomic loci interaction between adjacent regions and gap shows there is no interaction between the nearby region. We loaded the bed file on the top and left sides of the Hi-C plot, where the each line of green color indicates the gene of interest (*COA1*) and adjacent syntenic regions. chromosome or scaffold is mentioned on the bottom and right sides of the plot.

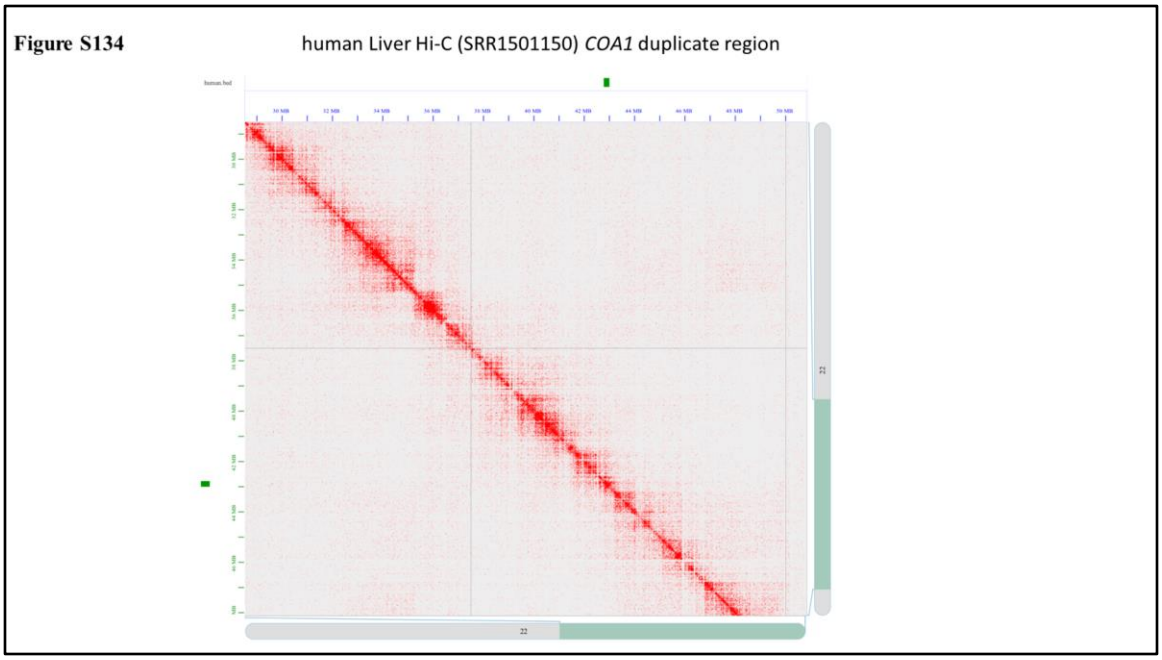


Figure S134: Hi-C figure generated using Juicebox of human liver data (SRR1501150) where red dots show the genomic loci interaction between adjacent regions and gap shows there is no interaction between the nearby region. We loaded the bed file on the top and left sides of the Hi-C plot, where the green color indicates the gene of interest (*COA1* duplicate) and adjacent syntenic regions. chromosome or scaffold is mentioned on the bottom and right sides of the plot.

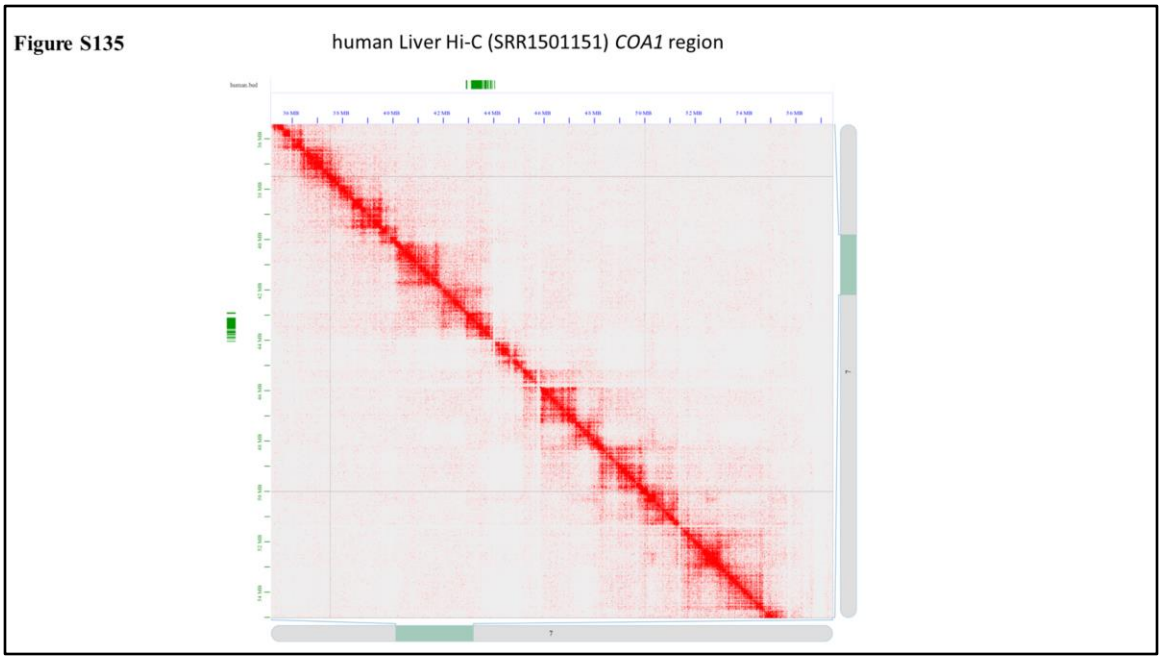


Figure S135: Hi-C figure generated using Juicebox of human liver data (SRR1501151) where red dots show the genomic loci interaction between adjacent regions and gap shows there is no interaction between the nearby region. We loaded the bed file on the top and left sides of the Hi-C plot, where the each line of green color indicates the gene of interest (*COA1*) and adjacent syntenic regions. chromosome or scaffold is mentioned on the bottom and right sides of the plot.

Figure S136 human Liver Hi-C (SRR1501151) *COA1* duplicate region

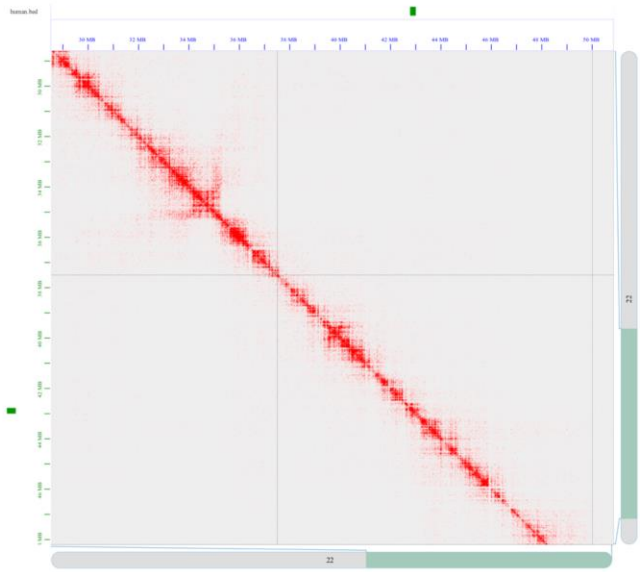


Figure S136: Hi-C figure generated using Juicebox of human liver data (SRR1501151) where red dots show the genomic loci interaction between adjacent regions and gap shows there is no interaction between the nearby region. We loaded the bed file on the top and left sides of the Hi-C plot, where the green color indicates the gene of interest (*COA1* duplicate) and adjacent syntenic regions. chromosome or scaffold is mentioned on the bottom and right sides of the plot.

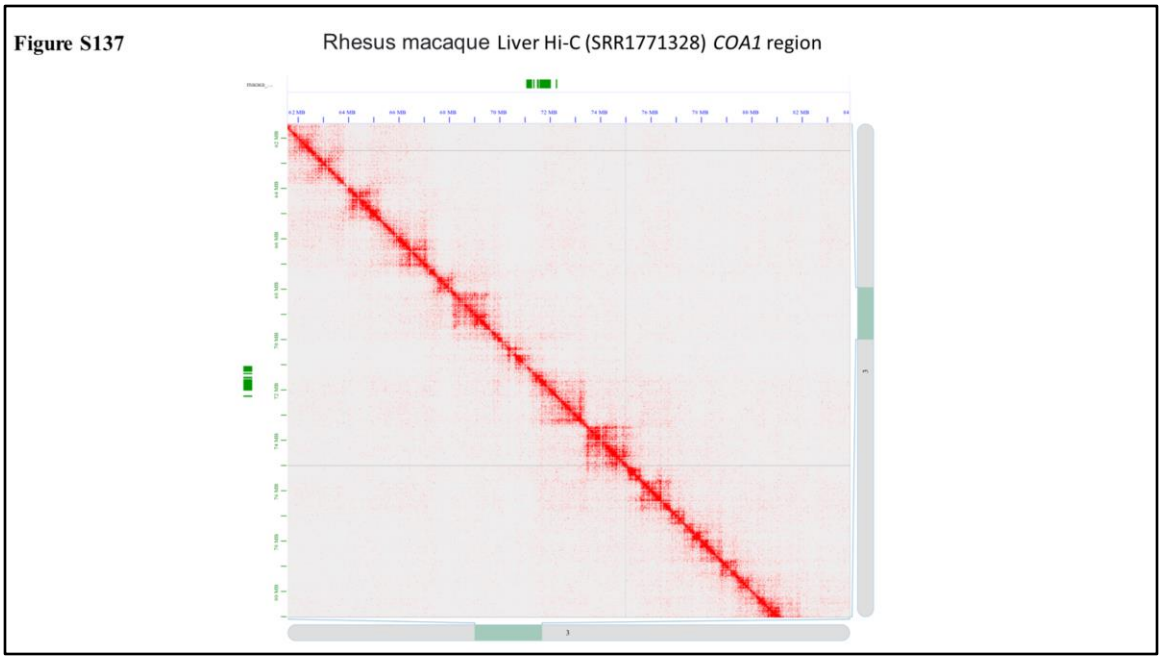


Figure S137: Hi-C figure generated using Juicebox of Rhesus macaque liver data (SRR1771328) where red dots show the genomic loci interaction between adjacent regions and gap shows there is no interaction between the nearby region. We loaded the bed file on the top and left sides of the Hi-C plot, where each line of the green color indicates the gene of interest (*COA1)* and adjacent syntenic regions. chromosome or scaffold is mentioned on the bottom and right sides of the plot.

Figure S138

Rhesus macaque Liver Hi-C (SRR1771328) duplicate *COA1* region

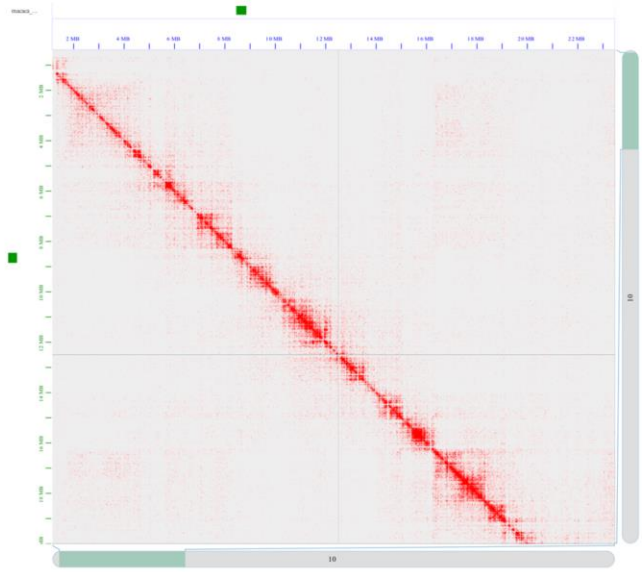


Figure S138: Hi-C figure generated using Juicebox of Rhesus macaque liver data (SRR1771328) where red dots show the genomic loci interaction between adjacent regions and gap shows there is no interaction between the nearby region. We loaded the bed file on the top and left sides of the Hi-C plot, where the green color indicates the gene of interest (*COA1* duplicate*)* and adjacent syntenic regions. chromosome or scaffold is mentioned on the bottom and right sides of the plot.

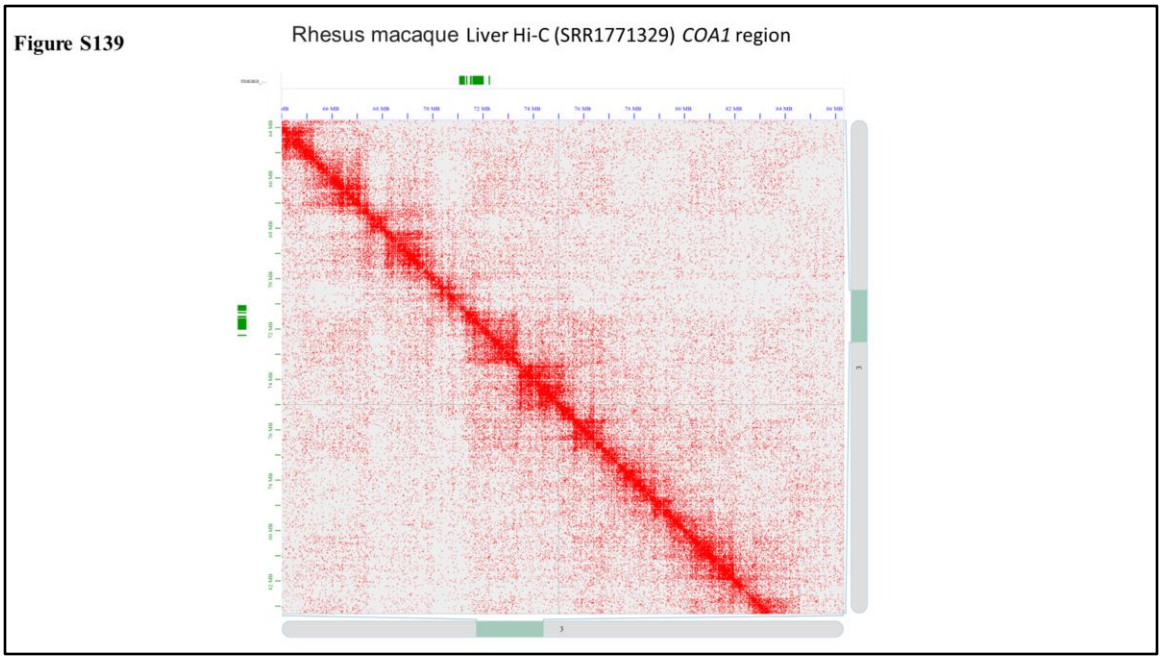


Figure S139: Hi-C figure generated using Juicebox of Rhesus macaque liver data (SRR1771329) where red dots show the genomic loci interaction between adjacent regions and gap shows there is no interaction between the nearby region. We loaded the bed file on the top and left sides of the Hi-C plot, where the each line of green color indicates the gene of interest (*COA1)* and adjacent syntenic regions. chromosome or scaffold is mentioned on the bottom and right sides of the plot.

Figure S140

Rhesus macaque Liver Hi-C (SRR1771329) duplicate *COA1* region

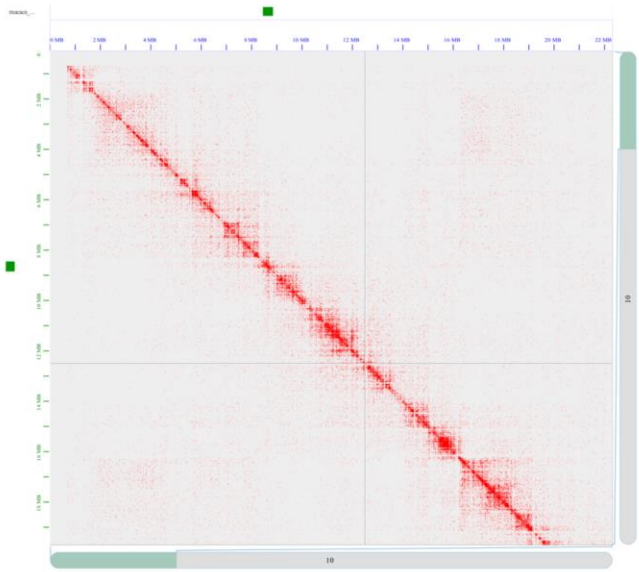


Figure S140: Hi-C figure generated using Juicebox of Rhesus macaque liver data (SRR1771329) where red dots show the genomic loci interaction between adjacent regions and gap shows there is no interaction between the nearby region. We loaded the bed file on the top and left sides of the Hi-C plot, where the green color indicates the gene of interest (*COA1* duplicate*)* and adjacent syntenic regions. chromosome or scaffold is mentioned on the bottom and right sides of the plot.

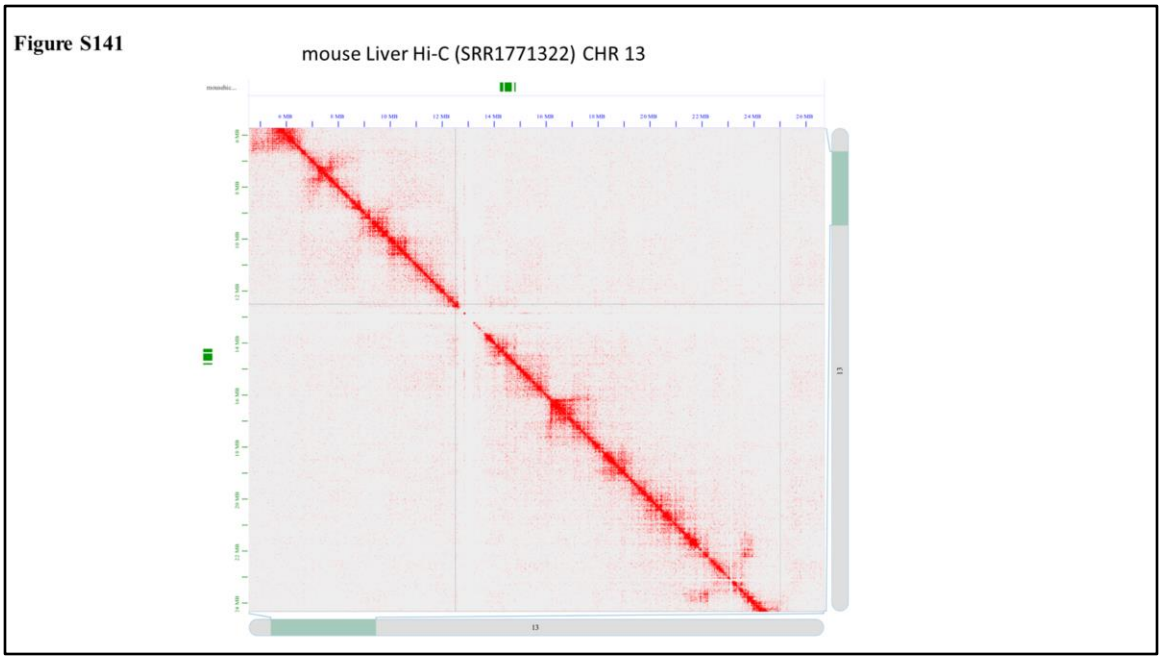


Figure S141: Hi-C figure generated using Juicebox of mouse liver data (SRR1771322) where red dots show the genomic loci interaction between adjacent regions and gap shows there is no interaction between the nearby region. We loaded the bed file on the top and left sides of the Hi-C plot, where each line of the green color indicates the gene of interest (*HECW1)* and adjacent syntenic regions. chromosome or scaffold is mentioned on the bottom and right sides of the plot.

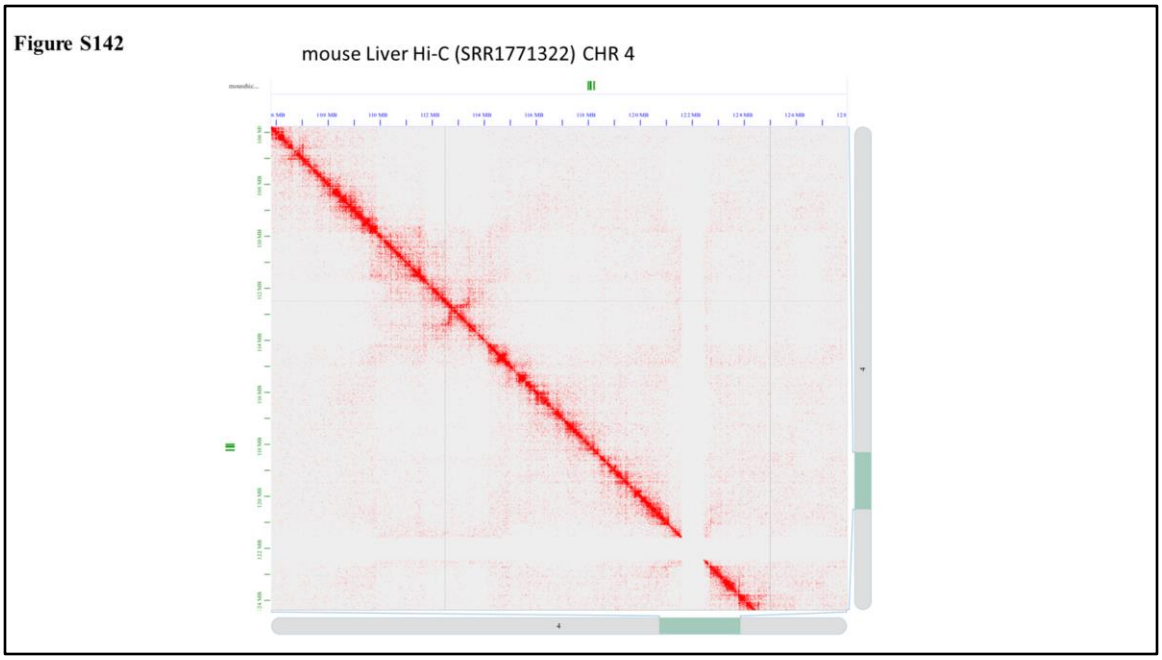


Figure S142: Hi-C figure generated using Juicebox of mouse liver data (SRR1771322) where red dots show the genomic loci interaction between adjacent regions and gap shows there is no interaction between the nearby region. We loaded the bed file on the top and left sides of the Hi-C plot, where each line of the green color indicates the gene of interest (*COA1* remnants*)* and adjacent syntenic regions. chromosome or scaffold is mentioned on the bottom and right sides of the plot.

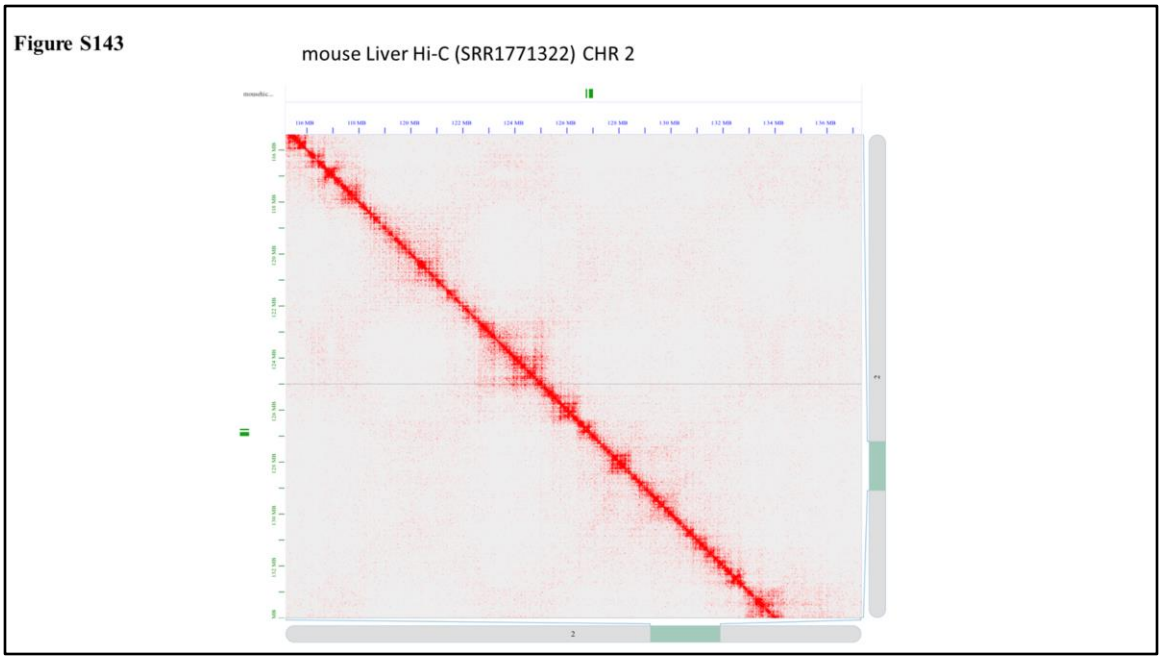


Figure S143: Hi-C figure generated using Juicebox of mouse liver data (SRR1771322) where red dots show the genomic loci interaction between adjacent regions and gap shows there is no interaction between the nearby region. We loaded the bed file on the top and left sides of the Hi-C plot, where each line of the green color indicates the gene of interest (*BLVRA)* and adjacent syntenic regions. chromosome or scaffold is mentioned on the bottom and right sides of the plot.

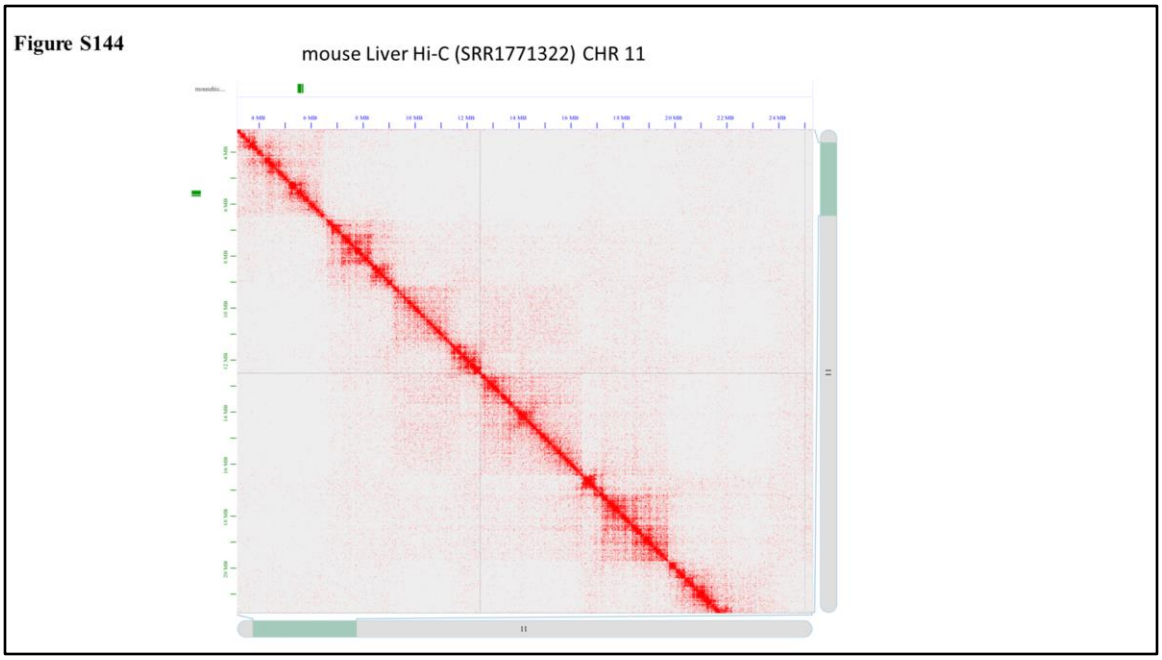


Figure S144: Hi-C figure generated using Juicebox of mouse liver data (SRR1771322) where red dots show the genomic loci interaction between adjacent regions and gap shows there is no interaction between the nearby region. We loaded the bed file on the top and left sides of the Hi-C plot, where each line of the green color indicates the gene of interest (*MRPS24)* and adjacent syntenic regions. chromosome or scaffold is mentioned on the bottom and right sides of the plot.

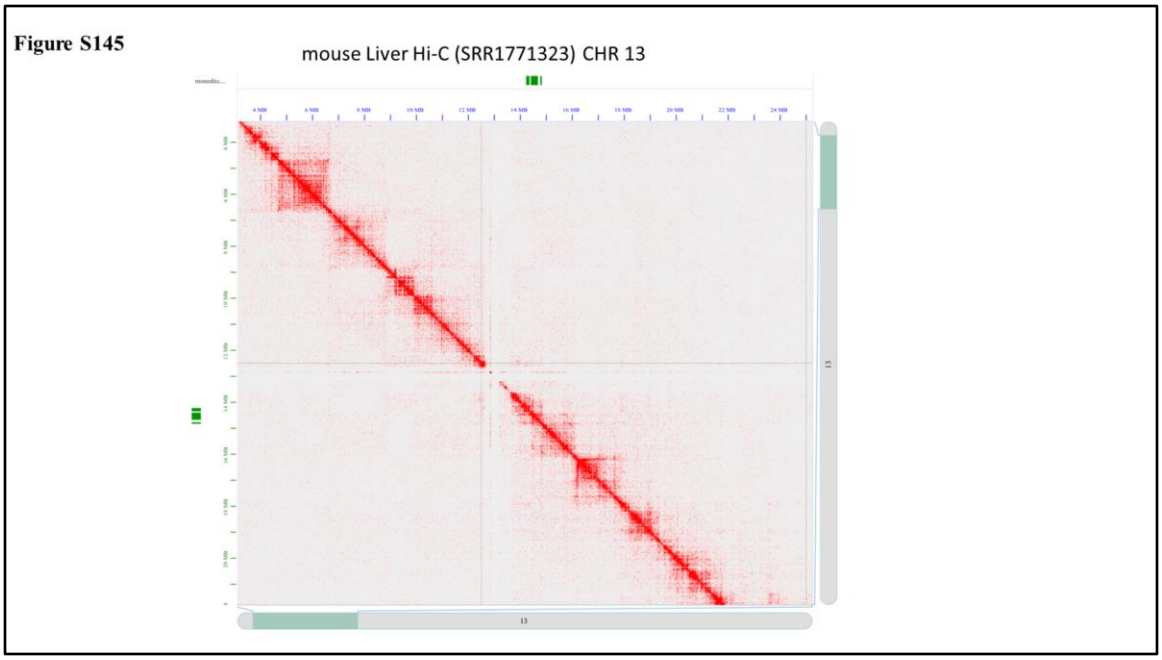


Figure S145: Hi-C figure generated using Juicebox of mouse liver data (SRR1771323) where red dots show the genomic loci interaction between adjacent regions and gap shows there is no interaction between the nearby region. We loaded the bed file on the top and left sides of the Hi-C plot, where each line of the green color indicates the gene of interest (*HECW1)* and adjacent syntenic regions. chromosome or scaffold is mentioned on the bottom and right sides of the plot.

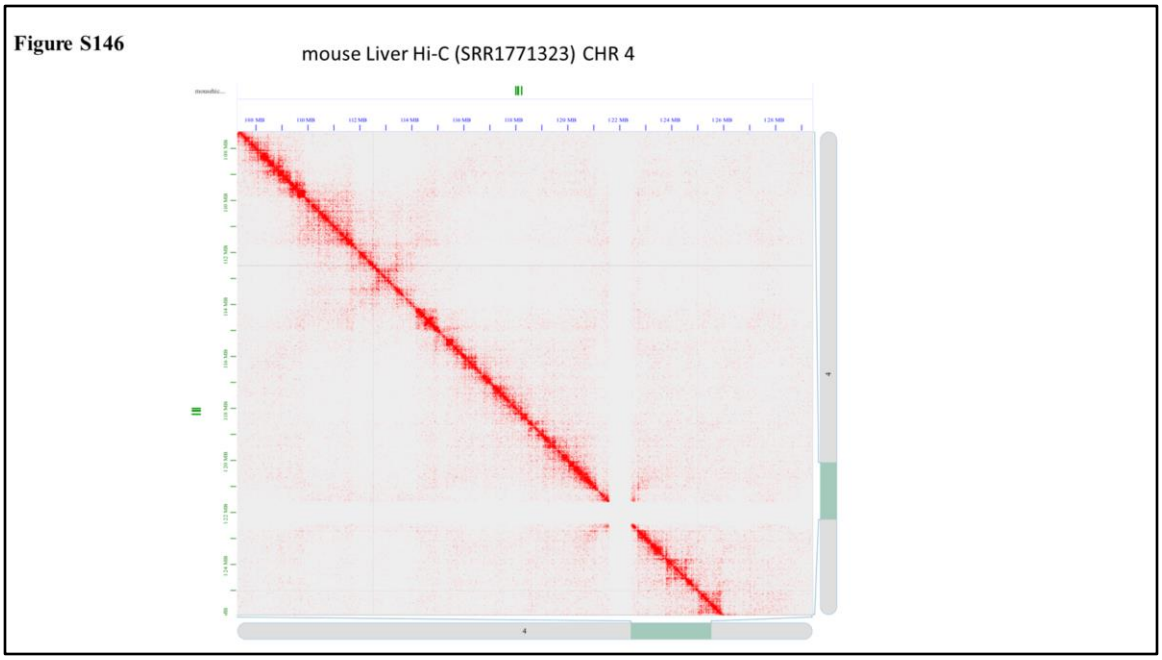


Figure S146: Hi-C figure generated using Juicebox of mouse liver data (SRR1771323) where red dots show the genomic loci interaction between adjacent regions and gap shows there is no interaction between the nearby region. We loaded the bed file on the top and left sides of the Hi-C plot, where each line of the green color indicates the gene of interest (*COA1* remnants*)* and adjacent syntenic regions. chromosome or scaffold is mentioned on the bottom and right sides of the plot.

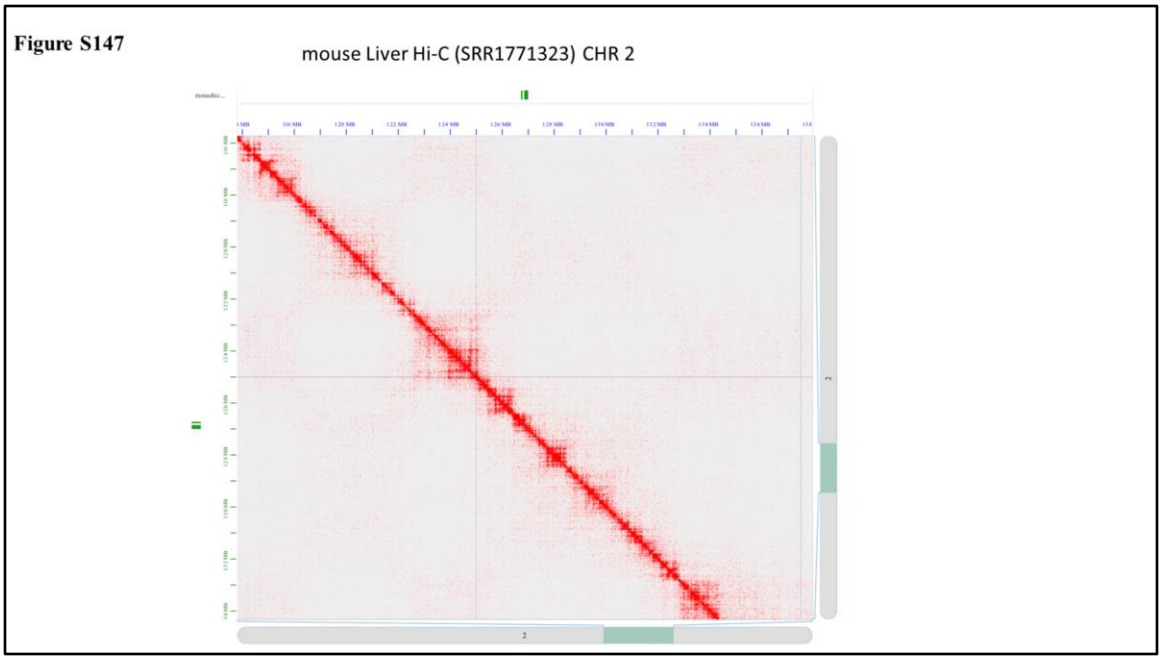


Figure S147: Hi-C figure generated using Juicebox of mouse liver data (SRR1771323) where red dots show the genomic loci interaction between adjacent regions and gap shows there is no interaction between the nearby region. We loaded the bed file on the top and left sides of the Hi-C plot, where each line of the green color indicates the gene of interest (*BLVRA)* and adjacent syntenic regions. chromosome or scaffold is mentioned on the bottom and right sides of the plot.

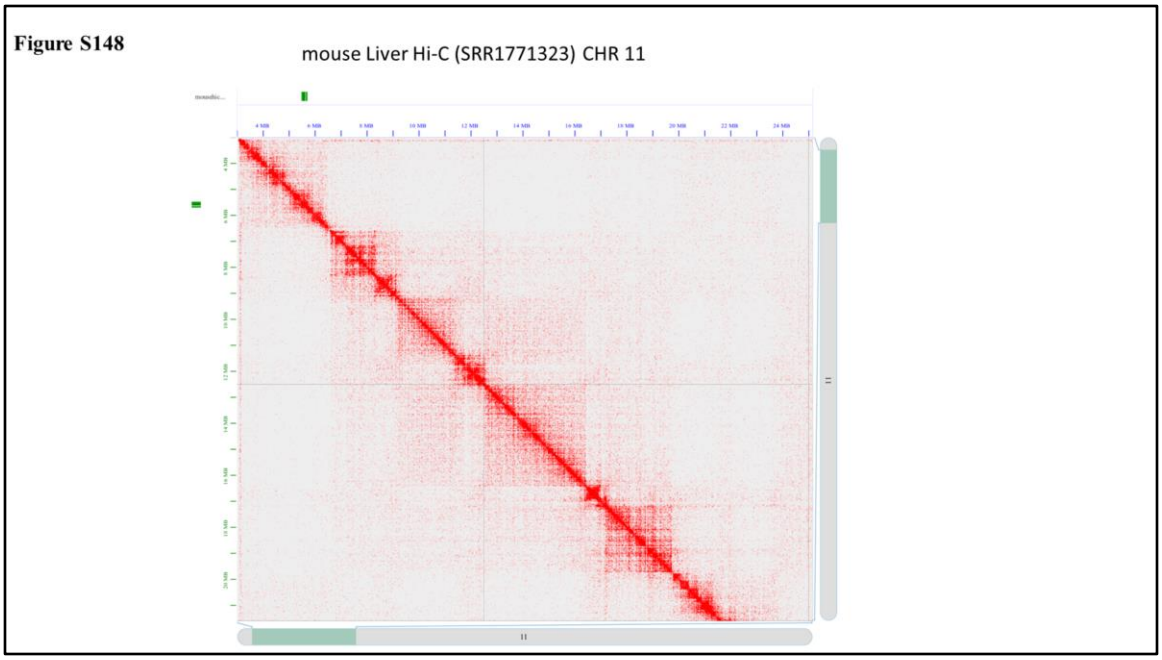


Figure S148: Hi-C figure generated using Juicebox of mouse liver data (SRR1771323) where red dots show the genomic loci interaction between adjacent regions and gap shows there is no interaction between the nearby region. We loaded the bed file on the top and left sides of the Hi-C plot, where each line of the green color indicates the gene of interest (*MRPS24)* and adjacent syntenic regions. chromosome or scaffold is mentioned on the bottom and right sides of the plot.

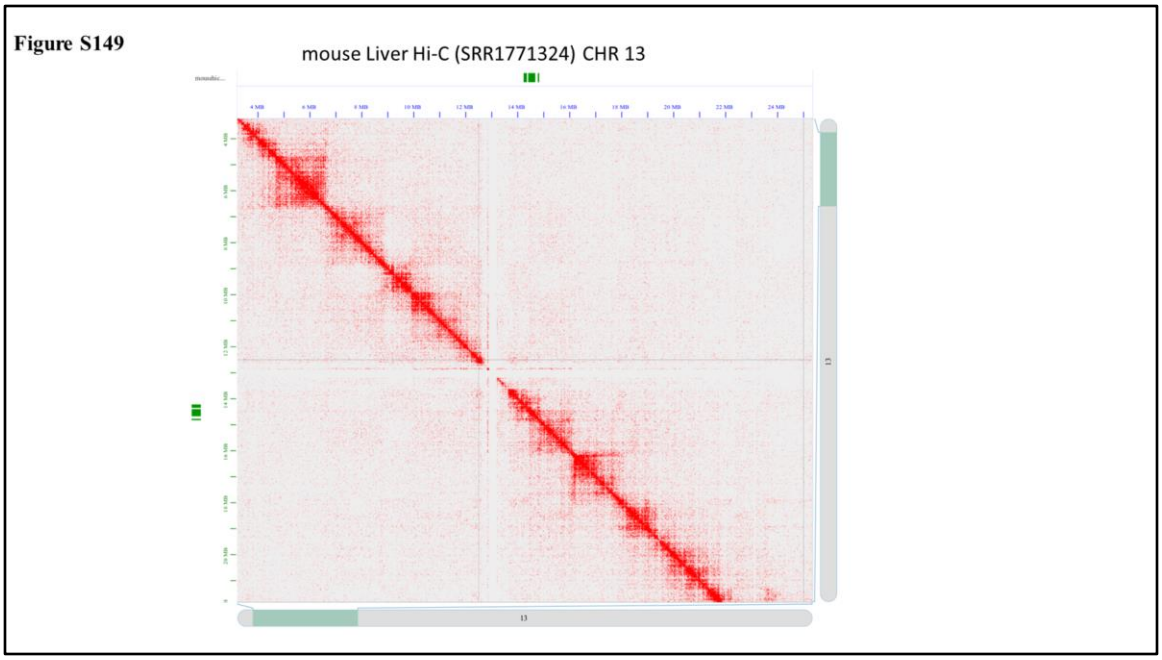


Figure S149: Hi-C figure generated using Juicebox of mouse liver data (SRR1771324) where red dots show the genomic loci interaction between adjacent regions and gap shows there is no interaction between the nearby region. We loaded the bed file on the top and left sides of the Hi-C plot, where each line of the green color indicates the gene of interest (*HECW1)* and adjacent syntenic regions. chromosome or scaffold is mentioned on the bottom and right sides of the plot.

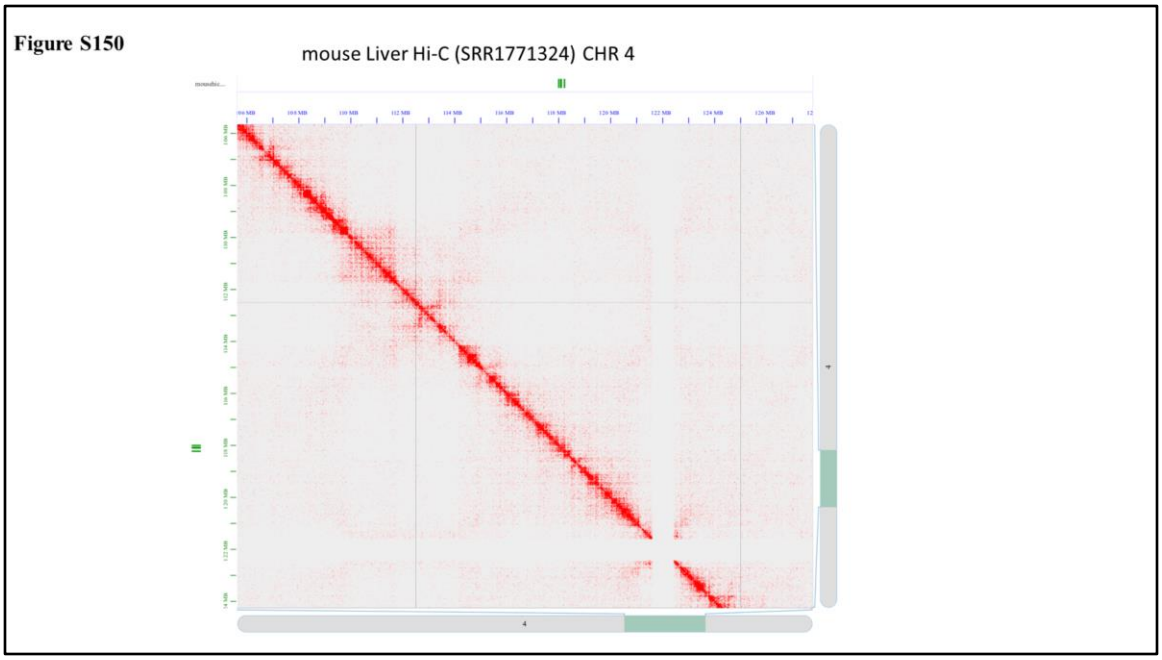


Figure S150: Hi-C figure generated using Juicebox of mouse liver data (SRR1771324) where red dots show the genomic loci interaction between adjacent regions and gap shows there is no interaction between the nearby region. We loaded the bed file on the top and left sides of the Hi-C plot, where each line of the green color indicates the gene of interest (*COA1* remnants*)* and adjacent syntenic regions. chromosome or scaffold is mentioned on the bottom and right sides of the plot.

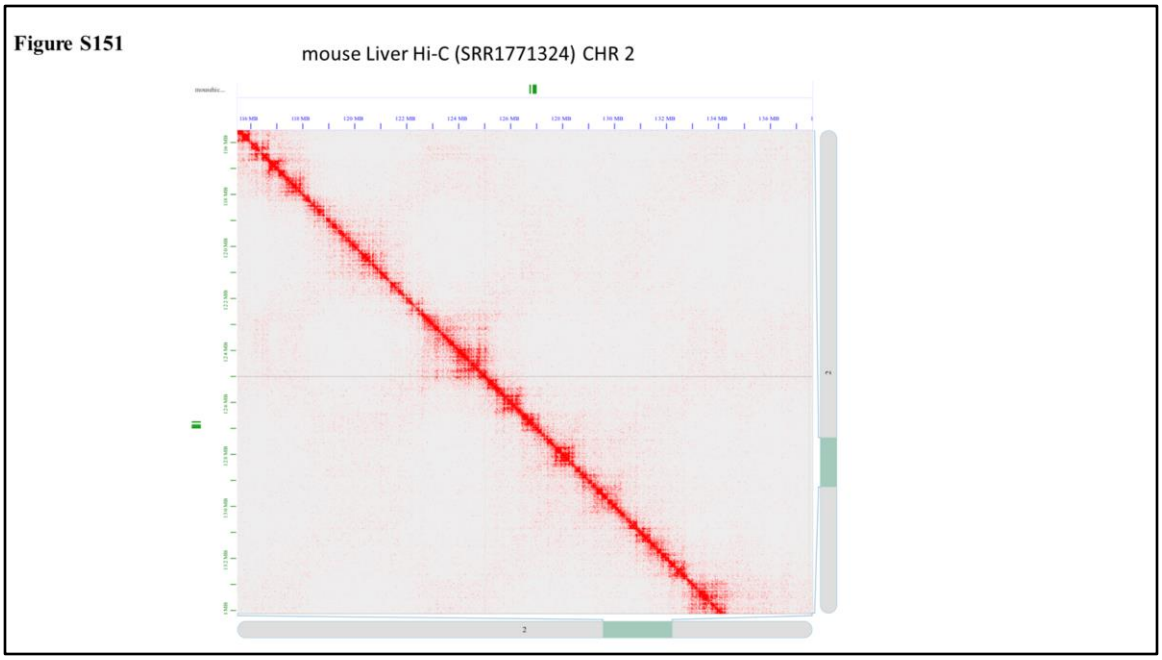


Figure S151: Hi-C figure generated using Juicebox of mouse liver data (SRR1771324) where red dots show the genomic loci interaction between adjacent regions and gap shows there is no interaction between the nearby region. We loaded the bed file on the top and left sides of the Hi-C plot, where each line of the green color indicates the gene of interest (*BLVRA)* and adjacent syntenic regions. chromosome or scaffold is mentioned on the bottom and right sides of the plot.

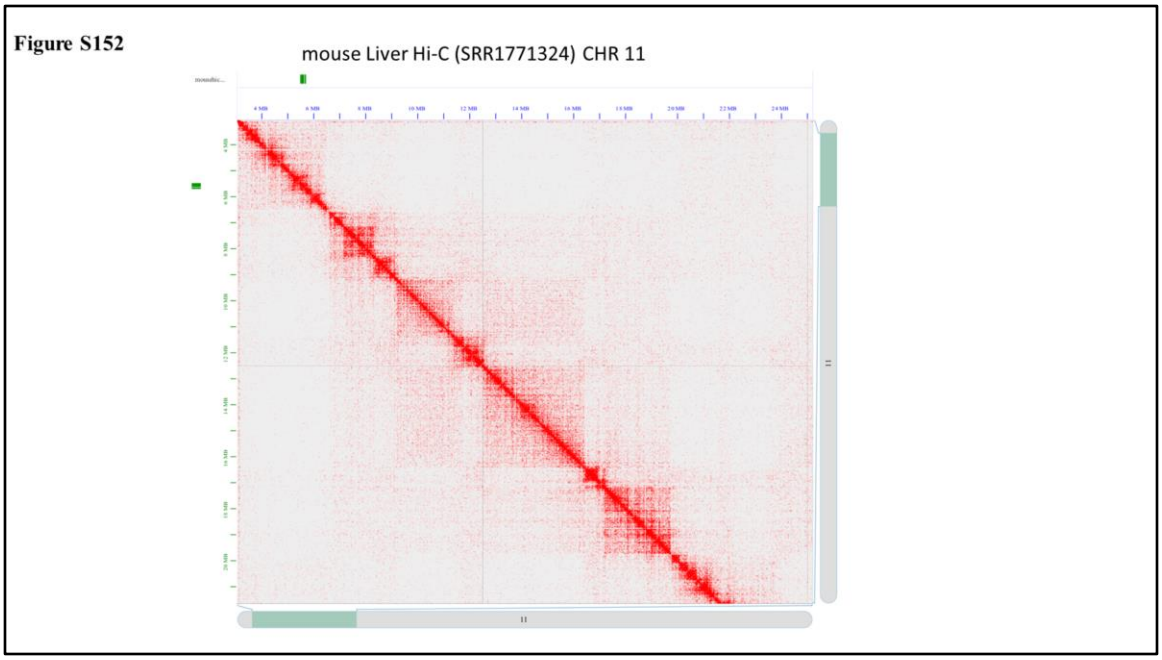


Figure S152: Hi-C figure generated using Juicebox of mouse liver data (SRR1771324) where red dots show the genomic loci interaction between adjacent regions and gap shows there is no interaction between the nearby region. We loaded the bed file on the top and left sides of the Hi-C plot, where each line of the green color indicates the gene of interest (*MRPS24)* and adjacent syntenic regions. chromosome or scaffold is mentioned on the bottom and right sides of the plot.

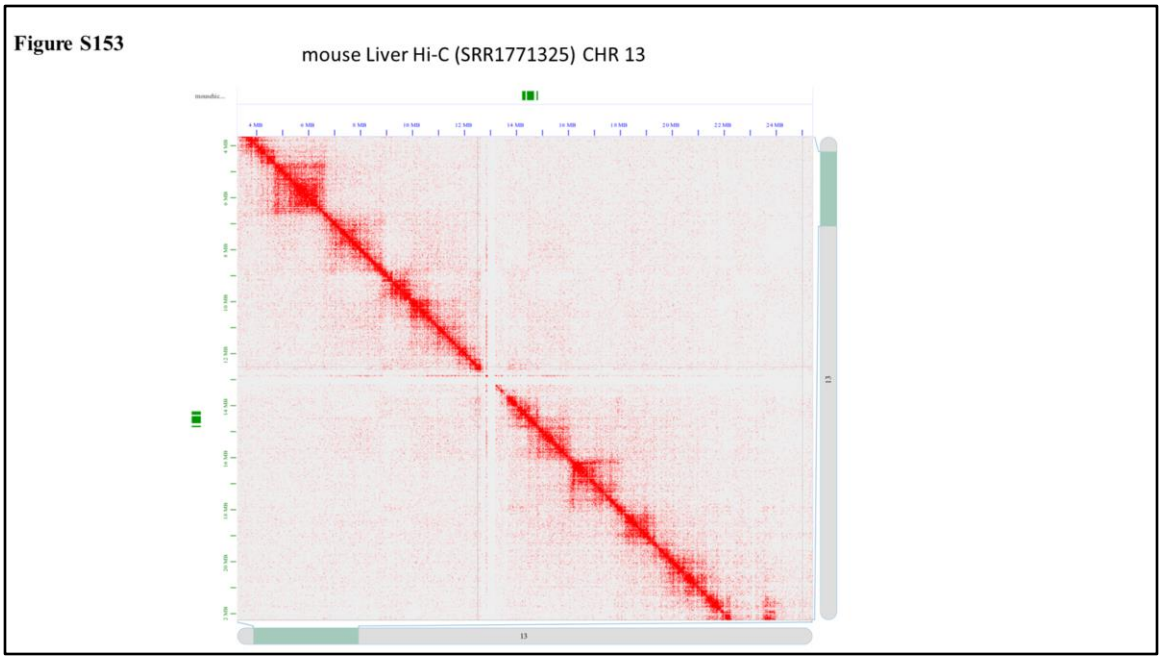


Figure S153: Hi-C figure generated using Juicebox of mouse liver data (SRR1771325) where red dots show the genomic loci interaction between adjacent regions and gap shows there is no interaction between the nearby region. We loaded the bed file on the top and left sides of the Hi-C plot, where each line of the green color indicates the gene of interest (*HECW1)* and adjacent syntenic regions. chromosome or scaffold is mentioned on the bottom and right sides of the plot.

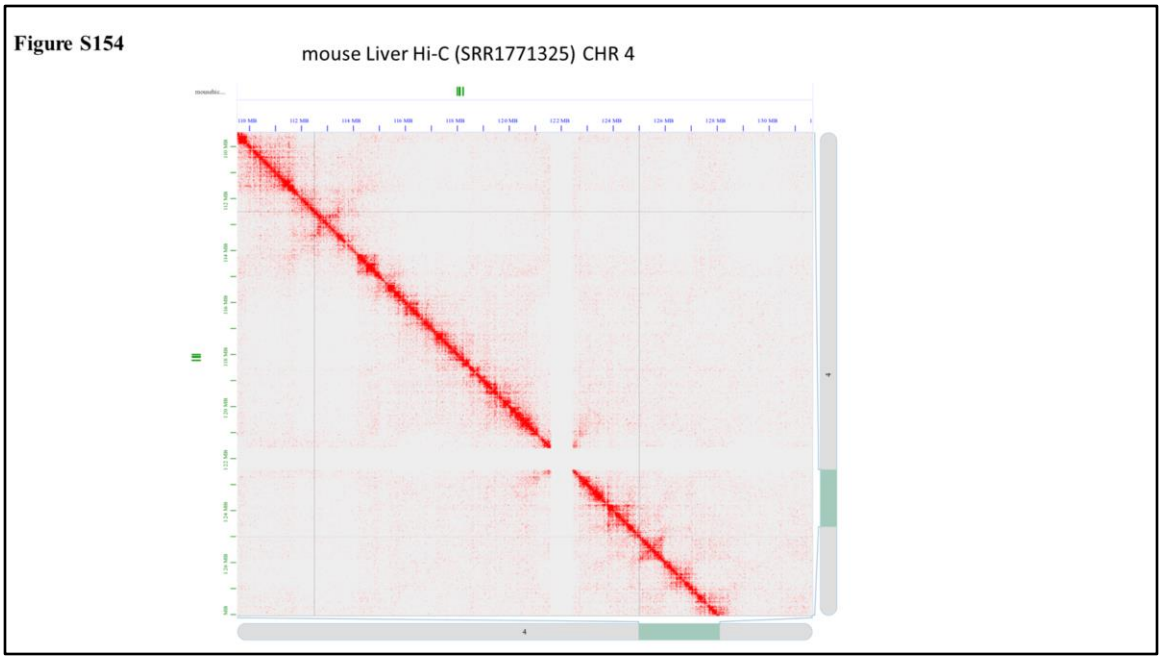


Figure S154: Hi-C figure generated using Juicebox of mouse liver data (SRR1771325) where red dots show the genomic loci interaction between adjacent regions and gap shows there is no interaction between the nearby region. We loaded the bed file on the top and left sides of the Hi-C plot, where each line of the green color indicates the gene of interest (*COA1* remnants*)* and adjacent syntenic regions. chromosome or scaffold is mentioned on the bottom and right sides of the plot.

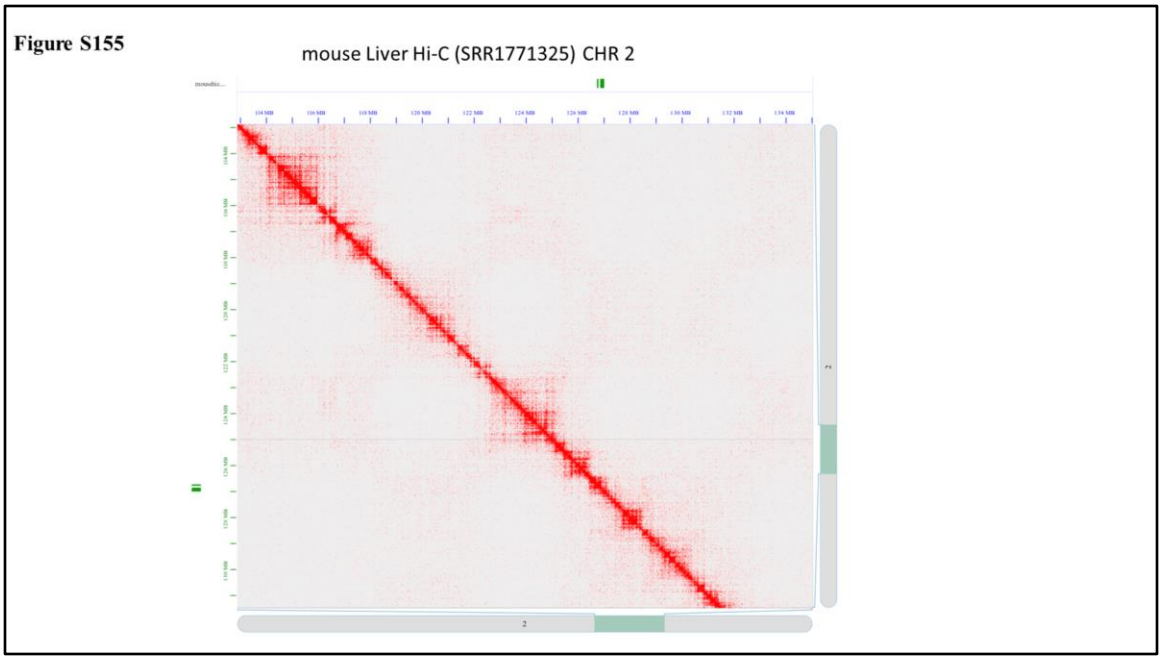


Figure S155: Hi-C figure generated using Juicebox of mouse liver data (SRR1771325) where red dots show the genomic loci interaction between adjacent regions and gap shows there is no interaction between the nearby region. We loaded the bed file on the top and left sides of the Hi-C plot, where each line of the green color indicates the gene of interest (*BLVRA)* and adjacent syntenic regions. chromosome or scaffold is mentioned on the bottom and right sides of the plot.

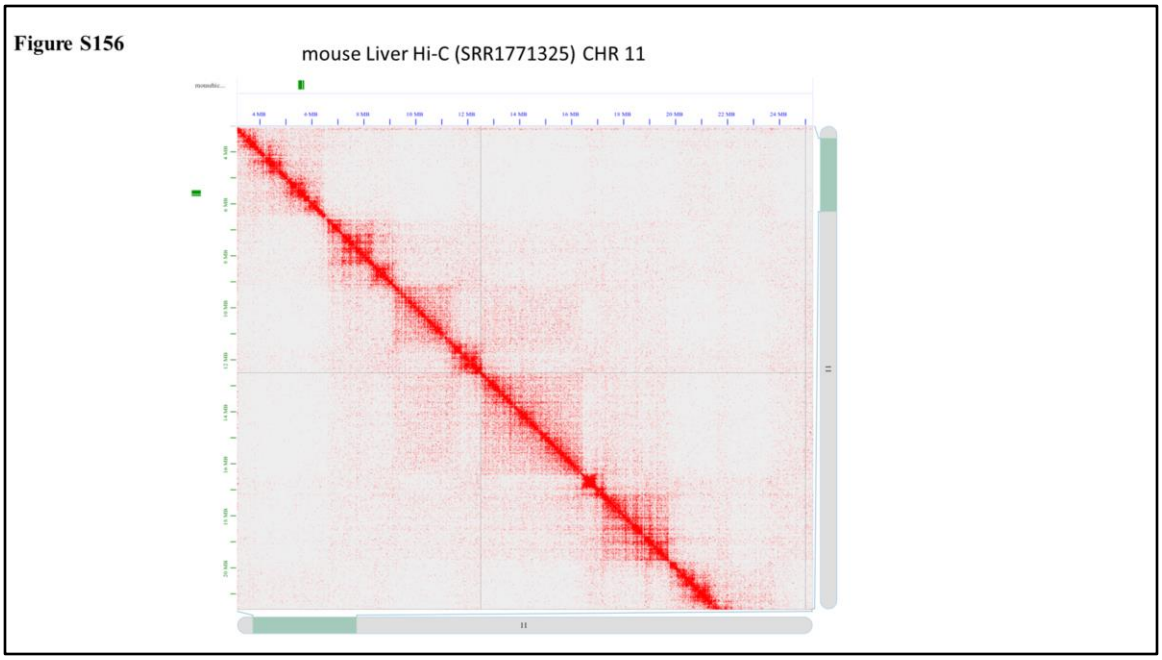


Figure S156: Hi-C figure generated using Juicebox of mouse liver data (SRR1771325) where red dots show the genomic loci interaction between adjacent regions and gap shows there is no interaction between the nearby region. We loaded the bed file on the top and left sides of the Hi-C plot, where each line of the green color indicates the gene of interest (*MRPS24)* and adjacent syntenic regions. chromosome or scaffold is mentioned on the bottom and right sides of the plot.

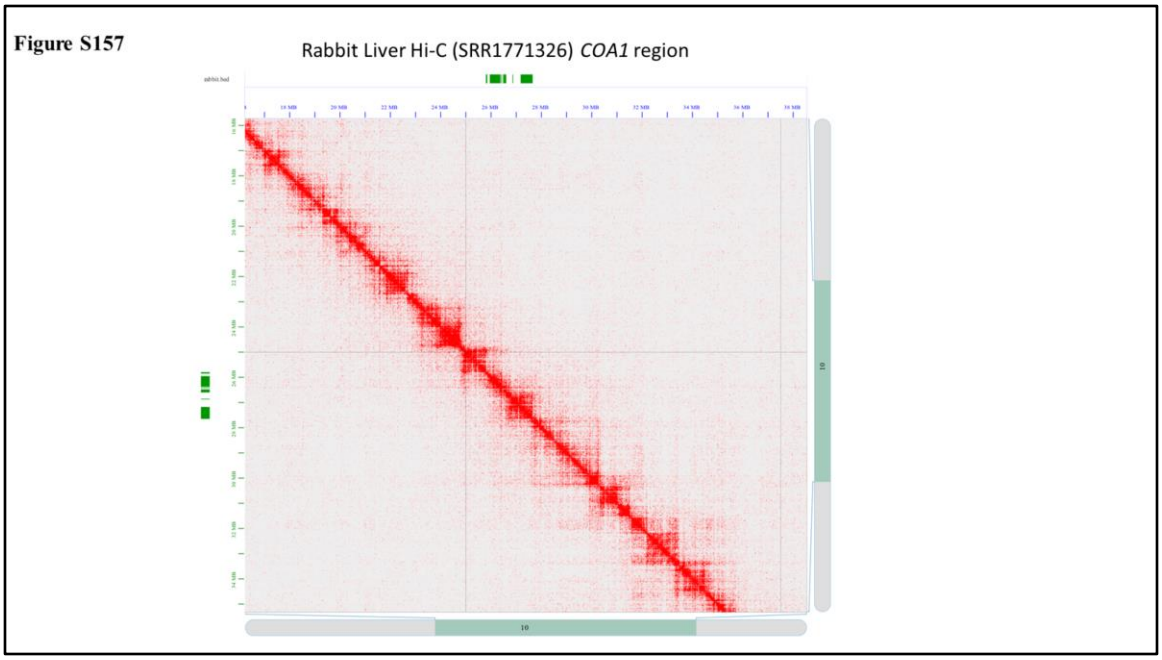


Figure S157: Hi-C figure generated using Juicebox of Rabbit liver data (SRR1771326) where red dots show the genomic loci interaction between adjacent regions and gap shows there is no interaction between the nearby region. We loaded the bed file on the top and left sides of the Hi-C plot, where each line of the green color indicates the gene of interest (*COA1)* and adjacent syntenic regions. chromosome or scaffold is mentioned on the bottom and right sides of the plot.

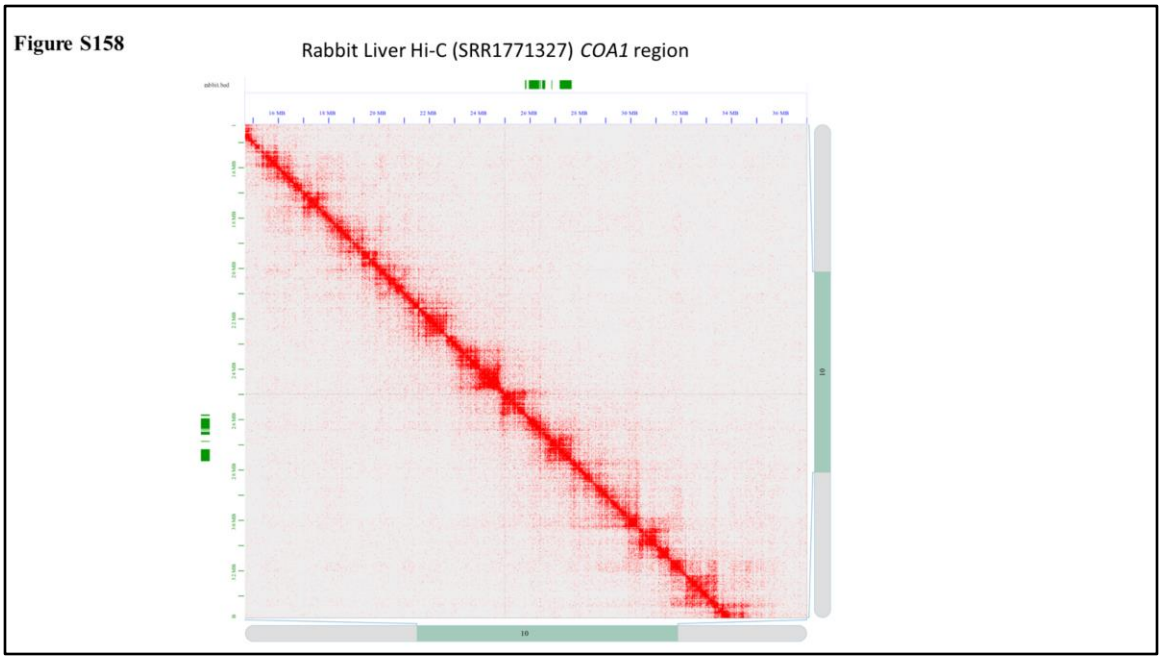


Figure S158: Hi-C figure generated using Juicebox of Rabbit liver data (SRR1771327) where red dots show the genomic loci interaction between adjacent regions and gap shows there is no interaction between the nearby region. We loaded the bed file on the top and left sides of the Hi-C plot, where each line of the green color indicates the gene of interest (*COA1)* and adjacent syntenic regions. chromosome or scaffold is mentioned on the bottom and right sides of the plot.

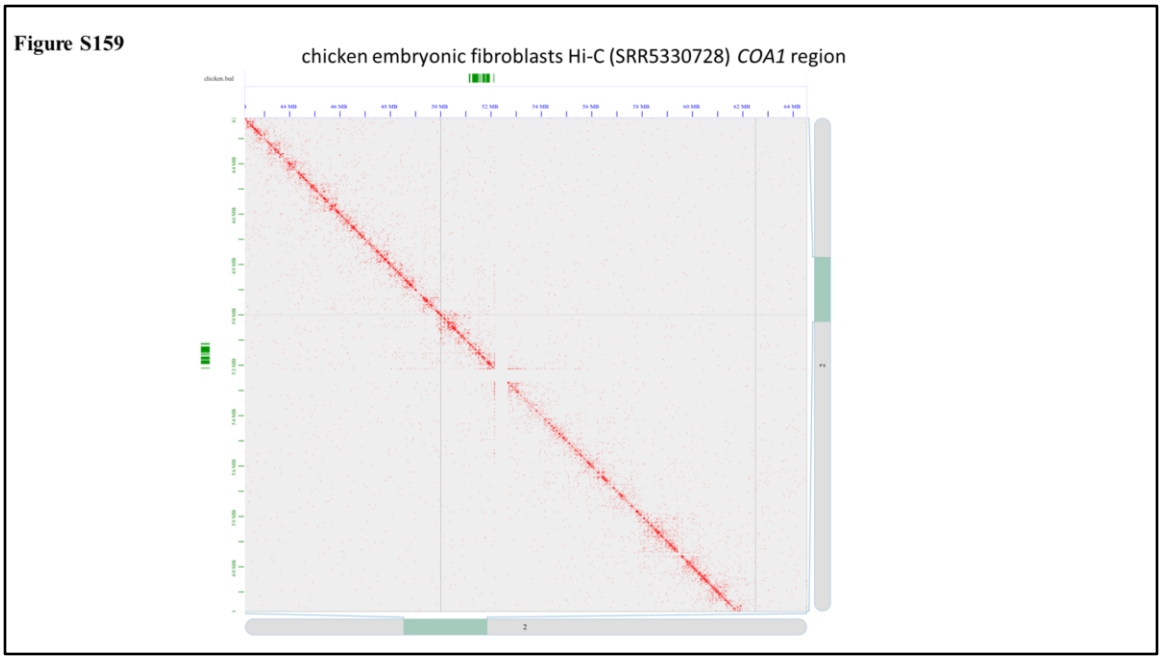


Figure S159: Hi-C figure generated using Juicebox of chicken embryonic fibroblasts (SRR5330728) where red dots show the genomic loci interaction between adjacent regions and gap shows there is no interaction between the nearby region. We loaded the bed file on the top and left sides of the Hi-C plot, where each line of the green color indicates the gene of interest (*COA1)* and adjacent syntenic regions. chromosome or scaffold is mentioned on the bottom and right sides of the plot.

Figure S160

chicken embryonic fibroblasts Hi-C (SRR5330729) *COA1* region

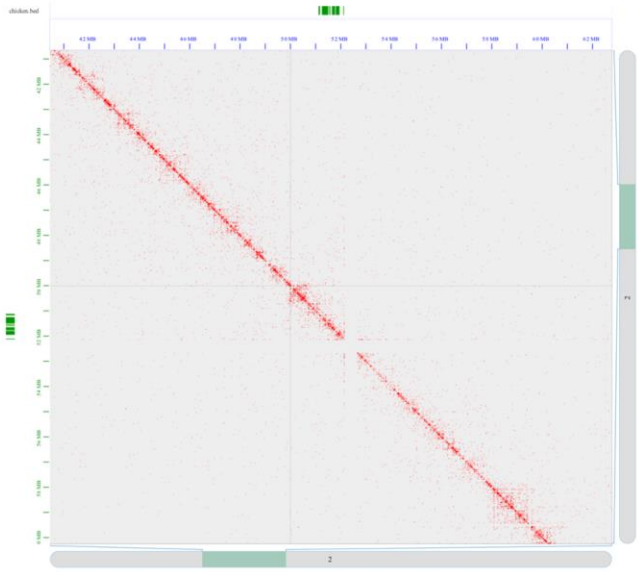


Figure S160: Hi-C figure generated using Juicebox of chicken embryonic fibroblasts (SRR5330729) where red dots show the genomic loci interaction between adjacent regions and gap shows there is no interaction between the nearby region. We loaded the bed file on the top and left sides of the Hi-C plot, where each line of the green color indicates the gene of interest (*COA1)* and adjacent syntenic regions. chromosome or scaffold is mentioned on the bottom and right sides of the plot.

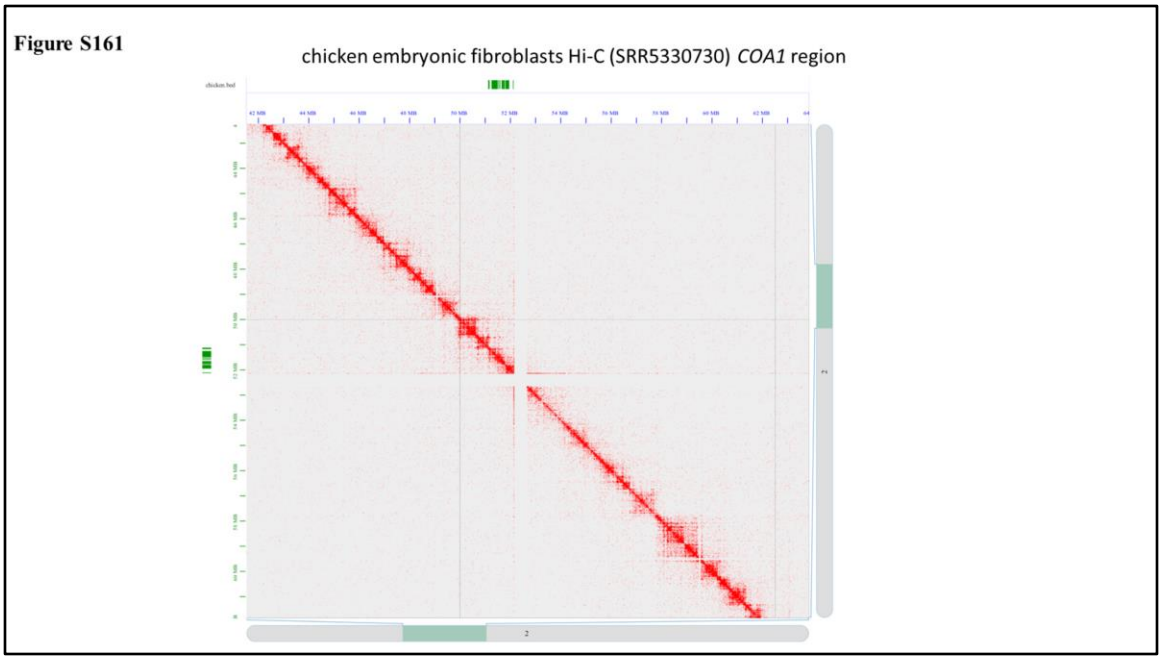


Figure S161: Hi-C figure generated using Juicebox of chicken embryonic fibroblasts (SRR5330730) where red dots show the genomic loci interaction between adjacent regions and gap shows there is no interaction between the nearby region. We loaded the bed file on the top and left sides of the Hi-C plot, where each line of the green color indicates the gene of interest (*COA1)* and adjacent syntenic regions. chromosome or scaffold is mentioned on the bottom and right sides of the plot.

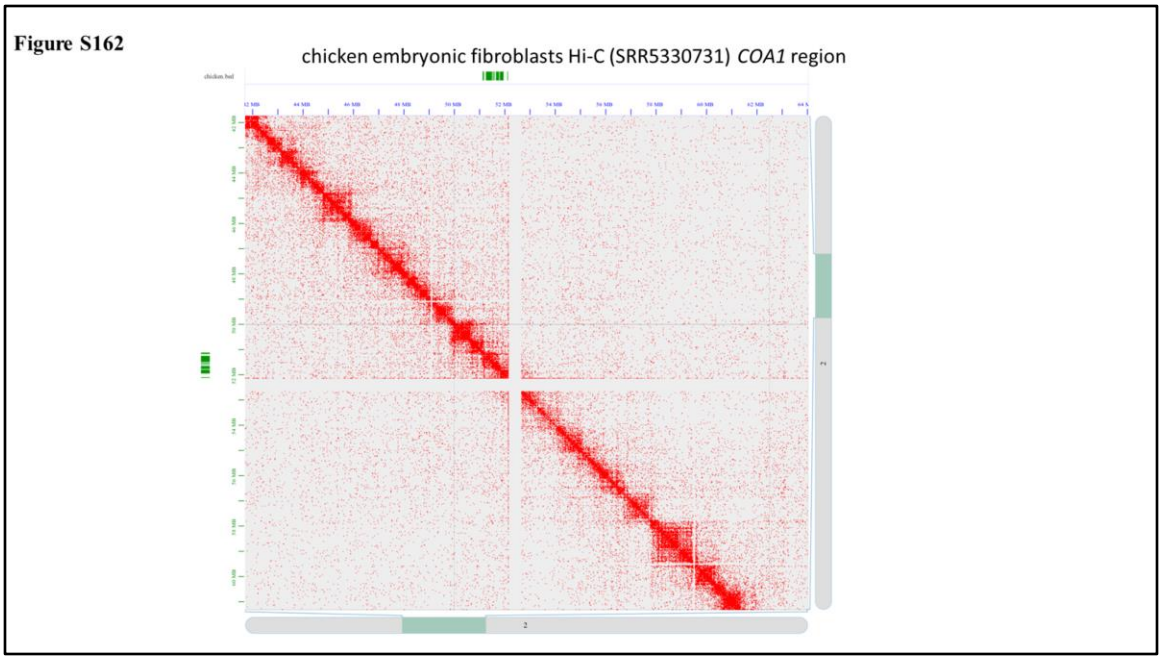


Figure S162: Hi-C figure generated using Juicebox of chicken embryonic fibroblasts (SRR5330731) where red dots show the genomic loci interaction between adjacent regions and gap shows there is no interaction between the nearby region. We loaded the bed file on the top and left sides of the Hi-C plot, where each line of the green color indicates the gene of interest (*COA1)* and adjacent syntenic regions. chromosome or scaffold is mentioned on the bottom and right sides of the plot.

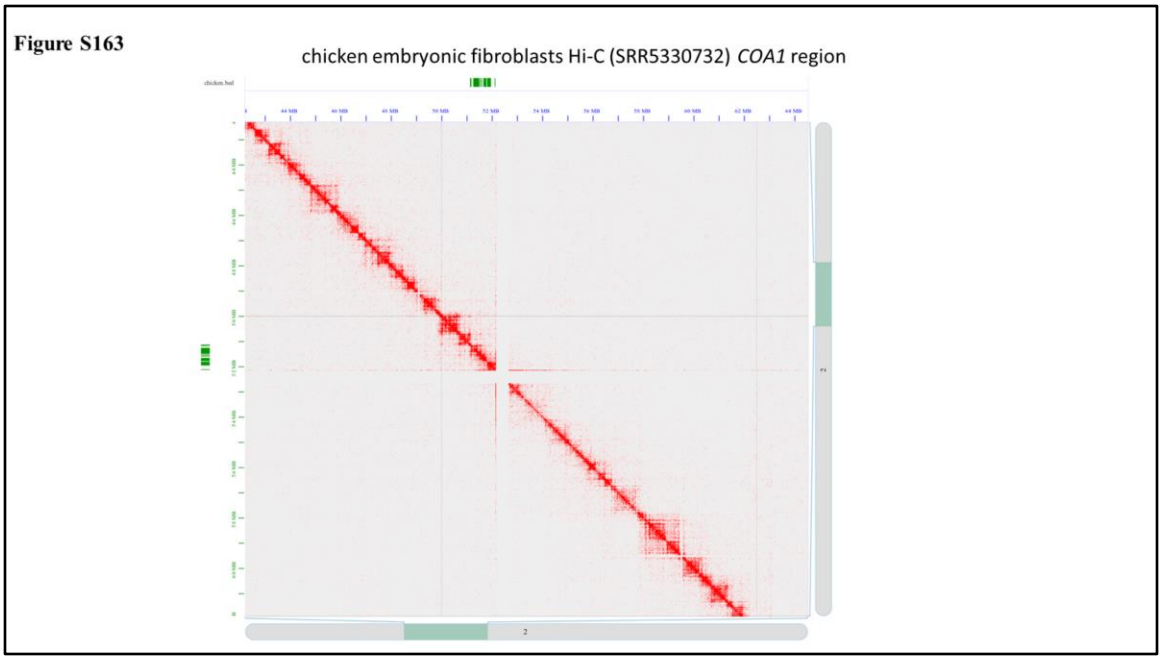


Figure S163: Hi-C figure generated using Juicebox of chicken embryonic fibroblasts (SRR5330732) where red dots show the genomic loci interaction between adjacent regions and gap shows there is no interaction between the nearby region. We loaded the bed file on the top and left sides of the Hi-C plot, where each line of the green color indicates the gene of interest (*COA1)* and adjacent syntenic regions. chromosome or scaffold is mentioned on the bottom and right sides of the plot.

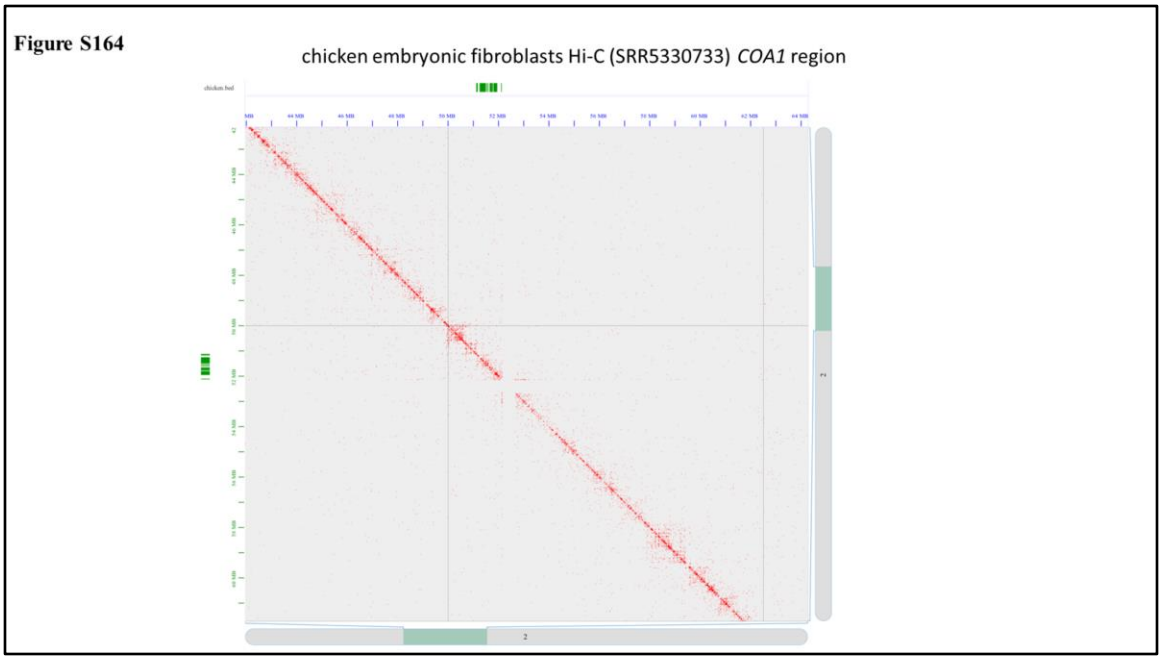


Figure S164: Hi-C figure generated using Juicebox of chicken embryonic fibroblasts (SRR5330733) where red dots show the genomic loci interaction between adjacent regions and gap shows there is no interaction between the nearby region. We loaded the bed file on the top and left sides of the Hi-C plot, where each line of the green color indicates the gene of interest (*COA1)* and adjacent syntenic regions. chromosome or scaffold is mentioned on the bottom and right sides of the plot.

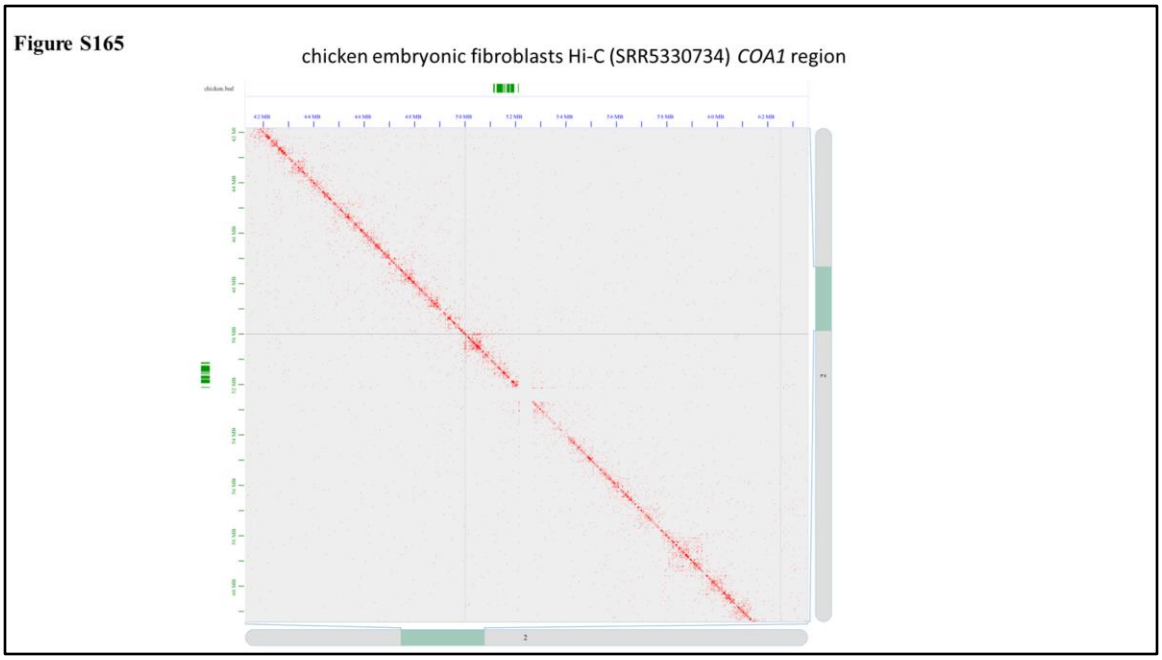


Figure S165: Hi-C figure generated using Juicebox of chicken embryonic fibroblasts (SRR5330734) where red dots show the genomic loci interaction between adjacent regions and gap shows there is no interaction between the nearby region. We loaded the bed file on the top and left sides of the Hi-C plot, where each line of the green color indicates the gene of interest (*COA1)* and adjacent syntenic regions. chromosome or scaffold is mentioned on the bottom and right sides of the plot.

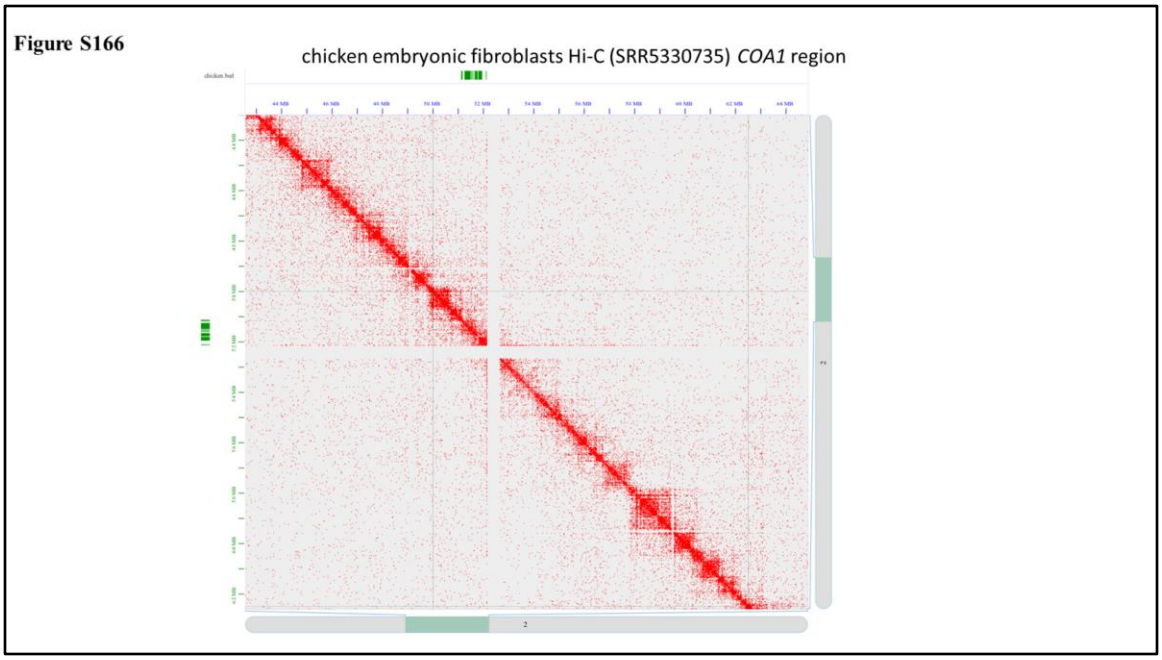


Figure S166: Hi-C figure generated using Juicebox of chicken embryonic fibroblasts (SRR5330735) where red dots show the genomic loci interaction between adjacent regions and gap shows there is no interaction between the nearby region. We loaded the bed file on the top and left sides of the Hi-C plot, where each line of the green color indicates the gene of interest (*COA1)* and adjacent syntenic regions. chromosome or scaffold is mentioned on the bottom and right sides of the plot.

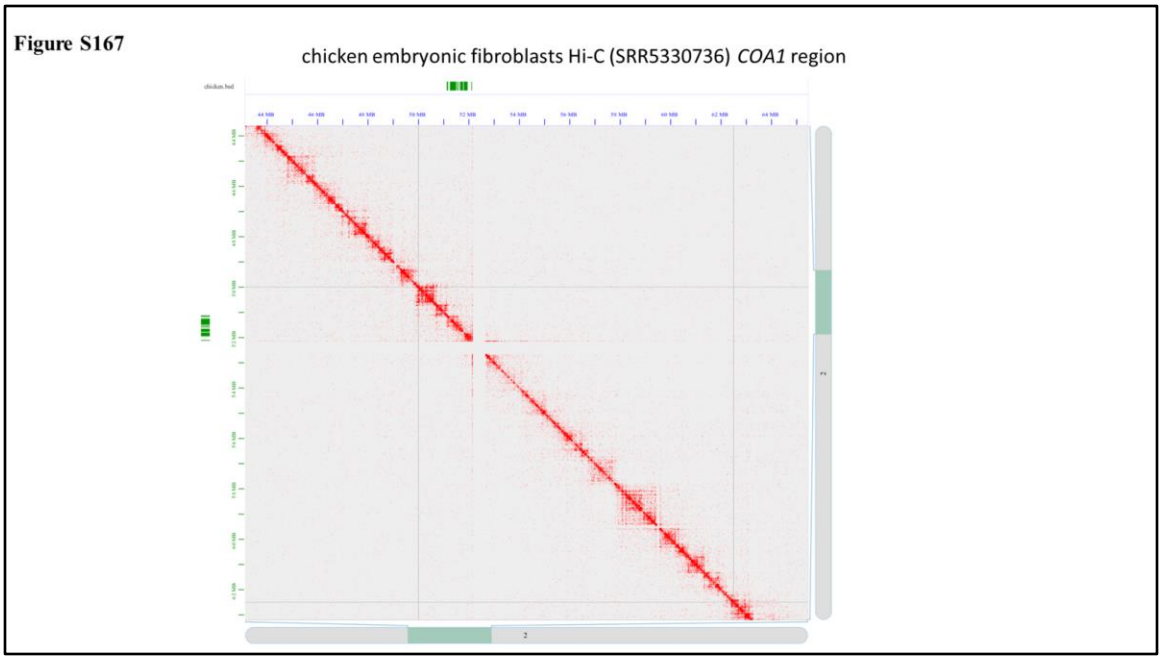


Figure S167: Hi-C figure generated using Juicebox of chicken embryonic fibroblasts (SRR5330736) where red dots show the genomic loci interaction between adjacent regions and gap shows there is no interaction between the nearby region. We loaded the bed file on the top and left sides of the Hi-C plot, where each line of the green color indicates the gene of interest (*COA1)* and adjacent syntenic regions. chromosome or scaffold is mentioned on the bottom and right sides of the plot.

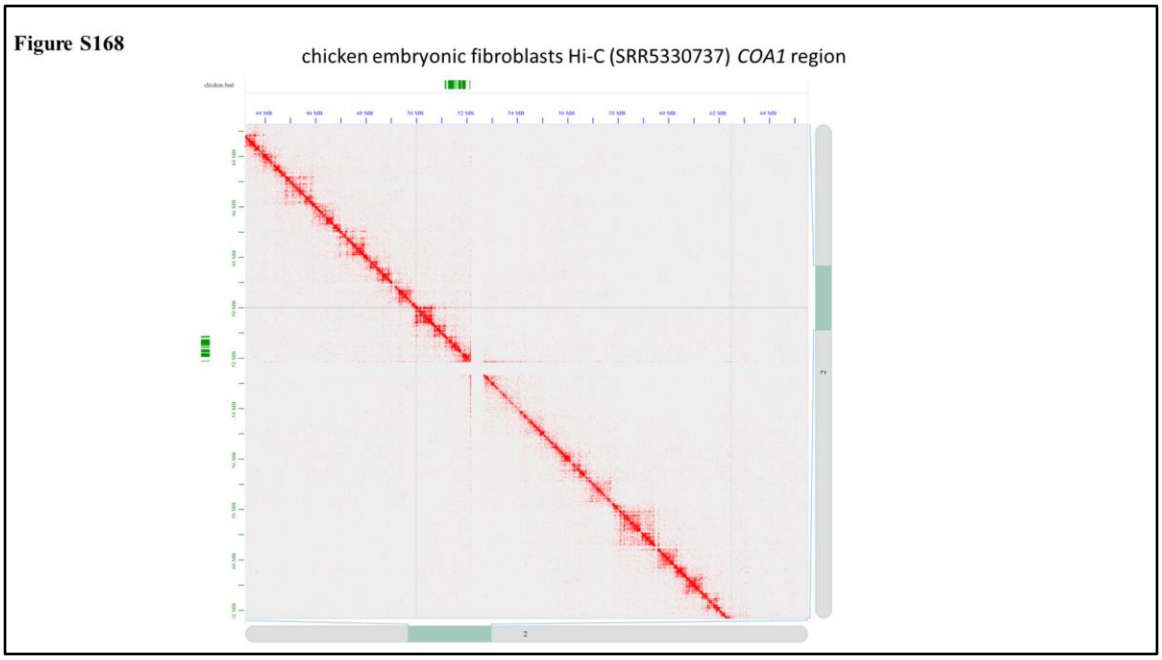


Figure S168: Hi-C figure generated using Juicebox of chicken embryonic fibroblasts (SRR5330737) where red dots show the genomic loci interaction between adjacent regions and gap shows there is no interaction between the nearby region. We loaded the bed file on the top and left sides of the Hi-C plot, where each line of the green color indicates the gene of interest (*COA1)* and adjacent syntenic regions. chromosome or scaffold is mentioned on the bottom and right sides of the plot.

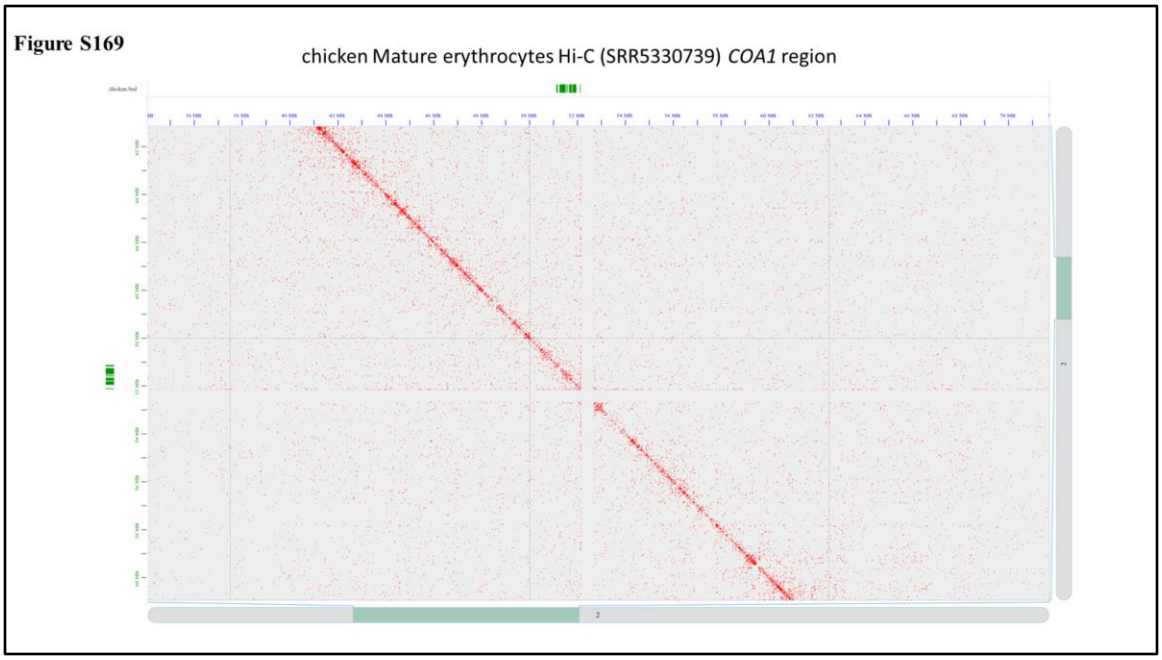


Figure S169: Hi-C figure generated using Juicebox of chicken Mature erythrocytes (SRR5330739) where red dots show the genomic loci interaction between adjacent regions and gap shows there is no interaction between the nearby region. We loaded the bed file on the top and left sides of the Hi-C plot, where each line of the green color indicates the gene of interest (*COA1)* and adjacent syntenic regions. chromosome or scaffold is mentioned on the bottom and right sides of the plot.

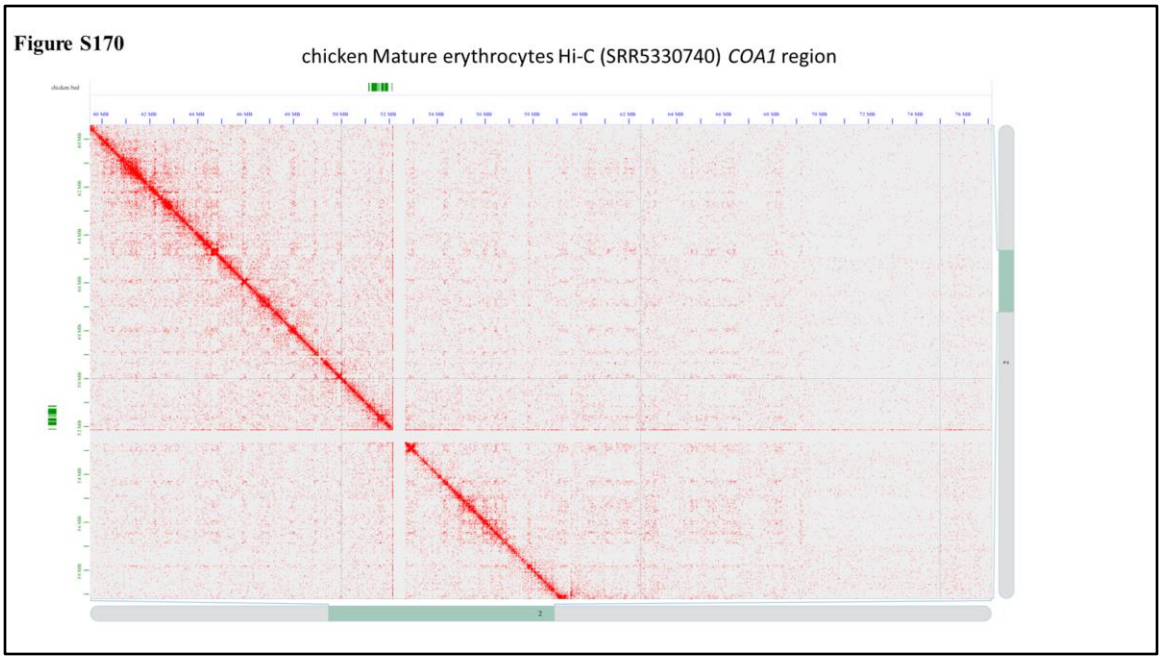


Figure S170: Hi-C figure generated using Juicebox of chicken Mature erythrocytes (SRR5330740) where red dots show the genomic loci interaction between adjacent regions and gap shows there is no interaction between the nearby region. We loaded the bed file on the top and left sides of the Hi-C plot, where each line of the green color indicates the gene of interest (*COA1)* and adjacent syntenic regions. chromosome or scaffold is mentioned on the bottom and right sides of the plot.

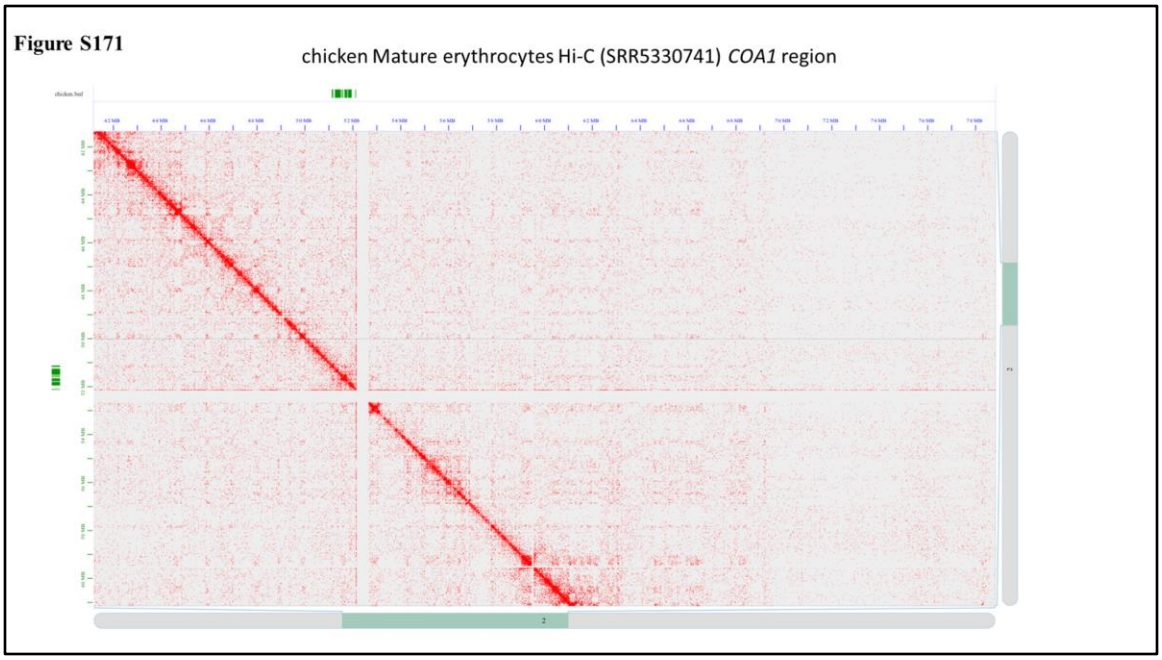


Figure S171: Hi-C figure generated using Juicebox of chicken Mature erythrocytes (SRR5330741) where red dots show the genomic loci interaction between adjacent regions and gap shows there is no interaction between the nearby region. We loaded the bed file on the top and left sides of the Hi-C plot, where each line of the green color indicates the gene of interest (*COA1)* and adjacent syntenic regions. chromosome or scaffold is mentioned on the bottom and right sides of the plot.

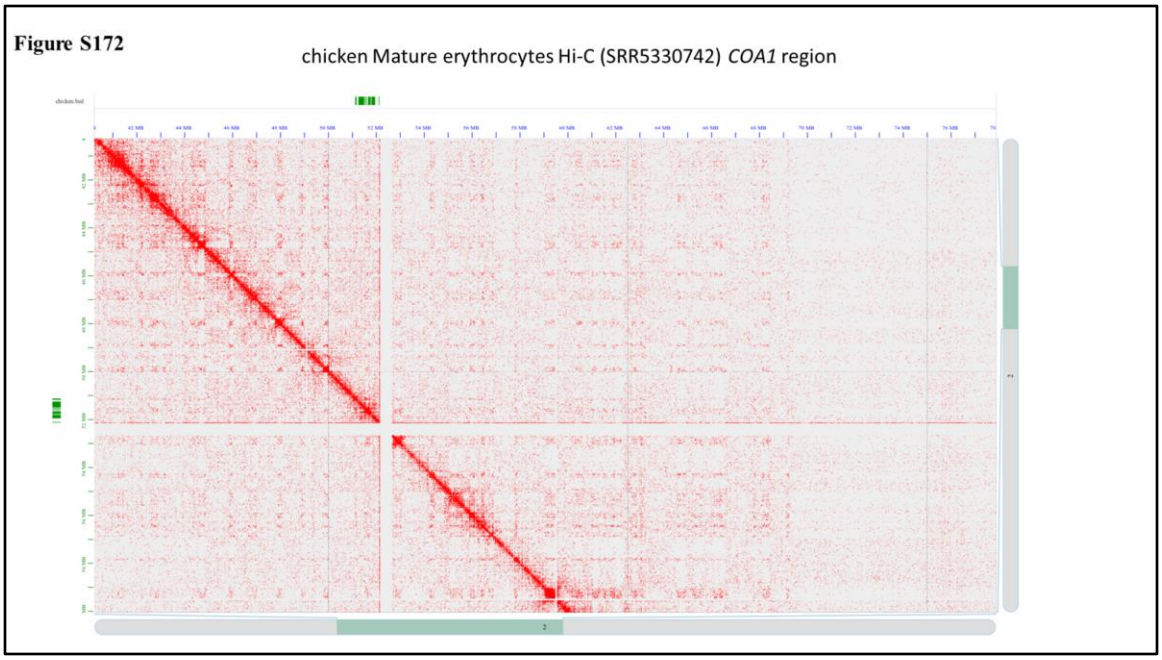


Figure S172: Hi-C figure generated using Juicebox of chicken Mature erythrocytes (SRR5330742) where red dots show the genomic loci interaction between adjacent regions and gap shows there is no interaction between the nearby region. We loaded the bed file on the top and left sides of the Hi-C plot, where each line of the green color indicates the gene of interest (*COA1)* and adjacent syntenic regions. chromosome or scaffold is mentioned on the bottom and right sides of the plot.

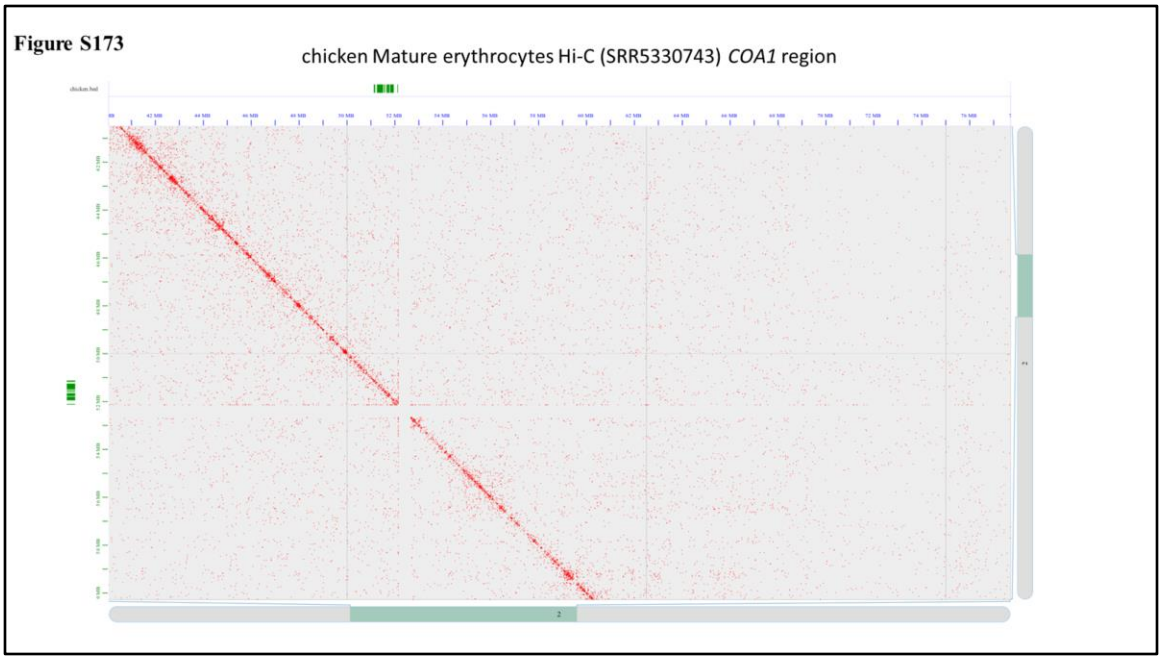


Figure S173: Hi-C figure generated using Juicebox of chicken Mature erythrocytes (SRR5330743) where red dots show the genomic loci interaction between adjacent regions and gap shows there is no interaction between the nearby region. We loaded the bed file on the top and left sides of the Hi-C plot, where each line of the green color indicates the gene of interest (*COA1)* and adjacent syntenic regions. chromosome or scaffold is mentioned on the bottom and right sides of the plot.

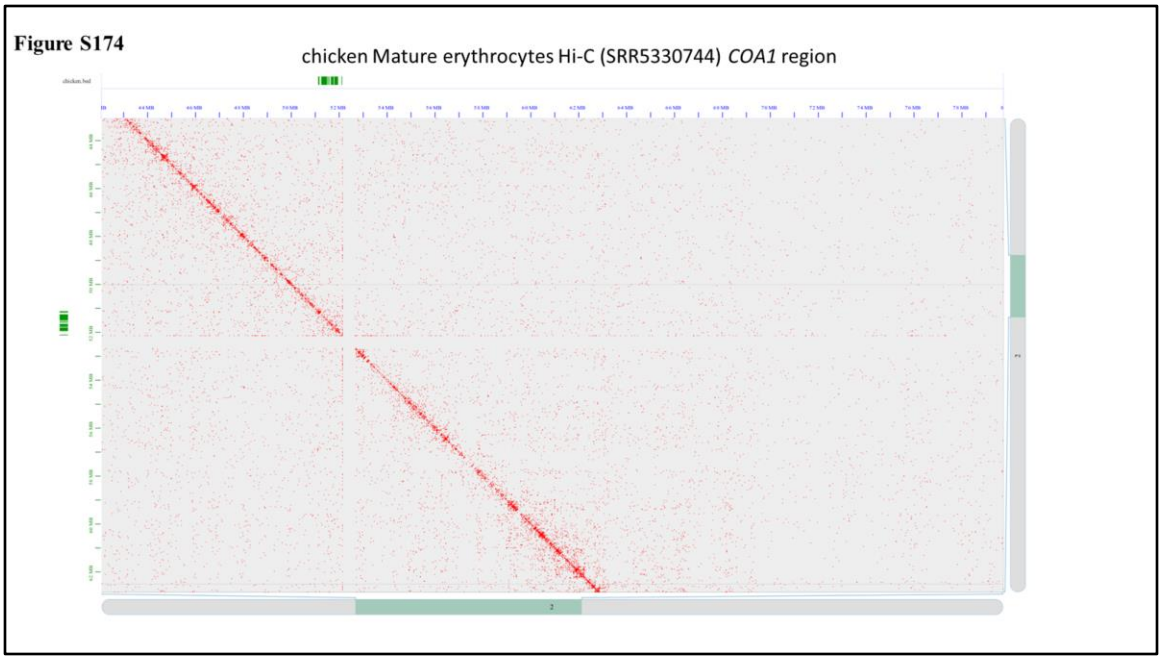


Figure S174: Hi-C figure generated using Juicebox of chicken Mature erythrocytes (SRR5330744) where red dots show the genomic loci interaction between adjacent regions and gap shows there is no interaction between the nearby region. We loaded the bed file on the top and left sides of the Hi-C plot, where each line of the green color indicates the gene of interest (*COA1)* and adjacent syntenic regions. chromosome or scaffold is mentioned on the bottom and right sides of the plot.

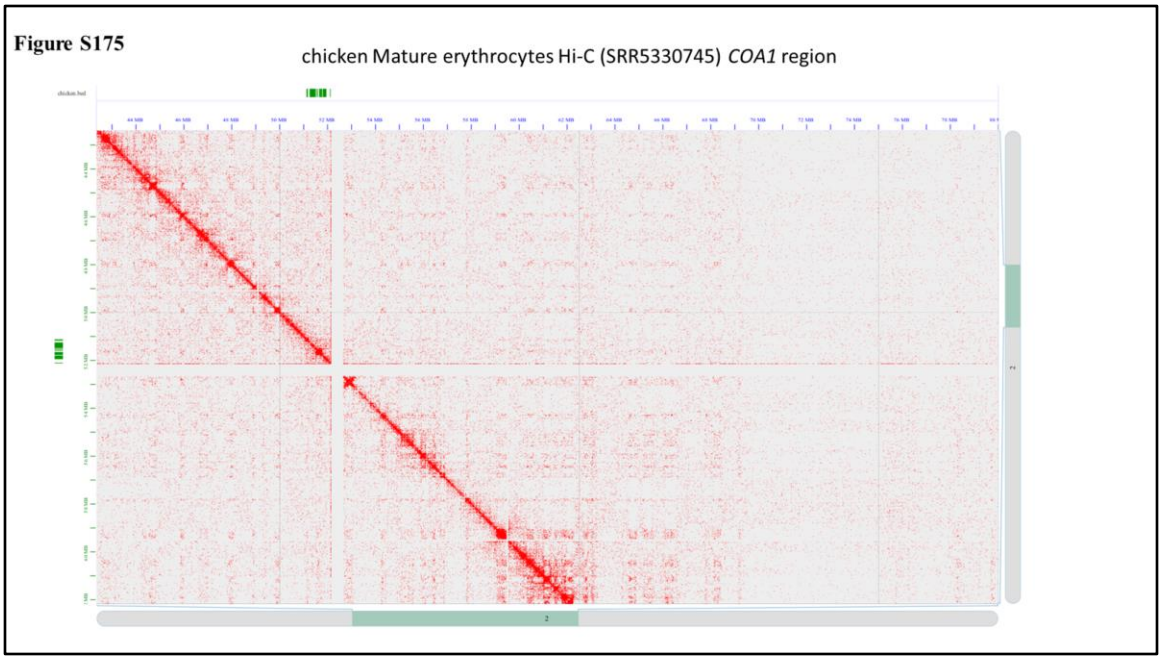


Figure S175: Hi-C figure generated using Juicebox of chicken Mature erythrocytes (SRR5330745) where red dots show the genomic loci interaction between adjacent regions and gap shows there is no interaction between the nearby region. We loaded the bed file on the top and left sides of the Hi-C plot, where each line of the green color indicates the gene of interest (*COA1)* and adjacent syntenic regions. chromosome or scaffold is mentioned on the bottom and right sides of the plot.

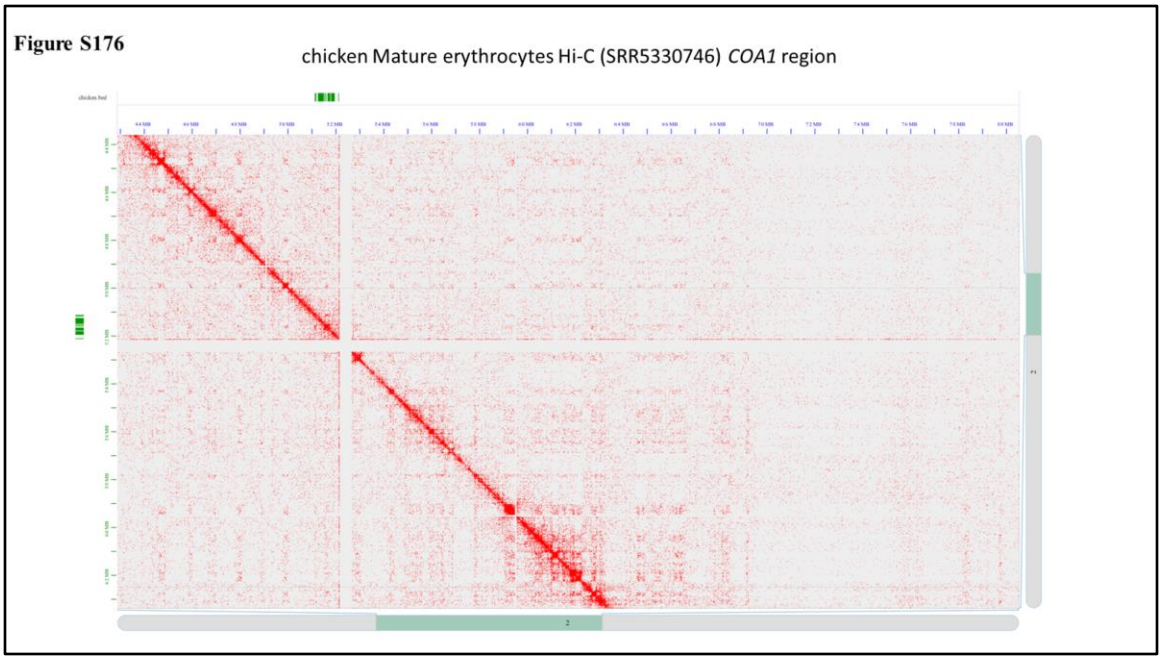


Figure S176: Hi-C figure generated using Juicebox of chicken Mature erythrocytes (SRR5330746) where red dots show the genomic loci interaction between adjacent regions and gap shows there is no interaction between the nearby region. We loaded the bed file on the top and left sides of the Hi-C plot, where each line of the green color indicates the gene of interest (*COA1)* and adjacent syntenic regions. chromosome or scaffold is mentioned on the bottom and right sides of the plot.

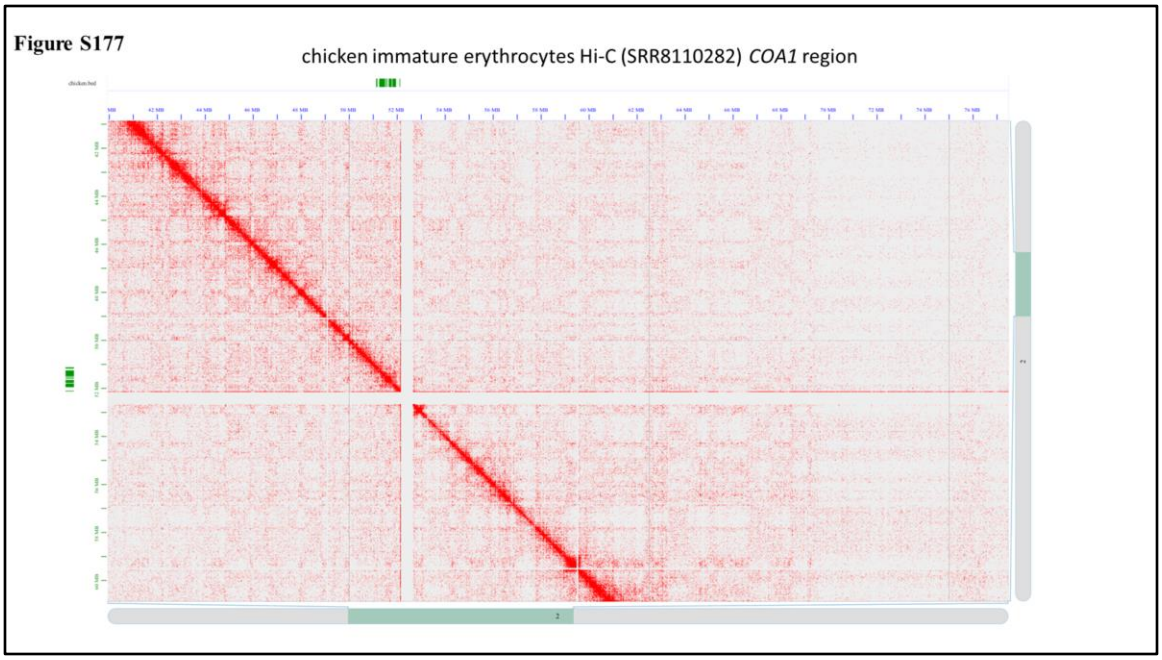


Figure S177: Hi-C figure generated using Juicebox of chicken immature erythrocytes data (SRR8110282) where red dots show the genomic loci interaction between adjacent regions and gap shows there is no interaction between the nearby region. We loaded the bed file on the top and left sides of the Hi-C plot, where each line of the green color indicates the gene of interest (*COA1)* and adjacent syntenic regions. chromosome or scaffold is mentioned on the bottom and right sides of the plot.

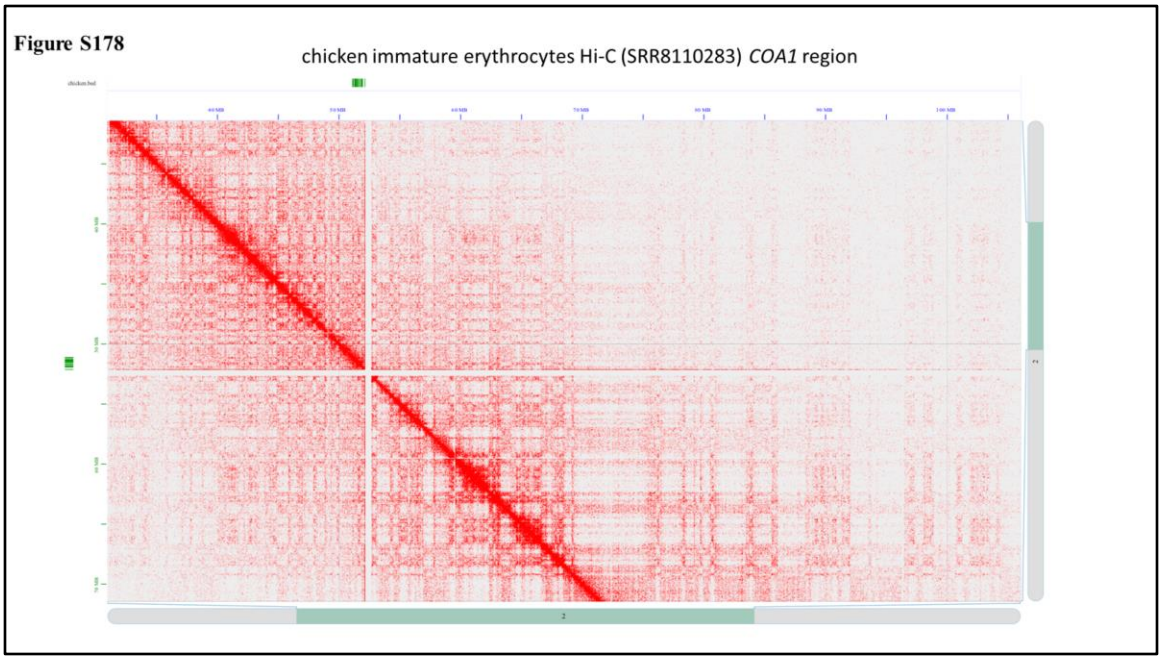


Figure S178: Hi-C figure generated using Juicebox of chicken immature erythrocytes data (SRR8110283) where red dots show the genomic loci interaction between adjacent regions and gap shows there is no interaction between the nearby region. We loaded the bed file on the top and left sides of the Hi-C plot, where each line of the green color indicates the gene of interest (*COA1)* and adjacent syntenic regions. chromosome or scaffold is mentioned on the bottom and right sides of the plot.

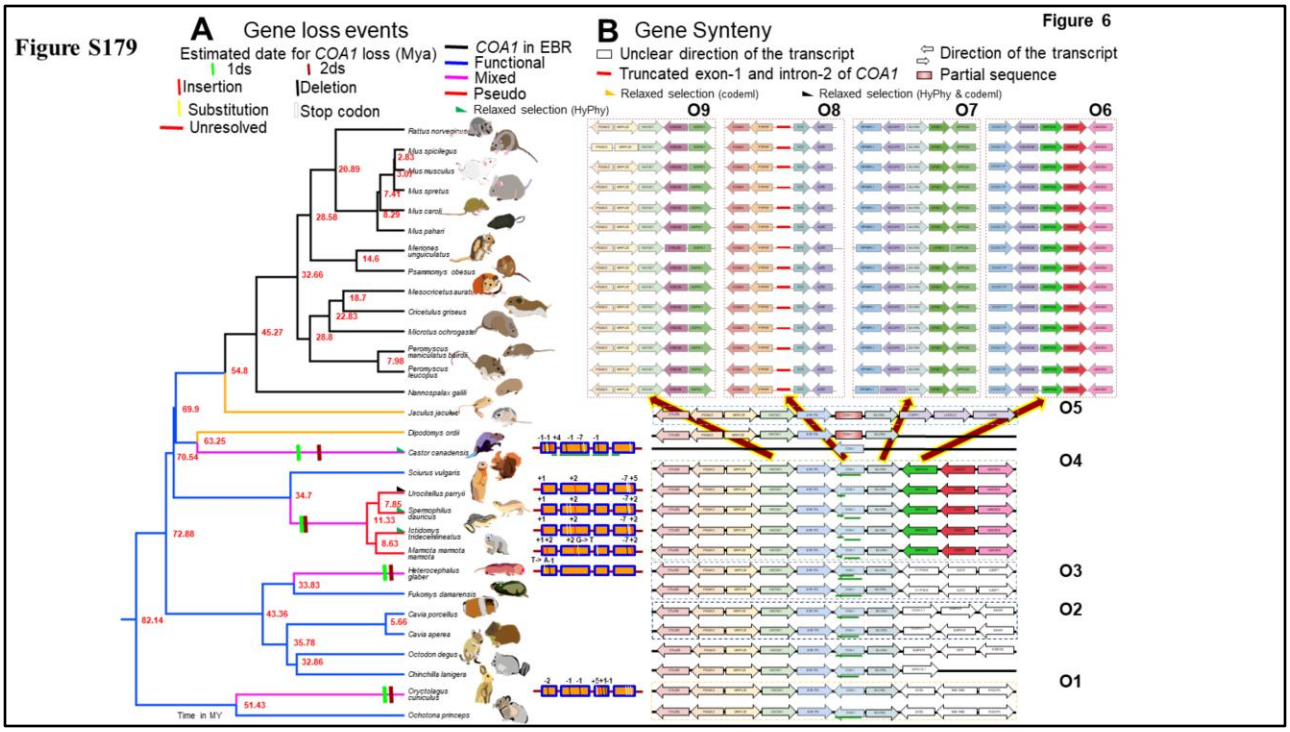


Figure S179: Recurrent loss of *COA1* gene in rodent species. (**A**) Gene loss events in seven rodent species through four events are represented exon-wise beside the time-calibrated phylogenetic tree obtained from the TimeTree website. Blue branches correspond to functional copies of *COA1*, and black branches correspond to the Evolutionary Breakpoint Region (EBR), pink banches correcsponds to mixed branches and red branches represnts pseudo branches. (**B**). Gene order in the genomic region flanking the *COA1* gene in rodent genomes. Arrows depict the direction of gene transcription relative to the *COA1* gene for consistency

across species. Boxes represent the genes located on short scaffolds with unknown orientation. Each dotted box contains one type of gene order, and the brown arrows highlighted in yellow emerging from gene order O5 depict the EBR event that leads to gene orders O6, O7, O8, and O9. Gene order O8 and O7 contain partial remains of the *COA1* gene and a functional *BLVRA* gene, respectively. A solid red line within gene order O8 depicts the partial exon one and intron-2 of *COA1* located between the *PTPRF* and *HYI* genes. The gene order O6 and O9 correspond to the regions on the left and right flanks of the region containing *STK17A*, *COA1*, and *BLVRA*.

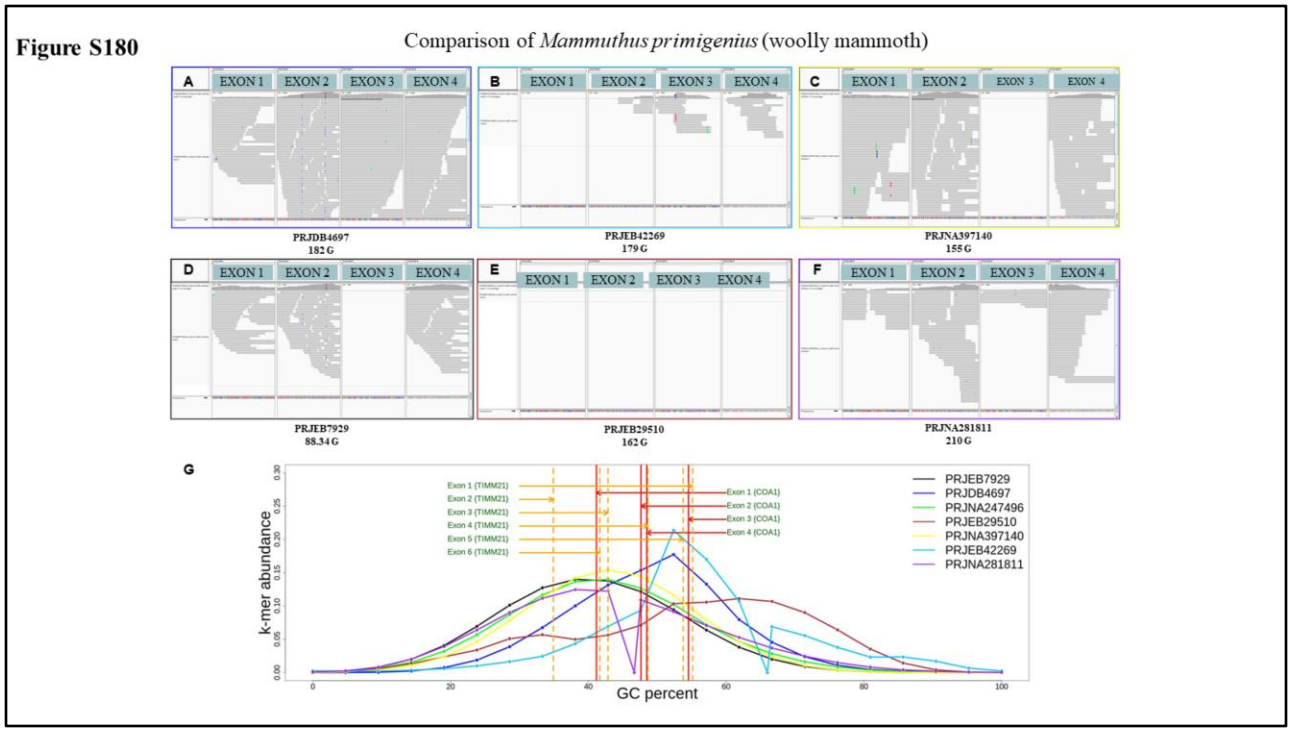


Figure S180: Comparison of different sequencing datasets of woolly mammoth (*Mammuthus primigenius*) for *COA1* gene exons. Gray rectangles show the reads mapped to each exon. Panels **A** to **F** shows the reads with sequence supporting each exon from different woolly mammoth SRA bio projects [PRJDB4697 (182 Gb), PRJEB42269 (179 Gb), PRJNA397140 (155 Gb), PRJEB7929 (88.34 Gb), PRJEB29510 (162 Gb), and PRJNA281811 (210 Gb)]. Panel **G** indicates the GC percentage vs. K-mer abundance of different project IDs mentioned in different colors. The vertical dotted lines in orange denote the GC percentage of *TIMM21* exons, and vertical solid lines in red indicate the GC percentage of *COA1* exons.

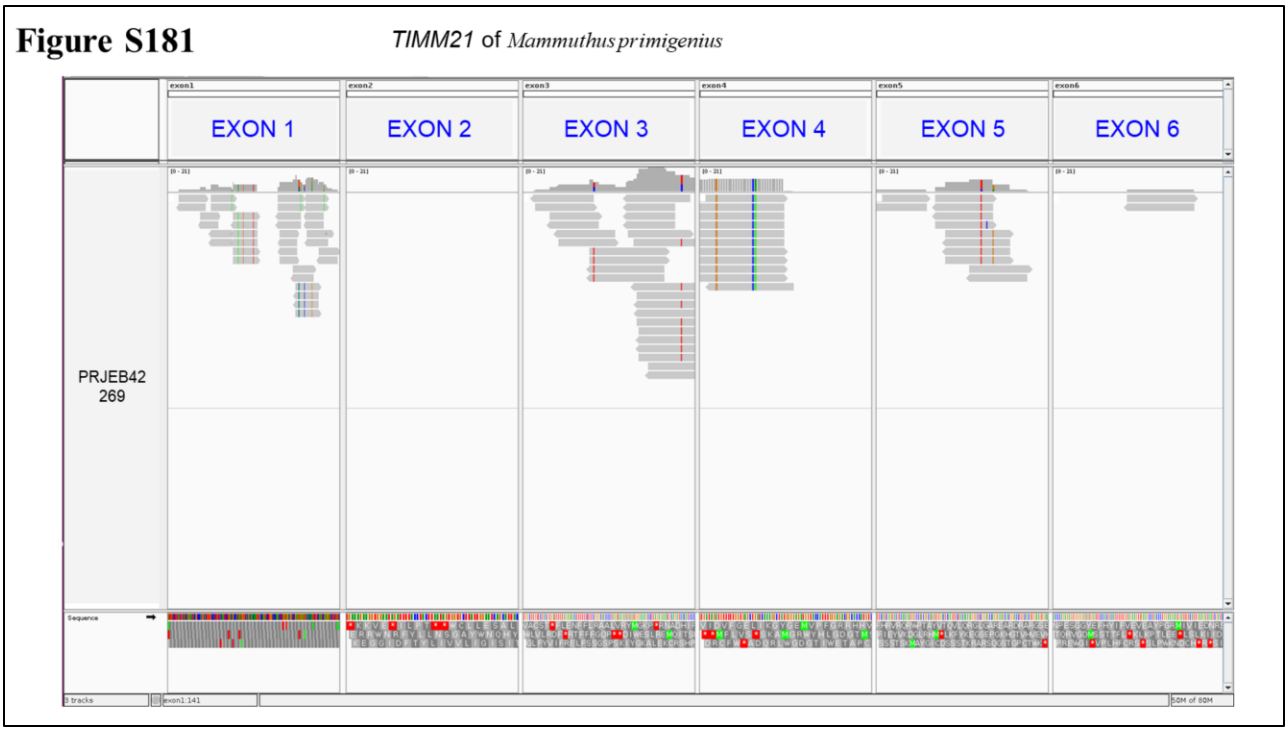


Figure S181: Exon-wise screenshot of *TIMM21* in *Mammuthus primigenius* (woolly mammoth) from Illumina dataset PRJEB42269. Gray boxes are showing reads for particular exons.

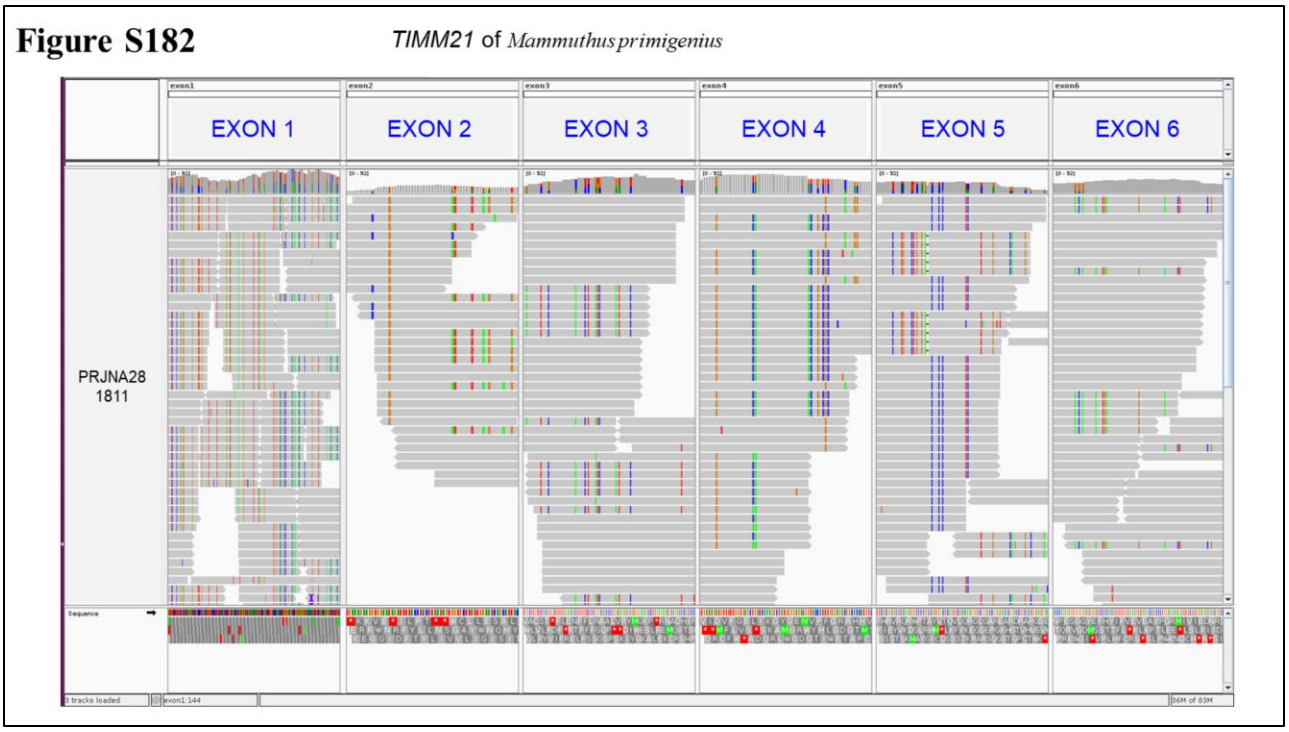


Figure S182: Exon-wise screenshot of *TIMM21* in *Mammuthus primigenius* (woolly mammoth) from Illumina dataset PRJ281811. Gray boxes are showing reads for particular exons.

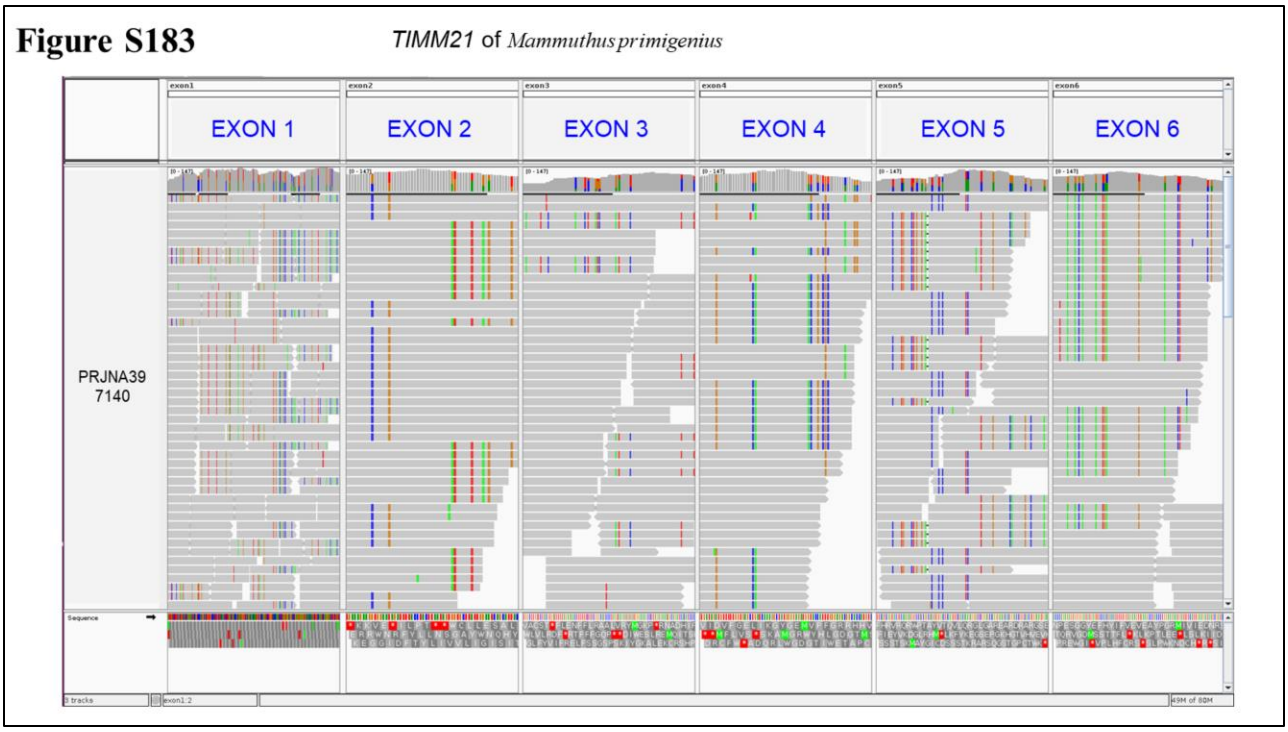


Figure S183: Exon-wise screenshot of *TIMM21* in *Mammuthus primigenius* (woolly mammoth) from Illumina dataset PRJNA397140. Gray boxes are showing reads for particular exons.

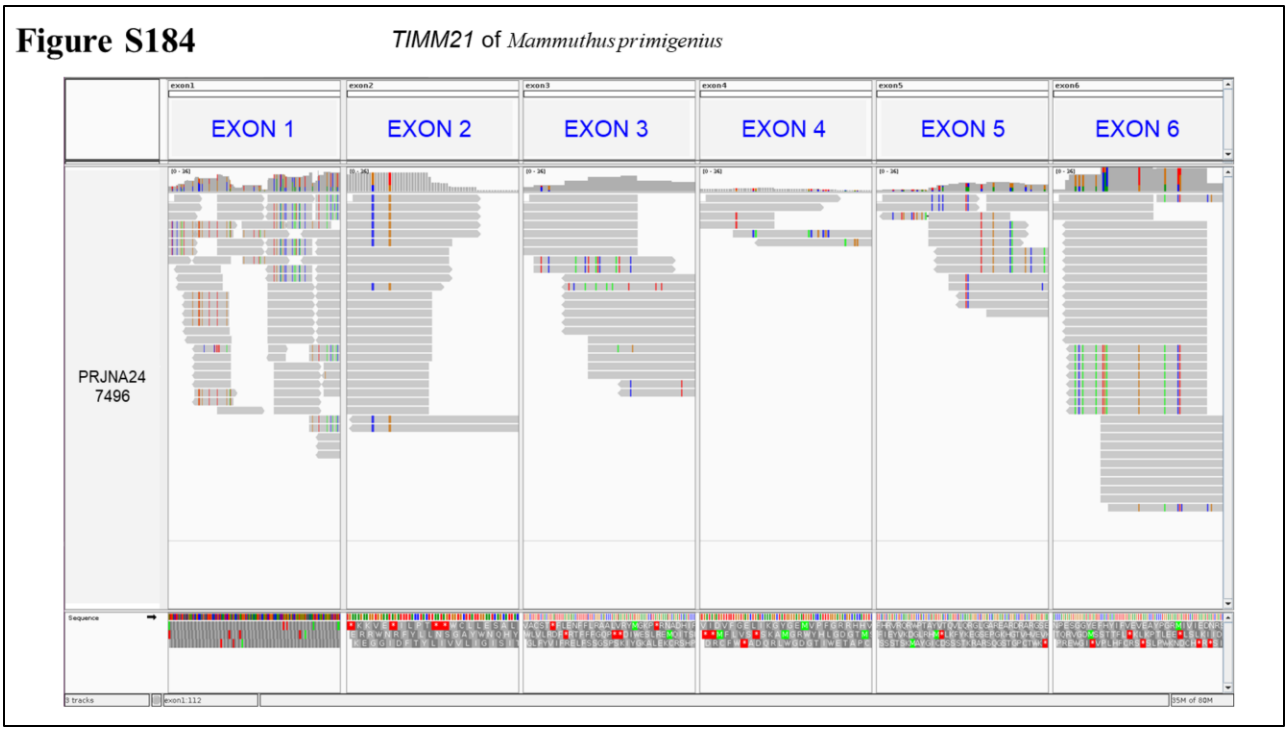


Figure S184: Exon-wise screenshot of *TIMM21* in *Mammuthus primigenius* (woolly mammoth) from Illumina dataset PRJNA247496. Gray boxes are showing reads for particular exons.

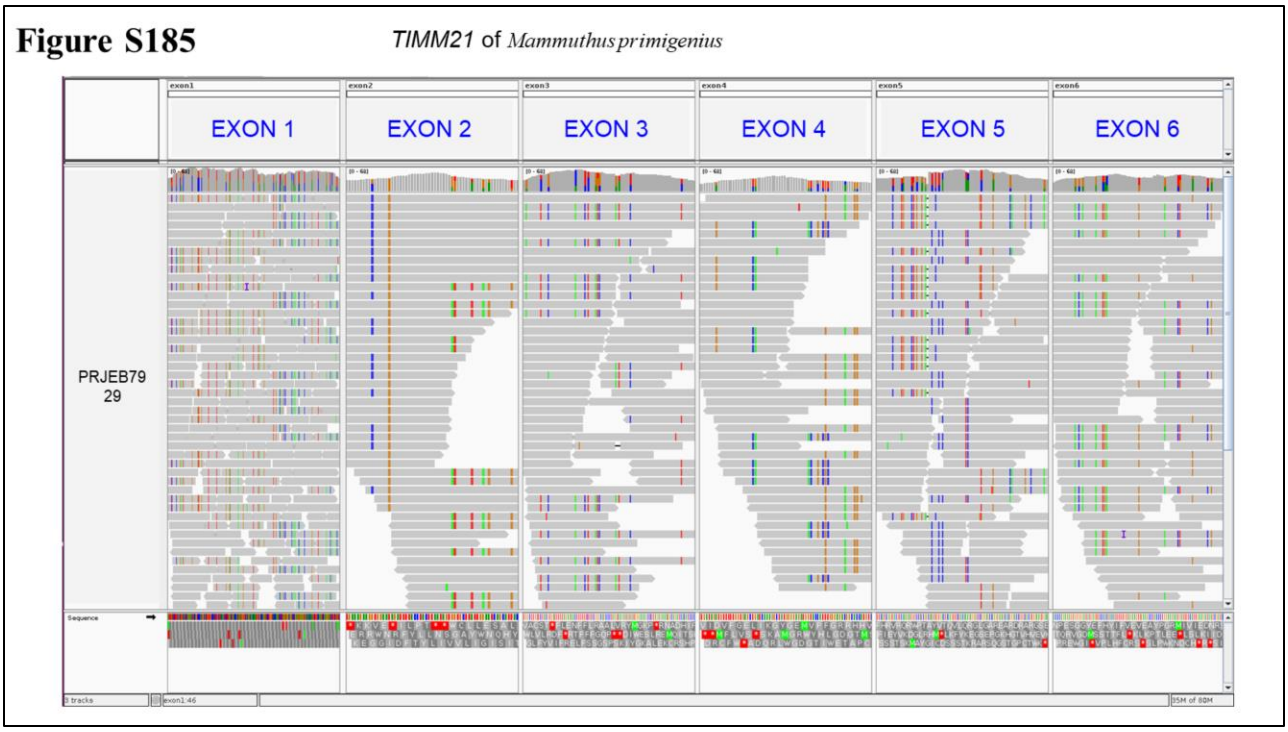


Figure S185: Exon-wise screenshot of *TIMM21* in *Mammuthus primigenius* (woolly mammoth) from Illumina dataset PRJEB7929. Gray boxes are showing reads for particular exons.

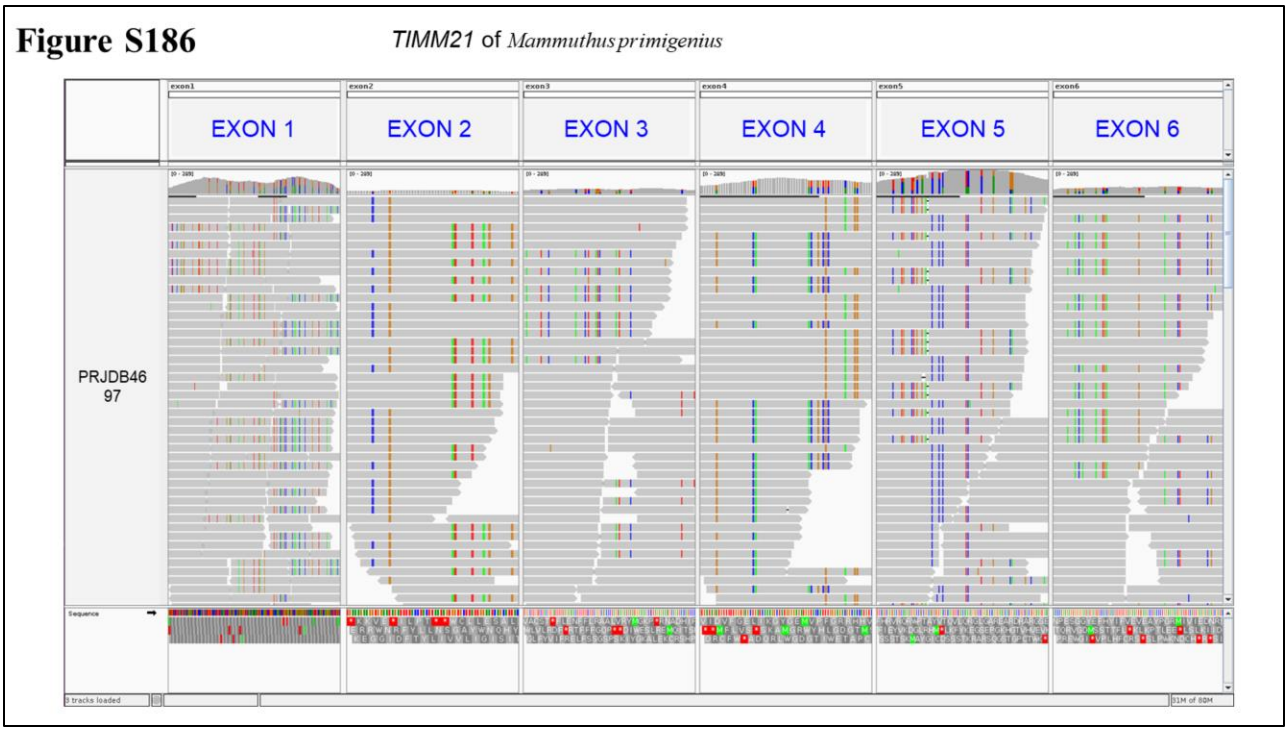


Figure S186: Exon-wise screenshot of *TIMM21* in *Mammuthus primigenius* (woolly mammoth) from Illumina dataset PRJDB4697. Gray boxes are showing reads for particular exons.

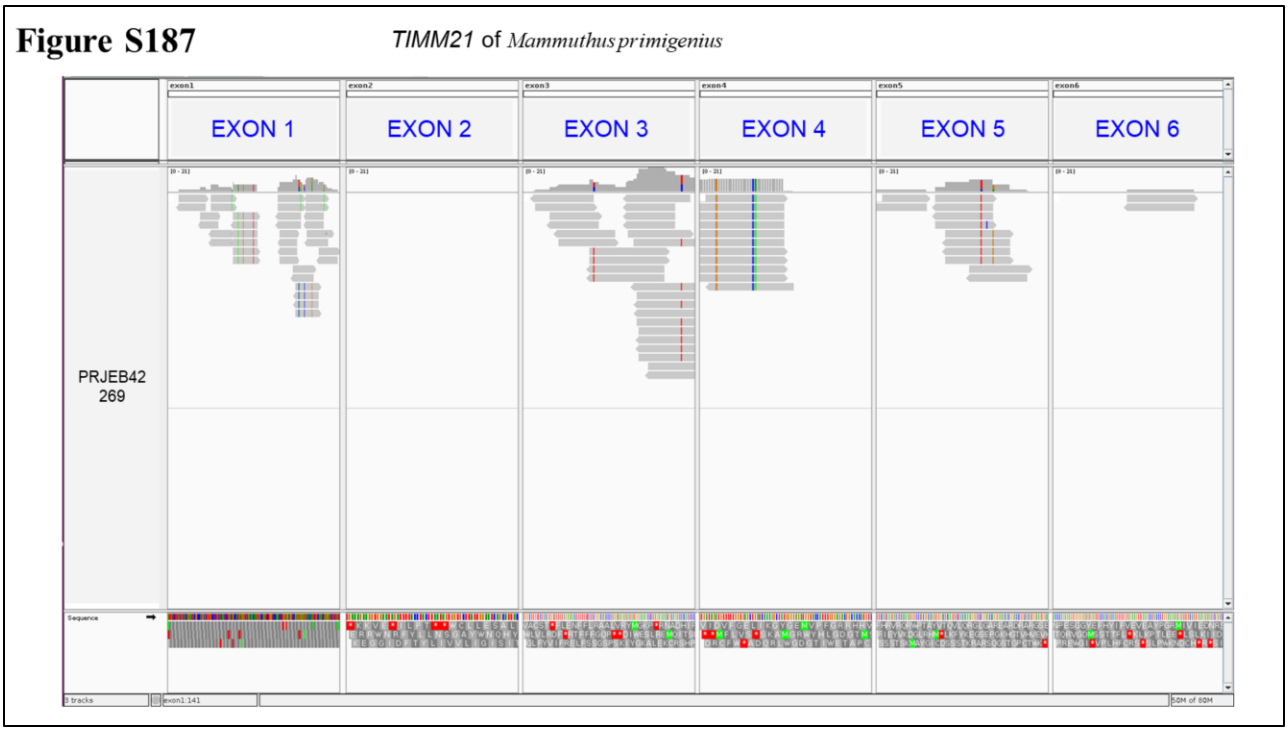


Figure S187: Exon-wise screenshot of *TIMM21* in *Mammuthus primigenius* (woolly mammoth) from Illumina dataset PRJEB42269. Gray boxes are showing reads for particular exons.

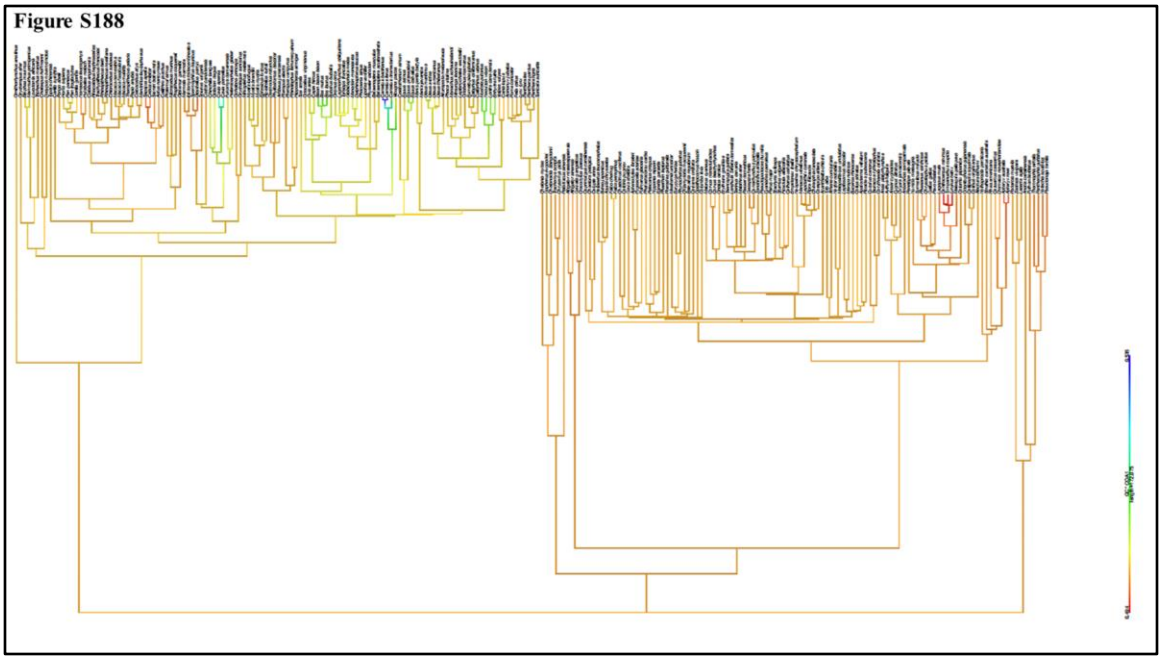


Figure S188: Ratio of GC* in *COA1* across the 240 species. The colour gradient represents the value of GC* for each species are shown with phylogenetic tree downloaded from the TimeTree website.

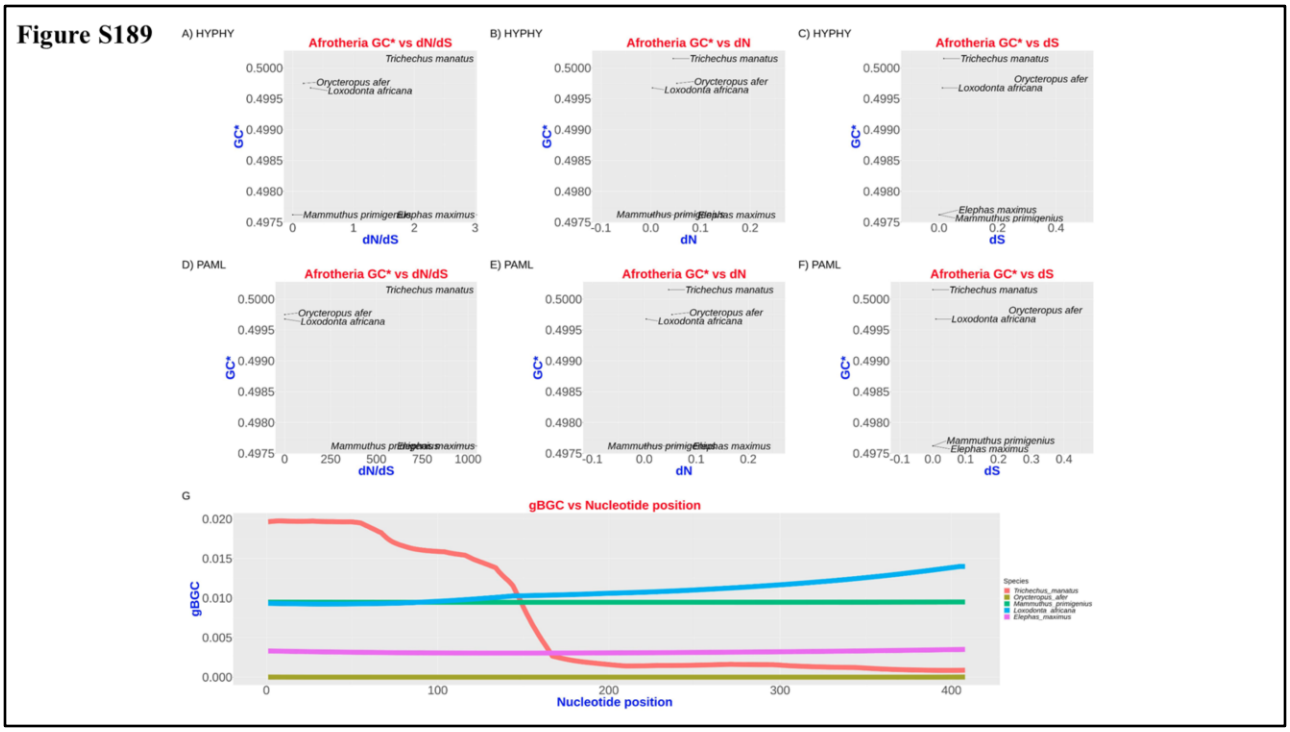


Figure S189: Values of GC* in Afrotheria are plotted with dN and dS which is calculated using mapNH. dN and dS were calculated using PAML and HyPhy both. gBGC across nucleotide position were calculated using PhastBias.

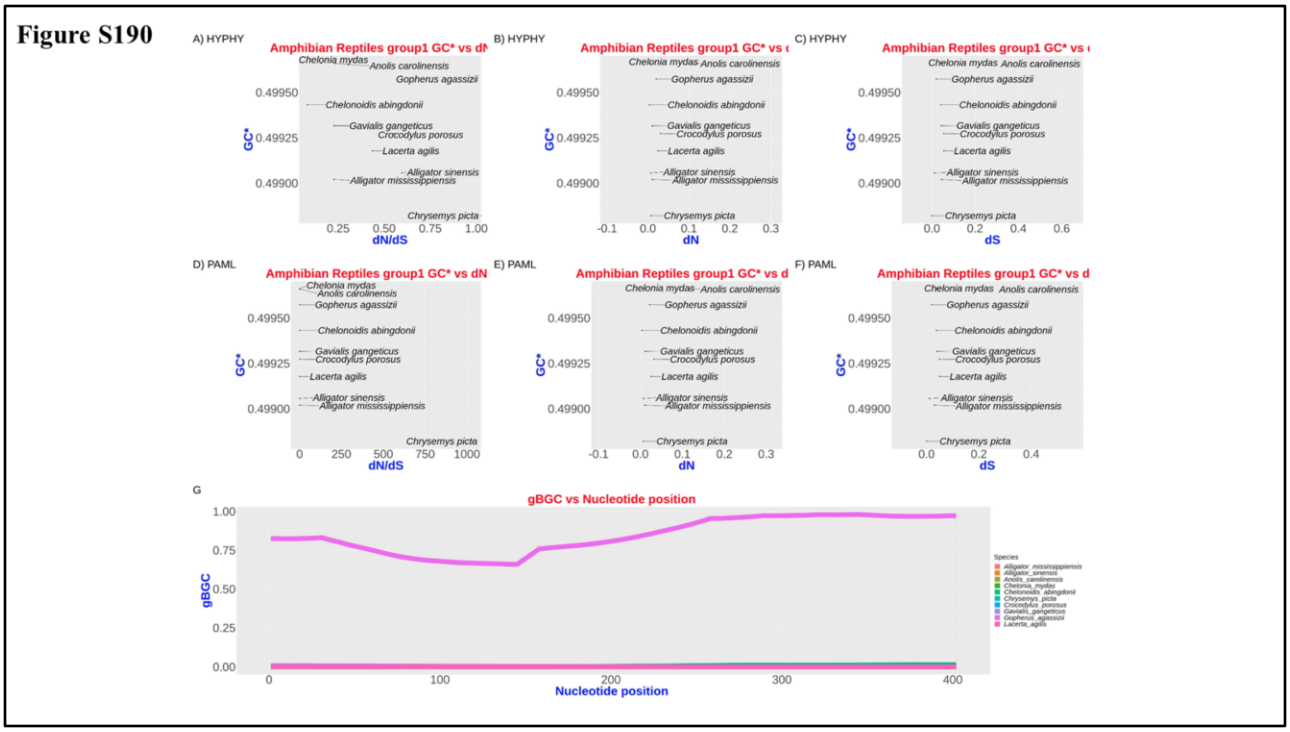


Figure S190: Values of GC* in Amphibian and Reptiles are plotted with dN and dS which is calculated using mapNH. dN and dS were calculated using PAML and HyPhy both. gBGC across nucleotide position were calculated using PhastBias. *Gopherus agassizii* has a significant gBGC across *COA1*.

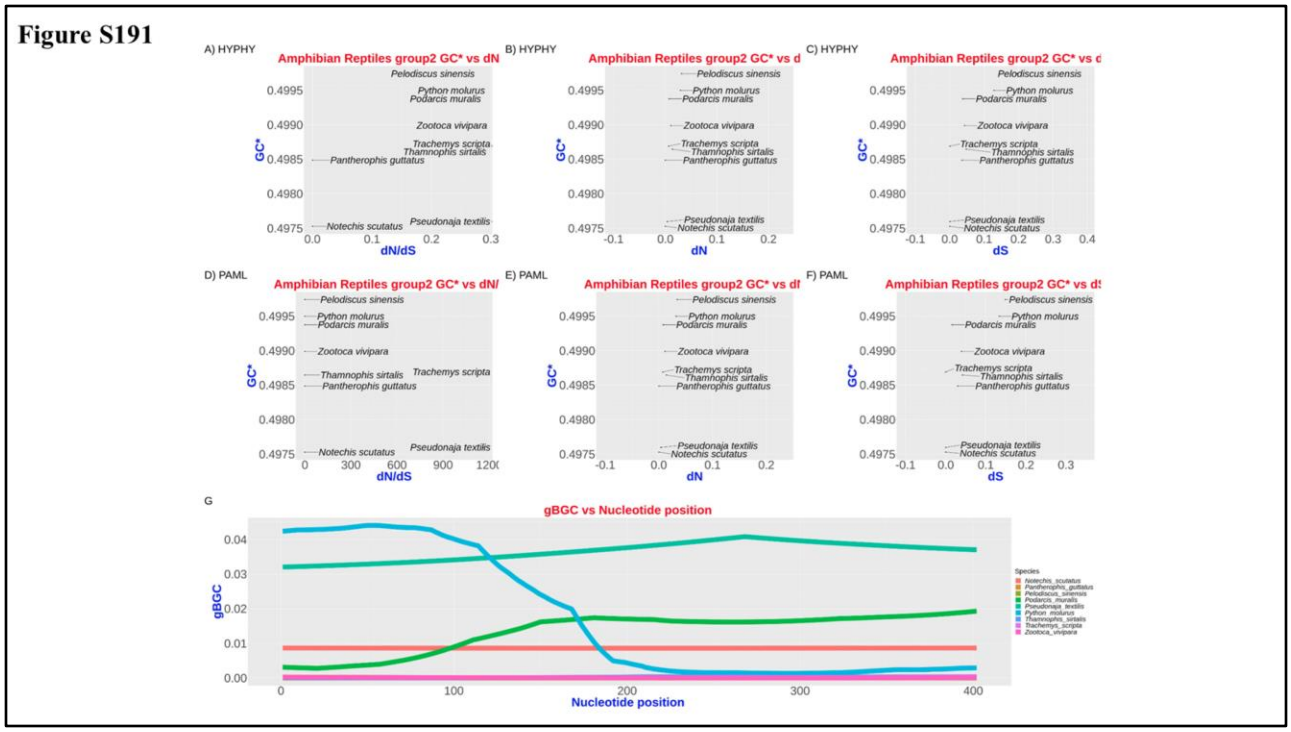


Figure S191: Values of GC* in amphibians and reptiles are plotted with dN and dS which is calculated using mapNH. dN and dS were calculated using PAML and HyPhy both. gBGC across nucleotide position were calculated using PhastBias.

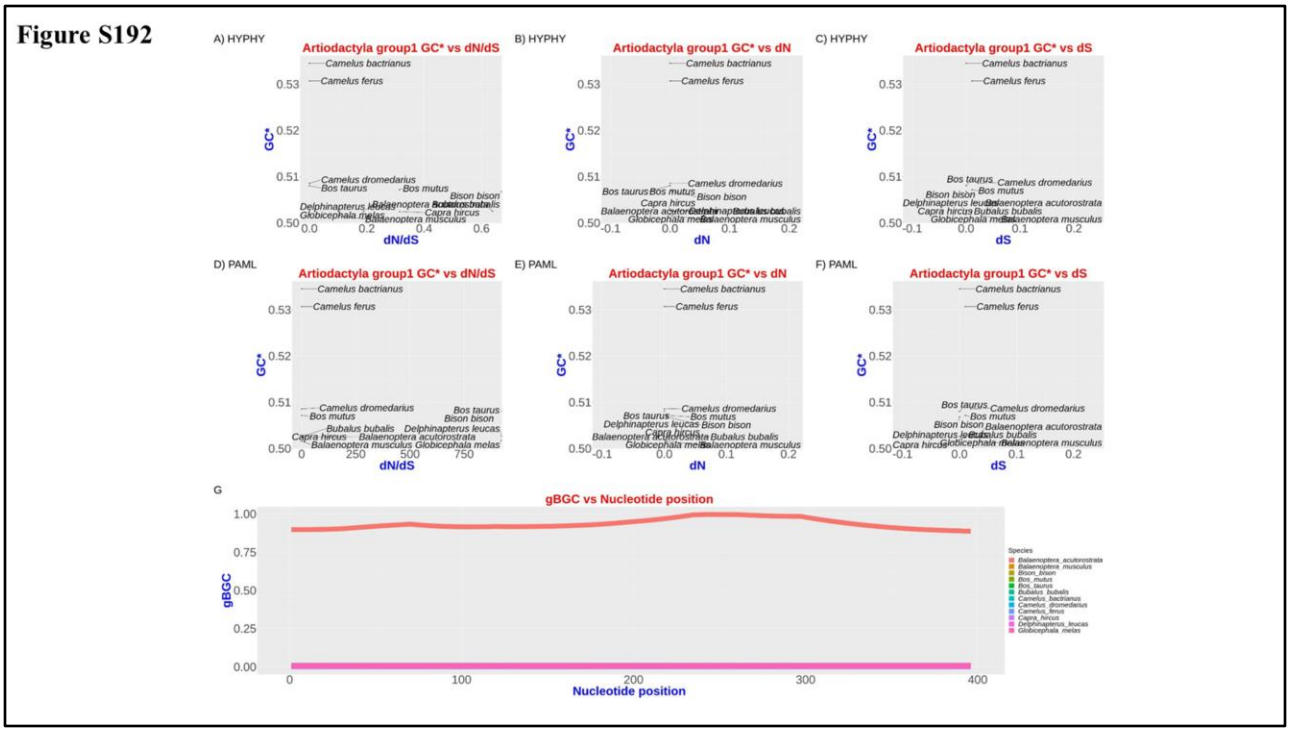


Figure S192: Values of GC* in Artiodactyla are plotted with dN and dS which is calculated using mapNH. dN and dS were calculated using PAML and HyPhy both. gBGC across nucleotide position were calculated using PhastBias. *Balaenoptera acutorostrata* has a significant gBGC across *COA1*.

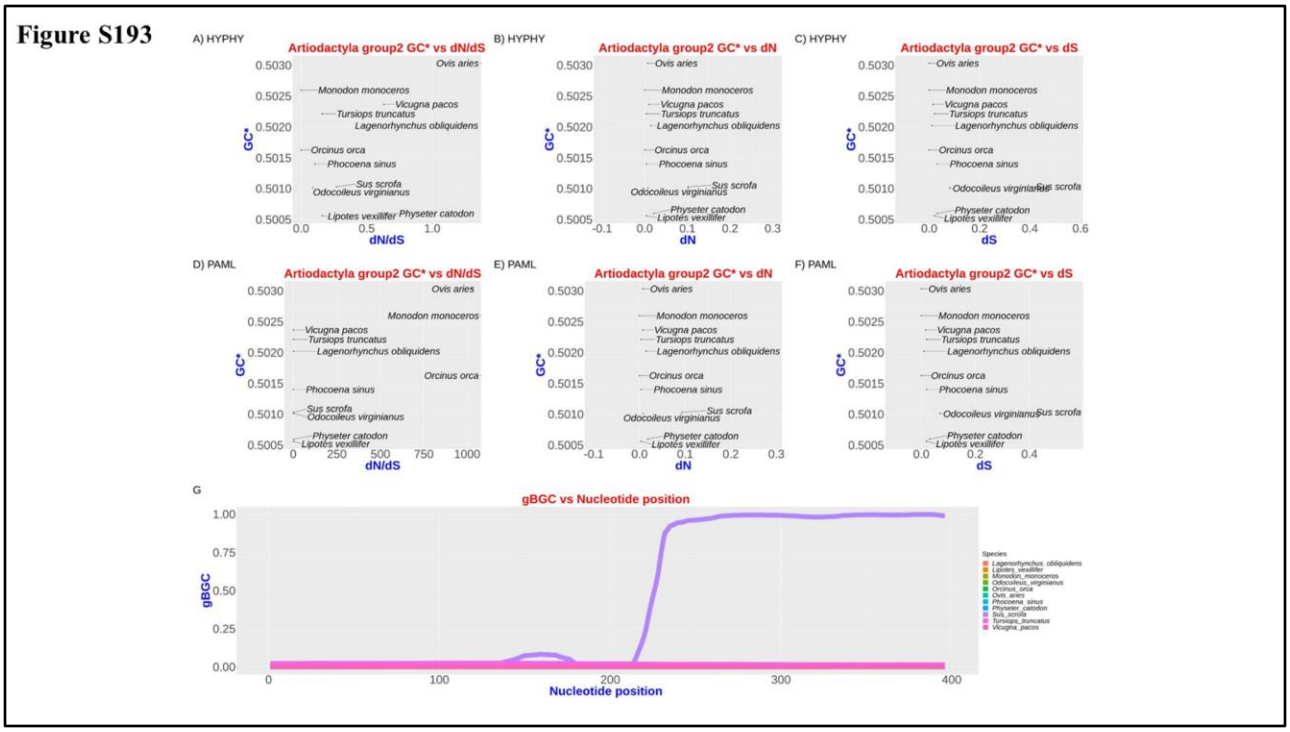


Figure S193: Values of GC* in Artiodactyla are plotted with dN and dS which is calculated using mapNH. dN and dS were calculated using PAML and HyPhy both. gBGC across nucleotide position were calculated using PhastBias. *Sus scrofa* have a significant gBGC in *COA1*.

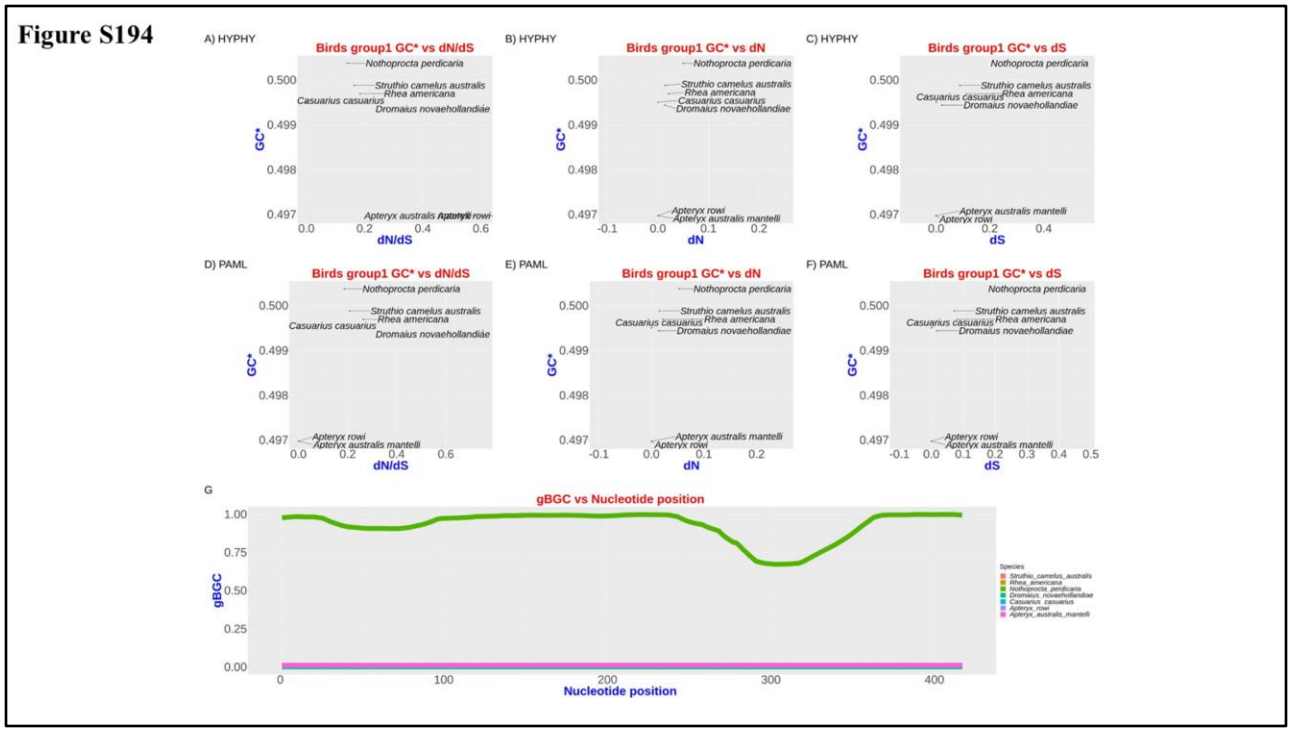


Figure S194: Values of GC* in birds are plotted with dN and dS which is calculated using mapNH. dN and dS were calculated using PAML and HyPhy both. gBGC across nucleotide position were calculated using PhastBias. *Nothoprocta perdicaria* has a significant gBGC across *COA1*.

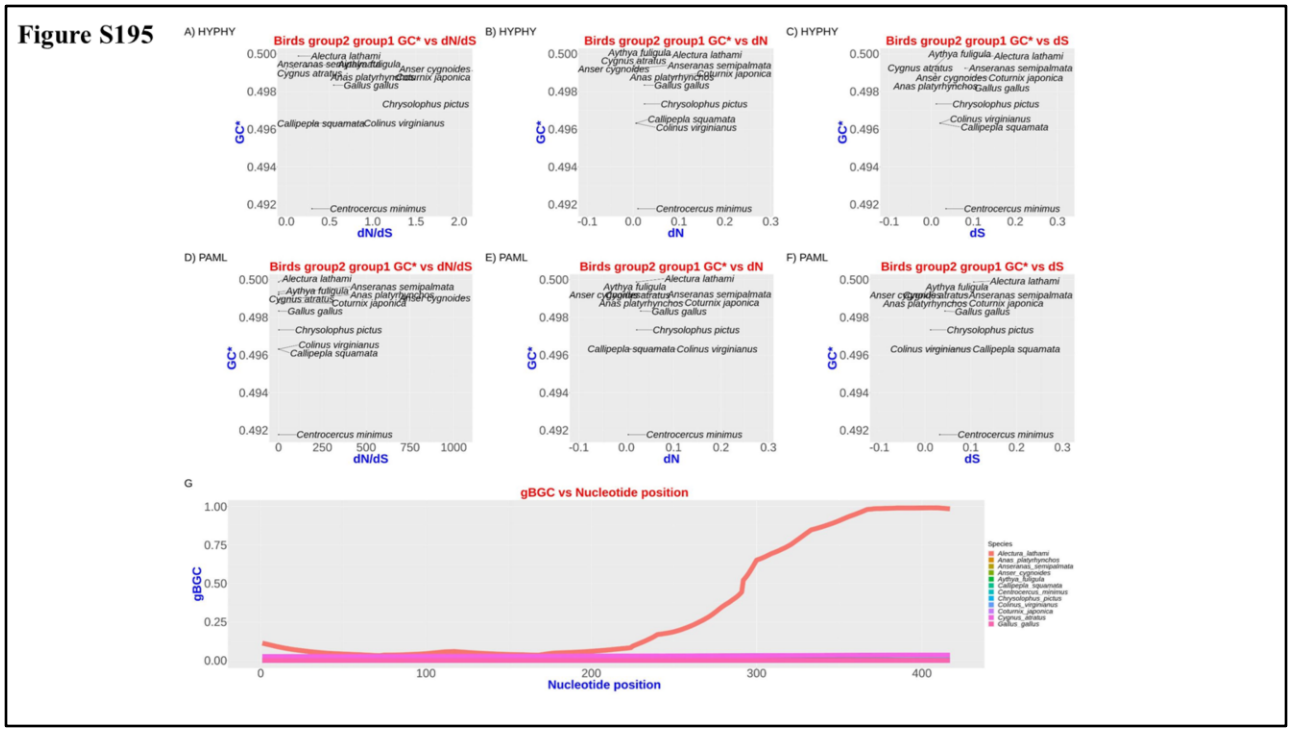


Figure S195: Values of GC* in birds are plotted with dN and dS which is calculated using mapNH. dN and dS were calculated using PAML and HyPhy both. gBGC across nucleotide position were calculated using PhastBias. *Alectura lathami* has a significant gBGC in *COA1*.

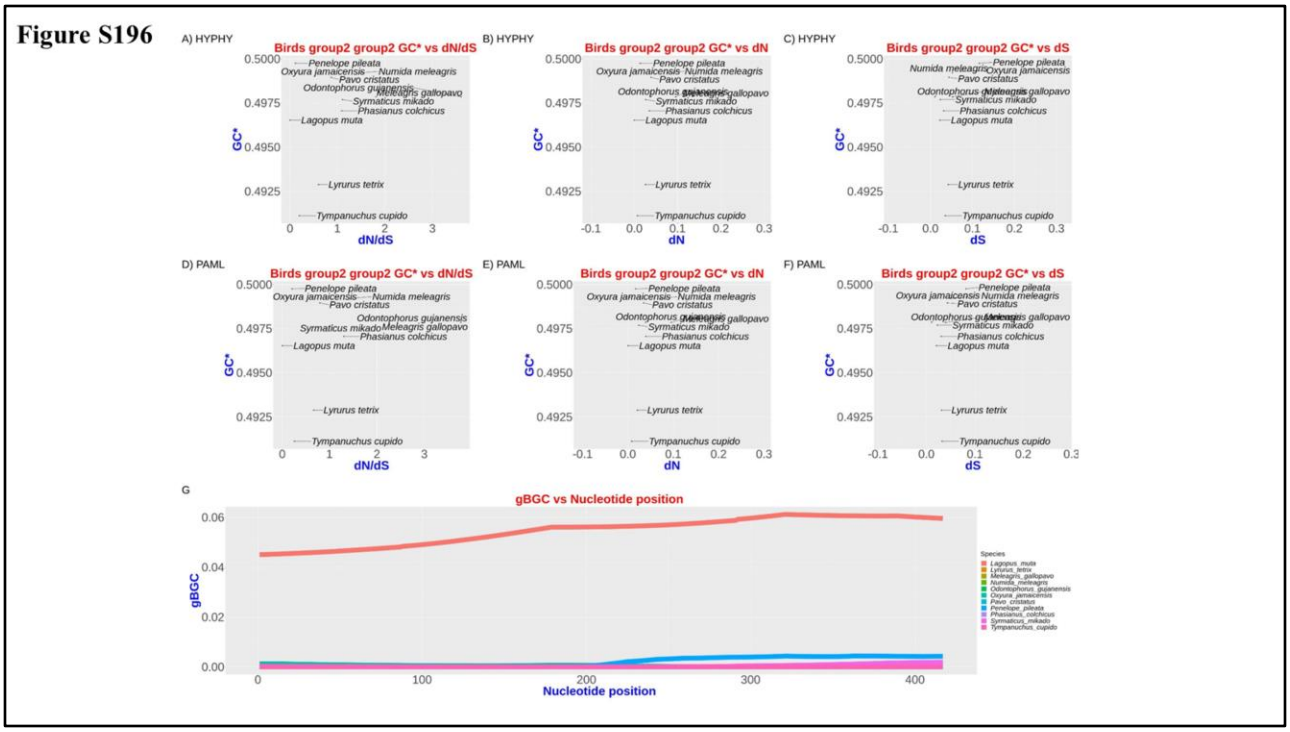


Figure S196: Values of GC* in birds are plotted with dN and dS which is calculated using mapNH. dN and dS were calculated using PAML and HyPhy both. gBGC across nucleotide position were calculated using PhastBias.

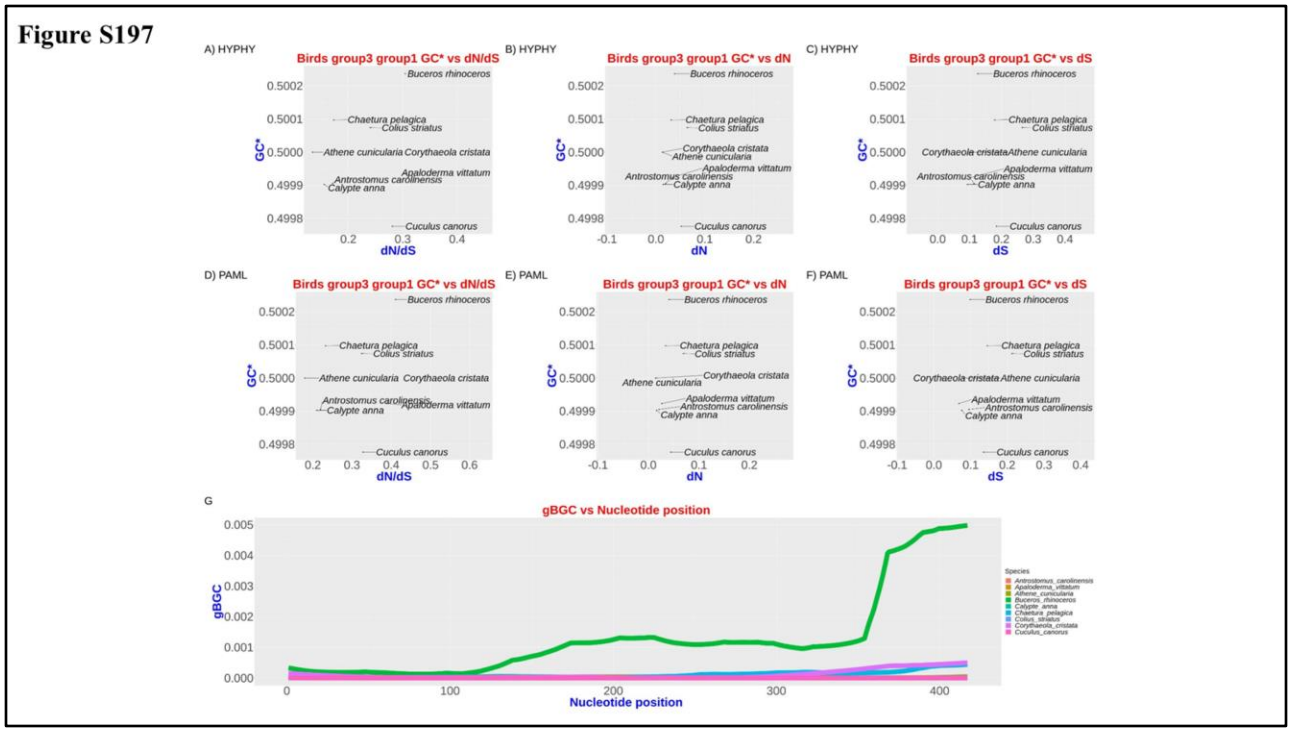


Figure S197: Values of GC* in birds are plotted with dN and dS which is calculated using mapNH. dN and dS were calculated using PAML and HyPhy both. gBGC across nucleotide position were calculated using PhastBias.