



OPEN

Recurrent erosion of *COA1*/*MITRAC15* exemplifies conditional gene dispensability in oxidative phosphorylation

Sagar Sharad Shinde, Sandhya Sharma, Lokdeep Teekas, Ashutosh Sharma & Nagarjun Vijay[✉]

Skeletal muscle fibers rely upon either oxidative phosphorylation or the glycolytic pathway with much less reliance on oxidative phosphorylation to achieve muscular contractions that power mechanical movements. Species with energy-intensive adaptive traits that require sudden bursts of energy have a greater dependency on glycolytic fibers. Glycolytic fibers have decreased reliance on OXPHOS and lower mitochondrial content compared to oxidative fibers. Hence, we hypothesized that gene loss might have occurred within the OXPHOS pathway in lineages that largely depend on glycolytic fibers. The protein encoded by the *COA1*/*MITRAC15* gene with conserved orthologs found in budding yeast to humans promotes mitochondrial translation. We show that gene disrupting mutations have accumulated within the *COA1* gene in the cheetah, several species of galliform birds, and rodents. The genomic region containing *COA1* is a well-established evolutionary breakpoint region in mammals. Careful inspection of genome assemblies of closely related species of rodents and marsupials suggests two independent *COA1* gene loss events co-occurring with chromosomal rearrangements. Besides recurrent gene loss events, we document changes in *COA1* exon structure in primates and felids. The detailed evolutionary history presented in this study reveals the intricate link between skeletal muscle fiber composition and the occasional dispensability of the chaperone-like role of the *COA1* gene.

Skeletal muscles control numerous locomotor functions in vertebrates¹. The hundreds of different muscles in the body consist of highly organized heterogeneous bundles of fibers. These fibers are classified based on the contractile properties, power source, and myosin component into type-1, 2A, 2B, and 2X². Muscles with aerobic type-1 and 2A fibers rely on the oxidative phosphorylation (OXPHOS) pathway for sustained locomotion³. Locomotion that requires sudden energy bursts depends on the predominantly anaerobic 2B and 2X fibers⁴. OXPHOS is the energy-releasing electron transport chain (ETC) coupled with the energy-requiring chemiosmosis^{5,6}. A chain of mitochondrial inner membrane-embedded proteins encoded by mitochondrial and nuclear genes forms four large complexes that transport electrons through redox reactions. The energy released results in a proton gradient, which uses a fifth membrane-embedded complex to generate ATP through chemiosmosis. The individual complexes are known to assemble into supercomplexes⁷. Optimization of the OXPHOS pathway leads to improved locomotor performance⁸. However, degeneration of locomotor abilities leads to a relaxed selective constraint on the OXPHOS pathway^{9,10}. The special locomotory needs of galliform birds, rodents, marsupials, and felids lead to a greater reliance on anaerobic fast fibers for sudden bursts of energy^{11–13}.

The ability to fly is a distinctive feature of birds except for lineages that have become entirely flightless or retain only a limited flying capacity^{14–17}. The large amount of energy required for flight has necessitated a high metabolic rate in birds¹⁸. Increased ATP generation fulfills these energy demands through metabolic adaptations in the OXPHOS pathway¹⁹. Avian flight is possible through a combination of flight muscles that consist of anaerobic white (fast glycolytic), intermediate/red-pink (fast oxidative), and aerobic red (slow oxidative) fibers^{20–22}. Birds with strong flight abilities, such as long-distance migrants and small passerines, contain fast oxidative fibers²³. However, galliform species have mostly glycolytic fibers to allow short bursts of activity²⁴. Hence, the OXPHOS pathway is under stronger selective constraint in non-galliform bird species than galliform birds due to the functional specialization of mitochondria to different muscle fibers²⁵.

Computational Evolutionary Genomics Lab, Department of Biological Sciences, IISER Bhopal, Bhauri, Madhya Pradesh, India. ✉ email: nagarjun@iiserb.ac.in

The cheetah (*Acinonyx jubatus*) epitomizes the relevance of speed and acceleration²⁶. In general, felids are adept at sprinting and can accelerate more rapidly than canids but cannot sustain them for a prolonged period²⁷. The predominance of type-2X fibers in felids provides the ability of rapid acceleration^{28–30}. Compared to canids, felids have a greater reliance on glycolytic fibers. In smaller mammals such as rodents, the higher relative speed results from faster constricting by the higher proportion of glycolytic fibers (mostly 2X and 2B)³¹. For instance, rodent limbs have more abundant type 2B fibers compared to larger mammals (including humans, which have no type 2B fibers in the limbs)^{32,33}. Marsupial species also attain high relative speeds with glycolytic fibers (2B and 2X)^{34,35}. The higher proportion of fast glycolytic fibers in felids, rodents, and marsupials results in relaxed selection on the OXPHOS pathway genes in these species. Glycolytic fibers have decreased reliance on OXPHOS and lower mitochondrial content than oxidative fibers^{25,36}.

Functional studies implicate *COA1* (also known as *MITRAC15*) in promoting mitochondrial translation and complex I and IV biogenesis^{37,38}. However, overexpression of other genes easily compensates for the mild effect of *COA1* gene knockout^{39,40}. Notably, the *COA1* gene was also identified as a positively selected gene in a genome-wide screen in primates⁴¹ and suggests that *COA1* can contribute to fitness increases through its role as a chaperone despite its mild phenotype. Our study evaluates whether the protein encoded by the *COA1* gene, a mitochondrial complex I translation factor with a chaperone-like role, is dispensable when the OXPHOS pathway is under relaxed selective constraints. The integration of the mitochondria in the host cell after endosymbiosis has involved an initial recruitment phase for OXPHOS proteins that increased the number of components followed by loss and length reduction made possible by mutual functional compensation^{42,43}. We hypothesized that the OXPHOS pathway might have experienced reduced purifying selection in felids, rodents, marsupials, and galliform birds based on an increased proportion of glycolytic fibers. Duplicate copies or alternative metabolic pathways compensate for gene function and decide gene dispensability⁴⁴. Hence, to evaluate our hypothesis, we aim to (1) investigate whether *COA1* has any homologs that could compensate its function, (2) screen the genomes of vertebrate species to identify and track the evolutionary history of *COA1* orthologs, (3) identify evidence of gene disruptive changes within the *COA1* locus using several types of high-throughput datasets and (4) reconstruct the sequence of events associated with the potential erosion of the *COA1* locus due to chromosomal rearrangement (CR) events at the evolutionary breakpoint region (EBR) spanning the *COA1* gene. We extensively screened publicly available genomes and transcriptomes of more than 365 vertebrate species to establish recurrent loss of the widely conserved *COA1* gene.

Results

***COA1* is a distant homolog of *TIMM21*.** We identified that the *TIMM21* gene is a distant homolog of *COA1* based on Position-specific iterative Basic local alignment search tool (PSI-Blast) and HMM-HMM-based lightning-fast iterative sequence search (HHblits) iterative profile-profile search of the Uniprot database. Clustering *COA1* and *TIMM21* homologous sequences from the Pfam 34.0 database using CLANS finds distinct swarms (see Fig. 1A). The *COA1* homologs form separate clusters for bacteria, fungi, plants, and animals genes. The inset in Fig. 1A shows the sub-clustering of fungi, plants, and animals within the *TIMM21* swarm (see Supplementary Text S1, Supplementary Figs. S1–S9, and Supplementary File S1–S5). The pairwise alignment of the human *COA1* protein sequence with the *TIMM21* protein (from *Saccharomyces cerevisiae*) shows that regions with the most substantial homology include the membrane-spanning domain and cover > 100 residues (see Fig. 1B and Supplementary Fig. S10). In addition to the primary sequence-homology detected, both *TIMM21* and *COA1* are known to play prominent roles in the mitochondria and have comparable secondary structures (see Fig. 1C,D). The strong homology between these proteins also allows for homology-based modeling of the tertiary structure of the *COA1* protein using *TIMM21* as a model (see Supplementary Fig. S9). The de novo predicted structures of *TIMM21* and *COA1* from the AlphaFold⁴⁵ protein structure database are also highly similar (see Supplementary Figs. S11–S16). Despite the lack of well-conserved motifs, we found three well-matching columns (marked with a ']' sign in Fig. 1B) between residues 91 to 95 in *COA1*. Two consecutive conserved residues occur at positions 57–58, 64–65, and 67–68 of *COA1*. The similar sequence, structure, and function of *COA1* and *TIMM21* strongly support that these genes are homologs. The presence of distinct *COA1* and *TIMM21* orthologs in fungi, animals, plants, and protists suggests both homologs existed in the LECA (Last Eukaryotic Common Ancestor).

***COA1* gene duplication, pseudogenisation, and exon reorganization.** The *COA1* gene has undergone independent gene duplications followed by pseudogenisation and degeneration of the duplicated copy in primates, carnivores, and a few rodent species (see Supplementary Text S2A, Supplementary Tables S1–S3, and Supplementary Figs. S17–S48). However, intact transcriptionally active *COA1* gene copies are present in more than 300 vertebrate species. In the cheetah (*Acinonyx jubatus*), the open reading frame at the *COA1* locus identified based on conserved synteny is disrupted (see Fig. 2). The cheetah gene disrupting premature stop codon is due to a single base C → T substitution at the 27th base of exon-2 assembled at the *COA1* locus (see Supplementary Table S1 and Supplementary Fig. S49–S51). The duplicated copy of *COA1* also contains a premature stop codon at the 49th base of exon-2 caused by a single base insertion at the 11th base of exon-2 (see Supplementary Figs. S52–S54). The *COA1* gene transcripts are missing in the skin transcriptome of the cheetah (see Supplementary Figs. S55–S57). We evaluated the possibility of alternative splice isoforms and potential assembly errors in great detail to rule out the case of an intact reading frame. We find multiple lines of evidence support *COA1* gene loss in the cheetah (see Supplementary Text S2B and Supplementary Figs. S19 and S55–S77). Gene loss in the cheetah occurred between 2.63 and 3.06 MYA (Supplementary Table S4).

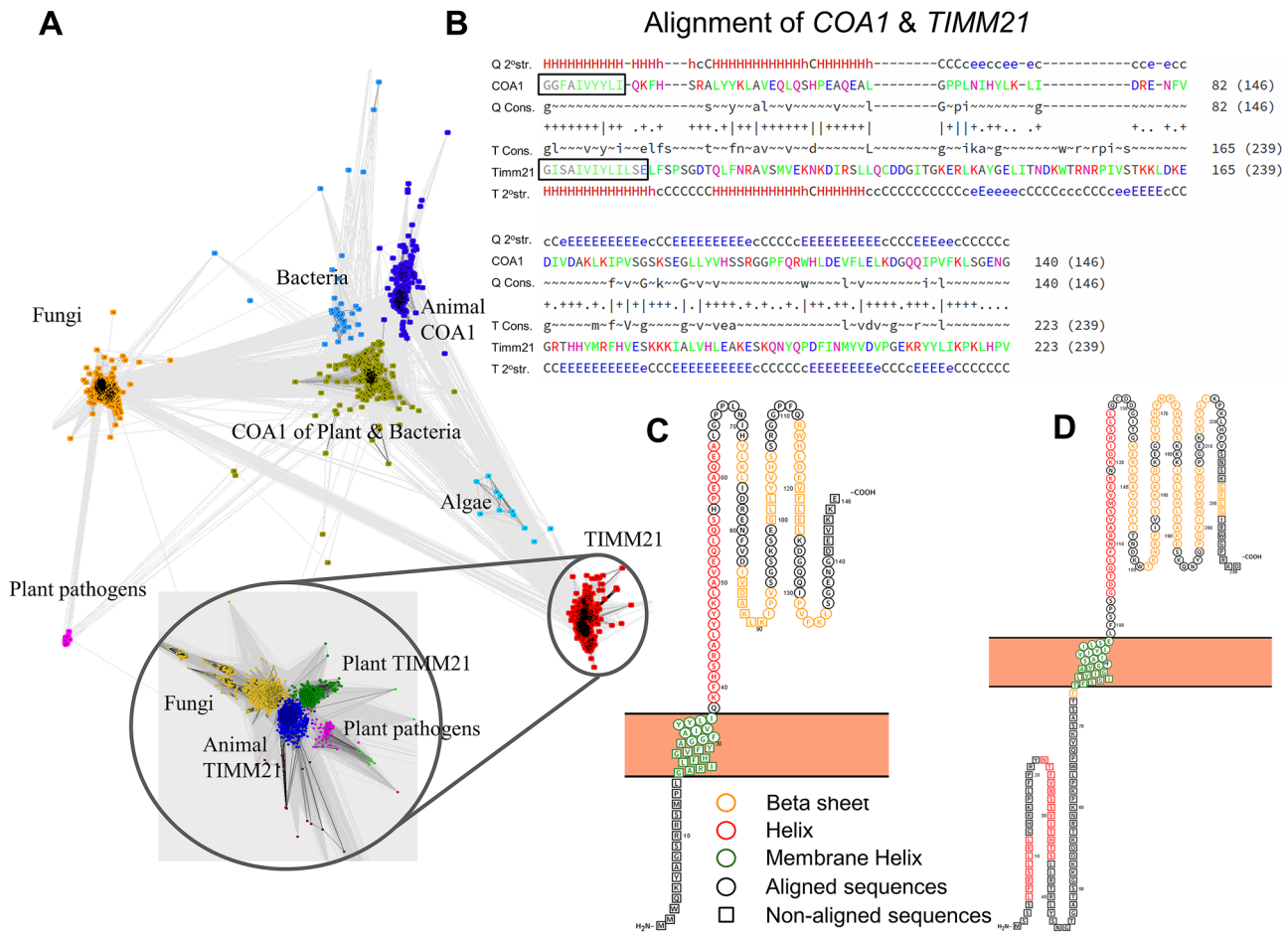


Figure 1. COA1/MITRAC15 and TIMM21 are distant homologs with similar amino acid profiles and secondary structures. (A) CLANS generated a cluster map of COA1 and TIMM21 using amino acid sequences downloaded from the Pfam 34.0 database. Different colors represent distinct clusters: COA1 is dark blue, TIMM21 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 (Supplementary Fig. S5). (B) The output of HHpred shows the alignment of human COA1 with yeast TIMM21. The region in the box highlights the predicted transmembrane helix. The ‘Q’ and ‘T’ refer to the query and template sequences in the alignment. The sequences, ‘2° str.’ and ‘Cons.’ denote the PSI-PRED secondary structure prediction and confidence values where H = helix, E = strand, C = loops, and CAPS = strong prediction in secondary structure prediction. The upper- and lower-case amino acids in the consensus sequences indicate high (> 60%) and moderate (> 40%) conservation, respectively. Symbols indicating the quality of the column-column match: ‘|’ very good, ‘+’ good, ‘?’ neutral, ‘_’ bad and ‘=’ very bad. (C) The PROTEUS2 predicted secondary structure of human (*Homo sapiens*) COA1 is displayed using PROTTER. (D) The PROTEUS2 predicted secondary structure of yeast (*Saccharomyces cerevisiae*) TIMM21 is displayed using PROTTER.

COA1 gene loss in galliform species. We found evidence of eight independent gene-disruption events in the COA1 gene in the galliform group (see Fig. 3A). The chicken (*Gallus gallus*) and Amazonian wood quail (*Odontophorus gujanensis*) have single-base G → T substitutions at the 69th base of exon-2 and the 72nd base of exon-4 in the COA1 gene, respectively (see Supplementary Table S5). These substitutions lead to (GAA → TAA) premature stop codons. Gene loss of COA1 is estimated to be around 32.97 MYA in chicken and approximately 19.84 MYA in the Amazonian wood quail (see Supplementary Tables S4 and S5). In the Indian peafowl (*Pavo cristatus*), two single-base deletions, one at 37th base of exon-1 and another at 31st base of exon-4, introduce frameshifts resulting in premature stop codons in exons 2, 3, and 4. The gene disrupting mutations identified in the Indian peafowl (*Pavo cristatus*) also occur in the green peafowl (*Pavo muticus*). Loss of the COA1 gene is estimated to have happened around 32.97 MYA in the peafowls (see Supplementary Tables S4 and S5). The exon-2 of pinnated grouse (*Tympanuchus cupido*) and helmeted guinea fowl (*Numida meleagris*) have independent 13 and 17 base deletions, which lead to several frameshift induced premature stop codons (see Supplementary Table S5). The 13-base deletion in the exon-2 of the pinnated grouse (*Tympanuchus cupido*) also occurs in Gunnison grouse (*Centrocercus minimus*), rock ptarmigan (*Lagopus muta*), and the black grouse (*Lyrurus tetrrix*)

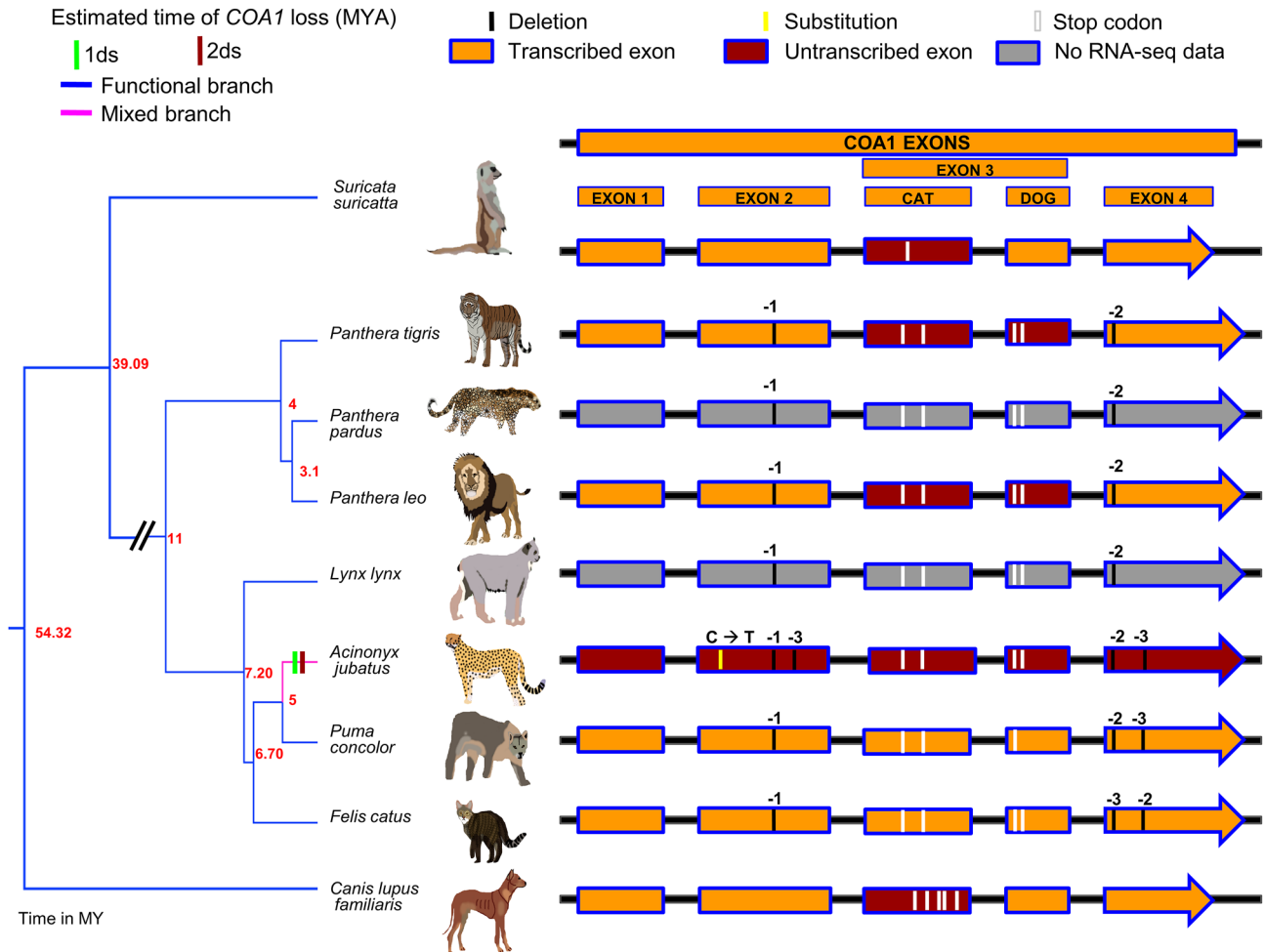


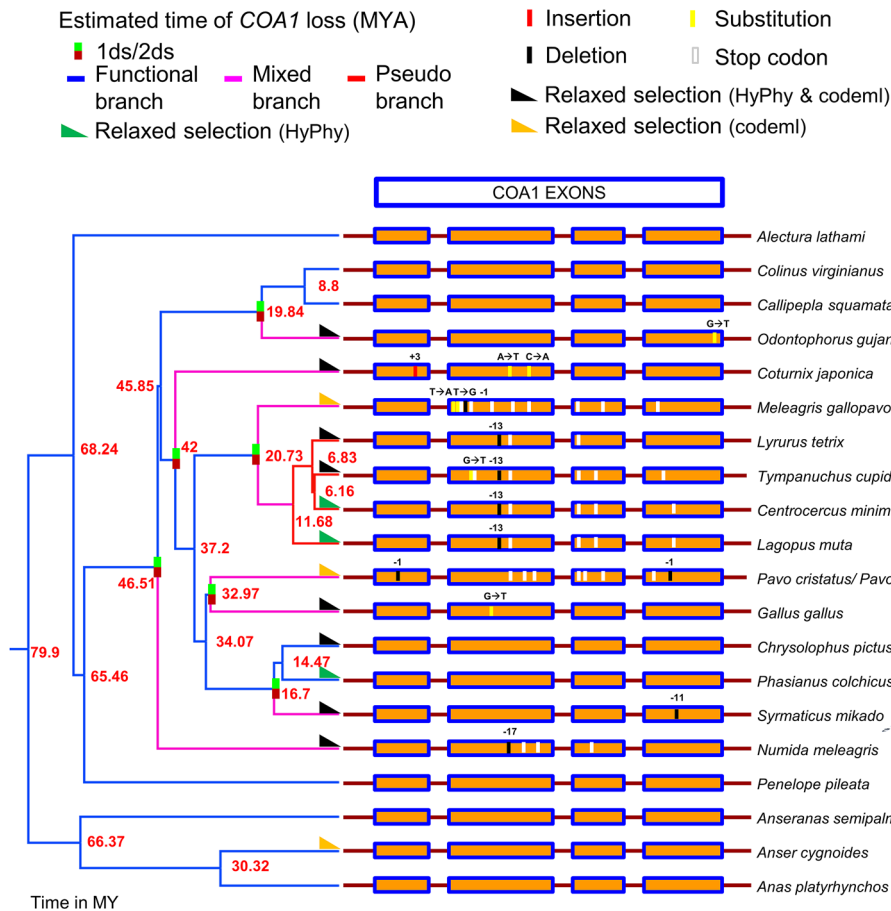
Figure 2. Loss of *COA1/MITRAC15* gene in feliforms. Gene loss event in cheetah (*Acinonyx jubatus*) is visualized using FigTree on a time-calibrated phylogenetic tree downloaded from the TimeTree website. The phylogenetic tree of the Feliformia group is obtained from the tree by Johnson 2006¹¹⁷. Blue branches in the tree represent functional branches, + pink-colored branches represent mixed (functional + pseudogenetic) branches. The method proposed by Meredith et al., 2009 was used to estimate the time of gene loss using two different substitution rate models (1ds and 2ds). Details of the estimated time of gene loss are provided in Supplementary Table S4. The evidence of transcription is shown in Supplementary Figs. S24–S42, S55–S57, S920–S924, S944–S949. The detailed figure showing gene order is in Supplementary Fig. S23.

(see Supplementary Table S5). The estimated time of gene loss in these four species is around 20.73 MYA, and for helmeted guineafowl is around 46.51 MYA (see Supplementary Table S4).

In turkey (*Meleagris gallopavo*), a two-base substitution at bases 7 and 8 and a single base deletion at the 37th base of exon-2 result in a frameshift in the *COA1* gene leading to premature stop codons. Gene loss in turkey is estimated to have occurred around 20.73 MYA (see Supplementary Tables S4 and S5). Two closely spaced single base substitutions (**AAC** → **TAA**) at 48th and 50th positions of exon-2 result in a premature stop codon in the Japanese quail (*Coturnix japonica*). The time of gene loss in the Japanese quail is estimated at around 42 MYA (see Supplementary Tables S4 and S5). The mikado pheasant (*Syrnaticus mikado*) has a frame-disrupting 11-base deletion in exon-4, and the time of gene loss is around 16.7 MYA (see Supplementary Tables S4 and S5). Other galliform species such as Australian brushturkey (*Alectura lathami*), northern bobwhite (*Colinus virginianus*), blue quail (*Callipepla squamata*), ring-necked pheasant (*Phasianus colchicus*), golden pheasant (*Chrysolophus pictus*), and white-crested guan (*Penelope pileata*) have intact *COA1* coding sequences. The coding region is intact in outgroup species such as swan goose (*Anser cygnoides*), duck (*Anas platyrhynchos*), and magpie goose (*Anseranas semipalmata*). Five genes upstream (*BLVRA*, *VOPPI*, *LANCL2*, *EGFR*, and *SEC61G*) and downstream (*STK17A*, *HECW1*, *MRPL32*, *PSMA2*, and *C7orf25*) from *COA1* retain a conserved order in birds (see Supplementary Table S6). We relied upon this conserved order to verify the 1-to-1 orthology of the *COA1* gene across species (see Supplementary Fig. S78).

Signatures of relaxed selection identified by molecular evolution tools (see “Materials and methods”) in galliform species with gene disrupting changes further support the loss of *COA1* in these lineages (see Supplementary Tables S7, S8, and S9). To evaluate the relevance of the gene disrupting mutations and signatures of relaxed selection identified in galliform species, the transcriptomes of Galloanserae species were screened to assess the

A Gene loss events



B Expression pattern

□ No data
 ■ Expression
 □ No Expression

Figure 3. Recurrent loss of *COA1/MITRAC15* gene in galliform species. **(A)** Gene loss events in eleven galliform species along a time-calibrated phylogenetic tree from the TimeTree website are visualized in FigTree. Blue branches in the tree represent functional branches, the pink-colored branches represent mixed (functional + pseudogenic) branches, and red-colored branches represent pseudo branches. The method proposed by Meredith et al., 2009 was used to estimate the time of gene loss using two different substitution rate models (1ds and 2ds). Details of the estimated time of gene loss are provided in Supplementary Table S4. Short colored bars depict the locations of the gene disrupting mutations on the four exons of *COA1* (Supplementary Table S5). **(B)** The *COA1* gene expression pattern in six tissues (brain, spleen, skin, liver, gonad, and blood) was assessed by screening RNA-seq datasets (Supplementary Figs. S466–S626). 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. A black right-angle triangle represents lineages detected to be under relaxed selection by both HyPhy RELAX and codeml. Those detected only by codeml are shown as orange right-angled triangles, and those detected only by HyPhy RELAX are shown as dark green right-angled triangles (Supplementary Tables S7, S8, and S9).

transcriptional status of *COA1*. We analyzed RNA-seq datasets of chicken tissues and found that the *COA1* gene is not transcribed despite screening more than 20 tissues (see Fig. 3B). Other Galloanserae species have RNA-seq data available for very few tissues. We evaluated the RNA-seq datasets from six tissues (Brain, Spleen, Skin, Liver, Gonad, and Blood) available in several species for the presence of *COA1* transcripts. Our search consistently revealed transcription of *COA1* gene in Anseriformes species in contrast to lack of transcription in galliform species except for Australian brushturkey (*Alectura lathamii*) and northern bobwhite (*Colinus virginianus*), which have intact *COA1* gene that is under strong purifying selection (see Fig. 3B).

Despite intact coding regions, the ring-necked pheasant (*Phasianus colchicus*) and golden pheasant (*Chrysolophus pictus*) *COA1* sequences have signatures of relaxed selection (see Supplementary Table S7). None of the four tissues (Brain, Spleen, Liver, and Gonad) for which RNA-seq data is available from the ring-necked pheasant shows any *COA1* transcripts. Similarly, the one tissue (Skin) for which RNA-seq data is available in the golden pheasant (*Chrysolophus pictus*) does not show *COA1* expression. The lack of gene expression and signatures of relaxed selection in the ring-necked pheasant (*Phasianus colchicus*) and golden pheasant (*Chrysolophus pictus*) suggests gene loss.

We evaluated the link between gene loss and the percent of glycolytic muscle fibers using comparative quantitative phylogenetic methods (see “Materials and methods”). Galliform species with gene loss tend to have >60%

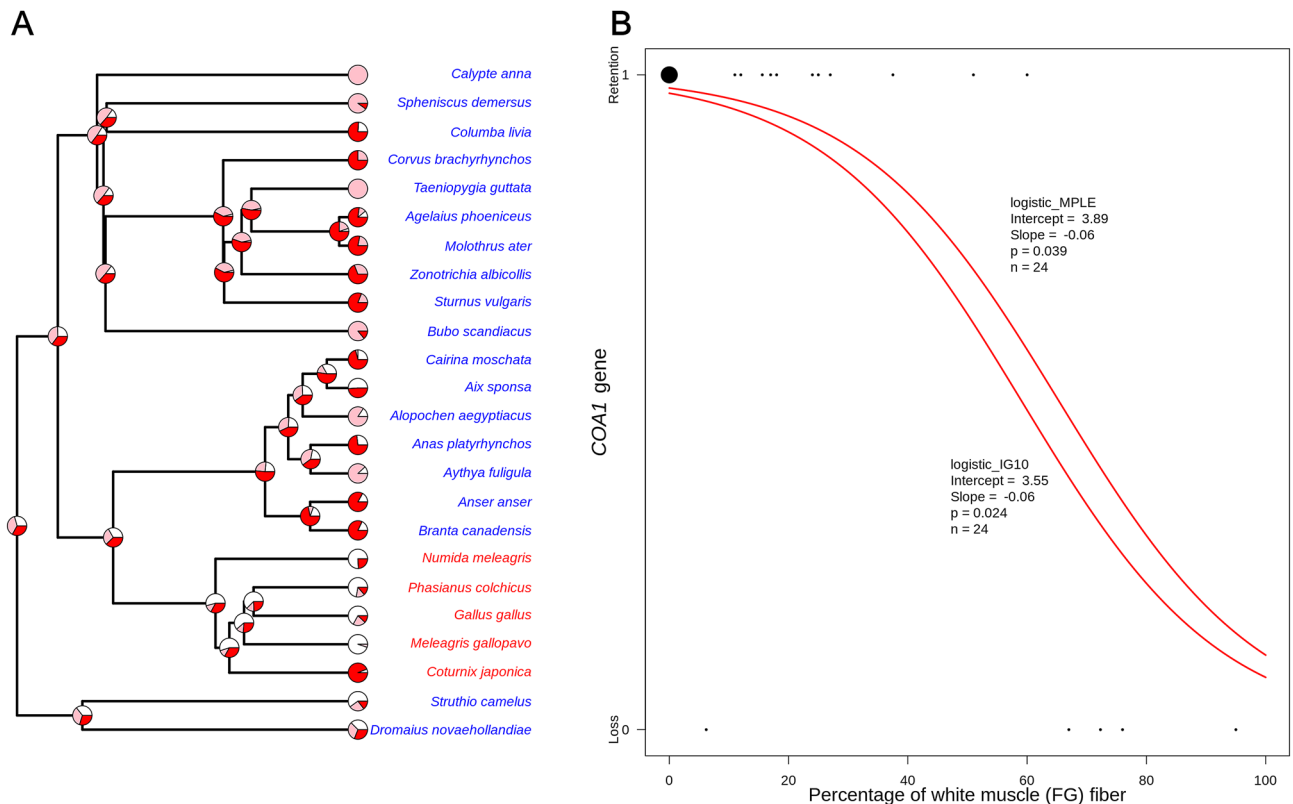


Figure 4. Phylogenetic correlation of skeletal muscle fiber types and loss/retention of *COA1/MITRAC15* gene in different bird species. **(A)** Proportion of white, pink, and red muscle fiber in pectoralis muscle depicted in pie charts (for detailed data, see Supplementary Table S10A). The ancestral state of muscle fiber type was estimated using the *phytools* package version 0.7–90¹¹⁸ in R version 4.1.0¹¹⁴. Species having lost *COA1* represented in red, species retaining *COA1* shown in blue. The tree is not time-calibrated and is downloaded from the TimeTree website and visualized in *phytools* package in R. **(B)** Two methods (logistic_MPLE and logistic_IG10) were used for phylogenetic regression analyses to evaluate the effect of the proportion of white fiber (fast glycolytic; FG) on the loss/retention of the *COA1* gene ($n = 24$). The 8 species without white fibers are represented by the big black dot, all others are shown as tiny individual dots along the FG.

glycolytic fiber in the pectoralis muscle (see Fig. 4A and Supplementary Table S10A). However, despite losing the *COA1* gene, the Japanese quail (*Coturnix japonica*) has less glycolytic muscle fiber than other galliform birds. The Japanese quail is also the only galliform species with flapping flight⁴⁶. Closer inspection of related species such as the common quail (*Coturnix coturnix*) and red-legged partridge (*Alectoris rufa*) suggests the *COA1* was potentially lost in an ancestral species with a higher glycolytic muscle fiber content (see Supplementary Table S10A). The lineage leading to the Japanese quail subsequently reduced the white fiber in the pectoralis muscle. However, the quail is unlikely to regain the gene due to Dollo's law of irreversibility⁴⁷.

Our results recovered a significant negative correlation between the proportion of white-fast-glycolytic muscle fiber type in pectoralis muscle and *COA1* gene retention (see Fig. 4B and Supplementary Table S10B). This negative correlation suggests that a high proportion of white muscle fiber may have resulted in the loss of *COA1*. Notably, non-galliform flightless birds retain an intact *COA1* gene, are transcriptionally active, and lack signatures of relaxed selection (see Supplementary Text S3 and Supplementary Table S11). Hence, the loss of the *COA1* gene is not associated with flying ability but rather with the proportion of white muscle fiber.

***COA1* occurs in an evolutionary breakpoint region.** We find evidence of *COA1* gene disrupting mutations and lack of gene expression in multiple RNA-seq datasets despite a conserved gene order in the rabbit (*Oryctolagus cuniculus*), naked mole-rat (*Heterocephalus glaber*), and four Sciuridae species (*Urocyon parryii*, *Spermophilus dauricus*, *Ictidomys tridecemlineatus*, *Marmota marmota marmota*). The gene disrupting mutations identified in the rabbit *COA1* gene includes a two-base pair deletion at the 22nd codon of exon-1 and single base pair deletions in exon-2 at the 13th and 37th codons, respectively. Gene disrupting changes in the third exon consist of a five-base insertion between the 11th and 12th codon, one base insertion at the 17th codon, and one base deletion in the 23rd codon (see Fig. 5 and Supplementary Table S12). These six gene-disrupting changes result in premature stop codons in exon-2 and exon-4. Gene loss in the rabbit is estimated to have occurred between 4.64 MYA and 6.39 MYA (see Fig. 5 and Supplementary Table S4). The lack of *COA1* RNA-seq reads in tissues such as the brain, liver, and testis that express *COA1* in closely related species supports the loss of the *COA1* gene in the naked mole-rat. Besides the lack of a start codon, a single gene disrupting mutation is found in

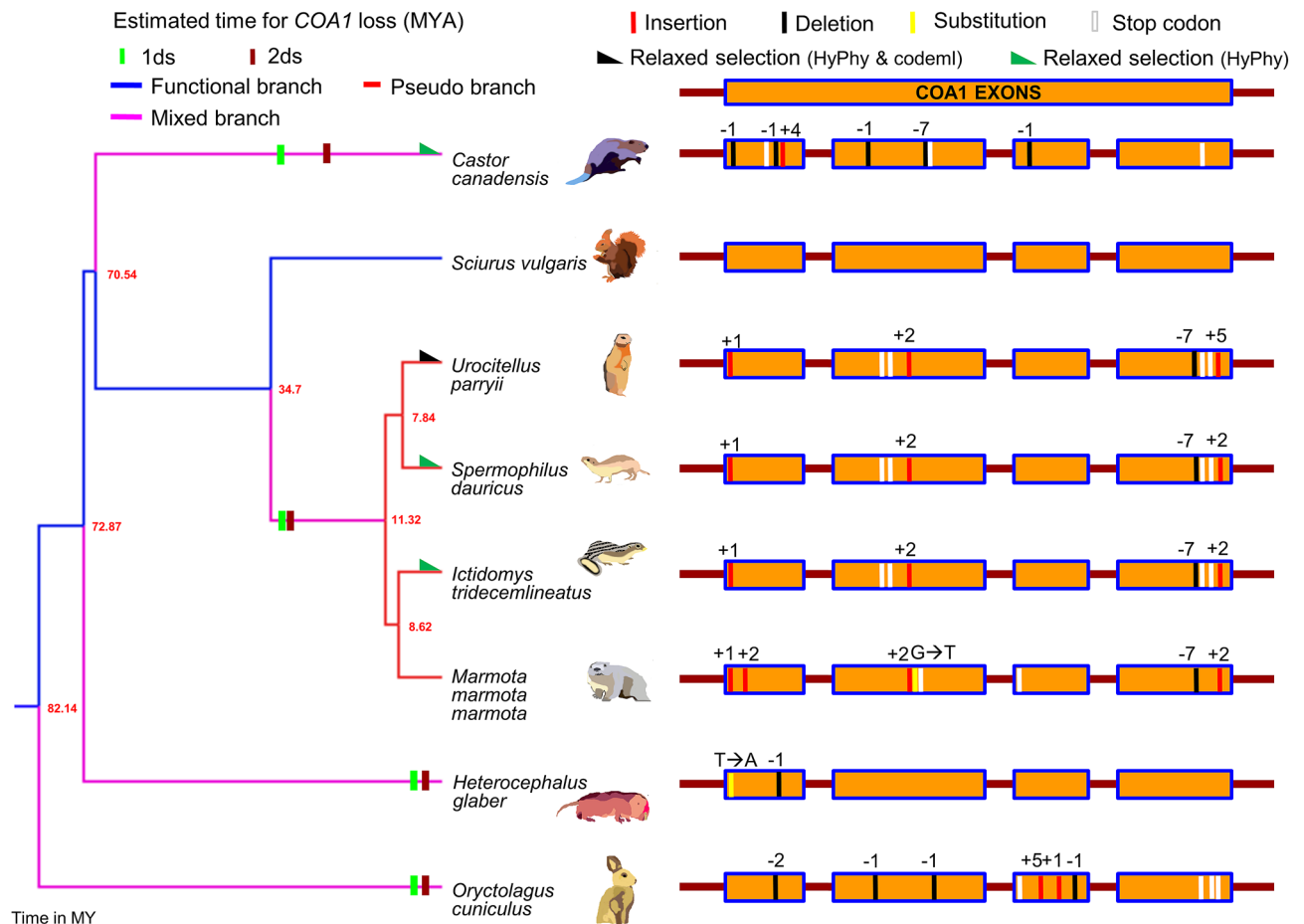


Figure 5. Recurrent loss of *COA1/MITRAC15* gene in rodent species. Gene loss events in seven rodent species through four events are represented exon-wise. Blue branches in the tree represent functional branches, pink-colored branches represent mixed (functional + pseudogenetic) branches, and red-colored branches represent pseudo branches. The gene disrupting mutations are shown for these species in *COA1* exon-wise except for *Sciurus vulgaris*, where *COA1* is functional. Details of the estimated time of gene loss are provided in Supplementary Table S4. The events are provided in detail in Supplementary Table S12. A black right-angle triangle represents lineages detected to be under relaxed selection by both HyPhy RELAX and codeml. Those detected only by codeml are shown as orange right-angled triangles, and those detected only by HyPhy RELAX are shown as dark green right-angled triangles (Supplementary Tables S7, S8, and S9). The time-calibrated phylogenetic tree was obtained from the TimeTree website and visualized in FigTree.

the naked mole-rat *COA1* gene and consists of a single base deletion at the 21st codon of exon-1. Gene loss in the naked mole-rat is estimated between 4.53 MYA and 6.12 MYA (see Supplementary Tables S4, S12, and Fig. 5).

The presence of common gene disrupting changes such as the one base pair insertion at second codon of exon-1, two base pair insertion at 25th codon of exon-2, seven base pair deletion between 25 and 26th codon of exon-4, and a 2-base insertion at 33rd codon of exon-4 supports a shared gene loss in four Sciuridae species (*Urocitellus parryii*, *Spermophilus dauricus*, *Ictidomys tridecemlineatus*, *Marmota marmota marmota*). The *COA1* gene of alpine marmot has additional gene disrupting changes consisting of a 2-base insertion between the 8th and 9th codon of exon-1 and a single nucleotide substitution at the 26th codon of exon-2. The 2-base insertion at the 33rd codon of exon-4 has extended to a five-base pair insertion in the Daurian ground squirrel (*Spermophilus dauricus*). The estimated time of gene loss for this shared event is between 30.62 MYA and 31.74 MYA (see Supplementary Tables S4, S12, and Fig. 5). The presence of intact open reading frames robustly expressed at syntenic locations in closely related species (~30–50 million years) strongly supports at least three independent *COA1* gene loss events (i.e., one in rabbit, one in the four Sciuridae species, and one in the naked mole-rat; see Fig. 5).

Multiple gene-disrupting mutations in the *COA1* gene of the North American beaver (*Castor canadensis*) suggest a fourth independent gene loss event. Gene-disrupting mutations in the beaver result in at least two premature stop codons. In the first exon, single-base deletions occur in the 3rd and 20th codon, a four-base insertion occurs between 33rd and 34th codon. The second exon has a single-base deletion in the 33rd codon and a seven-base pair deletion between 29 and 30th codons. A single base deletion occurs at the 12th codon of exon-3 (see Supplementary Table S12 and Fig. 5). The Illumina sequencing raw reads support the gene disrupting mutations identified in the genome assembly (Supplementary File S6), and duplicate copies do not occur.

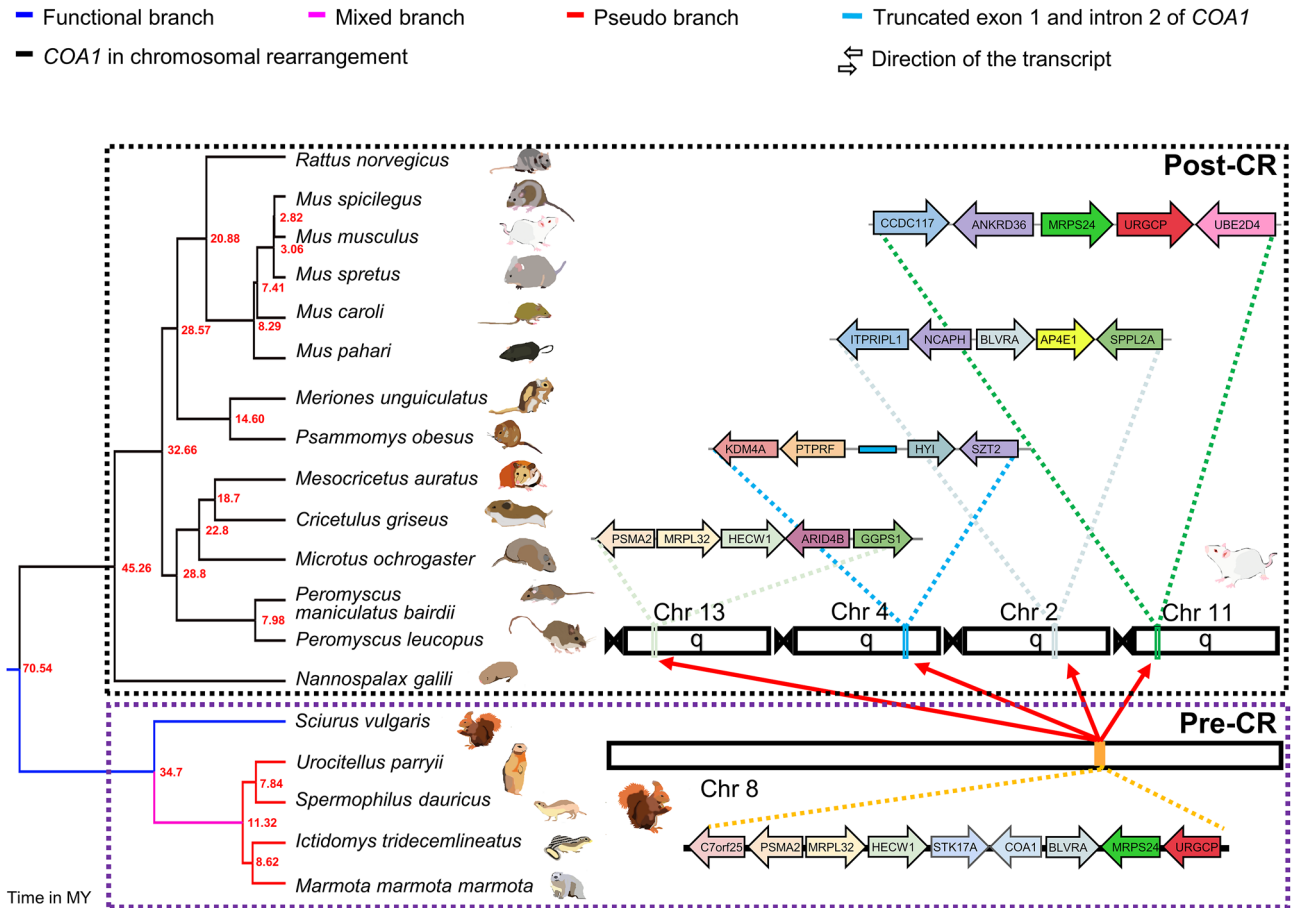


Figure 6. Recurrent loss of *COA1/MITRAC15* gene in rodent species. Gene loss events in fourteen rodent species are shown to occur through a chromosomal rearrangement (CR) event. Blue branches in the tree represent functional branches, the pink-colored branches represent mixed (functional + pseudogenic), and red-colored branches represent pseudogenic branches. The black branches correspond to the species that have lost the gene due to the rearrangement at the evolutionary breakpoint region (EBR). Gene order in the genomic region flanking the *COA1* gene in rodent species is shown with arrows that 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. The black (post-CR) and purple (pre-CR) dotted boxes represent post- and pre-chromosomal rearrangement gene orders, respectively. The *Mus musculus* chromosomes are used to represent the post-chromosomal rearrangement. Chromosomes 4 and 2 of *Mus musculus* contain partial remains of the *COA1* gene and a functional *BLVRA* gene, respectively. A solid sky-blue colored line depicts the partial exon-1 and intron 2 of *COA1* located between the *PTPRF* and *HYI* genes. The pre-chromosomal rearrangement is represented by the *COA1* located at chromosome 8 of *Sciurus vulgaris*. The gene order conservation in each of the rodent species is shown in Supplementary Fig. S179. The time-calibrated phylogenetic tree was obtained from the TimeTree website and visualized in FigTree.

The loss of the *COA1* gene in the beaver is estimated to have happened sometime between 26.30 MYA and 32.40 MYA (see Supplementary Table S4 and Fig. 5).

Repetitive elements and lack of long-read sequencing data in most rodent species prevent genome assembly verification. Hence, we have screened newly available unannotated genomes of several closely related rodent species. These genome assemblies were verified using long-read sequencing data or cloned fragments that cover parts of the genome when available (see Supplementary Text S4A, Supplementary Table S13, and Supplementary Figs. S79–S178). Gaps in the genome assembly also hamper the identification of the correct gene order. Previous reports that examined genome assemblies and EST data have claimed loss of the *STK17A* gene in mice due to a CR spanning this genomic region corresponding to chr7p13 in the human genome⁴⁸. Detailed examination of gene order flanking the *COA1* locus in several rodent genomes revealed the occurrence of this previously reported CR event (see Fig. 6 and Supplementary Fig. S179). Example gene orders for pre-CR (European red squirrel) and post-CR (mouse) species illustrate the movement of genes to four different chromosomes (see Fig. 6). Rather than a simple rearrangement, this appears to be a scattering of short genomic regions to other parts of the genome. We refer to this complex event that involved the movement of genes to more than two new chromosomes as a "gene-scattering" rearrangement.

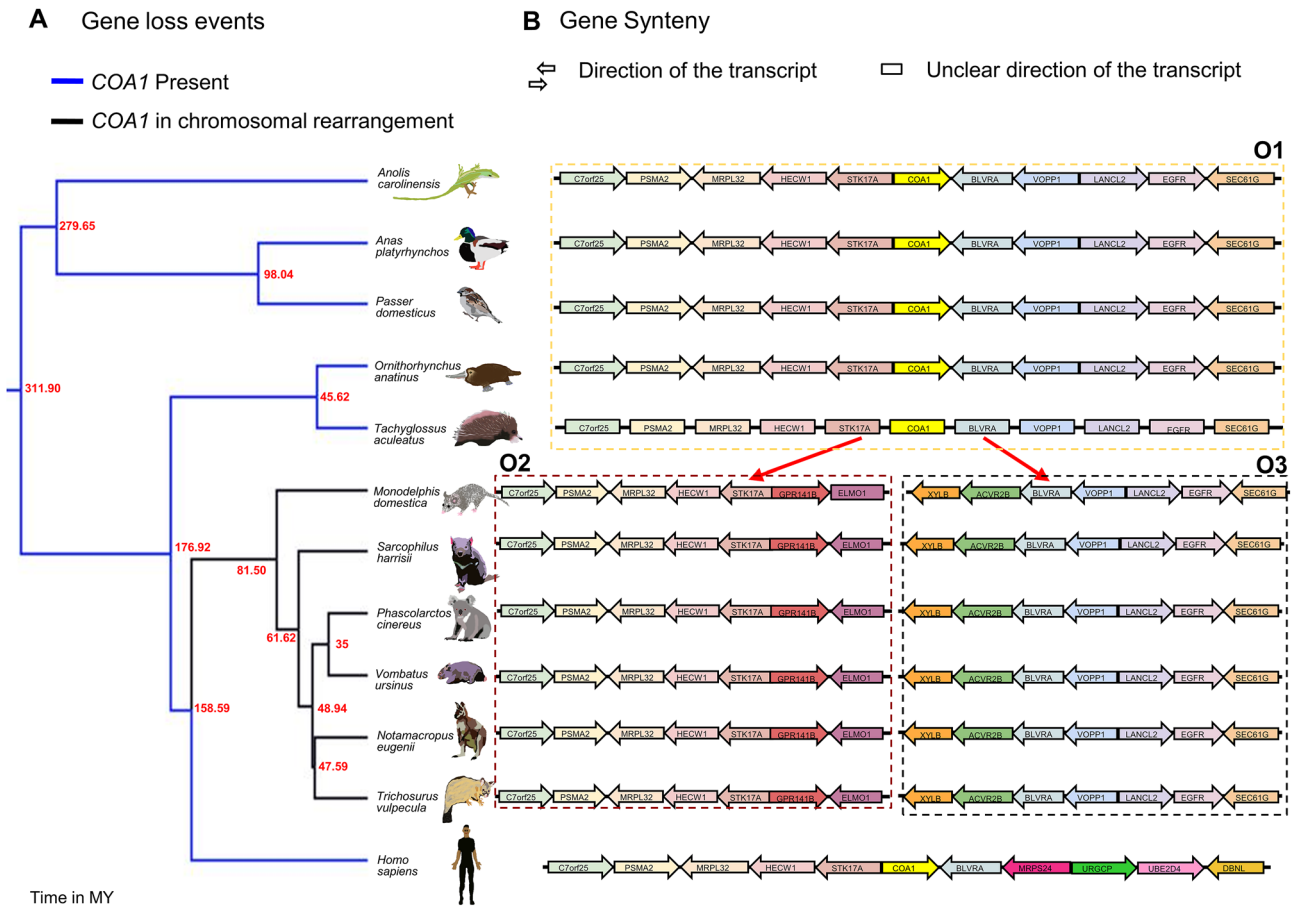


Figure 7. The genomic region spanning the *COA1/MITRAC15* gene coincides with an evolutionary breakpoint (EBR). (A) The phylogenetic relationship between marsupial species along with a few outgroup species. The time-calibrated phylogenetic tree is downloaded from the TimeTree website and visualized in FigTree. (B) The gene order in the region flanking the *COA1* gene. The arrows show the direction of gene transcription relative to the *COA1* gene for consistency across species. Each dotted box contains one type of gene order, and the red arrows from gene order O1 depict the chromosomal rearrangement (CR) event that leads to gene orders O2 and O3 in the six marsupial species. The outgroup species have the pre-CR gene order O1. In the post-CR species, we see gene orders O2 and O3. A functional *COA1* can be identified in the outgroup species but is presumably lost in marsupial species as it is missing in the genome assembly and raw read datasets.

Identifying gene loss events coinciding with Evolutionary Breakpoint Regions (EBRs) is notoriously challenging and has motivated nuanced inferences in bird⁴⁹ and rodent species⁵⁰ (see Supplementary Text S4B, Supplementary Tables S14–S17, and Supplementary Figs. S180–S228). Nonetheless, more than a dozen rodent species share the putative combined loss of *STK17A* and *COA1* (see Fig. 6). Based on the presence of adjacent genes, the rearranged regions could be tracked down to four different chromosomes (see Fig. 6 and Supplementary Figs. S179, O6, and O9). Genes on the left flank of *STK17A-COA1-BLVRA* consist of *PSMA2*, *MRPL32*, and *HECW1* in gene orders O1 to O5 (see Supplementary Fig. S179). After the CR, the same sequence of genes can be found in gene order O9 and occur adjacent to *ARID4B* and *GGPS1*. Genes on the right flank of *STK17A-COA1-BLVRA* consist of *MRPS24*, *URGCP*, and *UBE2D4* in gene order O4. Several other gene orders (O1–O5) occur on the right flank in various species. The sequence of genes found on the right flank in gene order O4 is also found sequentially in gene order O6 and occurs adjacent to *ANKRD36* and *CCDC117* after the CR. We found that the *BLVRA* gene has translocated to an entirely new location and does not co-occur with either the left or right flank. However, the new location of the *BLVRA* gene between the *NCAPH* and *ITPRIPL1* genes on the left flank and *AP4E1* and *SPPL2A* genes on the right flank is consistently conserved across all 14 post-CR species and corresponds to gene order O7. The remnants of *COA1* occur in a gene desert region between *PTRPF* and *HYI* genes in most post-CR species (see Supplementary Fig. S179, Supplementary Text S4C and Supplementary File S7).

Comparison of gene order in marsupial species with various outgroup species (including the platypus and short-beaked echidna from the order Monotremata) identified the presence of an independent CR event spanning the *COA1* locus (see Fig. 7). In contrast to the rodent-specific EBR, we found that the *STK17A* gene is intact in post-CR (gene order O2 and O3 in Fig. 7) marsupial species. However, an extensive search of marsupial genomes, transcriptomes, and sequencing datasets (including high coverage Pacbio datasets for the Koala) failed to find any evidence of *COA1* orthologs or their remnants. Lack of sequencing reads from *COA1* in marsupial species

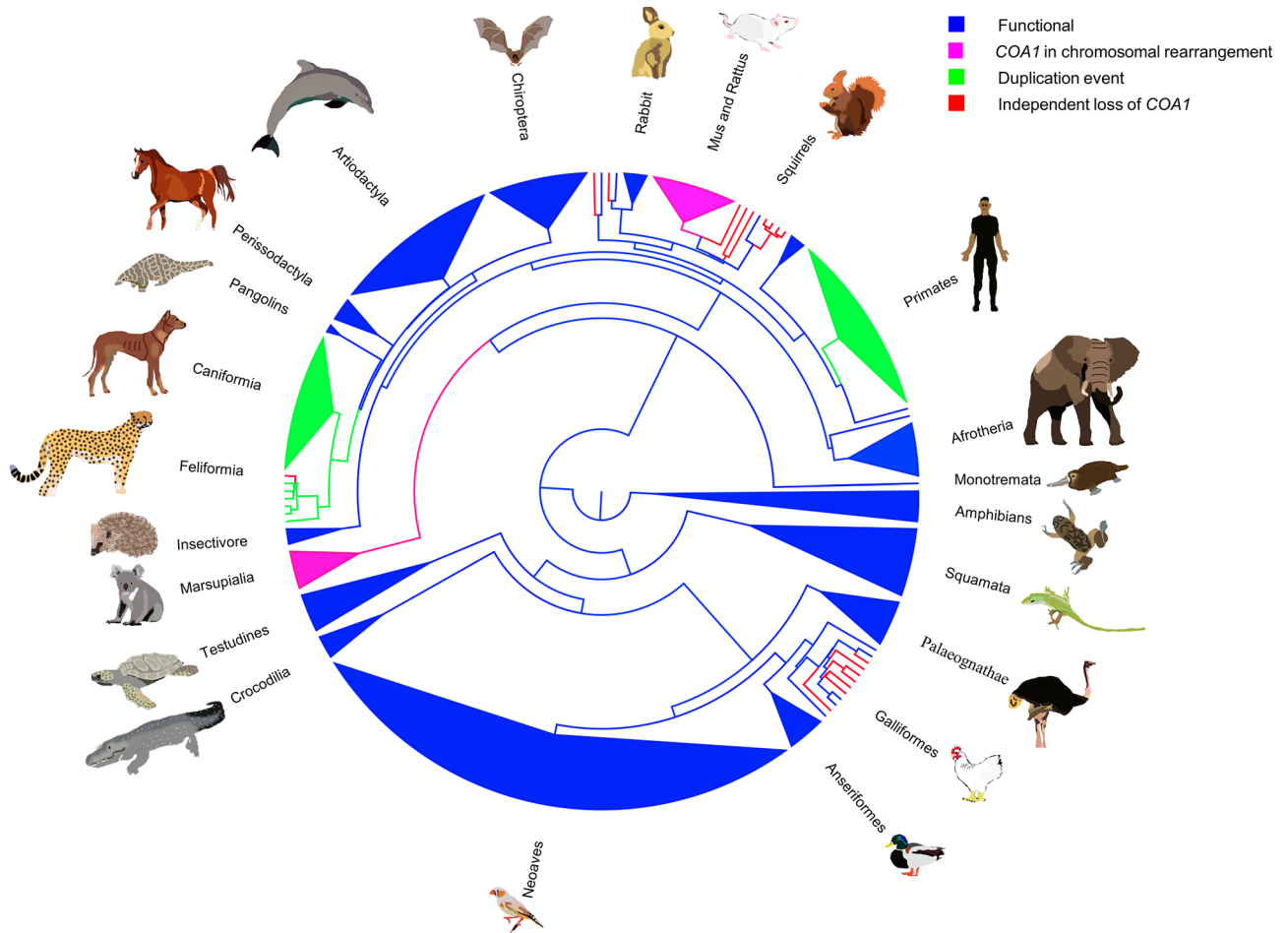


Figure 8. The status of the *COA1/MITRAC15* gene in various clades is shown in the phylogenetic tree obtained from the TimeTree website and visualized in FigTree. The total number of species used is 276. Different colors represent different types of events. Blue represents the intact *COA1* gene, and pink represents the *COA1* gene loss due to the chromosomal rearrangement (CR) at the evolutionary breakpoint region (EBR), green represents the duplication event of the *COA1* gene, and red represents the independent loss of the *COA1* gene.

suggests either complete erosion of the gene or drastic change in sequence composition that eludes homology detection tools or sequencing with currently available technologies. We found the *COA1* gene is intact in most species, is independently lost in multiple lineages, has undergone duplication and pseudogenisation in different clades, and is potentially lost due to CRs (see Fig. 8).

Discussion

Our search of the sequence databases identified that *COA1* and *TIMM21* are distant homologs with representative genes found in animals, plants, fungi, and protists. Earlier studies have noted that *COA1* resembles *TIMM21*, a subunit of the TIM23 complex⁵¹. Hence, it is plausible that *COA1* and *TIMM21* are the results of ancient gene duplication. *TIMM21* gene duplicates that interact with the mitochondrial import apparatus and respiratory chain complexes occur in *Arabidopsis*⁵². The *COA1* gene has undergone duplication events in carnivores, primates, and a few rodent species. The prevalence of such duplication events suggests that a higher *COA1* protein dosage is not harmful or that sophisticated regulatory machinery exists to maintain the correct dosage. Genes with duplicated copies have greater flexibility for subfunctionalization or neofunctionalization⁵³. The ancient origin of *COA1* and *TIMM21* has resulted in considerable sequence-level divergence and changes in the protein structure, suggesting that they may not be able to compensate for each other completely.

In contrast to gene duplication, the origin of new splice-isoforms increases the transcriptome complexity without increasing the gene count. The evolution of phenotypic novelty through alternative splicing has received greater attention thanks to the availability of large-scale transcriptomic and proteomic datasets in diverse species⁵⁴. While positive selection has a role in specific examples of alternative splicing^{55,56}, the vast majority of splicing is probably noisy, and neutral processes may explain its evolution⁵⁷. Alternative splicing also reduces premature protein truncation due to purifying selection⁵⁸. In the case of felid species, the alternative splicing of the third exon (see Fig. 2) may have evolved in response to the gene-disrupting changes. Verifying the relevance of the alternative splicing observed at the transcriptional level would require further scrutiny of the protein level isoforms of the *COA1* gene in felid and canid species. In primate species, the potential addition of the

extra coding-exon occurs by a shift of the start codon into the untranslated region. Such changes at the reading frame termini occur when the gene is under relaxed selective constraints⁵⁹. Acquisition of novel protein-coding sequences through changes in the exon length is also known to occur⁶⁰. We speculate that drastic lineage-specific changes in purifying selection have allowed for changes in intron–exon structure resulting in the evolution of new splice-isoforms of *COA1*.

The fast glycolytic (FG) fibers use carbohydrates, while fast oxidative glycolytic fibers (FOG) use lipids. Flight degenerate birds have a greater proportion of FG fibers, while the flying birds have mostly FOG fibers. Overall, flight degenerate birds such as galliform species utilize carbohydrates as substrates compared to lipids by flying species⁶¹. This switch of the primary energy source from lipids to carbohydrates has been linked to convergent amino acid changes in the *ATGL* and *ACOT7* genes of flight-degenerate birds¹⁴. This difference in the substrate used also leads to differing levels of reactive oxygen species (ROS).

In the OXPHOS pathway, the electrons from the electron transport chain end up in an oxygen molecule which is usually reduced to produce water at complex IV. However, a small fraction (0.1–2%) of the electrons leak at complexes I and III resulting in partial reduction of oxygen to form superoxide (O_2^-)⁶², a precursor of most other ROS⁶³. The $FADH_2/NADH$ or F/N ratio (ratio of electrons entering the respiratory chain via $FADH_2$ and $NADH$) correlates with ROS production^{64,65}. The breakdown of glucose results in a lower F/N ratio compared to the breakdown of fatty acids^{64,66}. Both the F/N ratio and radical formation increase with an increase in the fatty acid length^{65,66}. The generation of internal ROS in the eukaryotic cell has given rise to numerous adaptations to deal with damage caused by ROS⁶⁷. Thus, lipid utilizing FOG muscle fibers have a higher ROS level than carbohydrate using FG fibers. The lower ROS levels in galliform species that predominantly have FG fibers would result in relaxed selective constraints on the machinery involved in coping with ROS. Hence, *COA1* may have a role in ROS detoxification.

Radical formation due to catabolism of substrates with a high F/N ratio is countered by adaptations such as peroxisomes, uncoupling proteins, mitophagy, carnitine shuttles, complex adjustments, and supercomplex formation^{64,66,68}. One intriguing possibility is that changes in substrates lead to varying selection pressure on these processes. For instance, uncoupling protein 1 (*UCP1*) or thermogenin expressed in the inner mitochondrial membrane facilitates the regulated leakage of protons to generate heat in brown adipose tissue⁶⁹. The *UCP1* gene is absent in all birds⁷⁰ and some mammals^{71,72} despite its presence in fishes⁷³, amphibians⁷⁴, and marsupials⁷⁵. The integration of *UCP1* in the thermogenic pathway is considered a eutherian-mammal-specific adaptation unrelated to its ancestral innate immune functions⁷⁶. Repeated loss of *UCP1* in vertebrate lineages appears to result from changing functional roles^{71,77}. Genes involved in other processes countering radical formation could also have experienced changing selection regimes. The higher-order organization of the mitochondrial electron transport chain forms supercomplexes (SCs)⁷⁸. The ratio of SCs is known to vary between species and tissues⁷ and in response to metabolic demands of the cell⁷⁹. While the SCs are highly stable in reptiles, they are unstable in birds and mammals⁸⁰. We surmise that the loss of *COA1* in various bird and mammal species might reflect a change in the ratio of SCs.

Our computational analysis of more than 300 vertebrate genomes has found that the *COA1* gene is intact and transcribed in most (~90%) species, except for the cheetah, galliform, rodent, and marsupial species. Notably, the detailed investigation of the *COA1* gene in non-galliform bird species that are flightless or have a limited flying ability has found an intact transcribed gene. These non-galliform flightless birds tend to have a greater reliance on the hind limb muscles and have comparatively less white fiber in the skeletal muscle of the pectoralis. Therefore, the *COA1* gene loss appears to be associated with skeletal muscle fiber composition changes towards a higher proportion of white fiber in the pectoralis. Similar to galliform birds, the increase in the proportion of white fiber in the skeletal muscles of rodents compared to primates and felids compared to canids appears to be a result of changing locomotor needs. The prominent role of mitochondria in skeletal muscles is evident from locomotory disorders of the muscle tissue caused by defects in mitochondrial genes⁸¹. While our data support the loss of the *COA1* gene following relaxed selection on the OXPHOS pathway, other implications of gene loss also exist (see Supplementary Text S5).

EBRs are genomic regions that have undergone one or more structural changes resulting in altered karyotypes between lineages⁸². Recurrent non-random structural changes at the same genomic regions in multiple lineages potentially occur due to repeat elements^{83,84}, chromosome fragile sites^{85–87}, nucleotide composition, methylation level⁸⁸, and chromatin state^{89,90}. However, the prevalence of EBRs and their relevance to evolutionary processes has been the focus of considerable debate^{91–93}. Several lineage-specific gene loss events near EBRs in rodents are due to CRs^{48,94}. Notably, one of these lost genes, *STK17A*, is adjacent to the *COA1* gene. This genomic region containing *STK17A* and *COA1* genes has been implicated in a large-scale GWAS for longevity⁹⁵. The co-occurrence of a rearrangement with putative *COA1* gene loss in rodents and marsupials is very intriguing. However, rodent genomes have mutational hotspots with high lineage-specific GC-biased gene conversion (gBGC), resulting in a substantial gene sequence divergence⁵⁰. Such highly diverged orthologs can be challenging to identify due to difficulties in sequencing high GC regions. The overall magnitude of gBGC is relatively low in *COA1*, especially in rodents. Moreover, we find remnants of *COA1* in several post-CR rodent species that suggest actual gene loss. Several pre-CR rodent species have also independently accumulated gene disrupting mutations in the *COA1* gene. Hence, the *COA1* gene appears to be under relaxed selective constraint even before the rearrangement.

Conclusions

COA1 is a distant homolog of the *TIMM21* gene that has undergone recurrent gene loss in several galliform species (16.7–46.51 MYA) and rodent species (4.53–32.40 MYA). The gene loss occurs in species that rely primarily on glycolytic muscle fibers to achieve short bursts of activity. We show that *COA1* and the adjacent *STK17A* gene are located at an EBR and are missing from the genomes of several rodent species following CR. The *COA1* gene

has undergone duplication in carnivores and primates, followed by pseudogenization of one copy and divergence of the other copy in its intron–exon structure. The prevalence of repeated gene loss and duplication events in the history of *COA1* demonstrates the occasional context-dependent dispensability of this gene.

Materials and methods

Finding homologs of *COA1*. The amino acid sequence of the human *COA1* gene was used as a query in PSI-BLAST⁹⁶ against the non-redundant protein sequence database with eight iterations to identify homologs. Similarly, the human *COA1* amino acid sequence was used as the query in the program HHblits of HHsuite^{97,98} with the flags "-e 1e-3 -n 8 -p 20 -Z 5000 -z 1 -b 1 -B 5000 -d UniRef30_2020_06". The output from HHblits was used as an input to the CLANS program⁹⁹ with an e-value cut-off of 1e-4 to cluster the blast hits using the MPI Bioinformatics Toolkit^{100,101}. We ran the CLANS java application for more than 50,000 rounds on the webserver output to ensure stable clusters. Manual inspection of gene annotations allowed identification of each of the groups. Subsequently, we performed the HHblits search again with different settings such as "-glob" for global alignments and "-loc" for local alignments. Manually curated multiple sequence alignment of *COA1* open reading frames from 24 primate species was also separately used as a query for better sensitivity. The amino acid sequence of *TIMM21* provides a consistent hit with different search settings and databases.

Validation of *COA1* annotation. Despite being a fast-evolving gene, orthologs of *COA1* can be identified based on gene synteny and sequence identity. However, identifying *COA1* orthologs between distantly related species is challenging¹⁰². We screened the genome assemblies and annotations available on NCBI and Ensembl for *COA1* (*C7orf44* or *MITRAC15*) protein-coding transcripts. The *COA1* gene orthologs have been annotated in almost 300 vertebrate species (see Supplementary Table S18). However, the number of exons and the open reading frame (ORF) length vary between species. We validated the annotation of the *COA1* gene relying upon gene synteny in the genomic vicinity of the *COA1* gene, multiple sequence alignments, and RNA-seq data (see Supplementary Text S6, Supplementary Figs. S229–S462, and Supplementary Tables S6, S13, S18–S22).

Verification of *COA1* gene disrupting changes in raw read data. We have extended a previously published 5-pass strategy¹⁰³ to verify gene loss events by including Hi-C datasets for assembly validation. Briefly, to verify the correctness of the genome assembly nucleotide sequence, we used the *COA1* gene sequence of multiple species as a query for a blastn search of the short-read sequence database (Supplementary Table S23). Manual inspection of the blast search results ensured concordance between gene sequence and short-read data (Supplementary File S8). We also verified the correctness of genome assembly in the vicinity of the *COA1* gene by evaluating Pacbio long-read data when available (see Supplementary Figs. S463–S465). We used Hi-C datasets for assembly validation in chicken, naked-mole rat, mouse, human, macaque, rabbit, dog, and cheetah (see Supplementary Figs. S105–S178, Supplementary Text S6, Supplementary Tables S6, S13, S18–S22, and Supplementary Figs. S229–S462).

Assessing the transcriptional status of *COA1*. We analyzed transcriptomic datasets for evidence of transcription of *COA1* locus. The RNA-seq reads were mapped to the genome assemblies using the STAR read mapper¹⁰⁴. We visualized the resulting bam files using the Integrative Genomics Viewer (IGV) browser^{105,106}. For consistent representation across tissues and species, we used three different views: (1) Positions of all four exons of *COA1* identified using blast search are shown as a bed record below the RNA-seq bam files, (2) Zoomed-in views of each of the four exons are presented in four panels within a single screenshot and (3) Zoomed-in view of the first and last exons of *COA1* are shown along with the adjacent genes on both sides. The adjacent genes in the IGV screenshot act as positive controls (Supplementary Text S7, Supplementary Figs. S18, S24–S43, S48, S55–S57, S466–S954, and Supplementary Tables S18, S24).

Molecular evolutionary analyses. We performed molecular evolutionary analyses to identify signatures of relaxed selection using codeml from the PAML package and program RELAX from the HyPhy package. We also used the codeml program to estimate the time of gene loss using the method from Meredith et al.¹⁰⁷. We evaluated the role of GC content range and K-mer abundance in determining the gene coverage in sequencing datasets and quantified the magnitude of gBGC. We used computational prediction methods to find changes in RNA binding sites (Supplementary Text S8, Supplementary Figs. S189–S222, S955–S957, and Supplementary Tables S1–S5, S7–S9, S12, S14–S17, S25–S26).

Comparative phylogenetic logistic regression. The pectoralis muscle of birds helps in the movement of humerus bone around the shoulder and provides the power required for the wing's downstroke during flight. This muscle comprises around 8–11% of total body mass in adult flying birds¹⁰⁸ and is regarded as a quintessential locomotory muscle in flying birds. Fortunately, data on the muscle fiber composition of 43 bird species are compiled in a single study¹⁰⁹. Quantitative data is available for only 28 of these 43 species. Hence, we performed an additional search of the literature to identify other papers that measure muscle fiber composition using a comparable methodology and report the information in a similar quantitative format. Based on this search of literature^{23,109–113}, we have compiled a dataset (Supplementary Tables S10 and S27) of the area percentage of red, pink, and white muscle fiber in the avian pectoralis muscle (see Supplementary Text S9). We were able to retrieve muscle fiber composition for major pectoralis muscle for a total of 24 bird species whose *COA1* gene status is also available (see Supplementary Table S27). We considered the *COA1* gene's status as a binary variable, with 0 implying loss of the functional copy of the gene while 1 representing retention of the gene. To test whether

retention/loss of the *COA1* gene is dependent on muscle fiber type, we did a phylogenetically corrected logistic regression analysis on birds in R version 4.1.0¹¹⁴ using the phylolm package version 2.6.2^{115,116}.

Data availability

All data associated with this study are available in the Supplementary Materials. The data is also available in an easy-to-view format on the Github repository: https://github.com/ceglabsagarshinde/COA1_GENE.

Received: 18 June 2021; Accepted: 15 December 2021

Published online: 24 December 2021

References

- Weeks, O. I. Vertebrate skeletal muscle: Power source for locomotion. *Bioscience* **39**, 791–799 (1989).
- Talbot, J. & Maves, L. Skeletal muscle fiber type: using insights from muscle developmental biology to dissect targets for susceptibility and resistance to muscle disease. *Wiley Interdiscip. Rev. Dev. Biol.* **5**, 518–534 (2016).
- Shen, Y. Y. *et al.* Adaptive evolution of energy metabolism genes and the origin of flight in bats. *Proc. Natl. Acad. Sci. USA* **107**, 8666–8671 (2010).
- Hirt, M. R., Jetz, W., Rall, B. C. & Brose, U. A general scaling law reveals why the largest animals are not the fastest. *Nat. Ecol. Evol.* **1**, 1116–1122 (2017).
- Hatefi, Y. The mitochondrial electron transport and oxidative phosphorylation system. *Annu. Rev. Biochem.* **54**, 1015–1069 (1985).
- Mitchell, P. Coupling of phosphorylation to electron and hydrogen transfer by a chemi-osmotic type of mechanism. *Nature* **191**, 144–148 (1961).
- Schägger, H. & Pfeiffer, K. Supercomplexes in the respiratory chains of yeast and mammalian mitochondria. *EMBO J.* **19**, 1777 (2000).
- Conley, K. E. Mitochondria to motion: Optimizing oxidative phosphorylation to improve exercise performance. *J. Exp. Biol.* **219**, 243–249 (2016).
- Björnerfeldt, S., Webster, M. T. & Vilà, C. Relaxation of selective constraint on dog mitochondrial DNA following domestication. *Genome Res.* **16**, 990–994 (2006).
- Shen, Y. Y., Shi, P., Sun, Y. B. & Zhang, Y. P. Relaxation of selective constraints on avian mitochondrial DNA following the degeneration of flight ability. *Genome Res.* **19**, 1760–1765 (2009).
- Verdighione, R. & Cassandro, M. Characterization of muscle fiber type in the pectoralis major muscle of slow-growing local and commercial chicken strains. *Poult. Sci.* **92**, 2433–2437 (2013).
- Goto, M. *et al.* Distribution of muscle fibers in skeletal muscles of the cheetah (*Acinonyx jubatus*). *Mamm. Biol.* **78**, 127–133 (2013).
- Armstrong, R. & Phelps, R. Muscle fiber type composition of the rat hindlimb. *Am. J. Anat.* **171**, 259–272 (1984).
- Pan, S. *et al.* Convergent genomic signatures of flight loss in birds suggest a switch of main fuel. *Nat. Commun.* **10**, 1 (2019).
- Harshman, J. *et al.* Phylogenomic evidence for multiple losses of flight in ratite birds. *Proc. Natl. Acad. Sci. USA* **105**, 13462–13467 (2008).
- Sayol, F., Steinbauer, M. J., Blackburn, T. M., Antonelli, A. & Faurby, S. Anthropogenic extinctions conceal widespread evolution of flightlessness in birds. *Sci. Adv.* **6**, eabb6095 (2020).
- Sackton, T. B. *et al.* Convergent regulatory evolution and loss of flight in paleognathous birds. *Science (80-.)*. **364**, 74–78 (2019).
- Holmes, D. J. & Austad, S. N. Birds as animal models for the comparative biology of aging: A prospectus. *J. Gerontol. Ser. A Biol. Sci. Med. Sci.* **50** (1995).
- Das, J. The role of mitochondrial respiration in physiological and evolutionary adaptation. *BioEssays* **28**, 890–901 (2006).
- Butler, P. J. The physiological basis of bird flight. *Philos. Trans. R. Soc. B: Biol. Sci.* **371** (2016).
- Ogata, T. & Yamasaki, Y. Ultra-high-resolution scanning electron microscopy of mitochondria and sarcoplasmic reticulum arrangement in human red, white, and intermediate muscle fibers. *Anat. Rec.* **248**, 214–223 (1997).
- Barnard, E. A., Lyles, J. M. & Pizzey, J. A. Fibre types in chicken skeletal muscles and their changes in muscular dystrophy. *J. Physiol.* **331**, 333–354 (1982).
- Welch, K. C. & Altshuler, D. L. Fiber type homogeneity of the flight musculature in small birds. *Comp. Biochem. Physiol. B Biochem. Mol. Biol.* **152**, 324–331 (2009).
- Dial, K. P. Evolution of Avian locomotion: Correlates of flight style, locomotor modules, nesting biology, body size, development, and the origin of flapping flight. *Auk* **120**, 941–952 (2003).
- Picard, M., Hepple, R. T. & Burelle, Y. Mitochondrial functional specialization in glycolytic and oxidative muscle fibers: Tailoring the organelle for optimal function. *Am. J. Physiol. Cell Physiol.* **302**, 629–641 (2012).
- Wilson, A. M. *et al.* Locomotion dynamics of hunting in wild cheetahs. *Nature* **498**, 185–189 (2013).
- Bailey, I., Myatt, J. P. & Wilson, A. M. Group hunting within the Carnivora: Physiological, cognitive and environmental influences on strategy and cooperation. *Behav. Ecol. Sociobiol.* **67**, 1–17 (2013).
- Kohn, T. A., Burroughs, R., Hartman, M. J. & Noakes, T. D. Fiber type and metabolic characteristics of lion (*Panthera leo*), caracal (*Caracal caracal*) and human skeletal muscle. *Comp. Biochem. Physiol. A Mol. Integr. Physiol.* **159**, 125–133 (2011).
- Williams, T. M. *et al.* Skeletal muscle histology and biochemistry of an elite sprinter, the African cheetah. *J. Comp. Physiol. - B Biochem. Syst. Environ. Physiol.* **167**, 527–535 (1997).
- Hyatt, J. P. K., Roy, R. R., Stuart, R. & Talmadge, R. J. Myosin heavy chain composition of tiger (*Panthera tigris*) and cheetah (*Acinonyx jubatus*) hindlimb muscles. *J. Exp. Zool. Part A Ecol. Genet. Physiol.* **313**, 45–57 (2010).
- Schiaffino, S. & Reggiani, C. Fiber types in mammalian skeletal muscles. *Physiol. Rev.* **91**, 1447–1531 (2011).
- Kohn, T. A. & Myburgh, K. H. Regional specialization of rat quadriceps myosin heavy chain isoforms occurring in distal to proximal parts of middle and deep regions is not mirrored by citrate synthase activity. *J. Anat.* **210**, 8–18 (2007).
- Kohn, T. A. Insights into the skeletal muscle characteristics of three southern African antelope species. *Biol. Open* **3**, 1037–1044 (2014).
- Zhong, W. W. H., Lucas, C. A., Kang, L. H. D. & Hoh, J. F. Y. Electrophoretic and immunochemical evidence showing that marsupial limb muscles express the same fast and slow myosin heavy chains as eutherians. *Electrophoresis* **22**, 1016–1020 (2001).
- Zhong, W. W. H., Lucas, C. A. & Hoh, J. F. Y. Myosin isoforms and fibre types in limb muscles of Australian marsupials: Adaptations to hopping and non-hopping locomotion. *J. Comp. Physiol. B Biochem. Syst. Environ. Physiol.* **178**, 47–55 (2008).
- Mishra, P., Varuzhanyan, G., Pham, A. H. & Chan, D. C. Mitochondrial dynamics is a distinguishing feature of skeletal muscle fiber types and regulates organellar compartmentalization. *Cell Metab.* **22**, 1033–1044 (2015).
- Wang, C. *et al.* MITRAC15/COA1 promotes mitochondrial translation in a ND2 ribosome–nascent chain complex. *EMBO Rep.* **21**, 1 (2020).

38. Formosa, L. E. *et al.* Dissecting the roles of mitochondrial complex I intermediate assembly complex factors in the biogenesis of complex I. *Cell Rep.* **31**, 541 (2020).
39. Pierrel, F. *et al.* Coa1 links the Mss51 post-translational function to Cox1 cofactor insertion in cytochrome c oxidase assembly. *EMBO J.* **26**, 4335–4346 (2007).
40. Hess, D. C. *et al.* Computationally driven, quantitative experiments discover genes required for mitochondrial biogenesis. *PLoS Genet.* **5**, e1000407 (2009).
41. Van Der Lee, R., Wiel, L., Van Dam, T. J. P. & Huynen, M. A. Genome-scale detection of positive selection in nine primates predicts human-virus evolutionary conflicts. *Nucleic Acids Res.* **45**, 10634–10648 (2017).
42. van der Sluis, E. O. *et al.* Parallel structural evolution of mitochondrial ribosomes and OXPHOS complexes. *Genome Biol. Evol.* **7**, 1235–1251 (2015).
43. Speijer, D., Hammond, M. & Lukeš, J. Comparing early eukaryotic integration of mitochondria and chloroplasts in the light of internal ROS challenges: Timing is of the essence. *MBio* **11**, 1 (2020).
44. Gu, Z. *et al.* Role of duplicate genes in genetic robustness against null mutations. *Nature* **421**, 63–66 (2003).
45. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
46. Nespolo, R. F. *et al.* Aerobic power and flight capacity in birds: a phylogenetic test of the heart-size hypothesis. *J. Exp. Biol.* **221**, 1 (2018).
47. Dollo, L. The laws of evolution. *Bulletin de la Société belge de géologie, de paléontologie et d'hydrologie.* **7**, 164–166 (1893).
48. Fitzgerald, J. & Bateman, J. F. Why mice have lost genes for COL21A1, STK17A, GPR145 and AHRI: Evidence for gene deletion at evolutionary breakpoints in the rodent lineage. *Trends Genet.* **20**, 408–412 (2004).
49. Botero-Castro, F., Figuet, E., Tilak, M. K., Nabholz, B. & Galtier, N. Avian genomes revisited: Hidden genes uncovered and the rates versus traits paradox in birds. *Mol. Biol. Evol.* **34**, 3123–3131 (2017).
50. Hargreaves, A. D. *et al.* Genome sequence of a diabetes-prone rodent reveals a mutation hotspot around the ParaHox gene cluster. *Proc. Natl. Acad. Sci. U. S. A.* **114**, 7677–7682 (2017).
51. Mick, D. U. *et al.* MITRAC links mitochondrial protein translocation to respiratory-chain assembly and translational regulation. *Cell* **151**, 1528–1541 (2012).
52. Murcha, M. W., Kubiszewski-Jakubiak, S., Wang, Y. & Whelan, J. Evidence for interactions between the mitochondrial import apparatus and respiratory chain complexes via Tim21-like proteins in Arabidopsis. *Front. Plant Sci.* **5**, 1 (2014).
53. Taylor, J. S. & Raes, J. Duplication and divergence: The evolution of new genes and old ideas. *Annu. Rev. Genet.* **38**, 615–643 (2004).
54. Bush, S. J., Chen, L., Tovar-Corona, J. M. & Urrutia, A. O. Alternative splicing and the evolution of phenotypic novelty. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **372**, 1–7 (2017).
55. Parker, D. J., Gardiner, A., Neville, M. C., Ritchie, M. G. & Goodwin, S. F. The evolution of novelty in conserved genes; Evidence of positive selection in the Drosophila fruitless gene is localised to alternatively spliced exons. *Heredity (Edinb.)* **112**, 300–306 (2014).
56. Ramensky, V. E., Nurtudinov, R. N., Neverov, A. D., Mironov, A. A. & Gelfand, M. S. Positive selection in alternatively spliced exons of human genes. *Am. J. Hum. Genet.* **83**, 94–98 (2008).
57. Pickrell, J. K., Pai, A. A., Gilad, Y. & Pritchard, J. K. Noisy splicing drives mRNA isoform diversity in human cells. *PLoS Genet.* **6**, 1–11 (2010).
58. Xing, Y. & Lee, C. J. Negative selection pressure against premature protein truncation is reduced by alternative splicing and diploidy. *Trends Genet.* **20**, 472–475 (2004).
59. Shinde, S. S., Teekas, L., Sharma, S. & Vijay, N. Signatures of relaxed selection in the CYP8B1 gene of birds and mammals. *J. Mol. Evol.* **87**, 209–220 (2019).
60. Kishida, T., Suzuki, M. & Takayama, A. Evolution of the alternative AQP2 gene: Acquisition of a novel protein-coding sequence in dolphins. *Mol. Phylogenet. Evol.* **118**, 54–57 (2018).
61. Scanes, C. G. *Sturkie's Avian Physiology. Sturkie's Avian Physiology: Sixth Edition* (Elsevier, 2015). doi:<https://doi.org/10.1016/C2012-0-02488-X>.
62. Li, X. *et al.* Targeting mitochondrial reactive oxygen species as novel therapy for inflammatory diseases and cancers. *J. Hematol. Oncol.* **6**, 19 (2013).
63. Turrens, J. Mitochondrial formation of reactive oxygen species. *J. Physiol.* **552**, 335–344 (2003).
64. Speijer, D. How the mitochondrion was shaped by radical differences in substrates. *BioEssays* **36**, 634–643 (2014).
65. Speijer, D. Oxygen radicals shaping evolution: Why fatty acid catabolism leads to peroxisomes while neurons do without it. *BioEssays* **33**, 88–94 (2011).
66. Speijer, D., Manjeri, G. R. & Szklarczyk, R. How to deal with oxygen radicals stemming from mitochondrial fatty acid oxidation. *Philos. Trans. R. Soc. B Biol. Sci.* **369** (2014).
67. Speijer, D. Alternating terminal electron-acceptors at the basis of symbiogenesis: How oxygen ignited eukaryotic evolution. *BioEssays* **39**, 1600174 (2017).
68. Speijer, D. Can all major ROS forming sites of the respiratory chain be activated by high FADH2/NADH ratios?. *BioEssays* **41**, 1800180 (2019).
69. Krauss, S., Zhang, C. Y. & Lowell, B. B. The mitochondrial uncoupling-protein homologues. *Nat. Rev. Mol. Cell Biol.* **6**, 248–261 (2005).
70. Newman, S. A., Mezentseva, N. V. & Badyaev, A. V. Gene loss, thermogenesis, and the origin of birds. *Ann. N. Y. Acad. Sci.* **1289**, 36–47 (2013).
71. Mcgaugh, S. & Schwartz, T. S. Here and there, but not everywhere: Repeated loss of uncoupling protein 1 in amniotes. *Biol. Lett.* **13**, 1 (2017).
72. Emre, Y. *et al.* Avian UCP: The killjoy in the evolution of the mitochondrial uncoupling proteins. *J. Mol. Evol.* **65**, 392–402 (2007).
73. Jastroch, M., Wuertz, S., Kloas, W. & Klingenspor, M. Uncoupling protein 1 in fish uncovers an ancient evolutionary history of mammalian nonshivering thermogenesis. *Physiol. Genomics* **22**, 150–156 (2005).
74. Hughes, D. A., Jastroch, M., Stoneking, M. & Klingenspor, M. Molecular evolution of UCP1 and the evolutionary history of mammalian non-shivering thermogenesis. *BMC Evol. Biol.* **9**, 4 (2009).
75. Polymeropoulos, E. T., Jastroch, M. & Frappell, P. B. Absence of adaptive nonshivering thermogenesis in a marsupial, the fat-tailed dunnart (*Sminthopsis crassicaudata*). *J. Comp. Physiol. B Biochem. Syst. Environ. Physiol.* **182**, 393–401 (2012).
76. Jastroch, M. Uncoupling protein 1 controls reactive oxygen species in brown adipose tissue. *Proc. Natl. Acad. Sci. USA* **114**, 7744–7746 (2017).
77. Gaudry, M. J. *et al.* Inactivation of thermogenic UCP1 as a historical contingency in multiple placental mammal clades. *Sci. Adv.* **3**, 2878 (2017).
78. Letts, J. A. & Sazanov, L. A. Clarifying the supercomplex: The higher-order organization of the mitochondrial electron transport chain. *Nat. Struct. Mol. Biol.* **24**, 800–808 (2017).
79. Greggio, C. *et al.* Enhanced respiratory chain supercomplex formation in response to exercise in human skeletal muscle. *Cell Metab.* **25**, 301–311 (2017).
80. Bundgaard, A., James, A. M., Harbour, M. E., Murphy, M. P. & Fago, A. Stable mitochondrial C1C1I2 supercomplex interactions in reptiles versus homeothermic vertebrates. *J. Exp. Biol.* **223**, 1 (2020).

81. Gan, Z., Fu, T., Kelly, D. P. & Vega, R. B. Skeletal muscle mitochondrial remodeling in exercise and diseases. *Cell Res.* **28**, 969–980 (2018).
82. Lemaitre, C. *et al.* Analysis of fine-scale mammalian evolutionary breakpoints provides new insight into their relation to genome organisation. *BMC Genomics* **10**, (2009).
83. Schibler, L. *et al.* High-resolution comparative mapping among man, cattle and mouse suggests a role for repeat sequences in mammalian genome evolution. *BMC Genomics* **7**, 194 (2006).
84. Farré, M. *et al.* Novel insights into chromosome evolution in birds, archosaurs, and reptiles. *Genome Biol. Evol.* **8**, 2442–2451 (2016).
85. Ruiz-Herrera, A. *et al.* Evolutionary breakpoints are co-localized with fragile sites and intrachromosomal telomeric sequences in primates. *Cytogenet. Genome Res.* **108**, 234–247 (2005).
86. Durkin, S. G. & Glover, T. W. Chromosome fragile sites. *Annu. Rev. Genet.* **41**, 169–192 (2007).
87. Ruiz-Herrera, A., Castresana, J. & Robinson, T. J. Is mammalian chromosomal evolution driven by regions of genome fragility?. *Genome Biol.* **7**, R115 (2006).
88. Carbone, L. *et al.* Evolutionary breakpoints in the gibbon suggest association between cytosine methylation and karyotype evolution. *PLoS Genet.* **5**, e1000538 (2009).
89. Huvet, M. *et al.* Human gene organization driven by the coordination of replication and transcription. *Genome Res.* **17**, 1278–1285 (2007).
90. Boteva, L. *et al.* Common fragile sites are characterized by faulty condensin loading after replication stress. *Cell Rep.* **32**, 1 (2020).
91. Trinh, P., McLysaght, A. & Sankoff, D. Genomic features in the breakpoint regions between syntenic blocks. *Bioinformatics* **20**, i318–i325 (2004).
92. Peng, Q., Pevzner, P. A. & Tesler, G. The fragile breakage versus random breakage models of chromosome evolution. *PLoS Comput. Biol.* **2**, 100–111 (2006).
93. Alekseyev, M. A. & Pevzner, P. A. Are there rearrangement hotspots in the human genome?. *PLoS Comput. Biol.* **3**, 2111–2121 (2007).
94. Capilla, L. *et al.* Mammalian comparative genomics reveals genetic and epigenetic features associated with genome reshuffling in rodentia. *Genome Biol. Evol.* **8**, 3703–3717 (2016).
95. Garagnani, P. *et al.* Whole-genome sequencing analysis of semi-supercentenarians. *Elife* **10**, 1 (2021).
96. Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).
97. Steinegger, M. *et al.* HH-suite3 for fast remote homology detection and deep protein annotation. *BMC Bioinformatics* **20**, 1 (2019).
98. Remmert, M., Biegert, A., Hauser, A. & Söding, J. HHblits: Lightning-fast iterative protein sequence searching by HMM-HMM alignment. *Nat. Methods* **9**, 173–175 (2012).
99. Frickey, T. & Lupas, A. CLANS: A Java application for visualizing protein families based on pairwise similarity. *Bioinformatics* **20**, 3702–3704 (2004).
100. Gabler, F. *et al.* Protein sequence analysis using the MPI bioinformatics toolkit. *Curr. Protoc. Bioinforma.* **72**, 108 (2020).
101. Zimmermann, L. *et al.* A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at its Core. *J. Mol. Biol.* **430**, 2237–2243 (2018).
102. Szklarczyk, R. *et al.* Iterative orthology prediction uncovers new mitochondrial proteins and identifies C12orf62 as the human ortholog of COX14, a protein involved in the assembly of cytochrome c oxidase. *Genome Biol.* **13**, 1 (2012).
103. Sharma, S., Shinde, S. S., Teekas, L. & Vijay, N. Evidence for the loss of plasminogen receptor KT gene in chicken. *Immunogenetics* **72**, 507–515 (2020).
104. Dobin, A. *et al.* STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* (2013).
105. Robinson, J. T. *et al.* Integrative genomics viewer. *Nat. Biotechnol.* **29**, 24–26 (2011).
106. Thorvaldsdottir, H., Robinson, J. T. & Mesirov, J. P. Integrative Genomics Viewer (IGV): High-performance genomics data visualization and exploration. *Brief. Bioinform.* **14**, 178–192 (2013).
107. Meredith, R. W., Gatesy, J., Murphy, W. J., Ryder, O. A. & Springer, M. S. Molecular decay of the tooth gene enamelin (ENAM) mirrors the loss of enamel in the fossil record of placental mammals. *PLoS Genet.* **5**, 1 (2009).
108. Cao, T. & Jin, J.-P. Evolution of flight muscle contractility and energetic efficiency. *Front. Physiol.* **11**, 1–14 (2020).
109. Rosser, B. W. C. & George, J. C. The avian pectoralis: Histochemical characterization and distribution of muscle fiber types. *Can. J. Zool.* **64**, 1174–1185 (1986).
110. Kaiser, C. E. & George, J. C. Interrelationship amongst the avian orders Galliformes, Columbiformes, and Anseriformes as evinced by the fiber types in the pectoralis muscle. *Can. J. Zool.* **51**, 887–892 (1973).
111. Wiskus, K. J., Addis, P. B. & Ma, R.-I. Distribution of β R, α R and α W Fibers in Turkey Muscles. *Poult. Sci.* **55**, 562–572 (1976).
112. Schroeder, K. L., Sylvain, N. J., Kirkpatrick, L. J. & Rosser, B. W. C. Fibre types in primary 'flight' muscles of the African Penguin (*Spheniscus demersus*). *Acta Zool.* **96**, 510–518 (2015).
113. Kiessling, K. Muscle structure and function in the goose, quail, pheasant, guinea hen, and chicken. *Comp. Biochem. Physiol. B.* **57**, 287–292 (1977).
114. R Core Team. *R: A language and environment for statistical computing*. Vienna, Austria (2021).
115. Ives, A. R. & Garland, T. Phylogenetic logistic regression for binary dependent variables. *Syst. Biol.* **59**, 9–26 (2010).
116. Ho, T., Si, L. & Ané, C. A linear-time algorithm for Gaussian and non-Gaussian trait evolution models. *Syst. Biol.* **63**, 397–408 (2014).
117. Johnson, W. *et al.* The late Miocene radiation of modern Felidae: A genetic assessment. *Science* **311**, 73–77 (2006).
118. Revell, L. J. phytools: An R package for phylogenetic comparative biology (and other things). *Methods Ecol. Evol.* **3**, 217–223 (2012).

Acknowledgements

We thank the Council of Scientific and Industrial Research for fellowship to SSS and the Ministry of Human Resource Development for fellowship to LT and SS and University Grants commission to AS. NV has been awarded the Innovative Young Biotechnologist Award 2018 from the Department of Biotechnology (Government of India).

Author contributions

S.S.S. and N.V. wrote the manuscript with major inputs from S.S., A.S., and L.T. S.S.S. analyzed the data with the assistance of S.S., A.S., and L.T. Specific analysis benefited from insightful inputs provided by L.T. All authors reviewed the manuscript.

Funding

This article was funded by Department of Biotechnology, Ministry of Science and Technology, India (Grant no. BT/11/IYBA/2018/03) and Science and Engineering Research Board (Grant no. ECR/2017/001430).

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-04077-y>.

Correspondence and requests for materials should be addressed to N.V.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021