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

Sagar Sharad Shinde[1], Sandhya Sharma[1], Lokdeep Teekas[1], Ashutosh Sharma[1], Nagarjun Vijay[1]

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

*Correspondence: nagarjun@iiserb.ac.in

26   **Abstract**

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

44   **Keywords:** Cytochrome C Oxidase Assembly Factor 1, *COA1*, *MITRAC15*, Chicken, gene
45   loss, rodent

46   **1. Introduction**

47   Skeletal muscles control numerous locomotor functions in vertebrates (Weeks, 1989). The
48   hundreds of different muscles in the body consist of highly organized heterogeneous bundles
49   of fibers. These muscle fibers are classified based on contractile properties, power source,
50   and myosin component into type-1, 2A, 2B, and 2X (Talbot and Maves, 2016). Muscles with
51   type-1 and type-2A fibers rely on the oxidative phosphorylation (OXPHOS) pathway, the
52   primary source of ATP needed for locomotion and other energy-intensive tasks (Shen et al.,
53   2010). The energy releasing electron transport chain (ETC) coupled with the energy-requiring
54   chemiosmosis is known as (OXPHOS) (Hatefi, 1985; Mitchell, 1961). A chain of
55   mitochondrial inner membrane-embedded proteins encoded by both mitochondrial and
56   nuclear genes form four large complexes that transport electrons through redox reactions. The
57   energy released during these reactions results in a proton gradient, which uses a fifth
58   membrane-embedded complex to generate ATP through chemiosmosis. Optimization of
59   crucial steps in the OXPHOS pathway leads to improved locomotor performance (Conley,
60   2016). Origin of novel energetically demanding phenotypes has been possible through
61   adaptations in the OXPHOS pathway (Doan et al., 2004; Garvin et al., 2015; Wu et al., 2000;
62   Zhang and Broughton, 2015). Multiple genes of the OXPHOS pathway are under positive
63   selection in mammalian species with high energy demanding adaptations such as powered
64   flight in bats (Shen et al., 2010), survival of polar bears in cold Arctic environment (Welch et
65   al., 2014), high altitude adaptation in yak (Qiu et al., 2012), hypoxia tolerance in cetaceans

2

66   (Tian et al., 2018), ecotype specific divergence in killer whales (Foote et al., 2011) and
67   evolution of large brains in anthropoid primates (Grossman et al., 2004).
68
69   The loss of energetically demanding phenotypes reduces the strength of purifying selection
70   acting on the OXPHOS pathway. For instance, the domestication of dogs (Björnerfeldt et al.,
71   2006) and degeneration of locomotor abilities in birds (Shen et al., 2009) resulted in relaxed
72   selective constraint on the OXPHOS pathway proteins. Among other carnivores and rodents,
73   the great diversity of functionally important locomotor habits have variations in energy
74   requirements and corresponding differences in the magnitude of purifying selection (Samuels
75   and Van Valkenburgh, 2008; Taylor, 1989). Even within the same species, mitochondrially
76   encoded protein components of the OXPHOS pathway are under stronger purifying selection
77   than those protein components encoded by the nuclear genome (Popadin et al., 2013). These
78   differing purifying selection levels are due to gene expression level differences between
79   nuclear and mitochondrial OXPHOS genes (Nabholz et al., 2013). Despite the movement of
80   most genes from the ancestral mitochondria to the nucleus in eukaryotes, a separate
81   mitochondrial organelle is well conserved with scarce exceptions (Karnkowska et al., 2016;
82   Sloan et al., 2018). Turnover in the content of mitochondrial protein complexes has mainly
83   occurred before the emergence of eukaryotes with few gene gain/loss events reported in
84   vertebrates (Adams and Palmer, 2003; Cardol, 2011; Gabaldón et al., 2005; Gabaldón and
85   Huynen, 2007; Huynen et al., 2013; van Esveld and Huynen, 2018). However, lineage-
86   specific gene loss from the mitochondria has occurred in nonbilaterian organisms (Lavrov
87   and Pett, 2016), other metazoan lineages (Gissi et al., 2008), and plants (Depamphilis et al.,
88   1997; Palmer et al., 2000). The duplication of mitochondrial genes in bird lineages followed
89   by gene loss and genomic rearrangement events is relatively unique (Akiyama et al., 2017;
90   Mackiewicz et al., 2019; San Mauro et al., 2006; Urantówka et al., 2020).

91   The proton gradient established by the ETC also powers the generation of heat in mammalian
92   Non-Shivering Thermogenesis (NST) (Nedergaard et al., 2001). Thermogenin or uncoupling
93   protein 1 (*UCP1*) expressed in the inner mitochondrial membrane facilitates the regulated
94   leakage of protons to generate heat in brown adipose tissue (Krauss et al., 2005). The *UCP1*
95   gene is absent in all birds (Newman et al., 2013) and some mammals (Emre et al., 2007;
96   Mcgaugh and Schwartz, 2017) despite its presence in fish (Jastroch et al., 2005), amphibians
97   (Hughes et al., 2009), and marsupials (Polymeropoulos et al., 2012). The integration of *UCP1*
98   in the thermogenic pathway is considered a eutherian-mammal-specific adaptation unrelated
99   to its ancestral innate immune functions (Jastroch, 2017). The exceptional repeated loss of
100   this mitochondrial membrane protein in vertebrate lineages appears to result from its
101   changing functional roles (Gaudry et al., 2017; Mcgaugh and Schwartz, 2017). In contrast to
102   *UCP1*, most OXPHOS pathway genes are highly conserved, and defective protein
103   components generally result in clinical phenotypes (Hock et al., 2020). The proteins
104   *TMEM186* and *COA1/MITRAC15* are chaperones interacting with the Mitochondrial
105   Complex I Assembly (MCIA) complex, and defects in these genes do not result in any
106   clinical phenotypes (Hock et al., 2020; Signes and Fernandez-Vizarra, 2018).

107   Functional studies have implicated a role for *COA1/MITRAC15* in promoting mitochondrial
108   translation and complex I and IV biogenesis (Wang et al., 2020). However, overexpression of
109   other genes easily compensates for the mild effect of *COA1/MITRAC15* gene knockout (Hess
110   et al., 2009; Pierrel et al., 2007). Notably, the *COA1/MITRAC15* gene was also identified as a
111   positively selected gene in a genome-wide screen in primates (Van Der Lee et al., 2017) and
112   suggests that despite its mild phenotype, *COA1/MITRAC15* can contribute to fitness increases
113   through its role as a chaperone. *COA1/MITRAC15* resembles *TIMM21*, a subunit of the
114   *TIM23* complex (Mick et al., 2012). Such *TIMM21* gene duplicates interacting with the
115   mitochondrial import apparatus and respiratory chain complexes occur in Arabidopsis
116   (Murcha et al., 2014). Diversification of the mitochondrial import system has benefitted from
117   gene duplication events that have contributed new members to the Translocase of the Inner
118   Membrane (TIM) and Translocase of the Outer Membrane (TOM) protein complexes
119   (Fukasawa et al., 2017). Hence, it is plausible that *COA1/MITRAC15* results from a
120   duplication of the *TIMM21* gene followed by divergence.

121   Divergence of conventional or class-2 myosin genes after duplication has led to the
122   diversification of the *MYH* gene family (Moore et al., 1993; Weiss and Leinwand, 1996).
123   These myosin genes have distinct functions defined by their contractile properties and
124   ATPase activity (Resnicow et al., 2010). While *MYH7* and *MYH2* expressing fibers rely upon
125   the OXPHOS pathway, *MYH1* and *MYH4* expressing fibers are dependent on the glycolytic
126   pathway. The protein encoded by the *MYH7* gene occurs in both cardiac muscles and the
127   slow-contracting type-1 fibers (Schiaffino and Reggiani, 2011). However, the *MYH* genes
128   expressed in type-2 fibers are restricted mainly to skeletal muscles. The fast-contracting type-
129   2 fibers power explosive movements like jumping and sprinting. Such rapid movements form
130   an essential component of hunting strategies used by terrestrial predators and the escape
131   strategy of the prey (Kohn, 2014; J. W. Wilson et al., 2013). Felids, small-bodied rodents,
132   marsupials, certain cervids, and galliform birds have exceptional adaptations for rapid
133   locomotion.

134   The world's fastest mammal, the cheetah (*Acinonyx jubatus*), epitomizes the relevance of
135   speed and acceleration (A. M. Wilson et al., 2013). In general, felids are adept at sprinting
136   and can accelerate more rapidly than canids but cannot sustain it for a prolonged period
137   (Bailey et al., 2013). The predominance of type-2X fibers in felid species provides the ability
138   to achieve rapid acceleration (Hyatt et al., 2010; Kohn et al., 2011; Williams et al., 1997).
139   Compared to canids, felids have a greater reliance on glycolytic fibers. Glycolytic fibers have
140   decreased reliance on OXPHOS and lower mitochondrial content than oxidative fibers
141   (Mishra et al., 2015; Picard et al., 2012). Hence, the OXPHOS pathway might be under
142   stronger selective constraint in canids than felids. Like these predators, prey species like
143   antelopes are fast sprinters but have the added advantage of resistance to fatigue. The high
144   speed of these species relies on type-2X fibers with high glycolytic capacities, and the added
145   resistance to fatigue is possible due to the remarkable oxidative ability of these fibers (Curry
146   et al., 2012). The use of both glycolytic and oxidative pathways suggests the OXPHOS
147   pathway in these antelope species and other cervids would be under strong purifying
148   selection.

4

149 Despite drastic variation in body size within mammals, the relative speed of locomotion is
150 thought to be largely independent of body mass, at least in small mammals (Iriarte-Díaz,
151 2002). The higher relative speed of small mammals results from faster constriction made
152 possible by the higher proportion of fast fibers (mostly 2X and 2B) in each muscle
153 (Schiaffino and Reggiani, 2011). For instance, rodent limb muscles are known to have more
154 abundant type 2B fibers compared to larger mammals (including humans, which have no type
155 2B fibers in the limb muscles) (Kohn, 2014; Kohn and Myburgh, 2007). Marsupial species
156 also have high relative speeds and possess muscle fibers equivalent to eutherian mammals
157 (Zhong et al., 2001). The smaller marsupial species have type-2B and 2X muscle fibers in
158 several important muscles (Zhong et al., 2008). The higher proportion of fast glycolytic fibers
159 in rodents and marsupials potentially results in relaxed selection on the OXPHOS pathway
160 genes in these species.

161 The ability to fly is a distinctive feature of birds except for lineages that have become entirely
162 flightless or retain only a limited flying capacity (Harshman et al., 2008; Pan et al., 2019;
163 Sackton et al., 2019; Sayol et al., 2020). The large amount of energy required for flight has
164 necessitated a high metabolic rate in birds (Holmes and Austad, 1995). Increased ATP
165 generation fulfills these energy demands through metabolic adaptations in the OXPHOS
166 pathway (Das, 2006). The set of flight muscles possessed by a bird species determine several
167 aspects of flight performance and strongly influences life history and ecology (DuBay et al.,
168 2020). Avian flight is possible through a combination of flight muscles that consist of white
169 (fast glycolytic), intermediate/red-pink (fast oxidative), and red (slow oxidative) muscle
170 fibers (Barnard et al., 1982; Butler, 2016; Ogata and Yamasaki, 1997). Birds with strong
171 flight abilities, such as long-distant migrants and small passerines, contain mostly fast
172 oxidative fibers (Welch and Altshuler, 2009). In contrast to this, Galliformes contain mostly
173 glycolytic fibers that only allow short bursts of activity (Dial, 2003). Hence, the OXPHOS
174 pathway is under stronger selective constraint in non-Galliform bird species than Galliform
175 birds due to the functional specialization of mitochondria to different muscle fibers (Picard et
176 al., 2012).

178 This study evaluates whether the protein encoded by the *COA1/MITRAC15* gene, a
179 mitochondrial complex I translation factor with a chaperone-like role, is dispensable when
180 the OXPHOS pathway is under relaxed selective constraints. We hypothesized that the
181 OXPHOS pathway might have experienced reduced purifying selection in felids, rodents,
182 marsupials, and galliform birds based on increased glycolytic muscle fibers in these species.
183 Duplicate copies or alternative metabolic pathways compensate for gene function and decide
184 gene dispensability (Gu et al., 2003). Hence, to evaluate our hypothesis, we aim to (1)
185 investigate whether *COA1/MITRAC15* has any homologs that could compensate its function,
186 (2) screen the genomes of vertebrate species to identify and track the evolutionary history of
187 *COA1/MITRAC15* orthologs, (3) identify evidence of gene disruptive changes within the
188 *COA1/MITRAC15* locus and (4) reconstruct the sequence of events associated with the
189 potential erosion of the *COA1/MITRAC15* locus due to chromosomal rearrangement events at
190 the evolutionary breakpoint region spanning the *COA1/MITRAC15* gene. We extensively

191 screened publicly available genomes and transcriptomes of more than 200 vertebrate species
192 to establish recurrent loss of the widely conserved *COA1/MITRAC15* gene.

193 2. Materials and methods

194 2.1 Finding homologs of *COA1/MITRAC15*

195 The amino acid sequence of the human *COA1/MITRAC15* gene was used as a query in PSI-
196 BLAST (Altschul et al., 1997) against the nr database with eight iterations to identify
197 homologs. Similarly, the human *COA1/MITRAC15* amino acid sequence was the query in the
198 program HHblits of HHsuite (Remmert et al., 2012; Steinegger et al., 2019) with the flags "-e
199 1e-3 -n 8 -p 20 -Z 5000 -z 1 -b 1 -B 5000 -d UniRef30_2020_06". The output from HHblits
200 was used as input to the CLANS program (Frickey and Lupas, 2004) with an e-value cut-off
201 of 1e-4 to cluster the blast hits using the MPI Bioinformatics Toolkit (Gabler et al., 2020;
202 Zimmermann et al., 2018). We ran the CLANS java application for more than 50,000 rounds
203 on the webserver output to ensure stable clusters. Manually inspection of gene annotations
204 allowed identification of each of the groups. Subsequently, we performed the HHblits search
205 again with different settings such as "-glob" to perform global alignments and "-loc" to
206 conduct local alignments. The PFAM database was the alternative to the Uniclust 30
207 database. Manually curated multiple sequence alignment of *COA1/MITRAC15* open reading
208 frames from 24 primate species was also separately used to query for better sensitivity. The
209 protein *TIMM21* provides a consistent hit with different search settings and databases.

210 To further verify whether the database matches are homologous, we evaluated the biological
211 function, secondary structure similarity, relationship among top hits, and occurrence of
212 conserved motifs. To obtain secondary structure predictions for the proteins
213 *COA1/MITRAC15* and *TIMM21*, we used the PROTEUS2 webserver (Montgomerie et al.,
214 2008). The HeliQuest webserver (Gautier et al., 2008) provided each predicted helix's
215 physicochemical properties and amino acid compositions. While the three-dimensional (3-D)
216 structure of the *COA1/MITRAC15* protein is not available yet, multiple structures of the
217 *TIMM21* protein are available in the Protein Data Bank (PDB). It is possible to use
218 comparative/homology modeling to predict the 3-D structure based on the protein structure of
219 a related protein (Webb and Sali, 2016). Hence, we used the comparative modeling approach
220 implemented in Modeller (v10.0) software to model the structure of *COA1/MITRAC15* based
221 on the homologous structures available in PDB. The Phyre2 (Kelley et al., 2015) and Expasy
222 Swiss-Model (Waterhouse et al., 2018) webservers also predicted homologous 3-D structures
223 of *COA1/MITRAC15*. All the top hits were from 3-D structures of the IMS (Inter Membrane
224 Space) domain of *TIMM21* protein. The IMS domain of *TIMM21,* whose 3-D structures are
225 available on PDB, contains only the part of the protein that occurs after the membrane-
226 spanning helix. To model the structure of *COA1/MITRAC15* using these existing 3-D
227 structures, we used the *COA1/MITRAC15* amino acid sequence that occurs after the
228 membrane-spanning domain. We visualized the structure of *TIMM21* and the predicted
229 *COA1/MITRAC15* structure using (UCSF Chimera v1.15) ChimeraX (Pettersen et al., 2021).

230 2.2 Validation of *COA1/MITRAC15* annotation

231   Despite being a fast-evolving gene, orthologs of *COA1/MITRAC15* can be identified based on
232   gene synteny and sequence identity. However, identifying *COA1/MITRAC15* orthologs
233   between distantly related species is challenging (Szklarczyk et al., 2012). We screened the
234   genome assemblies and annotations available on NCBI and Ensembl for *COA1* (C7orf44 or
235   *MITRAC15*) protein-coding transcripts. The *COA1/MITRAC15* gene orthologs have been
236   annotated in almost 300 vertebrate species (see **Supplementary Table S1**). However, the
237   number of exons and the length of the ORF is highly variable between species. We validated
238   the annotation of the *COA1/MITRAC15* gene relying upon gene synteny in the genomic
239   vicinity of the *COA1/MITRAC15* gene, multiple sequence alignments, and RNA-seq data.
240   Annotation across most species endorses the existence of four coding exons that produce a
241   ~130 to 140 amino acid (aa) protein. The *COA1/MITRAC15* annotation in the human genome
242   (see **Supplementary Figure S1**) has multiple isoforms with seven exons. The additional
243   three exons annotated in the human genome upstream from the widely conserved four exons
244   need further investigation. Bird species such as *Nipponia nippon*, *Cuculus canorus*, *Pterocles*
245   *gutturalis*, *Gavia stellate*, *Buceros rhinoceros silvestris*, *Anser cygnoides domesticus*, *Anas*
246   *platyrhynchos* (corrected in XM_027451320.2), and *Fulmarus glacialis* have annotation for a
247   fifth exon upstream from the widely conserved four exons. Annotation for multiple isoforms
248   of the *COA1/MITRAC15* gene also exists in *Athene cunicularia*, *Tyto alba*, *Calidris pugnax*,
249   *Serinus canaria*, *Corvus moneduloides*, *Corvus brachyrhynchos*, *Egretta garzetta*, *Aquila*
250   *chrysaetos*, *Pipra filicauda*, *Corvus cornix*, *Cygnus atratus,* and *Parus major*. We examined
251   RNA-seq datasets of multiple species to evaluate the expression of the isoforms. RNA-seq
252   data in *Colius striatus* and *Eurypyga helias* (which had partial sequences annotated) allowed
253   reconstruction of full-length ORFs. In addition to bird genomes, the *COA1/MITRAC15* gene
254   ortholog is annotated in lizards (*Zootoca vivipara*, *Podarcis muralis*, *Lacerta agilis*, *Anolis*
255   *carolinensis*, *Gekko japonicus*, *Thamnophis sirtalis*, *Pantherophis guttatus*, *Notechis*
256   *scutatus*, *Pseudonaja textilis* and *Python bivittatus*), turtles (*Trachemys scripta elegans*,
257   *Chelonia mydas*, *Chelonoidis abingdonii*, *Chrysemys picta*, *Gopherus evgoodei* and
258   *Pelodiscus sinensis*), alligators (*Gavialis gangeticus*, *Alligator sinensis*, *Alligator*
259   *mississippiensis* and *Crocodylus porosus*), Even-toed ungulates (*Bos taurus*, *Sus scrofa*,
260   *Odocoileus virginianus texanus*, *Bison bison bison*, *Bos indicus x Bos taurus*, *Bos mutus*,
261   *Bubalus bubalis*, *Capra hircus*, *Ovis aries*, *Vicugna pacos*, *Camelus ferus*, *Camelus*
262   *bactrianus*, *Camelus dromedarius*, *Neophocaena asiaeorientalis asiaeorientalis*,
263   *Balaenoptera acutorostrata scammoni*, *Lipotes vexillifer*, *Lagenorhynchus obliquidens*,
264   *Globicephala melas*, *Orcinus orca*, *Tursiops truncatus*, *Phocoena sinus*, *Monodon*
265   *monoceros*, *Delphinapterus leucas*, *Physeter catodon* and *Balaenoptera musculus*), Odd-toed
266   ungulates (*Equus caballus*, *Equus asinus*, *Equus przewalskii* and *Ceratotherium simum*
267   *simum*), Pangolins (Manis pentadactyla and Manis javanica), *Galeopterus variegatus*, *Tupaia*
268   *chinensis* and Primates (*Homo sapiens*, *Macaca mulatta*, *Pan troglodytes*, *Chlorocebus*
269   *sabaeus*, *Callithrix jacchus*, *Colobus angolensis palliatus*, *Cercocebus atys*, *Macaca*
270   *fascicularis*, *Macaca nemestrina*, *Papio anubis*, *Theropithecus gelada*, *Mandrillus*
271   *leucophaeus*, *Trachypithecus francoisi*, *Rhinopithecus bieti*, *Rhinopithecus roxellana*,
272   *Piliocolobus tephrosceles*, *Gorilla gorilla*, *Pan paniscus*, *Pongo abelii*, *Nomascus*
273   *leucogenys*, *Hylobates moloch*, *Saimiri boliviensis*, *Sapajus apella*, *Cebus imitator*, *Aotus*

7

274   *nancymaae*, *Carlito syrichta*, *Propithecus coquereli*, *Microcebus murinus* and *Otolemur*
275   *garnettii*).

276   We screened the synteny pattern of the candidate *COA1/MITRAC15* gene in Galliformes and
277   Anseriformes using five upstream genes (*STK17A*, *HECW1*, *TNS3*, *PSMA2*, *MRPL32*) and
278   the five downstream genes (*BLVRA*, *VOPP1*, *LANCL2*, *EGFR*, *SEC61G*). The chicken
279   (*Gallus gallus*) has a chromosome level assembly, and the gene occurs on Chromosome 2,
280   and its region is syntenic with human (*Homo sapiens*) chromosome 2 (**Supplementary**
281   **Figure S2-S3**). The gene synteny is mostly conserved in these species and is present on the
282   same scaffold/chromosome. The blast search of the genome using the query gene sequence of
283   closely related species identified genes missing in the annotation. *Anas platyrhynchos* has
284   chromosome level assembly with the same gene order as *Gallus gallus* (**Supplementary**
285   **Figure S4**). *Anser cygnoides* and *Anseranas semipalmata* also contain this conserved gene
286   order. *Anas platyrhynchos*, *Numida meleagris*, *Coturnix japonica*, *Meleagris gallopavo* show
287   syntenic blocks aligning with the human chromosome 7 (**Supplementary Figure S5-S8**).
288   Synteny-based verification was done clade-wise in birds (see **Supplementary Table S2**),
289   rodents (**Supplementary Table S3**), carnivores (**Supplementary Table S4**), and primates
290   (**Supplementary Table S5**). Gene order and synteny relationships for representative species
291   from each of the clades are in **Supplementary Figure S9-S230**.

292   Vertebrate species have a conserved *COA1/MITRAC15* gene intron/exon organizational
293   structure. However, two lineages (primates and carnivores) with evidence of intron/exon
294   organization changes have also had *COA1/MITRAC15* gene duplication events. To ensure
295   that the observed differences were not a result of incorrect annotation, alignment artifacts, or
296   duplicated copies, we compared the *COA1/MITRAC15* gene organization across diverse
297   vertebrate species. Subsequently, we validated the annotations from NCBI and Ensembl
298   using RNA-seq datasets. Sequencing read haplotypes from the functional and pseudogenised
299   copy can be distinguished as their sequences have diverged.

2.3 Verification of *COA1/MITRAC15* gene disrupting changes in raw read data

301   We used a previously published 5-pass strategy to verify gene loss events (S. Sharma et al.,
302   2020). Briefly, to verify the correctness of the genome assembly nucleotide sequence, we
303   used the *COA1/MITRAC15* gene sequence of multiple species as a query for a blastn search
304   of the raw short-read database. The details of short-read datasets (both DNA and RNA) used
305   to validate gene sequence are in **Supplementary Table S6**. Manual inspection of the blast
306   search results ensured concordance between gene sequence and raw read data. All the blast
307   output files are in **Supplementary File S1**. In the chicken genome, we also verified the
308   correctness of genome assembly in the vicinity of the *COA1/MITRAC15* gene by evaluating
309   Pacbio long-read data (see **Supplementary Figure S231-S234**).

2.4 Assessing the transcriptional status of *COA1/MITRAC15*

311   We analyzed transcriptomic datasets for evidence of transcription of *COA1/MITRAC15* locus.
312   The RNA-seq reads were mapped to the genome assemblies using the STAR read mapper
313   (Dobin et al., 2013). We visualized the resulting bam files using the IGV browser (Robinson

314   et al., 2011; Thorvaldsdottir et al., 2013). For consistent representation across tissues and
315   species, we used three different views: (1) Positions of all four exons of *COA1/MITRAC15*
316   identified using blast search are shown as a bed record below the RNA-seq bam files, (2)
317   Zoomed-in views of each of the four exons are presented in four panels within a single
318   screenshot and (3) Zoomed-in view of the first and last exons of *COA1/MITRAC15* are shown
319   along with the adjacent genes on both sides. The adjacent genes in the IGV screenshot act as
320   positive controls.

321   No evidence for transcription of *COA1/MITRAC15* gene in chicken exists in the RNA-seq
322   data from 23 tissues consisting of blood, bone marrow, breast muscle, bursa, cerebellum,
323   cerebrum, comb, eye, fascia, gallbladder, gizzard, gonad, heart, immature egg, kidney, liver,
324   lung, mature egg, pancreas, shank, skin, spleen, uterus (**Supplementary Figure S235-S304**).
325   Among other Galliformes species, we found no evidence for expression of the
326   *COA1/MITRAC15* gene. (The spleen and gonad of the peacock, the skin of golden pheasant,
327   gonad, spleen, brain, muscle, liver, and heart of ring-necked pheasant, bursa, gonad spleen,
328   blood and uterus of helmeted guineafowl, breast muscle, gonad, spleen, brain, liver, heart,
329   and bursa of turkey, kidney, liver, muscle, lung, and heart of Japanese quail, the blood of
330   *Colinus virginianus* and blood of *Syrmaticus Mikado*, see **Supplementary Figure S305-**
331   **S373**). The only Galliform species to have a transcribed *COA1/MITRAC15* gene was
332   *Alectura lathami* (blood tissue: **Supplementary Figure S374-S376**).

333   In contrast to Galliformes, the *COA1/MITRAC15* gene is intact in Anseriformes species.
334   However, the *COA1/MITRAC15* gene annotation in duck (*Anas platyrhynchos platyrhynchos*)
335   contains two isoforms. The more extended isoform codes for a 265 amino acid protein and
336   consists of five exons. The shorter isoform (139 amino acid) is orthologous to the Galliform
337   ORF. Upon closer inspection of the first exon, only 24 of the 372 bases have RNA-seq read
338   support **(Supplementary Figure S377)**. Hence, this additional exon might be an annotation
339   artifact or part of the untranslated region. The last four annotated exons, which correspond to
340   the intact 139 amino acid encoding sequence, were found to be robustly expressed in the
341   gonad, spleen, liver, brain, and skin (**Supplementary figure S378-S385**). A similar
342   annotation of the fifth exon in *Anser cygnoides domesticus* appears to be an artifact. The
343   gonad, liver, and spleen express the last four exons (see **Supplementary Figure S386-S392**).
344   The RNA-seq data from blood tissue for magpie goose (*Anseranas semipalmata*) and
345   southern screamer (*Chauna torquata*) also supported the transcription of the
346   *COA1/MITRAC15* gene (**Supplementary Figure S393-S396**).

347   Having verified the expression of the *COA1/MITRAC15* gene in multiple Anseriformes
348   species, we screened additional bird RNA-seq datasets to evaluate the transcriptional activity
349   of the intact ORF found in these species. Many other bird genomes have annotations for
350   multiple isoforms of the *COA1/MITRAC15* gene, like the duck genomes. These isoforms
351   range in length from 136 to 265 amino acids and 4 to 7 exons. Based on careful examination
352   of multiple RNA-seq datasets across several closely related species and sequence homology,
353   we found that in most cases, the four-exon transcript coding for a 139 amino acid protein was
354   the only correct annotation. However, in some rare cases, additional exons have robust
355   expression and require further investigation. In the Corvidae group, annotation exists for

356 transcripts of lengths 170 and 139 aa. The first exon of the longer transcript lacked
357 expression.

358 In comparison, all four transcripts of the shorter transcript are present in the blood tissue of
359 western Jackdaw (*Corvus monedula*) as well as gonad, brain, spleen, and liver of hooded
360 crow (*Corvus cornix*) (**Supplementary Figure S397- S402**). The common canary (*Serinus*
361 *canaria*) has three transcripts with 177, 154, and 139 aa (**Supplementary Figure S403-**
362 **S404**). We checked the expression using liver and skin tissue and found support for all three
363 transcripts. However, upon closer inspection, the transcript with 139 aa was strongly
364 expressed, and the other two transcripts are potentially artifacts. Great tit (*Parus major*) has
365 two transcripts of lengths 169, 139 aa. While the kidney and liver express both transcripts, the
366 first exon has feeble expression and appears artefactual (**Supplementary Figure S405-S406**).

367 The golden eagle (*Aquila chrysaetos*) has four annotated transcripts with lengths of 219, 180,
368 159, and 139 aa. Transcript of 219 aa length contains six exons, transcripts of length 180 aa,
369 and 159 aa have five exons, and 139 aa transcript contains four exons. We found that exon 1
370 showed negligible expression, and exons 2 to 6 have high expression levels. However, exon 1
371 and 2 both have an in-frame stop codon (**Supplementary Figure S407-S411**). Hence, we
372 consider that the 139 aa long transcript expressed in the liver and muscle is correct. Red-
373 throated loon (*Gavia stellata*) has a single five exon transcript of length 155 aa annotated. We
374 discovered a lack of expression in the first exon compared to the last four exons that are
375 orthologous to the transcript of length 139 aa (**Supplementary Figure S412-S413**).

376 The ruff (*Calidris pugnax*) genome annotates three transcripts with lengths of 233, 229, and
377 139 aa. Transcript one and two contain seven exons each, and the third transcript contains
378 four exons. Exons 1 and 2 lack expression in the first two transcripts, and the third exon did
379 not have any start codon explaining the transcript. The last four exons have transcripts and
380 are orthologous to other species' *COA1*/*MITRAC15* gene (**Supplementary Figure S414-**
381 **S418**). In the little egret (*Egretta garzetta*), transcripts of lengths 212 and 203 are annotated
382 and contain five exons. We found evidence of expression of *COA1*/*MITRAC15* in blood
383 tissue (**Supplementary Figure S419-S420**). Although the first exon has a lower level of
384 expression than the last four exons, the consistent occurrence of the fifth exon across many
385 species suggests it might be part of the untranslated region. We annotated and verified the
386 expression of *COA1*/*MITRAC15* in *Phalacrocorax carbo, Phaethon lepturus, Opisthocomus*
387 *hoazin, Leptosomus discolor* (**Supplementary Figure S421-S428**). *Eurypyga helias* has an
388 unverified transcript length of 121 aa. Hence, we screened the genome and RNA-seq data and
389 found its transcript length is 139 aa (**Supplementary Figure S429-S431**). We verified the
390 *COA1*/*MITRAC15* gene expression using RNA-seq data in *Strigops habroptilus* as it had less
391 than 100 percent RNA-seq coverage (**Supplementary Figure S432-S433**). We also
392 examined the RNA-seq data from few other bird species to verify the *COA1*/*MITRAC15* gene
393 (see **Supplementary Figure S434-S481**). Bird species share this conserved gene order
394 (**Supplementary Figure S482**). The Anolis lizard (*Anolis carolinensis*) liver also expresses
395 the *COA1*/*MITRAC15* gene (**Supplementary Figure S483-S485**).

10

396 RNA-seq datasets from the European rabbit (*Oryctolagus cuniculus*) heart and liver showed
397 no evidence of transcription of *COA1/MITRAC15* (see **Supplementary Figure S486-S489**).
398 In contrast to the rabbit, intact *COA1/MITRAC15* gene is present in the Royle's pika
399 (*Ochotona roylei*) and Daurian pika (*Ochotona dauurica*) with blood RNA-seq datasets
400 showing robust expression (see **Supplementary Figure S490**). Screening of RNA-seq
401 datasets from the root ganglion, spinal cord, ovary, liver, spleen, and testis in the naked mole-
402 rat (*Heterocephalus glaber*) revealed no transcription of *COA1/MITRAC15* locus (see
403 **Supplementary Figure S491**). The closely related Damaraland mole-rat (*Fukomys*
404 *damarensis*) has robust *COA1/MITRAC15* expression in the brain, liver, and testis (see
405 **Supplementary Figure S492-S497**). The Brazilian guinea pig (*Cavia aperea*), the guinea pig
406 (*Cavia porcellus*), and the long-tailed chinchilla (*Chinchilla lanigera*) were all found to
407 express the *COA1/MITRAC15* gene robustly (see **Supplementary Figure S498-S505**). The
408 thirteen-lined ground squirrel (*Ictidomys tridecemlineatus*), the Arctic ground squirrel
409 (*Urocitellus parryii*), the groundhog (*Marmota monax*), and the Himalayan marmot
410 (*Marmota himalayana*) do not express the *COA1/MITRAC15* locus (see **Supplementary**
411 **Figure S506-S520**). In contrast to these species, the Eurasian red squirrel (*Sciurus vulgaris*)
412 has an intact *COA1/MITRAC15* expressed in the skin (see **Supplementary Figure S521-**
413 **S522**). Despite gene disrupting mutations, the North American beaver (*Castor canadensis*)
414 *COA1/MITRAC15* locus is expressed in the blood and spleen (see **Supplementary Figure**
415 **S523-S524**). Other tissues such as the brain, liver, stomach, ovarian follicle, skeletal muscle,
416 and kidney do not show any expression at the *COA1/MITRAC15* locus (see **Supplementary**
417 **Figure S525-S530**). The expressed transcript might represent a new long non-coding RNA
418 that cannot produce a functional *COA1/MITRAC15* protein due to the presence of premature
419 stop codons.

420 Chromosomal rearrangement in rodent species has resulted in the movement of genes
421 flanking *COA1*/*MITRAC15* to new locations. The *BLVRA* gene is transcriptionally active in
422 the mouse (*Mus musculus*) liver and heart even though it has translocated to an entirely
423 different location between *AP4E1* and *NCAPH* (see **Supplementary Figure S531**). Genes on
424 the left flank consisting of *HECW1*, *PSMA2*, and *MRPL32* are now located beside *ARID4B*
425 and are expressed in the mouse (see **Supplementary Figure S532-S533**). The genes from the
426 right flank (*MRPS24* and *URGCP*) are also transcriptionally active in the mouse at their new
427 location beside *ANKRD36* (see **Supplementary Figure S534**). Remnants of
428 *COA1*/*MITRAC15* occur between the *PTPRF* and *HYI* genes. However, no transcriptionally
429 activity is seen in the mouse in the region between *PTPRF* and *HYI* genes (see
430 **Supplementary Figure S535**). The new gene order and gene expression patterns are shared
431 by rat (*Rattus norvegicus*) (see **Supplementary Figure S536-S540**), steppe mouse (*Mus*
432 *spicilegus*) (see **Supplementary Figure S541-S545**), Gairdner's shrewmouse (*Mus pahari*)
433 (see **Supplementary Figure S546-S550**), Ryukyu mouse (*Mus caroli*) (see **Supplementary**
434 **Figure S551-S555**), Algerian mouse (*Mus spretus*) (see **Supplementary Figure S556-S560**),
435 deer mouse (*Peromyscus maniculatus*) (see **Supplementary Figure S561-S565**), prairie vole
436 (*Microtus ochrogaster*) (see **Supplementary Figure S566-S570**), golden hamster
437 (*Mesocricetus auratus*) (see **Supplementary Figure S571-S575**), Mongolian gerbil or
438 Mongolian jird (*Meriones unguiculatus*) (see **Supplementary Figure S576-S579**), Chinese

11

439 hamster (Cricetulus griseus) (see **Supplementary Figure S580-S584**), Northern Israeli blind
440 subterranean mole rat (*Nannospalax galili*) (see **Supplementary Figure S585-S589**), white-
441 footed mouse (*Peromyscus leucopus*) (see **Supplementary Figure S590-S594**) and fat sand
442 rat (*Psammomys obesus*) (see **Supplementary Figure S595-S599**). The banner-tailed
443 kangaroo rat (*Dipodomys spectabilis*) (see **Supplementary Figure S600-S601**) has a
444 different gene order and appears to represent one of the pre-EBR species. However, we
445 cannot rule out the possibility of genome assembly errors.

446 The genome assemblies of rodents such as the mouse and rat are well-curated and represent
447 some of the highest-quality reference genomes (Rhie et al., 2021). To ensure that the
448 chromosomal rearrangements identified are correct, we evaluated the correctness of genome
449 assemblies of the mouse (see **Supplementary Figure S602-S608**) and white-footed mouse
450 (*Peromyscus leucopus*) (see **Supplementary Figure S609-S616**) using PacBio long-read
451 sequencing datasets. The mouse genome assembly has been finished to a very high quality
452 using artificial clones of genome fragments (Osoegawa et al., 2000). We further verified the
453 mouse genome assembly by visualizing the coverage of assembly fragments across the
454 genomic regions of interest (see **Supplementary Figure S618-S623**). Repeat regions occur at
455 the boundaries of the evolutionary breakpoint regions (see the last row of screenshots).
456 Although repeat regions are a major contributing factor for the misassembly of genomes, the
457 conserved gene orders across several species and concordance in the timing of the
458 chromosomal rearrangement and support from long-read data support the presence of a
459 genuine change in gene order.

460 The *COA1/MITRAC15* gene is intact and robustly expressed in the platypus
461 (*Ornithorhynchus anatinus*) heart and brain (see **Supplementary Figure S624-S627**). Gene
462 order in the short-beaked echidna (*Tachyglossus aculeatus*) matches the platypus and other
463 outgroup species (see **Supplementary Figure S628**). In contrast to the monotreme species,
464 all marsupial genomes analyzed have a different gene order following chromosomal
465 rearrangements. The gray short-tailed opossum (*Monodelphis domestica*) has the gene
466 *ACVR2B* beside the new location of right flank genes of *COA1/MITRAC15*. The left flank
467 genes are beside *GPR141B*. No traces of the *COA1/MITRAC15* gene are found either in the
468 genome assembly or raw read datasets. The opossum brain expresses these adjacent genes
469 with no transcripts in the intergenic regions (see **Supplementary Figure S629-S631**). The
470 gene order and transcriptional activity were the same in the tammar wallaby (*Notamacropus
471 eugenii*) (Uterus: see **Supplementary Figure S632-S633**), koala (*Phascolarctos cinereus*)
472 (Liver and PBMC: see **Supplementary Figure S634-S636**), the Tasmanian devil
473 (*Sarcophilus harrisii*) (Lung and Spleen: see **Supplementary Figure S637-S639**), and the
474 common brushtail (*Trichosurus vulpecula*) (Liver: see **Supplementary Figure S640-S642**).
475 Long-read sequencing data in the koala supports the correctness of genome assembly (see
476 **Supplementary Figure S643-S645**).

477 The NCBI annotation documents the presence of transcripts, and the *COA1/MITRAC15* gene
478 is remarkably well conserved in ungulate species (see **Supplementary Table S1**). Within
479 ungulate species, certain Cervid species have remarkable sprinting abilities that allow them to
480 escape from predators. However, in addition to sprinting ability, these species are resistant to

481 fatigue. Hence, the prediction from our hypothesis is that gene loss would not occur in Cervid
482 species. The white-tailed deer (*Odocoileus virginianus*) liver and retropharyngeal lymph node
483 and the red deer (*Cervus elaphus*) blood transcriptomes express *COA1/MITRAC15* (see
484 **Supplementary Figure S646-S649**).

485 The *COA1/MITRAC15* gene has undergone duplication within the primate lineage. We
486 screened the genomes of 27 primate species to track down when the gene duplication event
487 occurred. Based on the presence of the duplicate copies, the duplication event is estimated to
488 have happened in the last 43 million years (see **Supplementary Figure S650-S651**).
489 Subsequent duplications have also occurred in Nancy Ma's night monkey (*Aotus nancymaae*)
490 and a shared duplication in the black-capped squirrel monkey (*Saimiri boliviensis*) and the
491 Panamanian white-faced capuchin (*Cebus imitator*). Concurrent with the gene duplication,
492 the intron-exon structure of the *COA1/MITRAC15* gene has also changed (see
493 **Supplementary Figure S652**). The functional copy of the *COA1/MITRAC15* gene is
494 transcriptionally active in the gray mouse lemur (*Microcebus murinus*) (Kidney and Lung:
495 see **Supplementary Figure S653-S654**), the northern greater galago (*Otolemur garnettii*)
496 (Liver: see **Supplementary Figure S655**), Coquerel's sifaka (*Propithecus coquereli*) (see
497 **Supplementary Figure S656**), Nancy Ma's night monkey (*Aotus nancymaae*) (Liver, Heart,
498 and Kidney: see **Supplementary Figure S657-S659**), the common marmoset (*Callithrix
499 jacchus*) (Lung, Liver, and Kidney: see **Supplementary Figure S660-S661**), the Panamanian
500 white-faced capuchin (*Cebus imitator*) (Blood: see **Supplementary Figure S662-S664**), the
501 black-capped squirrel monkey (*Saimiri boliviensis*) (Ovary and Heart: see **Supplementary
502 Figure S665-S668**), the sooty mangabey (*Cercocebus atys*) (Liver: see **Supplementary
503 Figure S669-S670**), the olive baboon (*Papio anubis*) (Kidney and Heart: see **Supplementary
504 Figure S671-672**), the crab-eating macaque (*Macaca fascicularis*) (Blood and Liver: see
505 **Supplementary Figure S673-S674**), the golden snub-nosed monkey (*Rhinopithecus
506 roxellana*) (Heart and Blood: see **Supplementary Figure S675-S676**), human (*Homo
507 sapiens*) (Liver : see **Supplementary Figure S677-S682**) and the Philippine tarsier (*Carlito
508 syrichta*) (see **Supplementary Figure S683**).

509 The intron/exon structure of the *COA1/MITRAC15* gene has undergone several changes in the
510 carnivore lineage (see **Supplementary Figure S684-S685**). However, outgroup species such
511 as the horse (*Equus caballus*) and pangolin (*Manis javanica*) lack intron/exon structure (see
512 **Supplementary Figure S686-S687**). We screened the RNA-seq dataset of multiple carnivore
513 species to validate the annotation and evaluate the intron/exon structure changes. Alternative
514 exon usage was also carefully analyzed to quantify the transcriptional status of
515 *COA1/MITRAC15* in different carnivore species. The *COA1/MITRAC15* gene is
516 transcriptionally active in the meerkat (*Suricata suricatta*) (testis and liver: see
517 **Supplementary Figure S688-S690**), dog (*Canis lupus familiaris*) (spleen and skeletal
518 muscle: see **Supplementary Figure S691-S702**), ferret (*Mustela putorius furo*) (heart and
519 kidney: see **Supplementary Figure S703-S704**), Giant panda (*Ailuropoda melanoleuca*)
520 (heart and liver: see **Supplementary Figure S705-S706**), American black bear (*Ursus
521 americanus*) (liver, kidney, and the brain: see **Supplementary Figure S707-S708**), and
522 Weddell seal (*Leptonychotes weddellii*) (lung and muscle: see **Supplementary Figure S709-**

523   **S712**). Detailed investigation of the splice junctions and actual positions of splice sites in dog

524   transcriptome also supports the *COA1/MITRAC15* gene annotation.

525   Skipping of the dog-like-exon-3 occurs in the transcriptomes of tiger (*Panthera tigris*

526   *altaica*), lion (*Panthera leo persica*), cat (*Felis catus*), and puma (*Puma concolor*) (see

527   **Supplementary Figure S713-S738**). Although annotation for the *COA1/MITRAC15* locus

528   exists in the cheetah (*Acinonyx jubatus*), we found no transcripts in the skin RNA-seq data

529   (see **Supplementary Figure S739-S740**). Close inspection of the *COA1/MITRAC15* locus in

530   cheetah suggests gene loss. We further compared the splice isoforms found in canine and

531   felid species through sashimi plots of the *COA1/MITRAC15* locus. The sashimi plot shows

532   the links between the splice sites and the number of reads that are splice mapped between

533   these sites (see **Supplementary Figure S741-S745**). Changes in the splice enhancers and

534   splice silencer elements were also compared between cat and dog (see **Supplementary**

535   **Figure S746**).

536   Co-expressed genes tend to perform related functions and are lost together. Hence, to identify

537   the loss of genes related to *COA1/MITRAC15,* we identified the top 50 genes co-expressed

538   with human ortholog based on the correlation values in COXPRESdb ver. 7.3 (Obayashi et

539   al., 2019). The presence of orthologs of these co-expressed genes in the high-quality genomes

540   of chicken and mouse using ENSEMBL BioMart (**Supplementary Table S7**). None of these

541   co-expressed genes appear lost in Galliformes or rodents.

542

543   **2.5 Molecular evolutionary analyses**

544   <u>2.5.1 Relaxed selection signatures</u>

545   Molecular signatures of relaxation in the degree of purifying selection generally accompany

546   the loss of gene functionality and have been used as evidence of gene loss (Hecker et al.,

547   2017; Sharma and Hiller, 2018; Shinde et al., 2019). Based on the gene sequence of

548   *COA1/MITRAC15*, we could identify eleven Galliform species with gene-disrupting

549   mutations (see **Supplementary Table S8** and **S9**). Two other Galliform species

550   (*Chrysolophus pictus* and *Phasianus colchicus*) do not express the *COA1/MITRAC15* gene.

551   Hence, we looked for signatures of relaxed selection in each of the terminal branches leading

552   to each Galliform species. We quantified branch-specific selection patterns using the program

553   RELAX (Wertheim et al., 2015) from the HyPhy package and the codeml program from the

554   PAML (Yang, 2007) package. To test for relaxed selection in the terminal branches, we

555   labeled the focal species as the foreground and used the Anseriformes species as the

556   background species. We downloaded the phylogenetic tree with branch lengths from the

557   TimeTree website. Although we found some evidence of relaxed selection in some of the

558   Galliform species, the RELAX program also reported intensification of selection (see

559   **Supplementary Table S10**). None of the internal branches were under relaxed selection.

560   We used the same phylogenetic tree and multiple sequence alignment to obtain branch-

561   specific estimates of ω using the codeml program. The branch-specific estimates of ω are all

562   greater than 1 in *Odontophorus gujanensis*, *Coturnix japonica*, *Meleagris gallopavo*,

563 *Tympanuchus cupido*, *Pavo cristatus*, *Chrysolophus pictus*, *Phasianus colchicus,* and *Numida*
564 *meleagris*. In the case of Galliform species (*Alectura lathami*, *Callipepla squamata,* and
565 *Penelope pileata*) with intact *COA1/MITRAC15* gene, the values of ω are all less than 1.
566 Except for chicken (*Gallus gallus*), species with gene-disrupting changes are not under
567 purifying selection (see **Supplementary Table S10** and **S11**). We evaluated the internal
568 nodes leading to the terminal branches for signatures of relaxed selection to ascertain whether
569 gene loss had occurred in the common ancestor of the Galliform species with gene-disrupting
570 mutations. However, all the ancestral branches appear to be under purifying selection and
571 support the idea of recurrent lineage-specific gene loss suggested by the lineage-specific gene
572 disrupting mutations seen in the Galliform species. Based on this branch-by-branch analysis
573 of selection signatures, we could identify the approximate time frame in which gene loss
574 might have occurred. To get a more accurate estimate of the gene loss timing, we used the
575 method described by (Meredith et al., 2009).

576 We relied upon multiple sequence alignments of carnivores (see **Supplementary Table S12**),
577 rodents (see **Supplementary Table S13**), and primates (see **Supplementary Table S14**) to
578 identify gene disrupting mutations and changes in intron-exon structure. We evaluated each
579 taxonomic group for lineage-specific relaxed selection (see **Supplementary Table S15**).
580 Based on previous reports (Van Der Lee et al., 2017) of positive selection in primates, we
581 additionally identified positively selected sites among primate species (see **Supplementary**
582 **Table S16**).

583 <u>2.5.2 Time of gene loss</u>

584 Different ω values were estimated for both of these labels (see **Supplementary Table S17**).
585 The ω values for mixed(ω_m) and functional(ω_f) branches were estimated using two different
586 codon substitution models (F1X4 and F3X4) to ensure the robustness of the estimates. The
587 calculation of gene loss timing relies upon estimates of T_p (time for which the gene has been
588 pseudogenic) using the method proposed by Meredith et al. (2009) by considering ω_p as 1.
589 Based on the assumptions of 1ds and 2ds, we could get a confidence interval for the
590 estimated time of gene loss (see **Supplementary Table S17**). Gene loss timing was estimated
591 separately in rodents and carnivores (see **Supplementary Table S17**).

592 <u>2.5.3 GC content range and kmer abundance</u>

593 The GC content range (minimum and maximum possible values of GC% for a given amino
594 acid sequence) was calculated (see **Supplementary Table S18**) for *COA1/MITRAC15* and
595 *PDX1* amino acid sequences in rodent and primate species using the window-based tool
596 CodSeqGen (Al-Ssulami et al., 2020). The ContMap function in the R package phytools
597 extrapolates the evolution of GC content along the phylogeny for both genes (see
598 **Supplementary Figure S747-S749**). The program jellyfish (v2.2.8) (Marçais and Kingsford,
599 2011) was used to get the kmers (count command with the flags -C -m 21 -s 1000M and -t
600 16) and their abundance (dump command). The seqkit fx2tab (v0.10.1) (Shen et al., 2016)
601 option calculated the abundance of kmers at different GC content bins and the GC content of
602 each of the *COA1*/*MITRAC15* gene exons (see **Supplementary Table S19**).

603   2.5.4 Quantification of gBGC

604   We calculated the (gBGC) for *COA1/MITRAC15* gene sequences of more than 200 species
605   using the program mapNH(v1.3.0) implemented in the testNH package (Dutheil, 2008). In
606   mapNH, we used multiple sequence alignments of the *COA1/MITRAC15* gene and species
607   tree as input with the flag model=K80. A single gene-wide estimate of gBGC termed GC* is
608   obtained for each species (see **Supplementary Table S20**). These estimates of GC* (GC* >
609   0.9 is significant) help understand the evolution of gBGC along the phylogeny using the
610   ContMap function of the phytools package. Additionally, we also calculated the gBGC for
611   taxonomic group-wise alignments using the programs phastBias and phyloFit implemented in
612   the PHAST (v1.3) package (Capra et al., 2013; Hubisz et al., 2011). In the first step, we use
613   the phyloFit program to fit phylogenetic models to multiple sequence alignments using the
614   specified tree topology (--tree flag with species tree as argument) and substitution model (--
615   subst-mod flag with HKY85 model as argument). Next, the phastBias program with the –bgc
616   flag identified gBGC tracts using the ".mod" file output from phyloFit (see **Supplementary
617   Table S21**, see **Supplementary Figure S750-S778**). The gBGC tracts are positions along the
618   gene with posterior probability >0.5.

619   2.5.5 Computational prediction of RNA binding sites
620   The regulation of gene expression and splicing tends to be determined by the RNA binding
621   sites present within the exons or introns of a gene (Fu and Ares, 2014). A combination of
622   such splice enhancers and splice silencer elements work in concert to facilitate the expression
623   of different isoforms (Dassi, 2017). The *COA1/MITRAC15* gene has changed the intron-exon
624   organization and has acquired novel splice isoforms in felid species. These changes in
625   splicing could result from changes in the RNA binding motifs present within the exons or
626   introns of the gene. In contrast to felids, the splicing pattern in canid species matches the
627   ancestral state. Hence, we compared the *COA1/MITRAC15* gene sequences of canid and felid
628   species to identify differences in the RNA binding motifs. We used the RBPmap (Paz et al.,
629   2014) webserver to predict the RNA binding sites in each exon and intron separately (see
630   **Supplementary Table S22**).

631   **3. Results**

632   **3.1 *COA1/MITRAC15* is a distant homolog of *TIMM21***

633   We identified that the *TIMM21* gene is a distant homolog of *COA1/MITRAC15* based on PSI-
634   Blast and HHblits iterative profile-profile search of the uniport database. Of the 500 top
635   search results from HHblits, 59 have annotation as "Cytochrome C oxidase assembly factor"
636   or "Cytochrome C oxidase assembly protein" or "*COA1*", and 120 as "*TIMM21*" homologs.
637   The annotation of 13 proteins are "hypothetical", nine are "membrane" proteins, eight are
638   "DUF1783 domain-containing" proteins, and 27 proteins are from diverse proteins. The
639   remaining 264 of the 500 hits are "Uncharacterized". The large number of "Uncharacterized"
640   proteins identified are challenging to interpret. Hence, to trace the relationships between the
641   proteins identified as homologs of *COA1/MITRAC15*, we investigated the sequence identity-
642   based clusters established by CLANS (see **Fig. 1A**). The large group of red dots consists of

16

643 proteins annotated as *TIMM21,* and the collection of blue dots contains proteins annotated as
644 *COA1/MITRAC15*. Homologs of *COA1/MITRAC15* from bacterial species form two clusters,
645 a distinct light blue cluster consisting of predominantly Planctomycetes bacteria and a diffuse
646 bunch of brown dots that consists of largely proteobacterial species. The group of orange dots
647 consists of proteins annotated as *COA1/MITRAC15* in fungal genomes. The
648 *COA1/MITRAC15* homologs in plants consist of a yellow cluster consisting of *Arabidopsis*
649 *thaliana* homolog At2g20390 and the magenta cluster of *TIMM21*-like proteins containing
650 *Arabidopsis thaliana* homolog At2g37940. The distinct *COA1/MITRAC15* and *TIMM21*
651 groups found by the cluster analysis suggest that *TIMM21* is a very distant homolog of
652 *COA1/MITRAC15*.

653 The list of proteins identified as homologs of human *COA1/MITRAC15* (**Supplementary**
654 **File S2-S3**) and primate *COA1/MITRAC15* orthologs (**Supplementary File S4**) contain
655 several *TIMM21* like proteins. Iterative PSI-BLAST search identified *TIMM21* homologs
656 from the second iteration onwards and found an increasing number of *TIMM21* hits in each
657 subsequent iteration (see **Supplementary File S5**). The pairwise alignment of the human
658 *COA1/MITRAC15* protein sequence with the *TIMM21* protein with the best alignment (i.e.,
659 *TIMM21* from *Amblyomma cajennenseis*) shows that regions with the most substantial
660 homology include the membrane-spanning domain and covers >100 residues (see **Fig. 1B**).
661 In addition to the primary sequence-homology detected, both *TIMM21* and *COA1/MITRAC15*
662 are known to play prominent roles in the mitochondria and have comparable secondary
663 structures (see **Fig. 1C, 1D**). The strong homology between these proteins also allows for
664 homology-based modeling of the tertiary structure of the *COA1/MITRAC15* protein using
665 *TIMM21* as a model (see **Supplementary Figure S779-S783**). Despite the lack of well-
666 conserved motifs, we found three well-matching columns (marked with a '|' sign in **Fig. 1B**)
667 between residues 91 to 95 in *COA1/MITRAC15*. Two consecutive conserved residues occur
668 at residues 57-58, 64-65, and 67-68 of *COA1/MITRAC15*. The similar sequence, structure,
669 and function of *COA1/MITRAC15* and *TIMM21* strongly support that these genes are
670 homologs.

671 **3.2 *COA1/MITRAC15* gene duplication, pseudogenisation, and exon reorganization**

672 The sequence divergence between *COA1/MITRAC15* and *TIMM21* appears to result from
673 changes in the *COA1/MITRAC15* gene intron/exon organization. The *COA1/MITRAC15* gene
674 has undergone independent gene duplications followed by pseudogenisation and degeneration
675 of the duplicated copy in both primates and carnivores. Consequently, the functional and
676 pseudogene copies of *COA1/MITRAC15* have diverged considerably and formed distinct
677 haplotypes. For example, the blast search of sequencing raw read data from the human
678 genome with *COA1/MITRAC15* gene sequence as a query results in two distinct haplotypes.
679 One set of reads correspond to the intact *COA1/MITRAC15* gene in humans, and the other set
680 of reads are from the pseudogenic copy (see **Fig. 2A**). Comparative analysis of primate
681 genome assemblies suggests that the pseudogenic copy results from a duplication of
682 *COA1/MITRAC15* within the primate lineage (see **Supplementary Figure S651**). After
683 duplication of the *COA1/MITRAC15* gene in primates, an extension of the N-terminal region
684 has occurred in Cercopithecidae and Catarrhini and is transcriptionally active (see

17

685   **Supplementary Figure S652**). However, new world monkeys do not have this N-terminal
686   extension denoted as exon-1a. Both Cercopithecidae and Catarrhini have an additional start
687   codon in exon-1a upstream from the original start codon in the ancestral exon-1 denoted as
688   exon-1b in species with N-terminal extension. A striking difference between Cercopithecidae
689   and Catarrhini is the lack of the internal start codon in Cercopithecidae, where Catarrhini has
690   a start codon. Since proteome level data is not available for these species, we rely solely on
691   the RNA-seq datasets and start and stop codons within the expressed transcripts to evaluate
692   the exon/intron structure changes. Using these carefully annotated primate sequences of
693   *COA1/MITRAC15*, we verified (see **Supplementary Table S16**) a previous report (Van Der
694   Lee et al., 2017) of positive selection in this gene among primates.

695   Independent duplication of *COA1/MITRAC15* has occurred in carnivores (see
696   **Supplementary Figure S685**). However, similar to primates, the duplicated copy has
697   undergone pseudogenization and diverged from the functional gene sequence. For example,
698   sequencing raw read data in the tiger consist of two distinct haplotypes corresponding to the
699   intact and pseudogene copies (see **Fig. 2B**). While the intact copy is located at a genomic
700   region (*STK17A* & *HECW1* upstream and *BLVRA* & *VOPP1* downstream) with conserved
701   synteny across other mammals, the pseudogene copy occurs adjacent to the *PRR32* gene.
702   Outgroup species such as horse (*Equus caballus*) and pangolin (*Manis javanica*) have a
703   single copy of the *COA1/MITRAC15* gene with all raw reads supporting a single haplotype
704   (see **Supplementary Figure S686**). Both sub-orders (Caniformia and Feliformia) within
705   Carnivora share this duplication of the *COA1/MITRAC15* gene (see **Supplementary Figure
706   S685**).

707   The intact *COA1/MITRAC15* copy is expressed in diverse transcriptomes among Caniformia
708   species, while the pseudogene copy lacks expression. The first and second exons are
709   orthologous; however, the genomic location of the transcribed third exon is different between
710   Feliformia (cat-like-exon-3) and Caniformia species (dog-like-exon-3) (see **Fig. 3**). The final
711   exon of the *COA1/MITRAC15* gene in Feliformia extends to 163 base pairs (*Panthera tigris
712   altaica, Panthera leo, Panthera pardus,* and *Lynx lynx*) and 160 base pairs (*Puma concolor
713   and Felis catus*) compared to the 100 base pairs in Caniformia species. A single deletion
714   event causes the difference of three base pairs between these two groups of Feliformia at the
715   24th base of exon-4. The extended final exon shared by all Feliformia species results from a
716   two-base frameshift deletion before the erstwhile stop codon in exon-4. Despite the extended
717   last exon in Feliformia species, the full-length open reading frames of Feliformia (130/131
718   amino acids) and Caniformia (135 amino acids) are comparable.

719   The shorter reading frame in Feliformia results from the majority of *COA1/MITRAC15*
720   transcripts skipping the dog-like-exon-3, whose inclusion results in premature stop codons in
721   all the seven Feliformia species. The dog-like-exon-3 is present in all *COA1/MITRAC15*
722   transcripts of Caniformia species and does not contain gene-disrupting changes. A single base
723   deletion in all Feliformia species changes the end phase of exon-2 to maintain an intact
724   reading frame while skipping the dog-like-exon-3. Transcriptomes of the cat (*Felis catus*)
725   from the spleen (see **Supplementary Figure S744**) and puma (*Puma concolor*) from blood
726   (see **Supplementary Figure S745**) exhibit expression of a proto cat-like-exon-3 which gets

727　spliced into some of the *COA1/MITRAC15* transcripts. However, the majority of transcripts
728　skip this proto cat-like-exon-3 which contains premature stop codons. These changes in exon
729　splicing patterns between Caniformia and Feliformia species appear to result from changes in
730　splice factor binding sites at the *COA1/MITRAC15* locus (see **Supplementary Figure S746**).

731　Except for the cheetah (*Acinonyx jubatus*), intact transcribed open reading frames are
732　discernible in all carnivore species at the *COA1/MITRAC15* locus identified based on
733　conserved synteny across mammals (see **Fig. 3**). The gene disrupting premature stop codon in
734　the cheetah is due to a single base C->T substitution at the 27th base of exon-2 assembled at
735　the *COA1/MITRAC15* locus. The duplicated copy of *COA1/MITRAC15* also contains a
736　premature stop codon at the 49th base of exon-2 caused by a single base insertion at the 11th
737　base of exon-2. The *COA1/MITRAC15* gene transcripts are missing in the skin transcriptome
738　of the cheetah. Hence, multiple lines of evidence support *COA1/MITRAC15* gene loss in the
739　cheetah. Gene loss in the cheetah occurred between 2.98-3 MYA (**Supplementary Table
740　S17**).
741

742　In contrast to primates and carnivores, reads support multiple haplotypes of
743　*COA1/MITRAC15* only in the second exon of naked mole-rat (see **Fig. 2C**). Hence, the
744　duplicated copy of *COA1/MITRAC15* in naked mole-rat appears to have mostly degraded.
745　However, we cannot rule out the possibility that the reads from other haplotypes spanning the
746　remaining three exons are missing due to high GC content. The sequencing reads support the
747　presence of a single intact open reading frame in the red squirrel (see **Fig. 2D**) and platypus
748　(see **Fig. 2E**). Although a single haplotype occurs in the raw read dataset of chicken, this
749　haplotype has gene-disrupting changes (see **Fig. 2F**). The gene-disrupting modifications
750　identified in the chicken *COA1/MITRAC15* gene were investigated further by screening long-
751　read datasets, transcriptomes, and genomes of various Galliform species.

752　**3.3 *COA1/MITRAC15* gene loss in Galliform species**

753　We found evidence of eight independent gene-disruption events in the *COA1/MITRAC15*
754　gene in the galliform group (see **Fig. 4A**). The chicken (*Gallus gallus*) and Amazonian wood
755　quail (*Odontophorus gujanensis*) have single-base G to T substitutions at the 69th base of
756　exon-2 and the 72nd base of exon-4 in the *COA1/MITRAC15* gene, respectively (see
757　**Supplementary Table S9**). These substitutions lead to (**GAA→TAA**) premature stop
758　codons. Gene loss of *COA1/MITRAC15* is estimated to have occurred between 23 MYA and
759　29 MYA in chicken and between 17 MYA and 18 MYA in the Amazonian wood quail (see
760　**Supplementary Table S9** and **S17**). In the Indian peafowl (*Pavo cristatus*), two single-base
761　deletions, one at 37th base of exon-1 and another at 31st base of exon-4, result in premature
762　stop codons in exons 2, 3, and 4. The gene disrupting mutations identified in the Indian
763　peafowl (*Pavo cristatus*) also occur in the green peafowl (*Pavo muticus*). Loss of the
764　*COA1/MITRAC15* gene is estimated to have occurred between 20 MYA and 29 MYA in the
765　peafowls (see **Supplementary Table S17**). The exon-2 of Pinnated grouse (*Tympanuchus
766　cupido*) and Helmeted guineafowl (*Numida Meleagris*) have independent 13 and 17 base

767 deletions. Changes in the reading frame resulting from these deletions lead to several

768 premature stop codons (see **Supplementary Table S9**). The 13-base deletion in the exon-2 of

769 the Pinnated Grouse (*Tympanuchus cupido*) also occurs in Gunnison grouse (*Centrocercus*

770 *minimus*), Rock ptarmigan (*Lagopus muta*), and the black grouse (*Lyrurus tetrix*). The

771 estimated time of gene loss in these four species is between 18 MYA and 20 MYA, and for

772 Helmeted guineafowl, it is between 39 MYA and 40 MYA (see **Supplementary Table S17**).

773

774 In Turkey (*Meleagris gallopavo*), a two-base substitution at bases 7 & 8 and a single base

775 deletion at the 37th base of exon-2 result in a frameshift in the *COA1/MITRAC15* gene leading

776 to premature stop codons. Gene loss in Turkey is estimated to have occurred between 14

777 MYA and 18 MYA. Two closely spaced single base substitutions (**AA**C→**T**A**A**) at 48th and

778 50th positions of exon-2 result in a premature stop codon in the Japanese quail (*Coturnix*

779 *japonica*). The time of gene loss in the Japanese quail is estimated between 35 MYA and 36

780 MYA (see **Supplementary Table S17**). The Mikado pheasant (*Syrmaticus mikado*) has an 11-

781 base deletion in exon-4, and the time of gene loss is between 14 MYA and 16 MYA. Other

782 Galliform species such as Australian brushturkey (*Alectura lathami*), Blue quail (*Callipepla*

783 *squamata*), Ring-necked pheasant (*Phasianus colchicus*), Golden pheasant (*Chrysolophus*

784 *pictus*), and White-crested guan (*Penelope pileata*) have intact *COA1/MITRAC15* coding

785 sequences. The coding region is intact in outgroup species such as Swan goose (*Anser*

786 *cygnoides*), Duck (*Anas platyrhynchos*), and Magpie goose (*Anseranas semipalmata*). Five

787 genes upstream (*BLVRA*, *VOPP1*, *LANCL2*, *EGFR*, and *SEC61G*) and downstream (*STK17A*,

788 *HECW1*, *MRPL32*, *PSMA2*, and *C7orf25*) from *COA1/MITRAC15* retain a conserved order in

789 birds. We relied upon this conserved order to verify the 1 to 1 orthology of the

790 *COA1/MITRAC15* gene across species (see **Fig. 4B**).

791

792 Signatures of relaxed selection in Galliform species with gene disrupting changes further

793 support the loss of *COA1/MITRAC15* in these lineages (see **Supplementary Table S10**).

794 Despite intact coding regions, the Ring-necked pheasant (*Phasianus colchicus*) and Golden

795 pheasant (*Chrysolophus pictus*) *COA1/MITRAC15* sequences also have signatures of relaxed

796 selection (see **Supplementary Table S10**). None of the four tissues (Brain, Spleen, Liver, and

797 Gonad) for which RNA-seq data is available from the Ring-necked pheasant shows any

798 *COA1/MITRAC15* transcripts. Similarly, the one tissue (Skin) for which RNA-seq data is

799 available in the Golden pheasant (*Chrysolophus pictus*) does not show *COA1/MITRAC15*

800 expression. To evaluate the relevance of the gene disrupting mutations and signatures of

801 relaxed selection identified in galliform species, the transcriptomes of Galloanserae species

802 were screened to assess the transcriptional status of *COA1/MITRAC15*. We evaluated RNA-

803 seq datasets of several comparable tissues across species and found the *COA1/MITRAC15*

804 gene is not transcribed in chicken despite screening more than 20 tissues (see **Fig. 4C**). Other

805 Galloanserae species have RNA-seq data available for very few tissues. We evaluated the

806 RNA-seq datasets from six tissues (Brain, Spleen, Skin, Liver, Gonad, and Blood) available

807 in several species for the presence of *COA1/MITRAC15* transcripts. Our search consistently

808 revealed transcription of *COA1/MITRAC15* gene in Anseriformes species in contrast to lack

809   of transcription in Galliform species except for Australian brushturkey (*Alectura lathami*) and
810   northern bobwhite (*Colinus virginianus*), which have intact *COA1/MITRAC15* gene that is
811   under strong purifying selection (see **Fig. 4C**). The lack of gene expression and signatures of
812   relaxed selection in the Ring-necked pheasant (*Phasianus colchicus*) and Golden pheasant
813   (*Chrysolophus pictus*) suggests gene loss. The putative gene loss in both these species
814   occurred between 12 MYA and 13 MYA.
815
816   **3.4 Complete erosion of *COA1/MITRAC15* locus is challenging to prove**

817   Search for the *COA1/MITRAC15* gene in the mammoth genome demonstrated striking
818   heterogeneity in coverage of the four exons based on the Illumina ancient DNA sequencing
819   datasets analyzed (see **Fig. 5A-F**). Despite having comparable genome-wide coverage, we
820   could see that not all exons occur in all the datasets. For instance, the re-sequencing dataset
821   from PRJEB29510 (162 Gb) does not have reads from any of the four *COA1/MITRAC15*
822   exons. However, the datasets from PRJEB7929 (88.34 Gb) and PRJNA397140 (155 Gb)
823   have reads covering three exons each despite having much lower genome-wide coverage. The
824   third exon of *COA1/MITRAC15* was also missing or had fewer reads than the other three
825   exons in most of the datasets. The dataset from PRJEB42269 had no reads from the first exon
826   but had a few reads from exons three and four. We reasoned that this heterogeneity in the
827   coverage of various *COA1/MITRAC15* exons was mainly a result of the well-established
828   sequencing bias of Illumina that results in inadequate coverage of GC-rich regions (Chen et
829   al., 2013). Quantification of GC content in each of the four *COA1/MITRAC15* exons and
830   kmer abundance in different GC content bins in each of the mammoth Illumina re-sequencing
831   datasets explains most of the heterogeneity in coverage between datasets as well as exons
832   (see **Fig. 5G**). In contrast to the *COA1/MITRAC15* gene, we did not see heterogeneity in the
833   sequencing coverage of *TIMM21* exons despite comparable GC content for some of the exons
834   (see **Fig. 5G** and **Supplementary Figure S784-S791**).

835   The heterogeneity in sequencing coverage of *COA1/MITRAC15* exons demonstrates the
836   challenges of detecting its presence in Illumina sequencing datasets. GC-biased gene
837   conversion (gBGC) plays a defining role in the base composition for any particular gene or
838   genomic region. It preferentially fixes GC in AT/GC heterozygotes and increases the GC
839   content. The GC content of the *COA1/MITRAC15* exons can be driven to extreme values by
840   gBGC. The magnitude of gBGC also varies across the genome within a species as well as
841   between species. Therefore, *COA1/MITRAC15* orthologs from closely related species or even
842   duplicated copies of *COA1/MITRAC15* within the same species can have very different GC
843   content. Such differences in GC content can result in correspondingly different coverage of
844   the gene sequence in Illumina data and masquerade as a gene loss event (Botero-Castro et al.,
845   2017; Hargreaves et al., 2017).

846   A well-known example for high GC content impeding sequencing is the gene *PDX1*, which
847   has striking differences in GC content between closely related rodent species and requires
848   dedicated GC-rich DNA enrichment protocols for sequencing. We contrasted
849   *COA1/MITRAC15* with the *PDX1* genes of rodents by comparing the minimum (see

850 **Supplementary Table S20**) and maximum (see **Supplementary Table S20**) GC contents
851 possible given their amino acid sequence. Although *COA1/MITRAC15* had lower GC content
852 levels than *PDX1*, we could not rule out the possibility of gBGC affecting some of the exons.
853 The values of GC* across more than 200 vertebrate species with intact *COA1/MITRAC15*
854 reading frames suggested considerable heterogeneity between taxa (see **Supplementary**
855 **Figure S749**). In each taxonomic group, the prevalence of gBGC was separately quantified
856 (see **Supplementary Figure S750-S772**). Strong patterns of gBGC occur in the
857 *COA1/MITRAC15* sequence of several species (see **Supplementary Figure S750-S772**:
858 elephant (*Loxodonta africana*), kagu (*Rhynochetus jubatus*), blue-crowned manakin
859 (*Lepidothrix coronata*), Chilean tinamou (*Nothoprocta perdicaria*), American black bear
860 (*Ursus americanus*), North American river otter *(Lontra canadensis)*, meerkat (*Suricata*
861 *suricatta*), California sea lion (*Zalophus californianus*), little brown bat (*Myotis lucifugus*),
862 large flying fox (*Pteropus vampyrus*), southern pig-tailed macaque (*Macaca nemestrina*),
863 Brazilian guinea pig (*Cavia aperea*), sheep (*Ovis aries*), eastern brown snake (*Pseudonaja*
864 *textilis*) and the Goode's thornscrub tortoise (*Gopherus evgoodei*)). However, none of the
865 rodent species with intact *COA1/MITRAC15* show any striking gBGC patterns. The GC
866 content vs. kmer abundance plots of Pacbio, BGI-seq, and Illumina datasets spans the entire
867 range of GC contents seen in *COA1/MITRAC15* exons (see **Supplementary Figure S775**).
868 Since the GC content of individual *COA1/MITRAC15* exons differs between species groups
869 (see **Supplementary Figure S775-S778**), the high GC content of certain regions might result
870 in inadequate sequencing coverage of the *COA1/MITRAC15* gene in some species. Hence, the
871 lack of sequencing reads covering *COA1/MITRAC15* cannot serve as definitive evidence of
872 gene loss.

873 ### 3.5 *COA1/MITRAC15* occurs in an evolutionary breakpoint region

874 We find evidence of *COA1/MITRAC15* gene disrupting mutations and lack of gene
875 expression in multiple RNA-seq datasets despite a conserved gene order in the rabbit
876 (*Oryctolagus cuniculus*), naked mole-rat (*Heterocephalus glaber*), and four Sciuridae species
877 (*Urocitellus parryii*, *Spermophilus dauricus*, *Ictidomys tridecemlineatus*, *Marmota marmota*
878 *marmota*). The gene disrupting mutations identified in the rabbit *COA1/MITRAC15* gene
879 includes a two-base pair deletion at the 22nd codon of exon-1 and single base pair deletions in
880 exon-2 at the 13th and 37th codons. Gene disrupting changes in the third exon consist of a
881 five-base insertion between the 11th and 12th codon, one base insertion at the 17th codon, and
882 one base deletion in the 23rd codon (see **Fig. 6** and **Supplementary Table S13**). These six
883 gene-disrupting changes result in premature stop codons in exon-2 and exon-4. Gene loss in
884 the rabbit is estimated to have occurred between 12 MYA and 17 MYA (see **Fig. 6** and
885 **Supplementary Table S17**). The lack of *COA1/MITRAC15* RNA-seq reads in tissues such as
886 the brain, liver, and testis that express *COA1/MITRAC15* in closely related species supports
887 the loss of the *COA1/MITRAC15* gene in the naked mole-rat. Besides the lack of a start
888 codon, a single gene disrupting mutation is found in the naked mole-rat *COA1/MITRAC15*
889 gene and consists of a single base deletion at the 21st codon of exon-1. Gene loss in the naked
890 mole-rat is estimated between 7 MYA and 11 MYA (see **Supplementary Table S17** and **Fig.**
891 **6**).

22

892 The presence of common gene disrupting changes such as a one base pair insertion at second
893 codon of exon-1, two base pair insertion at 25th codon of exon-2, seven base pair deletion
894 between 25th and 26th codon of exon-4, and a 2-base insertion at 33rd codon of exon-4
895 supports a shared gene loss in four Sciuridae species (*Urocitellus parryii*, *Spermophilus*
896 *dauricus*, *Ictidomys tridecemlineatus*, *Marmota marmota marmota*). The *COA1/MITRAC15*
897 gene of alpine marmot has additional gene disrupting changes consisting of a 2-base insertion
898 between the 8th and 9th codon of exon-1 and a single nucleotide substitution at the 26th codon
899 of exon-2. The 2-base insertion at the 33rd codon of exon-4 has extended to a five-base pair
900 insertion in the Daurian ground squirrel (*Spermophilus dauricus*). The estimated time of gene
901 loss for this shared event is between 10 MYA and 30 MYA (see **Supplementary Table S17**
902 and **Fig. 6**).

903 The presence of intact open reading frames robustly expressed at syntenic locations in closely
904 related (~30 to 50 million years) species strongly supports at least three independent
905 *COA1/MITRAC15* gene loss events (see **Fig. 6**). Multiple gene-disrupting mutations in the
906 *COA1/MITRAC15* gene of the North American beaver (*Castor canadensis*) suggest a fourth
907 independent gene loss event. Gene-disrupting mutations in the beaver result in at least two
908 premature stop codons. In the first exon, single-base deletions occur in the 3rd and 20th codon,
909 a four-base insertion occurs between 33rd and 34th codon. The second exon has a single-base
910 deletion in the 33rd codon and a seven-base pair deletion between 29th and 30th codons. A
911 single base deletion occurs at the 12th codon of exon-3 (see **Supplementary Table S13** and
912 **Fig. 6**). The genome assembly of the North American beaver is fragmented, and the synteny
913 of the flanking regions cannot be verified. The Illumina sequencing raw reads support the
914 gene disrupting mutations identified in the genome assembly (**Supplementary File S1**), and
915 duplicate copies don't occur. The loss of the *COA1/MITRAC15* gene in the beaver is
916 estimated to have occurred sometime between 3 MYA and 23 MYA (see **Supplementary**
917 **Table S17** and **Fig. 6**).

918 The North American beaver is phylogenetically closely related to the Ord's kangaroo rat
919 (*Dipodomys ordii*) and the lesser Egyptian jerboa (*Jaculus jaculus*). The more contiguous
920 genome assemblies of the jerboa and kangaroo rat allow verification of a conserved gene
921 order likely to be shared by the North American beaver (see **Fig. 6**). The presence of
922 repetitive elements and lack of long-read sequencing data in most rodent species prevents
923 genome assembly verification. Hence, we have screened the genomes of several closely
924 related rodent species and verified the genome assemblies using long-read sequencing data or
925 cloned fragments that cover parts of the genome. Gaps in the genome assembly also hamper
926 the identification of the correct gene order. Previous reports that examined genome
927 assemblies and EST data have claimed loss of the *STK17A* gene in mice due to a
928 chromosomal rearrangement spanning this genomic region (Fitzgerald and Bateman, 2004).
929 Detailed examination of gene order flanking the *COA1/MITRAC15* locus in several rodent
930 genomes revealed the occurrence of this previously reported chromosomal rearrangement
931 event (see **Fig. 6**).

932 Identifying gene loss events coinciding with EBRs is notoriously challenging and has
933 motivated nuanced inferences in both bird (Botero-Castro et al., 2017) and rodent species

23

934   (Hargreaves et al., 2017). Nonetheless, more than a dozen rodent species share the putative
935   combined loss of *STK17A* and *COA1/MITRAC15* (see **Fig. 6**). Based on the presence of
936   adjacent genes, the rearranged regions could be tracked down to two different chromosomes
937   (see **Fig. 6, O6,** and **O9**). Genes on the left flank of *STK17A-COA1-BLVRA* consist of
938   *PSMA2*, *MRPL32,* and *HECW1* in gene orders O1 to O5. After the chromosomal
939   rearrangement, the same sequence of genes can be found in gene order O9 and occur adjacent
940   to *ARID4B* and *GGPS1.* Genes on the right flank of *STK17A-COA1-BLVRA* consist of
941   *MRPS24*, *URGCP,* and *UBE2D4* in gene order O4. Several other gene orders (O1 to O5)
942   occur on the right flank in various species. The sequence of genes found on the right flank in
943   gene order O4 is also found sequentially in gene order O6 and occurs adjacent to *ANKRD36*
944   and *CCDC117* after the chromosomal rearrangement.

945   We found that the *BLVRA* gene has translocated to an entirely new location and does not co-
946   occur with either the left or right flank. However, the new location of the *BLVRA* gene
947   between the *NCAPH* and *ITPRIPL1* genes on the left flank and *AP4E1* and *SPPL2A* genes on
948   the right flank is consistently conserved across all 14 post-EBR species and corresponds to
949   gene order O7. Both *COA1/MITRAC15* and *STK17A* are missing in the post-EBR rodent
950   genome assemblies. The search of the genome assemblies, sequencing raw read datasets, and
951   RNA-seq datasets also failed to find any evidence of an intact *COA1/MITRAC15* or *STK17A*
952   gene. All raw read and genome assembly hits for *STK17A* while using queries from pre-EBR
953   rodent genomes could be traced back to the *STK17B* gene that matches with the *STK17A* gene
954   at a short sequence stretch. The *STK17A* gene is lost or has sequence properties that prevent it
955   from being sequenced with currently available technologies. The exon-1 region of
956   *COA1/MITRAC15* occurs in a gene desert region between *PTPRF* and *HYI* genes in post-
957   EBR species. Using blast search of *COA1/MITRAC15* introns, we found strong support for
958   the existence of *COA1/MITRAC15* intron-2 close to the exon-1 hit. Pairwise genome
959   alignments provide support for the presence of *COA1/MITRAC15* gene remains at this
960   location (see **Supplementary File S6**). Notably, the *COA1/MITRAC15* remnants of a
961   truncated exon-1 and intron-2 occur in the gene desert located between *PTPRF* and *HYI*
962   genes only in post-EBR species. None of the pre-EBR species had any such remains. Hence,
963   the *COA1/MITRAC15* remnants between *PTPRF* and *HYI* genes are unlikely to have resulted
964   from duplicated copies of *COA1/MITRAC15*. The synteny of this region is well conserved
965   with *KDM4A* and *PTPRF* on the left flank and *HYI* and *SZT2* on the right side and
966   corresponds to gene order O8. Careful examination of this region in RNA-seq datasets found
967   no evidence of transcripts.

968   Comparison of gene order in marsupial species with various outgroup species (including the
969   platypus and short-beaked echidna from the order Monotremata) identified the presence of an
970   independent chromosomal rearrangement event spanning the *COA1/MITRAC15* locus (see
971   **Fig. 7**). In contrast to the rodent-specific EBR, we found that the *STK17A* gene is intact in
972   post-EBR (gene order O2 and O3 in **Fig. 7**) marsupial species. However, an extensive search
973   of marsupial genomes, transcriptomes, and raw sequencing read datasets (including high
974   coverage Pacbio datasets for the Koala) failed to find any evidence of *COA1/MITRAC15*
975   orthologs or its remnants. Lack of sequencing reads from *COA1/MITRAC15* in marsupial

24

976 species suggests either complete erosion of the gene or drastic change in sequence
977 composition that eludes sequencing with currently available technologies.

978 **4. Discussion**

979 Our search of the sequence databases identified that *COA1/MITRAC15* and *TIMM21* are
980 distant homologs with representative genes found in animals, plants, and fungi. The
981 occurrence of *COA1/MITRAC15* homologs in α-proteobacteria supports an ancestral role for
982 these genes (Kurland and Andersson, 2000). Endosymbiotic theories explain the origin of
983 eukaryotes and their mitochondria (Martin et al., 2015). Cells that lacked mitochondria never
984 attained the complexity seen in eukaryotes. Hence, true intermediates to this transition from
985 prokaryotes to eukaryotes are not available. The number of genes within mitochondria varies
986 from five to over a hundred in different eukaryotes (Bevan and Lang, 2004). Species with a
987 higher number of genes in the mitochondria provide a snapshot of the endocytosed bacteria-
988 like ancestral entity. The gene-rich mitochondrial genomes of Jakobid protists are models to
989 study the evolution of mitochondria (Burger et al., 2013). Although *TIMM21* homologs are
990 present in the genome of the Jakobid *Andalucia godoyi*, the *COA1/MITRAC15* gene is
991 missing (Gray et al., 2020). The single-copy homologs of *TIMM21* and *COA1/MITRAC15* in
992 bacterial species and Jakobid protist mitochondria suggest that duplication of *TIMM21* might
993 have occurred during the movement of *TIMM21* homologs from the mitochondria to the
994 nucleus. A sampling of more Jakobid genomes might resolve the timing of duplication of
995 *TIMM21* to *COA1/MITRAC15*.

996 The *COA1/MITRAC15* gene has undergone subsequent duplication events in carnivores and
997 primates. The prevalence of such duplication events suggests that either a higher
998 *COA1/MITRAC15* protein dosage is not harmful or sophisticated regulatory machinery to
999 maintain the correct dosage exists. Genes with duplicated copies have greater flexibility for
1000 subfunctionalization or neofunctionalisation (Taylor and Raes, 2004). In contrast to gene
1001 duplication, the origin of new splice-isoforms increases the transcriptome complexity without
1002 increasing the gene count. The evolution of phenotypic novelty through alternative splicing
1003 has received greater attention thanks to the availability of large-scale transcriptomic and
1004 proteomic datasets in diverse species (Bush et al., 2017). While positive selection has a role
1005 in specific examples of alternative splicing (Parker et al., 2014; Ramensky et al., 2008), the
1006 vast majority of splicing is probably noisy, and neutral processes may explain its evolution
1007 (Pickrell et al., 2010). Alternative splicing also reduces premature protein truncation due to
1008 purifying selection (Xing and Lee, 2004). In the case of felid species, the alternative splicing
1009 of the third exon (see **Fig. 2**) may have evolved in response to the gene-disrupting changes.
1010 Verifying the relevance of the alternative splicing observed at the transcriptional level would
1011 require further scrutiny of the protein level isoforms of the *COA1/MITRAC15* gene in felid
1012 and canid species. In primate species, the potential addition of the extra coding-exon occurs
1013 by a shift of the start codon into the untranslated region. Such changes at the reading frame
1014 termini occur when the gene is under relaxed selective constraints (Shinde et al., 2019).
1015 Acquisition of novel protein-coding sequences through changes in the exon length is also
1016 known to occur (Kishida et al., 2018). We speculate that drastic lineage-specific changes in

1017 purifying selection have allowed for changes in intron-exon structure resulting in the
1018 evolution of new splice-isoforms of *COA1/MITRAC15*.

1019 Gene loss can be dealt with through compensation from another gene (Xiong and Lei, 2020)
1020 or is associated with a biological pathway rewiring (Vijay, 2020). Large-scale changes in
1021 gene content are associated with major evolutionary transitions that drastically alter the
1022 fitness landscape. Prominent examples of such shifts are the origin of flight in birds
1023 (Meredith et al., 2014) and the movement of mammals from land to water seen in cetaceans
1024 (Huelsmann et al., 2019). Recurrent gene loss events following relaxed selective constraint in
1025 various other lineages have also been documented (Schneider et al., 2019; Sharma and Hiller,
1026 2018; Valente et al., 2020). Loss of genes in the Galliform lineage while being intact in the
1027 Anseriformes lineage has been linked to differences in the immune response of these clades
1028 (Barber et al., 2010; S. Sharma et al., 2020). The *COA1/MITRAC15* gene is not known to
1029 have any obvious immune functions, and its loss in Galliform birds appears to be a
1030 consequence of relaxed selection on the OXPHOS pathway. Our computational analysis of
1031 more than 200 vertebrate genomes has found that the *COA1/MITRAC15* gene is intact and
1032 transcribed in most species, except for the Cheetah, Galliformes, rodents, and marsupial
1033 species. Notably, the detailed investigation of the *COA1/MITRAC15* gene in other bird
1034 species that are flightless or have a limited ability to fly has found an intact transcribed gene.
1035 Therefore, the loss of the *COA1/MITRAC15* gene appears to be associated with changes in
1036 skeletal muscle fiber composition. The prominent role of mitochondria in skeletal muscles is
1037 evident from diseases of the muscle tissue caused by defects in mitochondrial genes (Gan et
1038 al., 2018).

1040 The correlation between recurrent gene loss and the presence of specific phenotypes has
1041 provided crucial insights into the evolution of these traits. Stomach loss in gnathostomes co-
1042 occurs with the loss of several genes that code for digestive enzymes (Castro et al., 2013).
1043 The loss of ketogenesis has occurred through the recurrent loss of the *HMGCS2* gene (Jebb
1044 and Hiller, 2018). Gene losses associated with dietary composition, the patterns of feeding,
1045 and gut microbiomes have also been identified (Hecker et al., 2019). Recurrent loss of Toll-
1046 like receptors (TLRs), which play prominent roles in the innate immune system, is associated
1047 with impaired ability to detect extracellular flagellin (V. Sharma et al., 2020). The repeated
1048 loss of the cortistatin gene is related to modifications in the circadian pathway (Valente et al.,
1049 2020). In the *COA1/MITRAC15* gene, we record the independent occurrence of gene
1050 disrupting changes in closely related species of Galliformes and rodents. However, we can
1051 rule out the possibility of a common regulatory mutation that initially resulted in the loss of
1052 gene expression followed by the independent accumulation of the gene disrupting changes
1053 that we observe. Our hypothesis predicts gene loss following changes in skeletal muscle fiber
1054 composition. The *COA1/MITRAC15* gene does not directly alter the muscle fiber composition
1055 and might have subsequently experienced relaxed selective constraint due to increased
1056 glycolytic muscle fibers. Hence, it is tempting to speculate that the independent gene
1057 disrupting changes reflect recurrent gene loss events. However, the mechanistic basis of
1058 changes in muscle fiber composition between species is yet to be understood. Identifying the

1059 genetic changes that determine muscle fiber composition and the sequence of events would
1060 provide greater clarity regarding when and why the *COA1/MITRAC15* gene loss occurred.

1061 Evolutionary Breakpoint Regions (EBRs) are genomic regions that have undergone one or
1062 more structural changes resulting in altered karyotypes between lineages (Lemaitre et al.,
1063 2009). Recurrent non-random structural changes at the same regions in multiple lineages
1064 potentially occur due to the presence of repeat elements (Farré et al., 2016; Schibler et al.,
1065 2006), chromosome fragile sites (Durkin and Glover, 2007; Ruiz-Herrera et al., 2006, 2005),
1066 nucleotide composition, methylation level (Carbone et al., 2009) and chromatin state (Boteva
1067 et al., 2020; Huvet et al., 2007). However, the prevalence of EBRs and their relevance to
1068 evolutionary processes has been the focus of considerable debate (Alekseyev and Pevzner,
1069 2007; Peng et al., 2006; Trinh et al., 2004). Several lineage-specific gene loss events near
1070 EBRs in rodents are due to chromosomal rearrangements (Capilla et al., 2016; Fitzgerald and
1071 Bateman, 2004). Notably, one of these lost genes, *STK17A,* is located adjacent to the
1072 *COA1/MITRAC15* gene. The co-occurrence of an EBR with putative *COA1/MITRAC15* gene
1073 loss in rodents and marsupials is very intriguing. However, rodent genomes have mutational
1074 hotspots with high lineage-specific gBGC resulting in a substantial gene sequence divergence
1075 (Hargreaves et al., 2017). Such highly diverged orthologs can be challenging to identify due
1076 to difficulties in sequencing high GC regions. In *COA1/MITRAC15*, the magnitude of gBGC
1077 is relatively low, especially in rodents. Moreover, we find remnants of *COA1/MITRAC15* in
1078 several post-EBR species that suggest actual gene loss, at least in rodents. Several pre-EBR
1079 rodent species have also independently accumulated gene disrupting mutations in the
1080 *COA1/MITRAC15* gene. Hence, the *COA1/MITRAC15* gene appears to be under relaxed
1081 selective constraint even before the occurrence of the EBR.

1082 Species with exceptionally large body sizes or extremely long lifespans have a greater
1083 number of cell divisions. An increment in the number of cell divisions enhances cancer risk.
1084 However, paradoxically, large-bodied animals like elephants and whales do not have a higher
1085 incidence of cancer (Peto et al., 1975; Tollis et al., 2017). Cancer resistance due to lineage-
1086 specific changes in gene content may explain this paradox (Caulin et al., 2015; Caulin and
1087 Maley, 2011; DeGregori, 2011). While specific genetic changes in mammalian species lead
1088 to cancer resistance (Tollis et al., 2019; Vazquez et al., 2018), the reasons for lower cancer
1089 incidence in birds compared to mammals are mostly unexplored (Møller et al., 2017).
1090 Interestingly, the *COA1/MITRAC15* gene is an oncogene with a role in colorectal cancer (Xue
1091 et al., 2020), and its loss could reduce cancer risk. Silencing of *COA1/MITRAC15* by
1092 miRNAs strongly suppresses Giant cell tumors of the bone (Fellenberg et al., 2016; Herr et
1093 al., 2017). Our discovery of *COA1/MITRAC15* gene loss in Galliformes sets a precedent for
1094 the indisputable identification of gene loss events in birds and might reveal other oncogenes
1095 which are lost. We also identify *COA1/MITRAC15* gene loss in the beaver and naked mole-
1096 rat genomes, species that are models to study longevity (Zhou et al., 2020). High-quality
1097 near-complete vertebrate genomes with very few errors will further aid in the large-scale
1098 identification of gene loss events across the vertebrate phylogeny (Rhie et al., 2021).

1099 ## 5. **Conclusions**

1100 *COA1/MITRAC15* is a distant homolog of the *TIMM21* gene that has undergone recurrent
1101 gene loss in several Galliform and rodent species. Gene loss events occurred between 15
1102 MYA and 46 MYA in Galliform species and between 2 MYA and 30 MYA in rodents. The
1103 gene loss event occurs in species that rely primarily on glycolytic muscle fibers to achieve
1104 short bursts of activity. We show that *COA1/MITRAC15* and the adjacent *STK17A* gene are
1105 located at an Evolutionary Breakpoint Region (EBR) and are missing from the genomes of
1106 several rodent species following chromosomal rearrangement events. Pseudogenic and
1107 functional copies of *COA1/MITRAC15* are present in carnivores and primates, with the
1108 functional copy diverging in its intron-exon structure. Prevalence of repeated gene loss and
1109 duplication events in the history of *COA1/MITRAC15* not only demonstrates the
1110 dispensability of this gene but also hints at its ability to provide fitness increases in a context-
1111 dependent manner.

Acknowledgment

Competing interest statement

1118 None to declare

Availability of data

1120 All data associated with this study are available in the Supplementary Materials.

References

1122 Adams, K.L., Palmer, J.D., 2003. Evolution of mitochondrial gene content: Gene loss and
1123 transfer to the nucleus. Mol. Phylogenet. Evol. 29, 380–395.
1124 https://doi.org/10.1016/S1055-7903(03)00194-5

1125 Akiyama, T., Nishida, C., Momose, K., Onuma, M., Takami, K., Masuda, R., 2017. Gene
1126 duplication and concerted evolution of mitochondrial DNA in crane species. Mol.
1127 Phylogenet. Evol. 106, 158–163. https://doi.org/10.1016/j.ympev.2016.09.026

1128 Al-Ssulami, A.M., Azmi, A.M., Hussain, M., 2020. CodSeqGen: A tool for generating
1129 synonymous coding sequences with desired GC-contents. Genomics 112, 237–242.
1130 https://doi.org/10.1016/j.ygeno.2019.02.002

1131 Alekseyev, M.A., Pevzner, P.A., 2007. Are there rearrangement hotspots in the human
1132 genome? PLoS Comput. Biol. 3, 2111–2121.
1133 https://doi.org/10.1371/journal.pcbi.0030209

1134 Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J.,
1135 1997. Gapped BLAST and PSI-BLAST: A new generation of protein database search
1136 programs. Nucleic Acids Res. https://doi.org/10.1093/nar/25.17.3389

1137 Bailey, I., Myatt, J.P., Wilson, A.M., 2013. Group hunting within the Carnivora:
1138   Physiological, cognitive and environmental influences on strategy and cooperation.
1139   Behav. Ecol. Sociobiol. https://doi.org/10.1007/s00265-012-1423-3

1140 Barber, M.R.W., Aldridge, J.R., Webster, R.G., Magor, K.E., Magor, K.E., 2010. Association
1141   of RIG-I with innate immunity of ducks to influenza. Proc. Natl. Acad. Sci. U. S. A.
1142   107, 5913–8. https://doi.org/doi: 10.1073/pnas.1001755107

1143 Barnard, E.A., Lyles, J.M., Pizzey, J.A., 1982. Fibre types in chicken skeletal muscles and
1144   their changes in muscular dystrophy. J. Physiol. 331, 333–354.
1145   https://doi.org/10.1113/jphysiol.1982.sp014375

1146 Bevan, R.B., Lang, B.F., 2004. Mitochondrial genome evolution: the origin of mitochondria
1147   and of eukaryotes. Springer, Berlin, Heidelberg, pp. 1–35.
1148   https://doi.org/10.1007/b96830

1149 Björnerfeldt, S., Webster, M.T., Vilà, C., 2006. Relaxation of selective constraint on dog
1150   mitochondrial DNA following domestication. Genome Res. 16, 990–994.
1151   https://doi.org/10.1101/gr.5117706

1152 Botero-Castro, F., Figuet, E., Tilak, M.K., Nabholz, B., Galtier, N., 2017. Avian genomes
1153   revisited: Hidden genes uncovered and the rates versus traits paradox in birds. Mol. Biol.
1154   Evol. 34, 3123–3131. https://doi.org/10.1093/molbev/msx236

1155 Boteva, L., Nozawa, R.S., Naughton, C., Samejima, K., Earnshaw, W.C., Gilbert, N., 2020.
1156   Common Fragile Sites Are Characterized by Faulty Condensin Loading after
1157   Replication Stress. Cell Rep. 32. https://doi.org/10.1016/j.celrep.2020.108177

1158 Burger, G., Gray, M.W., Forget, L., Lang, B.F., 2013. Strikingly bacteria-like and gene-rich
1159   mitochondrial genomes throughout jakobid protists. Genome Biol. Evol. 5, 418–438.
1160   https://doi.org/10.1093/gbe/evt008

1161 Bush, S.J., Chen, L., Tovar-Corona, J.M., Urrutia, A.O., 2017. Alternative splicing and the
1162   evolution of phenotypic novelty. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 372, 1–7.

1163 Butler, P.J., 2016. The physiological basis of bird flight. Philos. Trans. R. Soc. B Biol. Sci.
1164   https://doi.org/10.1098/rstb.2015.0384

1165 Capilla, L., Sánchez-Guillén, R.A., Farré, M., Paytuví-Gallart, A., Malinverni, R., Ventura,
1166   J., Larkin, D.M., Ruiz-Herrera, A., 2016. Mammalian comparative genomics reveals
1167   genetic and epigenetic features associated with genome reshuffling in rodentia. Genome
1168   Biol. Evol. 8, 3703–3717. https://doi.org/10.1093/gbe/evw276

1169 Capra, J.A., Hubisz, M.J., Kostka, D., Pollard, K.S., Siepel, A., 2013. A Model-Based
1170   Analysis of GC-Biased Gene Conversion in the Human and Chimpanzee Genomes.
1171   PLoS Genet. https://doi.org/10.1371/journal.pgen.1003684

1172 Carbone, L., Harris, R.A., Vessere, G.M., Mootnick, A.R., Humphray, S., Rogers, J., Kim,
1173   S.K., Wall, J.D., Martin, D., Jurka, J., Milosavljevic, A., de Jong, P.J., 2009.
1174   Evolutionary Breakpoints in the Gibbon Suggest Association between Cytosine
1175   Methylation and Karyotype Evolution. PLoS Genet. 5, e1000538.
1176   https://doi.org/10.1371/journal.pgen.1000538

1177 Cardol, P., 2011. Mitochondrial NADH:Ubiquinone oxidoreductase (complex I) in

1178   eukaryotes: A highly conserved subunit composition highlighted by mining of protein
1179   databases. Biochim. Biophys. Acta - Bioenerg. 1807, 1390–1397.
1180   https://doi.org/10.1016/j.bbabio.2011.06.015

1181   Castro, L.F.C., Goncalves, O., Mazan, S., Tay, B.-H., Venkatesh, B., Wilson, J.M., 2013.
1182   Recurrent gene loss correlates with the evolution of stomach phenotypes in gnathostome
1183   history. Proc. R. Soc. B Biol. Sci. 281, 20132669–20132669.
1184   https://doi.org/10.1098/rspb.2013.2669

1185   Caulin, A.F., Graham, T.A., Wang, L.-S., Maley, C.C., 2015. Solutions to Peto's paradox
1186   revealed by mathematical modelling and cross-species cancer gene analysis. Philos.
1187   Trans. R. Soc. B Biol. Sci. 370, 20140222. https://doi.org/10.1098/rstb.2014.0222

1188   Caulin, A.F., Maley, C.C., 2011. Peto's Paradox: Evolution's prescription for cancer
1189   prevention. Trends Ecol. Evol. https://doi.org/10.1016/j.tree.2011.01.002

1190   Chen, Y.-C., Liu, T., Yu, C.-H., Chiang, T.-Y., Hwang, C.-C., 2013. Effects of GC Bias in
1191   Next-Generation-Sequencing Data on De Novo Genome Assembly. PLoS One 8,
1192   e62856. https://doi.org/10.1371/journal.pone.0062856

1193   Conley, K.E., 2016. Mitochondria to motion: Optimizing oxidative phosphorylation to
1194   improve exercise performance. J. Exp. Biol. https://doi.org/10.1242/jeb.126623

1195   Curry, J.W., Hohl, R., Noakes, T.D., Kohn, T.A., 2012. High oxidative capacity and type IIx
1196   fibre content in springbok and fallow deer skeletal Muscle suggest fast sprinters with a
1197   resistance to fatigue. J. Exp. Biol. 215, 3997–4005. https://doi.org/10.1242/jeb.073684

1198   Das, J., 2006. The role of mitochondrial respiration in physiological and evolutionary
1199   adaptation. BioEssays. https://doi.org/10.1002/bies.20463

1200   Dassi, E., 2017. Handshakes and fights: The regulatory interplay of RNA-binding proteins.
1201   Front. Mol. Biosci. https://doi.org/10.3389/fmolb.2017.00067

1202   DeGregori, J., 2011. Evolved tumor suppression: Why are we so good at not getting cancer?
1203   Cancer Res. https://doi.org/10.1158/0008-5472.CAN-11-0342

1204   Depamphilis, C.W., Young, N.D., Wolfe, A.D., 1997. Evolution of plastid gene rps2 in a
1205   lineage of hemiparasitic and holoparasitic plants: Many losses of photosynthesis and
1206   complex patterns of rate variation. Proc. Natl. Acad. Sci. U. S. A. 94, 7367–7372.
1207   https://doi.org/10.1073/pnas.94.14.7367

1208   Dial, K.P., 2003. Evolution of Avian Locomotion: Correlates of Flight Style, Locomotor
1209   Modules, Nesting Biology, Body Size, Development, and The Origin of Flapping Flight.
1210   Auk 120, 941–952. https://doi.org/10.1093/auk/120.4.941

1211   Doan, J.W., Schmidt, T.R., Wildman, D.E., Uddin, M., Goldberg, A., Hüttemann, M.,
1212   Goodman, M., Weiss, M.L., Grossman, L.I., 2004. Coadaptive evolution in cytochrome
1213   c oxidase: 9 of 13 subunits show accelerated rates of nonsynonymous substitution in
1214   anthropoid primates. Mol. Phylogenet. Evol. 33, 944–950.
1215   https://doi.org/10.1016/j.ympev.2004.07.016

1216   Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson,
1217   M., Gingeras, T.R., 2013. STAR: Ultrafast universal RNA-seq aligner. Bioinformatics.

1218 DuBay, S.G., Wu, Y., Scott, G.R., Qu, Y., Liu, Q., Smith, J.H., Xin, C., Hart Reeve, A.,
1219   Juncheng, C., Meyer, D., Wang, J., Johnson, J., Cheviron, Z.A., Lei, F., Bates, J., 2020.
1220   Life history predicts flight muscle phenotype and function in birds. J. Anim. Ecol. 89,
1221   1262–1276. https://doi.org/10.1111/1365-2656.13190

1222 Durkin, S.G., Glover, T.W., 2007. Chromosome fragile sites. Annu. Rev. Genet.
1223   https://doi.org/10.1146/annurev.genet.41.042007.165900

1224 Dutheil, J., 2008. Detecting site-specific biochemical constraints through substitution
1225   mapping. J. Mol. Evol. 67, 257–265. https://doi.org/10.1007/s00239-008-9139-8

1226 Emre, Y., Hurtaud, C., Ricquier, D., Bouillaud, F., Hughes, J., Criscuolo, F., 2007. Avian
1227   UCP: The killjoy in the evolution of the mitochondrial uncoupling proteins. J. Mol.
1228   Evol. 65, 392–402. https://doi.org/10.1007/s00239-007-9020-1

1229 Farré, M., Narayan, J., Slavov, G.T., Damas, J., Auvil, L., Li, C., Jarvis, E.D., Burt, D.W.,
1230   Griffin, D.K., Larkin, D.M., 2016. Novel insights into chromosome evolution in birds,
1231   archosaurs, and reptiles. Genome Biol. Evol. 8, 2442–2451.
1232   https://doi.org/10.1093/gbe/evw166

1233 Fellenberg, J., Sähr, H., Kunz, P., Zhao, Z., Liu, L., Tichy, D., Herr, I., 2016. Restoration of
1234   miR-127-3p and miR-376a-3p counteracts the neoplastic phenotype of giant cell tumor
1235   of bone derived stromal cells by targeting COA1, GLE1 and PDIA6. Cancer Lett. 371,
1236   134–141. https://doi.org/10.1016/j.canlet.2015.10.039

1237 Fitzgerald, J., Bateman, J.F., 2004. Why mice have lost genes for COL21A1, STK17A,
1238   GPR145 and AHRI: Evidence for gene deletion at evolutionary breakpoints in the rodent
1239   lineage. Trends Genet. https://doi.org/10.1016/j.tig.2004.07.002

1240 Foote, A.D., Morin, P.A., Durban, J.W., Pitman, R.L., Wade, P., Willerslev, E., Gilbert,
1241   M.T.P., da Fonseca, R.R., 2011. Positive selection on the killer whale mitogenome. Biol.
1242   Lett. 7, 116–118. https://doi.org/10.1098/rsbl.2010.0638

1243 Frickey, T., Lupas, A., 2004. CLANS: A Java application for visualizing protein families
1244   based on pairwise similarity. Bioinformatics 20, 3702–3704.
1245   https://doi.org/10.1093/bioinformatics/bth444

1246 Fu, X.D., Ares, M., 2014. Context-dependent control of alternative splicing by RNA-binding
1247   proteins. Nat. Rev. Genet. https://doi.org/10.1038/nrg3778

1248 Fukasawa, Y., Oda, T., Tomii, K., Imai, K., 2017. Origin and Evolutionary Alteration of the
1249   Mitochondrial Import System in Eukaryotic Lineages. Mol. Biol. Evol. 34, 1574–1586.
1250   https://doi.org/10.1093/molbev/msx096

1251 Gabaldón, T., Huynen, M.A., 2007. From endosymbiont to host-controlled organelle: The
1252   hijacking of mitochondrial protein synthesis and metabolism. PLoS Comput. Biol. 3,
1253   2209–2218. https://doi.org/10.1371/journal.pcbi.0030219

1254 Gabaldón, T., Rainey, D., Huynen, M.A., 2005. Tracing the evolution of a large protein
1255   complex in the eukaryotes, NADH:ubiquinone oxidoreductase (Complex I). J. Mol.
1256   Biol. 348, 857–870. https://doi.org/10.1016/j.jmb.2005.02.067

1257 Gabler, F., Nam, S., Till, S., Mirdita, M., Steinegger, M., Söding, J., Lupas, A.N., Alva, V.,
1258   2020. Protein Sequence Analysis Using the MPI Bioinformatics Toolkit. Curr. Protoc.

1259   Bioinforma. 72, 108. https://doi.org/10.1002/cpbi.108

1260 Gan, Z., Fu, T., Kelly, D.P., Vega, R.B., 2018. Skeletal muscle mitochondrial remodeling in
1261   exercise and diseases. Cell Res. https://doi.org/10.1038/s41422-018-0078-7

1262 Garvin, M.R., Bielawski, J.P., Sazanov, L.A., Gharrett, A.J., 2015. Review and meta-analysis
1263   of natural selection in mitochondrial complex I in metazoans. J. Zool. Syst. Evol. Res.
1264   53, 1–17. https://doi.org/10.1111/jzs.12079

1265 Gaudry, M.J., Jastroch, M., Treberg, J.R., Hofreiter, M., Paijmans, J.L.A., Starrett, J., Wales,
1266   N., Signore, A. V., Springer, M.S., Campbell, K.L., 2017. Inactivation of thermogenic
1267   UCP1 as a historical contingency in multiple placental mammal clades. Sci. Adv. 3,
1268   e1602878. https://doi.org/10.1126/sciadv.1602878

1269 Gautier, R., Douguet, D., Antonny, B., Drin, G., 2008. HELIQUEST: a web server to screen
1270   sequences with specific -helical properties. Bioinformatics 24, 2101–2102.
1271   https://doi.org/10.1093/bioinformatics/btn392

1272 Gissi, C., Iannelli, F., Pesole, G., 2008. Evolution of the mitochondrial genome of Metazoa as
1273   exemplified by comparison of congeneric species. Heredity (Edinb).
1274   https://doi.org/10.1038/hdy.2008.62

1275 Gray, M.W., Burger, G., Derelle, R., Klimeš, V., Leger, M.M., Sarrasin, M., Vlček, Č.,
1276   Roger, A.J., Eliáš, M., Lang, B.F., 2020. The draft nuclear genome sequence and
1277   predicted mitochondrial proteome of Andalucia godoyi, a protist with the most gene-rich
1278   and bacteria-like mitochondrial genome. BMC Biol. 18. https://doi.org/10.1186/s12915-
1279   020-0741-6

1280 Grossman, L.I., Wildman, D.E., Schmidt, T.R., Goodman, M., 2004. Accelerated evolution
1281   of the electron transport chain in anthropoid primates. Trends Genet.
1282   https://doi.org/10.1016/j.tig.2004.09.002

1283 Gu, Z., Steinmetz, L.M., Gu, X., Scharfe, C., Davis, R.W., Li, W.H., 2003. Role of duplicate
1284   genes in genetic robustness against null mutations. Nature 421, 63–66.
1285   https://doi.org/10.1038/nature01198

1286 Hargreaves, A.D., Zhou, L., Christensen, J., Marlétaz, F., Liu, S., Li, F., Jansen, P.G., Spiga,
1287   E., Hansen, M.T., Pedersen, S.V.H., Biswas, S., Serikawa, K., Fox, B.A., Taylor, W.R.,
1288   Mulley, J.F., Zhang, G., Heller, R.S., Holland, P.W.H., 2017. Genome sequence of a
1289   diabetes-prone rodent reveals a mutation hotspot around the ParaHox gene cluster. Proc.
1290   Natl. Acad. Sci. U. S. A. 114, 7677–7682. https://doi.org/10.1073/pnas.1702930114

1291 Harshman, J., Braun, E.L., Braun, M.J., Huddleston, C.J., Bowie, R.C.K., Chojnowski, J.L.,
1292   Hackett, S.J., Han, K.L., Kimball, R.T., Marks, B.D., Miglia, K.J., Moore, W.S., Reddy,
1293   S., Sheldon, F.H., Steadman, D.W., Steppan, S.J., Witt, C.C., Yuri, T., 2008.
1294   Phylogenomic evidence for multiple losses of flight in ratite birds. Proc. Natl. Acad. Sci.
1295   U. S. A. 105, 13462–13467. https://doi.org/10.1073/pnas.0803242105

1296 Hatefi, Y., 1985. The Mitochondrial Electron Transport and Oxidative Phosphorylation
1297   System. Annu. Rev. Biochem. 54, 1015–1069.
1298   https://doi.org/10.1146/annurev.bi.54.070185.005055

1299 Hecker, N., Sharma, V., Hiller, M., 2019. Convergent gene losses illuminate metabolic and

1300   physiological changes in herbivores and carnivores. Proc. Natl. Acad. Sci. U. S. A. 116,
1301   3036–3041. https://doi.org/10.1073/pnas.1818504116

1302 Hecker, N., Sharma, V., Hiller, M., 2017. Transition to an Aquatic Habitat Permitted the
1303   Repeated Loss of the Pleiotropic KLK8 Gene in Mammals. Genome Biol. Evol. 9,
1304   3179–3188.

1305 Herr, I., Sähr, H., Zhao, Z., Yin, L., Omlor, G., Lehner, B., Fellenberg, J., 2017. MiR-127 and
1306   miR-376a act as tumor suppressors by in vivo targeting of COA1 and PDIA6 in giant
1307   cell tumor of bone. Cancer Lett. 409, 49–55.
1308   https://doi.org/10.1016/j.canlet.2017.08.029

1309 Hess, D.C., Myers, C.L., Huttenhower, C., Hibbs, M.A., Hayes, A.P., Paw, J., Clore, J.J.,
1310   Mendoza, R.M., Luis, B.S., Nislow, C., Giaever, G., Costanzo, M., Troyanskaya, O.G.,
1311   Caudy, A.A., 2009. Computationally Driven, Quantitative Experiments Discover Genes
1312   Required for Mitochondrial Biogenesis. PLoS Genet. 5, e1000407.
1313   https://doi.org/10.1371/journal.pgen.1000407

1314 Hock, D.H., Robinson, D.R.L., Stroud, D.A., 2020. Blackout in the powerhouse: clinical
1315   phenotypes associated with defects in the assembly of OXPHOS complexes and the
1316   mitoribosome. Biochem. J. 477, 4085–4132. https://doi.org/10.1042/BCJ20190767

1317 Holmes, D.J., Austad, S.N., 1995. Birds as animal models for the comparative biology of
1318   aging: A prospectus. Journals Gerontol. - Ser. A Biol. Sci. Med. Sci.
1319   https://doi.org/10.1093/gerona/50A.2.B59

1320 Hubisz, M.J., Pollard, K.S., Siepel, A., 2011. Phast and Rphast: Phylogenetic analysis with
1321   space/time models. Brief. Bioinform. 12, 41–51. https://doi.org/10.1093/bib/bbq072

1322 Huelsmann, M., Hecker, N., Springer, M.S., Gatesy, J., Sharma, V., Hiller, M., 2019. Genes
1323   lost during the transition from land to water in cetaceans highlight genomic changes
1324   associated with aquatic adaptations. Sci. Adv. 5, 6671–6696.
1325   https://doi.org/10.1126/sciadv.aaw6671

1326 Hughes, D.A., Jastroch, M., Stoneking, M., Klingenspor, M., 2009. Molecular evolution of
1327   UCP1 and the evolutionary history of mammalian non-shivering thermogenesis. BMC
1328   Evol. Biol. 9, 4. https://doi.org/10.1186/1471-2148-9-4

1329 Huvet, M., Nicolay, S., Touchon, M., Audit, B., D'Aubenton-Carafa, Y., Arneodo, A.,
1330   Thermes, C., 2007. Human gene organization driven by the coordination of replication
1331   and transcription. Genome Res. 17, 1278–1285. https://doi.org/10.1101/gr.6533407

1332 Huynen, M.A., Duarte, I., Szklarczyk, R., 2013. Loss, replacement and gain of proteins at the
1333   origin of the mitochondria. Biochim. Biophys. Acta - Bioenerg.
1334   https://doi.org/10.1016/j.bbabio.2012.08.001

1335 Hyatt, J.P.K., Roy, R.R., Stuart, R., Talmadge, R.J., 2010. Myosin heavy chain composition
1336   of tiger (Panthera tigris) and cheetah (Acinonyx jubatus) hindlimb muscles. J. Exp. Zool.
1337   Part A Ecol. Genet. Physiol. 313, 45–57. https://doi.org/10.1002/jez.574

1338 Iriarte-Díaz, J., 2002. Differential scaling of locomotor performance in small and large
1339   terrestrial mammals, Journal of Experimental Biology. The Company of Biologists.
1340   https://doi.org/10.1242/JEB.205.18.2897

1341   Jastroch, M., 2017. Uncoupling protein 1 controls reactive oxygen species in brown adipose
1342   tissue. Proc. Natl. Acad. Sci. U. S. A. https://doi.org/10.1073/pnas.1709064114

1343   Jastroch, M., Wuertz, S., Kloas, W., Klingenspor, M., 2005. Uncoupling protein 1 in fish
1344   uncovers an ancient evolutionary history of mammalian nonshivering thermogenesis.
1345   Physiol. Genomics 22, 150–156. https://doi.org/10.1152/physiolgenomics.00070.2005

1346   Jebb, D., Hiller, M., 2018. Recurrent loss of HMGCS2 shows that ketogenesis is not essential
1347   for the evolution of large mammalian brains. Elife. https://doi.org/10.7554/eLife.38906

1348   Karnkowska, A., Vacek, V., Zubáčová, Z., Treitli, S.C., Petrželková, R., Eme, L., Novák, L.,
1349   Žárský, V., Barlow, L.D., Herman, E.K., Soukal, P., Hroudová, M., Doležal, P., Stairs,
1350   C.W., Roger, A.J., Eliáš, M., Dacks, J.B., Vlček, Č., Hampl, V., 2016. A eukaryote
1351   without a mitochondrial organelle. Curr. Biol. 26, 1274–1284.
1352   https://doi.org/10.1016/j.cub.2016.03.053

1353   Kelley, L.A., Mezulis, S., Yates, C.M., Wass, M.N., Sternberg, M.J.E., 2015. The Phyre2
1354   web portal for protein modeling, prediction and analysis. Nat. Protoc. 10, 845–858.
1355   https://doi.org/10.1038/nprot.2015.053

1356   Kishida, T., Suzuki, M., Takayama, A., 2018. Evolution of the alternative AQP2 gene:
1357   Acquisition of a novel protein-coding sequence in dolphins. Mol. Phylogenet. Evol. 118,
1358   54–57. https://doi.org/10.1016/j.ympev.2017.09.012

1359   Kohn, T.A., 2014. Insights into the skeletal muscle characteristics of three southern African
1360   antelope species. Biol. Open 3, 1037–1044. https://doi.org/10.1242/bio.20149241

1361   Kohn, T.A., Burroughs, R., Hartman, M.J., Noakes, T.D., 2011. Fiber type and metabolic
1362   characteristics of lion (Panthera leo), caracal (Caracal caracal) and human skeletal
1363   muscle. Comp. Biochem. Physiol. - A Mol. Integr. Physiol. 159, 125–133.
1364   https://doi.org/10.1016/j.cbpa.2011.02.006

1365   Kohn, T.A., Myburgh, K.H., 2007. Regional specialization of rat quadriceps myosin heavy
1366   chain isoforms occurring in distal to proximal parts of middle and deep regions is not
1367   mirrored by citrate synthase activity. J. Anat. 210, 8–18. https://doi.org/10.1111/j.1469-
1368   7580.2007.00661.x

1369   Krauss, S., Zhang, C.Y., Lowell, B.B., 2005. The mitochondrial uncoupling-protein
1370   homologues. Nat. Rev. Mol. Cell Biol. https://doi.org/10.1038/nrm1592

1371   Kurland, C.G., Andersson, S.G.E., 2000. Origin and Evolution of the Mitochondrial
1372   Proteome. Microbiol. Mol. Biol. Rev. 64, 786–820.
1373   https://doi.org/10.1128/mmbr.64.4.786-820.2000

1374   Lavrov, D. V., Pett, W., 2016. Animal mitochondrial DNA as we do not know it: Mt-Genome
1375   Organization and Evolution in Nonbilaterian Lineages. Genome Biol. Evol. 8, 2896–
1376   2913. https://doi.org/10.1093/gbe/evw195

1377   Lemaitre, C., Zaghloul, L., Sagot, M.F., Gautier, C., Arneodo, A., Tannier, E., Audit, B.,
1378   2009. Analysis of fine-scale mammalian evolutionary breakpoints provides new insight
1379   into their relation to genome organisation. BMC Genomics 10.
1380   https://doi.org/10.1186/1471-2164-10-335

1381   Mackiewicz, P., Urantówka, A.D., Kroczak, A., Mackiewicz, D., Lavrov, D., 2019.

Resolving Phylogenetic Relationships within Passeriformes Based on Mitochondrial Genes and Inferring the Evolution of Their Mitogenomes in Terms of Duplications. Genome Biol. Evol. 11, 2824–2849. https://doi.org/10.1093/gbe/evz209

Marçais, G., Kingsford, C., 2011. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. Bioinformatics. https://doi.org/10.1093/bioinformatics/btr011

Martin, W.F., Garg, S., Zimorski, V., 2015. Endosymbiotic theories for eukaryote origin. Philos. Trans. R. Soc. B Biol. Sci. https://doi.org/10.1098/rstb.2014.0330

Mcgaugh, S., Schwartz, T.S., 2017. Here and there, but not everywhere: Repeated loss of uncoupling protein 1 in amniotes. Biol. Lett. 13. https://doi.org/10.1098/rsbl.2016.0749

Meredith, R.W., Gatesy, J., Murphy, W.J., Ryder, O.A., Springer, M.S., 2009. Molecular decay of the tooth gene enamelin (ENAM) mirrors the loss of enamel in the fossil record of placental mammals. PLoS Genet. 5. https://doi.org/10.1371/journal.pgen.1000634

Meredith, R.W., Zhang, G., Gilbert, M.T.P., Jarvis, E.D., Springer, M.S., 2014. Evidence for a single loss of mineralized teeth in the common avian ancestor. Science 346, 1254390. https://doi.org/https://doi.org/10.1007/s00239-019-09903-6

Mick, D.U., Dennerlein, S., Wiese, H., Reinhold, R., Pacheu-Grau, D., Lorenzi, I., Sasarman, F., Weraarpachai, W., Shoubridge, E.A., Warscheid, B., Rehling, P., 2012. MITRAC links mitochondrial protein translocation to respiratory-chain assembly and translational regulation. Cell 151, 1528–1541. https://doi.org/10.1016/j.cell.2012.11.053

Mishra, P., Varuzhanyan, G., Pham, A.H., Chan, D.C., 2015. Mitochondrial Dynamics Is a Distinguishing Feature of Skeletal Muscle Fiber Types and Regulates Organellar Compartmentalization. Cell Metab. 22, 1033–1044. https://doi.org/10.1016/j.cmet.2015.09.027

Mitchell, P., 1961. Coupling of phosphorylation to electron and hydrogen transfer by a chemi-osmotic type of mechanism. Nature 191, 144–148. https://doi.org/10.1038/191144a0

Møller, A.P., Erritzøe, J., Soler, J.J., 2017. Life history, immunity, Peto's paradox and tumours in birds. J. Evol. Biol. 30, 960–967. https://doi.org/10.1111/jeb.13060

Montgomerie, S., Cruz, J.A., Shrivastava, S., Arndt, D., Berjanskii, M., Wishart, D.S., 2008. PROTEUS2: a web server for comprehensive protein structure prediction and structure-based annotation. Nucleic Acids Res. 36, W202. https://doi.org/10.1093/nar/gkn255

Moore, L.A., Tidyman, W.E., Arrizubieta, M.J., Bandman, E., 1993. The evolutionary relationship of avian and mammalian myosin heavy-chain genes. J. Mol. Evol. 36, 21–30. https://doi.org/10.1007/BF02407303

Murcha, M.W., Kubiszewski-Jakubiak, S., Wang, Y., Whelan, J., 2014. Evidence for interactions between the mitochondrial import apparatus and respiratory chain complexes via Tim21-like proteins in Arabidopsis. Front. Plant Sci. 5. https://doi.org/10.3389/fpls.2014.00082

Nabholz, B., Ellegren, H., Wolf, J.B.W., 2013. High levels of gene expression explain the strong evolutionary constraint of mitochondrial protein-coding genes. Mol. Biol. Evol. 30, 272–284. https://doi.org/10.1093/molbev/mss238

35

1423 Nedergaard, J., Golozoubova, V., Matthias, A., Asadi, A., Jacobsson, A., Cannon, B., 2001.
1424   UCP1: The only protein able to mediate adaptive non-shivering thermogenesis and
1425   metabolic inefficiency. Biochim. Biophys. Acta - Bioenerg.
1426   https://doi.org/10.1016/S0005-2728(00)00247-4

1427 Newman, S.A., Mezentseva, N. V., Badyaev, A. V., 2013. Gene loss, thermogenesis, and the
1428   origin of birds. Ann. N. Y. Acad. Sci. 1289, 36–47. https://doi.org/10.1111/nyas.12090

1429 Obayashi, T., Kagaya, Y., Aoki, Y., Tadaka, S., Kinoshita, K., 2019. COXPRESdb v7: A
1430   gene coexpression database for 11 animal species supported by 23 coexpression
1431   platforms for technical evaluation and evolutionary inference. Nucleic Acids Res. 47,
1432   D55–D62. https://doi.org/10.1093/nar/gky1155

1433 Ogata, T., Yamasaki, Y., 1997. Ultra-high-resolution scanning electron microscopy of
1434   mitochondria and sarcoplasmic reticulum arrangement in human red, white, and
1435   intermediate muscle fibers. Anat. Rec. 248, 214–223.
1436   https://doi.org/10.1002/(SICI)1097-0185(199706)248:2<214::AID-AR8>3.0.CO;2-S

1437 Osoegawa, K., Tateno, M., Woon, P.Y., Frengen, E., Mammoser, A.G., Catanese, J.J.,
1438   Hayashizaki, Y., De Jong, P.J., 2000. Bacterial artificial chromosome libraries for
1439   mouse sequencing and functional analysis. Genome Res. 10, 116–128.
1440   https://doi.org/10.1101/gr.10.1.116

1441 Palmer, J.D., Adams, K.L., Cho, Y., Parkinson, C.L., Qiu, Y.L., Song, K., 2000. Dynamic
1442   evolution of plant mitochondrial genomes: Mobile genes and introns and highly variable
1443   mutation rates. Proc. Natl. Acad. Sci. U. S. A. https://doi.org/10.1073/pnas.97.13.6960

1444 Pan, S., Lin, Y., Liu, Q., Duan, J., Lin, Z., Wang, Y., Wang, X., Lam, S.M., Zou, Z., Shui, G.,
1445   Zhang, Y., Zhang, Z., Zhan, X., 2019. Convergent genomic signatures of flight loss in
1446   birds suggest a switch of main fuel. Nat. Commun. 10. https://doi.org/10.1038/s41467-
1447   019-10682-3

1448 Parker, D.J., Gardiner, A., Neville, M.C., Ritchie, M.G., Goodwin, S.F., 2014. The evolution
1449   of novelty in conserved genes; Evidence of positive selection in the Drosophila fruitless
1450   gene is localised to alternatively spliced exons. Heredity (Edinb). 112, 300–306.
1451   https://doi.org/10.1038/hdy.2013.106

1452 Paz, I., Kosti, I., Ares, M., Cline, M., Mandel-Gutfreund, Y., 2014. RBPmap: A web server
1453   for mapping binding sites of RNA-binding proteins. Nucleic Acids Res. 42, W361.
1454   https://doi.org/10.1093/nar/gku406

1455 Peng, Q., Pevzner, P.A., Tesler, G., 2006. The fragile breakage versus random breakage
1456   models of chromosome evolution. PLoS Comput. Biol. 2, 100–111.
1457   https://doi.org/10.1371/journal.pcbi.0020014

1458 Peto, H., Roe, F.J.C., Lee, P.N., Levy, L., Clack, J., 1975. Cancer and ageing in mice and
1459   men. Br. J. Cancer 32, 411–426. https://doi.org/10.1038/bjc.1975.242

1460 Pettersen, E.F., Goddard, T.D., Huang, C.C., Meng, E.C., Couch, G.S., Croll, T.I., Morris,
1461   J.H., Ferrin, T.E., 2021. <scp>UCSF ChimeraX</scp>□: Structure visualization for
1462   researchers, educators, and developers. Protein Sci. 30, 70–82.
1463   https://doi.org/10.1002/pro.3943

1464   Picard, M., Hepple, R.T., Burelle, Y., 2012. Mitochondrial functional specialization in
1465   glycolytic and oxidative muscle fibers: Tailoring the organelle for optimal function. Am.
1466   J. Physiol. - Cell Physiol. https://doi.org/10.1152/ajpcell.00368.2011

1467   Pickrell, J.K., Pai, A.A., Gilad, Y., Pritchard, J.K., 2010. Noisy splicing drives mRNA
1468   isoform diversity in human cells. PLoS Genet. 6, 1–11.

1469   Pierrel, F., Bestwick, M.L., Cobine, P.A., Khalimonchuk, O., Cricco, J.A., Winge, D.R.,
1470   2007. Coa1 links the Mss51 post-translational function to Cox1 cofactor insertion in
1471   cytochrome c oxidase assembly. EMBO J. 26, 4335–4346.
1472   https://doi.org/10.1038/sj.emboj.7601861

1473   Polymeropoulos, E.T., Jastroch, M., Frappell, P.B., 2012. Absence of adaptive nonshivering
1474   thermogenesis in a marsupial, the fat-tailed dunnart (Sminthopsis crassicaudata). J.
1475   Comp. Physiol. B Biochem. Syst. Environ. Physiol. 182, 393–401.
1476   https://doi.org/10.1007/s00360-011-0623-x

1477   Popadin, K.Y., Nikolaev, S.I., Junier, T., Baranova, M., Antonarakis, S.E., 2013. Purifying
1478   selection in mammalian mitochondrial protein-coding genes is highly effective and
1479   congruent with evolution of nuclear genes. Mol. Biol. Evol. 30, 347–355.
1480   https://doi.org/10.1093/molbev/mss219

1481   Qiu, Q., Zhang, G., Ma, T., Qian, W., Ye, Z., Cao, C., Hu, Q., Kim, J., Larkin, D.M., Auvil,
1482   L., Capitanu, B., Ma, J., Lewin, H.A., Qian, X., Lang, Y., Zhou, R., Wang, L., Wang,
1483   K., Xia, J., Liao, S., Pan, S., Lu, X., Hou, H., Wang, Y., Zang, X., Yin, Y., Ma, H.,
1484   Zhang, J., Wang, Z., Zhang, Yingmei, Zhang, D., Yonezawa, T., Hasegawa, M., Zhong,
1485   Y., Liu, W., Zhang, Yan, Huang, Z., Zhang, S., Long, R., Yang, H., Lenstra, J.A.,
1486   Cooper, D.N., Wu, Y., Wang, Jun, Shi, P., Wang, Jian, Liu, J., Wang, Junyis, 2012. The
1487   yak genome and adaptation to life at high altitude. Nat. Genet. 44, 946–949.
1488   https://doi.org/10.1038/ng.2343

1489   Ramensky, V.E., Nurtdinov, R.N., Neverov, A.D., Mironov, A.A., Gelfand, M.S., 2008.
1490   Positive Selection in Alternatively Spliced Exons of Human Genes. Am. J. Hum. Genet.
1491   83, 94–98. https://doi.org/10.1016/j.ajhg.2008.05.017

1492   Remmert, M., Biegert, A., Hauser, A., Söding, J., 2012. HHblits: Lightning-fast iterative
1493   protein sequence searching by HMM-HMM alignment. Nat. Methods 9, 173–175.
1494   https://doi.org/10.1038/nmeth.1818

1495   Resnicow, D.I., Deacon, J.C., Warrick, H.M., Spudich, J.A., Leinwand, L.A., 2010.
1496   Functional diversity among a family of human skeletal muscle myosin motors. Proc.
1497   Natl. Acad. Sci. U. S. A. 107, 1053–1058. https://doi.org/10.1073/pnas.0913527107

1498   Rhie, A., McCarthy, S.A., Fedrigo, O., Damas, J., Formenti, G., Koren, S., Uliano-Silva, M.,
1499   Chow, W., Fungtammasan, A., Kim, J., Lee, C., Ko, B.J., Chaisson, M., Gedman, G.L.,
1500   Cantin, L.J., Thibaud-Nissen, F., Haggerty, L., Bista, I., Smith, M., Haase, B.,
1501   Mountcastle, J., Winkler, S., Paez, S., Howard, J., Vernes, S.C., Lama, T.M., Grutzner,
1502   F., Warren, W.C., Balakrishnan, C.N., Burt, D., George, J.M., Biegler, M.T., Iorns, D.,
1503   Digby, A., Eason, D., Robertson, B., Edwards, T., Wilkinson, M., Turner, G., Meyer, A.,
1504   Kautt, A.F., Franchini, P., Detrich, H.W., Svardal, H., Wagner, M., Naylor, G.J.P.,
1505   Pippel, M., Malinsky, M., Mooney, M., Simbirsky, M., Hannigan, B.T., Pesout, T.,
1506   Houck, M., Misuraca, A., Kingan, S.B., Hall, R., Kronenberg, Z., Sović, I., Dunn, C.,
1507   Ning, Z., Hastie, A., Lee, J., Selvaraj, S., Green, R.E., Putnam, N.H., Gut, I., Ghurye, J.,

1508 Garrison, E., Sims, Y., Collins, J., Pelan, S., Torrance, J., Tracey, A., Wood, J., Dagnew,
1509   R.E., Guan, D., London, S.E., Clayton, D.F., Mello, C. V., Friedrich, S.R., Lovell, P. V.,
1510   Osipova, E., Al-Ajli, F.O., Secomandi, S., Kim, H., Theofanopoulou, C., Hiller, M.,
1511   Zhou, Y., Harris, R.S., Makova, K.D., Medvedev, P., Hoffman, J., Masterson, P., Clark,
1512   K., Martin, F., Howe, Kevin, Flicek, P., Walenz, B.P., Kwak, W., Clawson, H.,
1513   Diekhans, M., Nassar, L., Paten, B., Kraus, R.H.S., Crawford, A.J., Gilbert, M.T.P.,
1514   Zhang, G., Venkatesh, B., Murphy, R.W., Koepfli, K.-P., Shapiro, B., Johnson, W.E., Di
1515   Palma, F., Marques-Bonet, T., Teeling, E.C., Warnow, T., Graves, J.M., Ryder, O.A.,
1516   Haussler, D., O'Brien, S.J., Korlach, J., Lewin, H.A., Howe, Kerstin, Myers, E.W.,
1517   Durbin, R., Phillippy, A.M., Jarvis, E.D., 2021. Towards complete and error-free
1518   genome assemblies of all vertebrate species. Nature 592, 737–746.
1519   https://doi.org/10.1038/s41586-021-03451-0

1520 Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G.,
1521   Mesirov, J.P., 2011. Integrative genomics viewer. Nat. Biotechnol. 29, 24–6.

1522 Ruiz-Herrera, A., Castresana, J., Robinson, T.J., 2006. Is mammalian chromosomal evolution
1523   driven by regions of genome fragility? Genome Biol. 7, R115.
1524   https://doi.org/10.1186/gb-2006-7-12-r115

1525 Ruiz-Herrera, A., García, F., Giulotto, E., Attolini, C., Egozcue, J., Ponsà, M., Garcia, M.,
1526   2005. Evolutionary breakpoints are co-localized with fragile sites and intrachromosomal
1527   telomeric sequences in primates. Cytogenet. Genome Res. 108, 234–247.
1528   https://doi.org/10.1159/000080822

1529 Sackton, T.B., Grayson, P., Cloutier, A., Hu, Z., Liu, J.S., Wheeler, N.E., Gardner, P.P.,
1530   Clarke, J.A., Baker, A.J., Clamp, M., Edwards, S. V., 2019. Convergent regulatory
1531   evolution and loss of flight in paleognathous birds. Science (80-.). 364, 74–78.
1532   https://doi.org/10.1126/science.aat7244

1533 Samuels, J.X., Van Valkenburgh, B., 2008. Skeletal indicators of locomotor adaptations in
1534   living and extinct rodents. J. Morphol. 269, 1387–1411.
1535   https://doi.org/10.1002/jmor.10662

1536 San Mauro, D., Gower, D.J., Zardoya, R., Wilkinson, M., 2006. A hotspot of gene order
1537   rearrangement by tandem duplication and random loss in the vertebrate mitochondrial
1538   genome. Mol. Biol. Evol. 23, 227–234. https://doi.org/10.1093/molbev/msj025

1539 Sayol, F., Steinbauer, M.J., Blackburn, T.M., Antonelli, A., Faurby, S., 2020. Anthropogenic
1540   extinctions conceal widespread evolution of flightlessness in birds. Sci. Adv. 6,
1541   eabb6095. https://doi.org/10.1126/sciadv.abb6095

1542 Schiaffino, S., Reggiani, C., 2011. Fiber types in Mammalian skeletal muscles. Physiol. Rev.
1543   91, 1447–1531. https://doi.org/10.1152/physrev.00031.2010

1544 Schibler, L., Roig, A., Mahe, M.F., Laurent, P., Hayes, H., Rodolphe, F., Cribiu, E.P., 2006.
1545   High-resolution comparative mapping among man, cattle and mouse suggests a role for
1546   repeat sequences in mammalian genome evolution. BMC Genomics 7, 194.
1547   https://doi.org/10.1186/1471-2164-7-194

1548 Schneider, K., Adams, C.E., Elmer, K.R., 2019. Parallel selection on ecologically relevant
1549   gene functions in the transcriptomes of highly diversifying salmonids. BMC Genomics
1550   20, 1–23. https://doi.org/10.1186/s12864-019-6361-2

1551   Sharma, S., Shinde, S.S., Teekas, L., Vijay, N., 2020. Evidence for the loss of plasminogen
1552   receptor KT gene in chicken. Immunogenetics 72, 507–515.
1553   https://doi.org/10.1007/s00251-020-01186-2

1554   Sharma, V., Hecker, N., Walther, F., Stuckas, H., Hiller, M., Yeager, M., 2020. Convergent
1555   Losses of TLR5 Suggest Altered Extracellular Flagellin Detection in Four Mammalian
1556   Lineages. Mol. Biol. Evol. 37, 1847–1854. https://doi.org/10.1093/molbev/msaa058

1557   Sharma, V., Hiller, M., 2018. Loss of Enzymes in the Bile Acid Synthesis Pathway Explains
1558   Differences in Bile Composition among Mammals. Genome Biol. Evol. 10, 3211–3217.
1559   https://doi.org/https://doi.org/10.1093/gbe/evy243

1560   Shen, W., Le, S., Li, Y., Hu, F., 2016. SeqKit: A cross-platform and ultrafast toolkit for
1561   FASTA/Q file manipulation. PLoS One 11, e0163962.
1562   https://doi.org/10.1371/journal.pone.0163962

1563   Shen, Y.Y., Liang, L., Zhu, Z.H., Zhou, W.P., Irwin, D.M., Zhang, Y.P., 2010. Adaptive
1564   evolution of energy metabolism genes and the origin of flight in bats. Proc. Natl. Acad.
1565   Sci. U. S. A. 107, 8666–8671. https://doi.org/10.1073/pnas.0912613107

1566   Shen, Y.Y., Shi, P., Sun, Y.B., Zhang, Y.P., 2009. Relaxation of selective constraints on
1567   avian mitochondrial DNA following the degeneration of flight ability. Genome Res. 19,
1568   1760–1765. https://doi.org/10.1101/gr.093138.109

1569   Shinde, S.S., Teekas, L., Sharma, S., Vijay, N., 2019. Signatures of Relaxed Selection in the
1570   CYP8B1 Gene of Birds and Mammals. J. Mol. Evol. 87, 209–220.
1571   https://doi.org/https://doi.org/10.1007/s00239-019-09903-6

1572   Signes, A., Fernandez-Vizarra, E., 2018. Assembly of mammalian oxidative phosphorylation
1573   complexes I–V and supercomplexes. Essays Biochem.
1574   https://doi.org/10.1042/EBC20170098

1575   Sloan, D.B., Warren, J.M., Williams, A.M., Wu, Z., Abdel-Ghany, S.E., Chicco, A.J., Havird,
1576   J.C., 2018. Cytonuclear integration and co-evolution. Nat. Rev. Genet.
1577   https://doi.org/10.1038/s41576-018-0035-9

1578   Steinegger, M., Meier, M., Mirdita, M., Vöhringer, H., Haunsberger, S.J., Söding, J., 2019.
1579   HH-suite3 for fast remote homology detection and deep protein annotation. BMC
1580   Bioinformatics 20. https://doi.org/10.1186/s12859-019-3019-7

1581   Szklarczyk, R., Wanschers, B.F.J., Cuypers, T.D., Esseling, J.J., Riemersma, M., Van Den
1582   Brand, M.A.M., Gloerich, J., Lasonder, E., van den Heuvel, L.P., Nijtmans, L.G.,
1583   Huynen, M.A., 2012. Iterative orthology prediction uncovers new mitochondrial
1584   proteins and identifies C12orf62 as the human ortholog of COX14, a protein involved in
1585   the assembly of cytochrome c oxidase. Genome Biol. 13. https://doi.org/10.1186/gb-
1586   2012-13-2-r12

1587   Talbot, J., Maves, L., 2016. Skeletal muscle fiber type: using insights from muscle
1588   developmental biology to dissect targets for susceptibility and resistance to muscle
1589   disease. Wiley Interdiscip. Rev. Dev. Biol. https://doi.org/10.1002/wdev.230

1590   Taylor, J.S., Raes, J., 2004. Duplication and divergence: The evolution of new genes and old
1591   ideas. Annu. Rev. Genet. https://doi.org/10.1146/annurev.genet.38.072902.092831

1592   Taylor, M.E., 1989. Locomotor Adaptations by Carnivores, in: Carnivore Behavior, Ecology,
1593   and Evolution. Springer US, pp. 382–409. https://doi.org/10.1007/978-1-4757-4716-
1594   4_15

1595   Thorvaldsdottir, H., Robinson, J.T., Mesirov, J.P., 2013. Integrative Genomics Viewer (IGV):
1596   high-performance genomics data visualization and exploration. Brief. Bioinform. 14,
1597   178–192. https://doi.org/10.1093/bib/bbs017

1598   Tian, R., Xu, S., Chai, S., Yin, D., Zakon, H., Yang, G., 2018. Stronger selective constraint
1599   on downstream genes in the oxidative phosphorylation pathway of cetaceans. J. Evol.
1600   Biol. 31, 217–228. https://doi.org/10.1111/jeb.13213

1601   Tollis, M., Boddy, A.M., Maley, C.C., 2017. Peto's Paradox: How has evolution solved the
1602   problem of cancer prevention? BMC Biol. https://doi.org/10.1186/s12915-017-0401-7

1603   Tollis, M., Robbins, J., Webb, A.E., Kuderna, L.F.K., Caulin, A.F., Garcia, J.D., Bèrubè, M.,
1604   Pourmand, N., Marques-Bonet, T., O'Connell, M.J., Palsbøll, P.J., Maley, C.C., 2019.
1605   Return to the Sea, Get Huge, Beat Cancer: An Analysis of Cetacean Genomes Including
1606   an Assembly for the Humpback Whale (Megaptera novaeangliae). Mol. Biol. Evol. 36,
1607   1746–1763. https://doi.org/10.1093/molbev/msz099

1608   Trinh, P., McLysaght, A., Sankoff, D., 2004. Genomic features in the breakpoint regions
1609   between syntenic blocks. Bioinformatics 20, i318–i325.
1610   https://doi.org/10.1093/bioinformatics/bth934

1611   Urantówka, A.D., Kroczak, A., Mackiewicz, P., 2020. New view on the organization and
1612   evolution of Palaeognathae mitogenomes poses the question on the ancestral gene
1613   rearrangement in Aves. BMC Genomics 21, 1–25. https://doi.org/10.1186/s12864-020-
1614   07284-5

1615   Valente, R., Alves, L.Q., Nabais, M., Alves, F., Sousa-Pinto, I., Ruivo, R., Castro, L.F.C.,
1616   2020. Convergent Cortistatin losses parallel modifications in circadian rhythmicity and
1617   energy homeostasis in Cetacea and other mammalian lineages. Genomics.
1618   https://doi.org/10.1016/j.ygeno.2020.11.002

1619   Van Der Lee, R., Wiel, L., Van Dam, T.J.P., Huynen, M.A., 2017. Genome-scale detection of
1620   positive selection in nine primates predicts human-virus evolutionary conflicts. Nucleic
1621   Acids Res. 45, 10634–10648. https://doi.org/10.1093/nar/gkx704

1622   van Esveld, S.L., Huynen, M.A., 2018. Does mitochondrial DNA evolution in metazoa drive
1623   the origin of new mitochondrial proteins? IUBMB Life 70, 1240–1250.
1624   https://doi.org/10.1002/iub.1940

1625   Vazquez, J.M., Sulak, M., Chigurupati, S., Lynch, V.J., 2018. A Zombie LIF Gene in
1626   Elephants Is Upregulated by TP53 to Induce Apoptosis in Response to DNA Damage.
1627   Cell Rep. https://doi.org/10.1016/j.celrep.2018.07.042

1628   Vijay, N., 2020. Loss of inner kinetochore genes is associated with the transition to an
1629   unconventional point centromere in budding yeast. PeerJ 8.
1630   https://doi.org/10.7717/peerj.10085

1631   Wang, C., Richter□Dennerlein, R., Pacheu□Grau, D., Liu, F., Zhu, Y., Dennerlein, S.,
1632   Rehling, P., 2020. MITRAC15/COA1 promotes mitochondrial translation in a ND2

40

1633   ribosome–nascent chain complex. EMBO Rep. 21.
1634   https://doi.org/10.15252/embr.201948833

1635 Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F.T.,
1636   De Beer, T.A.P., Rempfer, C., Bordoli, L., Lepore, R., Schwede, T., 2018. SWISS-
1637   MODEL: Homology modelling of protein structures and complexes. Nucleic Acids Res.
1638   46, W296–W303. https://doi.org/10.1093/nar/gky427

1639 Webb, B., Sali, A., 2016. Comparative protein structure modeling using MODELLER. Curr.
1640   Protoc. Bioinforma. 2016, 5.6.1-5.6.37. https://doi.org/10.1002/cpbi.3

1641 Weeks, O.I., 1989. Vertebrate Skeletal Muscle: Power Source for Locomotion. Bioscience
1642   39, 791–799. https://doi.org/10.2307/1311185

1643 Weiss, A., Leinwand, L.A., 1996. The mammalian myosin heavy chain gene family. Annu.
1644   Rev. Cell Dev. Biol. https://doi.org/10.1146/annurev.cellbio.12.1.417

1645 Welch, A.J., Bedoya-Reina, O.C., Carretero-Paulet, L., Miller, W., Rode, K.D., Lindqvist, C.,
1646   2014. Polar bears exhibit genome-wide signatures of bioenergetic adaptation to life in
1647   the arctic environment. Genome Biol. Evol. 6, 433–450.
1648   https://doi.org/10.1093/gbe/evu025

1649 Welch, K.C., Altshuler, D.L., 2009. Fiber type homogeneity of the flight musculature in
1650   small birds. Comp. Biochem. Physiol. - B Biochem. Mol. Biol. 152, 324–331.
1651   https://doi.org/10.1016/j.cbpb.2008.12.013

1652 Wertheim, J.O., Murrell, B., Smith, M.D., Kosakovsky Pond, S.L., Scheffler, K., 2015.
1653   RELAX: Detecting Relaxed Selection in a Phylogenetic Framework. Mol. Biol. Evol.
1654   32, 820–832.

1655 Williams, T.M., Dobson, G.P., Mathieu-Costello, O., Morsbach, D., Worley, M.B., Phillips,
1656   J.A., 1997. Skeletal muscle histology and biochemistry of an elite sprinter, the African
1657   cheetah. J. Comp. Physiol. - B Biochem. Syst. Environ. Physiol. 167, 527–535.
1658   https://doi.org/10.1007/s003600050105

1659 Wilson, A.M., Lowe, J.C., Roskilly, K., Hudson, P.E., Golabek, K.A., McNutt, J.W., 2013.
1660   Locomotion dynamics of hunting in wild cheetahs. Nature 498, 185–189.
1661   https://doi.org/10.1038/nature12295

1662 Wilson, J.W., Mills, M.G.L., Wilson, R.P., Peters, G., Mills, M.E.J., Speakman, J.R., Durant,
1663   S.M., Bennett, N.C., Marks, N.J., Scantlebury, M., 2013. Cheetahs, Acinonyx jubatus,
1664   balance turn capacity with pace when chasing prey. Biol. Lett. 9.
1665   https://doi.org/10.1098/rsbl.2013.0620

1666 Wu, W., Schmidt, T.R., Goodman, M., Grossman, L.I., 2000. Molecular evolution of
1667   cytochrome c oxidase subunit I in primates: Is there coevolution between mitochondrial
1668   and nuclear genomes? Mol. Phylogenet. Evol. 17, 294–304.
1669   https://doi.org/10.1006/mpev.2000.0833

1670 Xing, Y., Lee, C.J., 2004. Negative selection pressure against premature protein truncation is
1671   reduced by alternative splicing and diploidy. Trends Genet.
1672   https://doi.org/10.1016/j.tig.2004.07.009

1673 Xiong, Y., Lei, F., 2020. SLC2A12 of SLC2 gene family in bird provides functional

1674   compensation for the loss of SLC2A4 gene in other vertebrates. Mol. Biol. Evol.
1675   https://doi.org/10.1093/molbev/msaa286

1676   Xue, Y., Li, P.-D., Tang, X.-M., Yan, Z.-H., Xia, S.-S., Tian, H.-P., Liu, Z.-L., Zhou, T.,
1677   Tang, X.-G., Zhang, G.-J., 2020. <p>Cytochrome C Oxidase Assembly Factor 1
1678   Homolog Predicts Poor Prognosis and Promotes Cell Proliferation in Colorectal Cancer
1679   by Regulating PI3K/AKT Signaling</p>. Onco. Targets. Ther. Volume 13, 11505–
1680   11516. https://doi.org/10.2147/OTT.S279024

1681   Yang, Z., 2007. PAML 4: Phylogenetic Analysis by Maximum Likelihood. Mol. Biol. Evol.
1682   24, 1586–1591. https://doi.org/10.1093/molbev/msm088

1683   Zhang, F., Broughton, R.E., 2015. Heterogeneous natural selection on oxidative
1684   phosphorylation genes among fishes with extreme high and low aerobic performance.
1685   BMC Evol. Biol. 15, 173. https://doi.org/10.1186/s12862-015-0453-7

1686   Zhong, W.W.H., Lucas, C.A., Hoh, J.F.Y., 2008. Myosin isoforms and fibre types in limb
1687   muscles of Australian marsupials: Adaptations to hopping and non-hopping locomotion.
1688   J. Comp. Physiol. B Biochem. Syst. Environ. Physiol. 178, 47–55.
1689   https://doi.org/10.1007/s00360-007-0198-8

1690   Zhong, W.W.H., Lucas, C.A., Kang, L.H.D., Hoh, J.F.Y., 2001. Electrophoretic and
1691   immunochemical evidence showing that marsupial limb muscles express the same fast
1692   and slow myosin heavy chains as eutherians. Electrophoresis 22, 1016–1020.
1693   https://doi.org/10.1002/1522-2683()22:6<1016::AID-ELPS1016>3.0.CO;2-K

1694   Zhou, X., Dou, Q., Fan, G., Zhang, Q., Sanderford, M., Kaya, A., Johnson, J., Karlsson, E.K.,
1695   Tian, X., Mikhalchenko, A., Kumar, S., Seluanov, A., Zhang, Z.D., Gorbunova, V., Liu,
1696   X., Gladyshev, V.N., 2020. Beaver and Naked Mole Rat Genomes Reveal Common
1697   Paths to Longevity. Cell Rep. 32. https://doi.org/10.1016/j.celrep.2020.107949

1698   Zimmermann, L., Stephens, A., Nam, S.Z., Rau, D., Kübler, J., Lozajic, M., Gabler, F.,
1699   Söding, J., Lupas, A.N., Alva, V., 2018. A Completely Reimplemented MPI
1700   Bioinformatics Toolkit with a New HHpred Server at its Core. J. Mol. Biol. 430, 2237–
1701   2243. https://doi.org/10.1016/j.jmb.2017.12.007

1702

1703   **Figure legends**

1704   **Figure 1**: *COA1/MITRAC15* and *TIMM21* are distant homologs with similar amino acid
1705   sequence profiles and secondary structures. (**A**) Cluster map of *COA1/MITRAC15* homologs
1706   identified using profile-profile search implemented in HHblits. The cluster of
1707   *COA1/MITRAC15*: blue, *TIMM21*: red, homologs of *COA1/MITRAC15* from species of
1708   fungi: orange, homologs of *COA1/MITRAC15* from bacterial species: light blue cluster and
1709   diffuse brown cluster, *COA1/MITRAC15* homologs in plants, represented by *Arabidopsis*
1710   *thaliana* homolog At2g20390: yellow, *TIMM21*-like proteins that exist as duplicated copies
1711   in plants, represented by *Arabidopsis thaliana* homolog At2g37940: magenta. (**B**) The output
1712   of HHpred showing the alignment of human *COA1/MITRAC15* with yeast *TIMM21*. The
1713   region in the box highlights the predicted transmembrane helix. (**C**) The predicted secondary
1714   structure of human (*Homo sapiens) COA1/MITRAC15*. (**D**) The predicted secondary structure

1715   of yeast (*Saccharomyces cerevisiae*) *TIMM21*.

1716   **Figure 2**: Comparison of haplotypes of *COA1/MITRAC15* gene inferred based on sequencing
1717   reads in different species visualized in IGV browser. (**A**) Two haplotypes of
1718   *COA1/MITRAC15* in humans (*Homo sapiens*) corresponding to the functional and
1719   pseudogene copies. (**B**) Two haplotypes of exon 1 to exon 4 of *COA1/MITRAC15* in tiger
1720   (*Panthera tigris*). (**C**) Two haplotypes in exon two and one haplotype of remaining exons of
1721   *COA1/MITRAC15* in naked mole-rat (*Heterocephalus glaber*). (**D**) The single haplotype of
1722   *COA1/MITRAC15* gene in chicken (*Gallus gallus*). (**E**) The single haplotype of
1723   *COA1/MITRAC15* in the platypus (*Ornithorhynchus anatinus*). (**F**) The single haplotype of
1724   *COA1/MITRAC15* in red squirrel (*Sciurus vulgaris*).

1725   **Figure 3**: Loss of *COA1/MITRAC15* gene in Feliform. (A) Gene loss event in *Acinonyx*
1726   *jubatus* besides a time-calibrated phylogenetic tree downloaded from the time tree website.
1727   (B) Gene order in the genomic region flanking the *COA1/MITRAC15* gene and its exons in
1728   genomes. Red and blue arrows depict the direction of gene transcription relative to the
1729   *COA1/MITRAC15* gene for consistency across species. Gray boxes represent the genes
1730   located on short scaffolds with unknown orientation.

1731   **Figure 4**: Recurrent loss of *COA1/MITRAC15* gene in Galliform species. (**A**) Gene loss
1732   events in ten Galliform species besides a time-calibrated phylogenetic tree downloaded from
1733   the time tree website. Blue branches in the tree represent functional branches, and the
1734   magenta-colored branches represent mixed (functional + pseudogenic) branches. The method
1735   proposed by Meredith et al., 2009 was used to estimate the time of gene loss using two
1736   different substitution rates (1ds and 2ds). Short colored bars depict the locations of the gene
1737   disrupting mutations on the four exons of *COA1/MITRAC15*. (**B**) Gene order in the genomic
1738   region flanking the *COA1/MITRAC15* gene in bird genomes. Red and blue arrows depict the
1739   direction of gene transcription relative to the *COA1/MITRAC15* gene for consistency across
1740   species. Gray boxes represent the genes located on short scaffolds with unknown orientation.
1741   (**C**) The gene expression pattern of the *COA1/MITRAC15* gene in six tissues (brain, spleen,
1742   skin, liver, gonad, and blood) was assessed by screening RNA-seq datasets. The red-colored
1743   blocks depict the robust expression of the *COA1/MITRAC15* gene, the black-colored blocks
1744   depict a lack of *COA1/MITRAC15* gene expression in that particular tissue, and the white-
1745   colored blocks represent a lack of data for that tissue.

1746   **Figure 5**: Comparison of different sequencing datasets of woolly mammoth (*Mammuthus*
1747   *primigenius*) for *COA1/MITRAC15* gene exons. Gray rectangles show the reads mapped to
1748   each exon. Panels **A** to **F** shows the reads with sequence supporting each exon from different
1749   woolly mammoth SRA bio projects [PRJDB4697 (182 Gb), PRJEB42269 (179 Gb),
1750   PRJNA397140 (155 Gb), PRJEB7929 (88.34 Gb), PRJEB29510 (162 Gb), and
1751   PRJNA281811 (210 Gb)]. Panel **G** indicates the GC percentage vs. K-mer abundance of
1752   different project IDs mentioned in different colors. The vertical dotted lines in orange denote
1753   the GC percentage of *TIMM21* exons, and vertical solid lines in red indicate the GC
1754   percentage of *COA1/MITRAC15* exons.

1755   **Figure 6**: Recurrent loss of *COA1/MITRAC15* gene in rodent species. (**A**) Gene loss events in
1756   seven rodent species through four events are represented exon-wise beside the pink-colored
1757   branches of the time-calibrated phylogenetic tree obtained from the time tree website. Blue
1758   branches correspond to functional copies of *COA1/MITRAC15,* and black branches
1759   correspond to the Evolutionary Breakpoint Region (EBR) (**B**). Gene order in the genomic
1760   region flanking the *COA1/MITRAC15* gene in rodent genomes. Arrows depict the direction of
1761   gene transcription relative to the *COA1/MITRAC15* gene for consistency across species.
1762   Boxes represent the genes located on short scaffolds with unknown orientation. Each dotted
1763   box contains one type of gene order, and the brown arrows highlighted in yellow emerging
1764   from gene order O5 depict the EBR event that leads to gene orders O6, O7, O8, and O9. Gene
1765   order O8 and O7 contain partial remains of the *COA1/MITRAC15* gene and a functional
1766   *BLVRA* gene, respectively. A solid red line within gene order O8 depicts the partial exon one
1767   and intron 2 of *COA1/MITRAC15* located between the *PTPRF* and *HYI* genes. The gene
1768   order O6 and O9 correspond to the regions on the left and right flanks of the region
1769   containing *STK17A*, *COA1/MITRAC15,* and *BLVRA*.

1770   **Figure 7**: The genomic region spanning the *COA1/MITRAC15* gene coincides with an
1771   evolutionary breakpoint (EBR). (**A**) The phylogenetic relationship between marsupial species
1772   along with few outgroup species. The phylogenetic tree is from the time tree website. (**B**) The
1773   gene order in the region flanking the *COA1/MITRAC15* gene. The arrows show the direction
1774   of gene transcription relative to the *COA1/MITRAC15* gene for consistency across species.
1775   Each dotted box contains one type of gene order, and the red arrows from gene order O1
1776   depict the EBR event that leads to gene orders O2 and O3 in the six marsupial species. The
1777   outgroup species have the Pre-EBR gene order O1. In the post-EBR gene orders O2 and O3,
1778   the *COA1/MITRAC15* gene occurs in the EBR, and the gene order of flanking genes is
1779   changed. A functional *COA1/MITRAC15* can be identified in the outgroup species but is
1780   presumably lost in marsupial species as it is missing in the genome assembly and raw read
1781   datasets.

Figure 1

Alignment of *COA1* & *TIMM21*
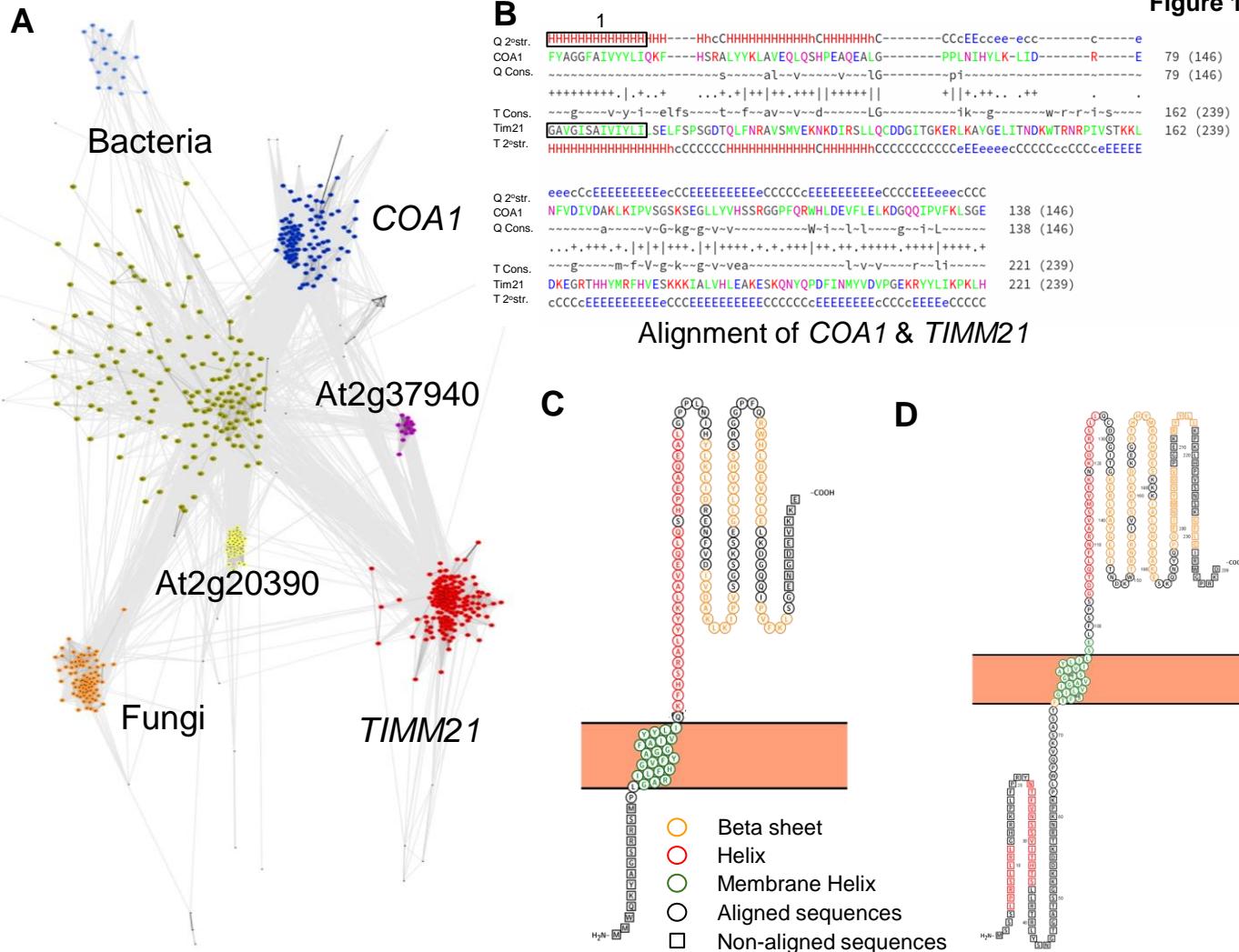
Figure 2

Haplotypes of *COA1*

Human

Tiger

Naked mole rat

Red squirrel

Platypus

Chicken
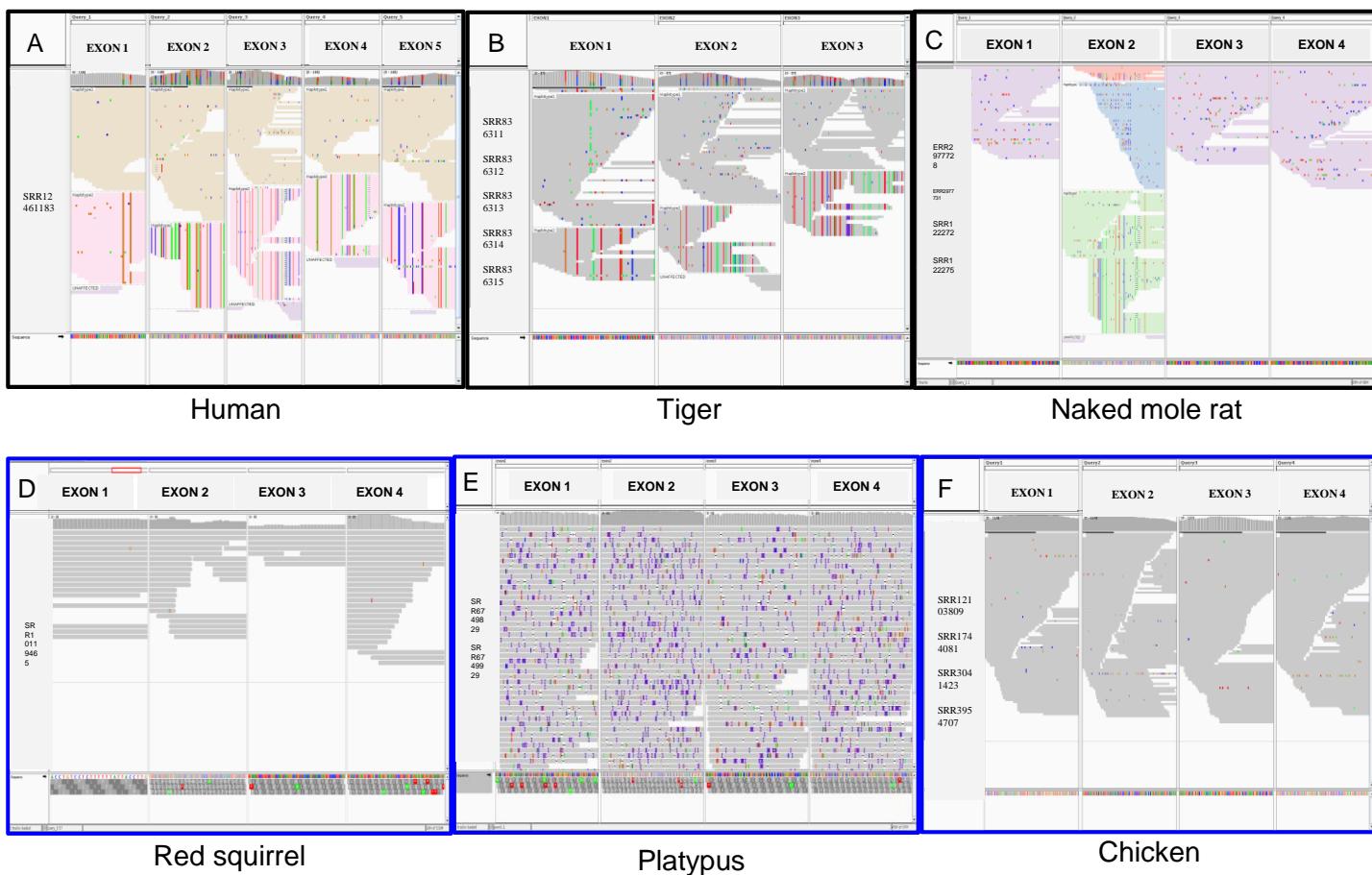
Figure 3

A Gene loss events

| Deletion | Substitution

Estimated date of COA1 loss (Mya)
■ 1ds ■ 2ds

■ Functional branches
■ Mixed branches

B Gene Synteny
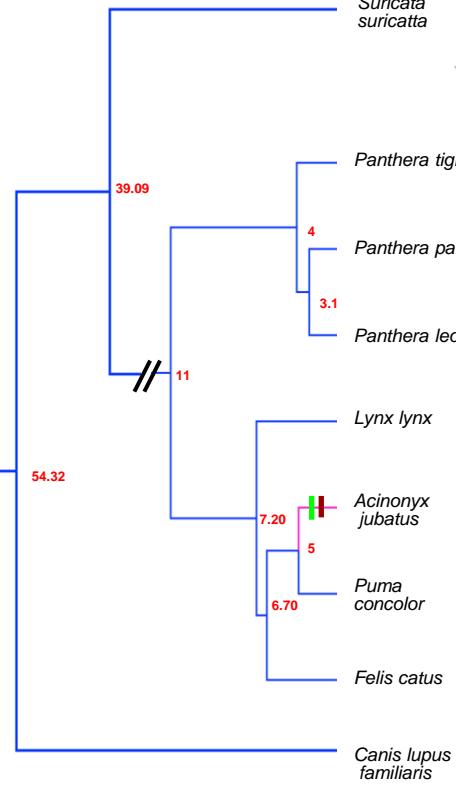
← Direction of the transcript
→ Direction of the transcript
▭ Unclear direction of the transcript

▯ Stop codon
▭ No RNA seq data
▬ Untranscribed exon
▬ Transcribed exon

HECW1 STK17A **COA1 EXONS** BLVRA VOPP1

EXON 3

EXON 1 EXON 2 CAT DOG EXON 4

Suricata suricata

Panthera tigris

Panthera pardus

Panthera leo

Lynx lynx

Acinonyx jubatus

Puma concolor

Felis catus

Canis lupus familiaris

Time in MY
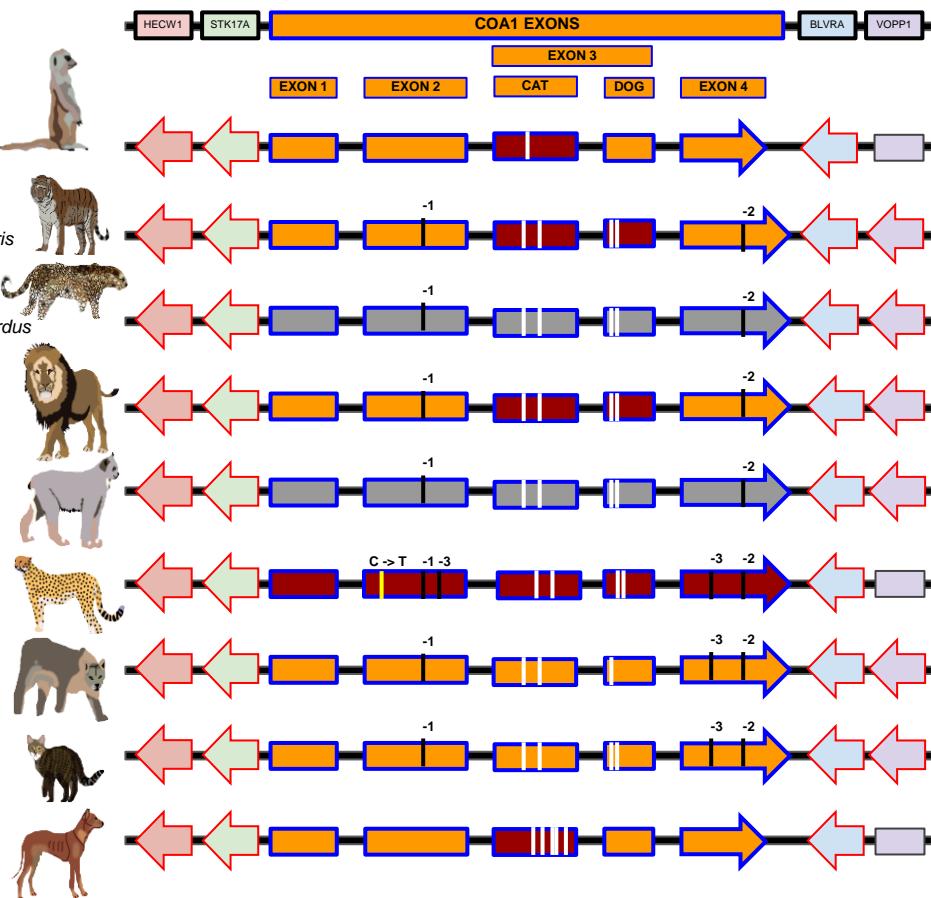
Figure 4

A Gene loss events

| Insertion | Deletion | Substitution

Estimated date of COA1 loss (Mya)
| 1ds | 2ds

Functional branches
Mixed branches
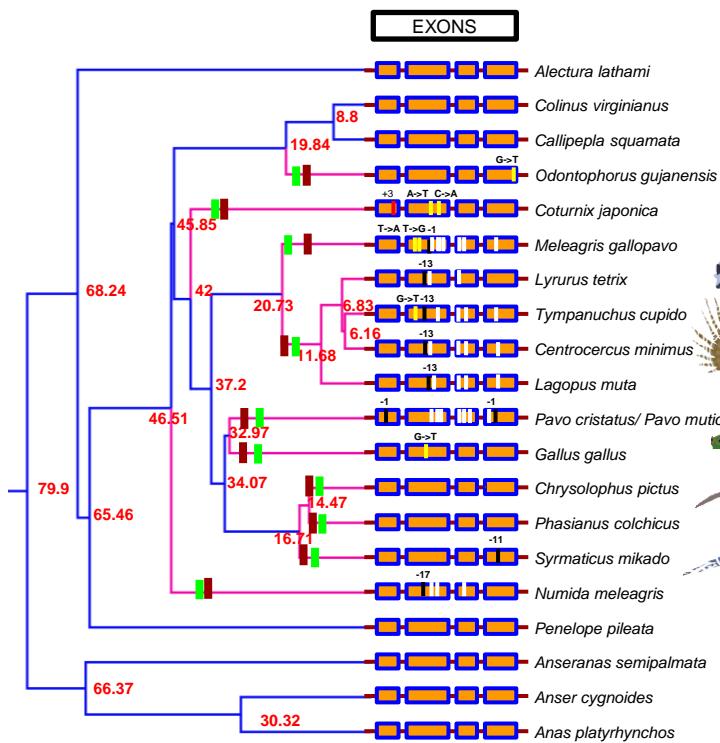
B Gene Synteny

Direction of the transcript

Unclear direction of the transcript

Stop codon
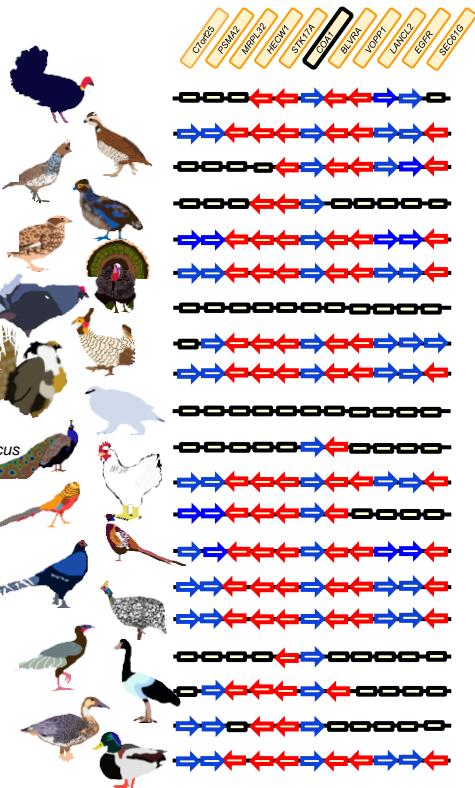
C Expression pattern

No data
Expression
No Expression

Tissue: Brain, Spleen, Skin, Liver, Gonad, Blood

Genes: C7orf25, PSM2, MRPL32, HECW1, STK17A, COA1, BLVRA, VOPP1, LANCL2, EGFR, SEC61G

EXONS

Species:
Alectura lathami
Colinus virginianus
Callipepla squamata
Odontophorus gujanensis
Coturnix japonica
Meleagris gallopavo
Lyrurus tetrix
Tympanuchus cupido
Centrocercus minimus
Lagopus muta
Pavo cristatus/ Pavo muticus
Gallus gallus
Chrysolophus pictus
Phasianus colchicus
Syrmaticus mikado
Numida meleagris
Penelope pileata
Anseranas semipalmata
Anser cygnoides
Anas platyrhynchos

Time in MY

Figure 5

Comparison of *Mammuthus primigenius* (woolly mammoth)

PRJDB4697
182 G

PRJEB42269
179 G

PRJNA397140
155 G

PRJEB7929
88.34 G

PRJEB29510
162 G

PRJNA281811
210 G
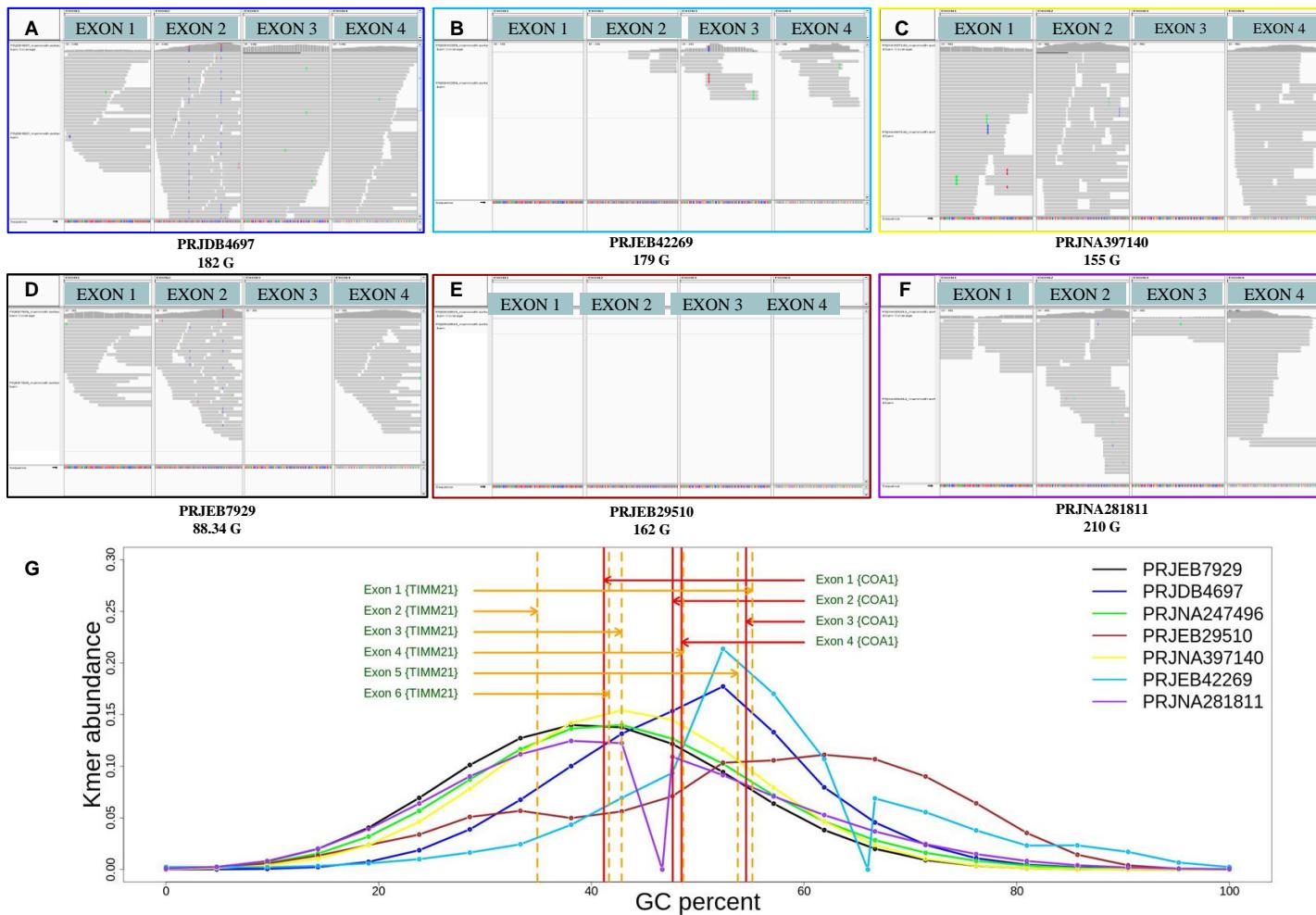
Figure 6

A Gene loss events

Estimated date for COA1 loss (Mya)

- 1ds (green)
- 2ds (dark red)
- Insertion (red)
- Deletion (black)
- Substitution (yellow)
- Stop codon (grey)

COA1 in EBR (black)
Functional (blue)
Mixed (magenta)
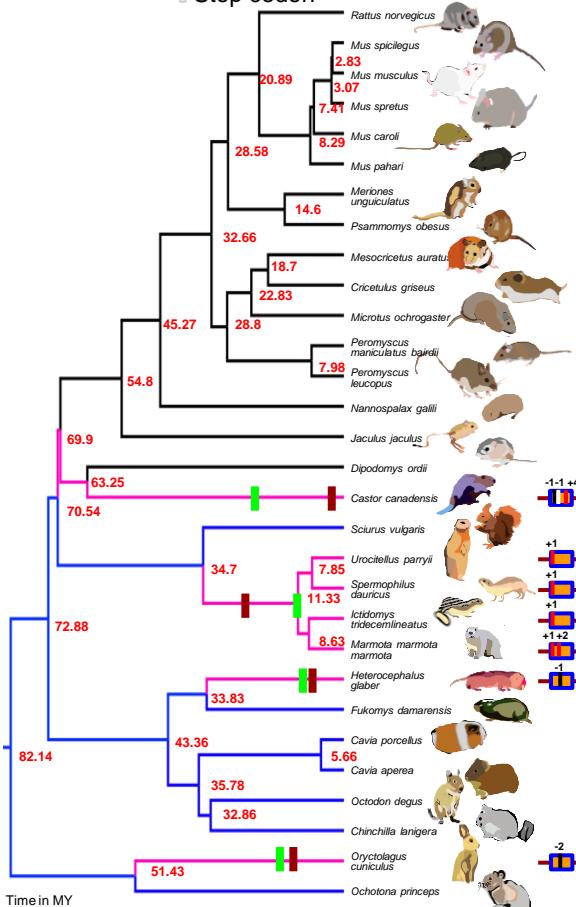
B Gene Synteny

- Unclear direction of the transcript
- Truncated Exon 1 and Intron 2 of *COA1* (red)
- Direction of the transcript
- Partial sequence (pink)
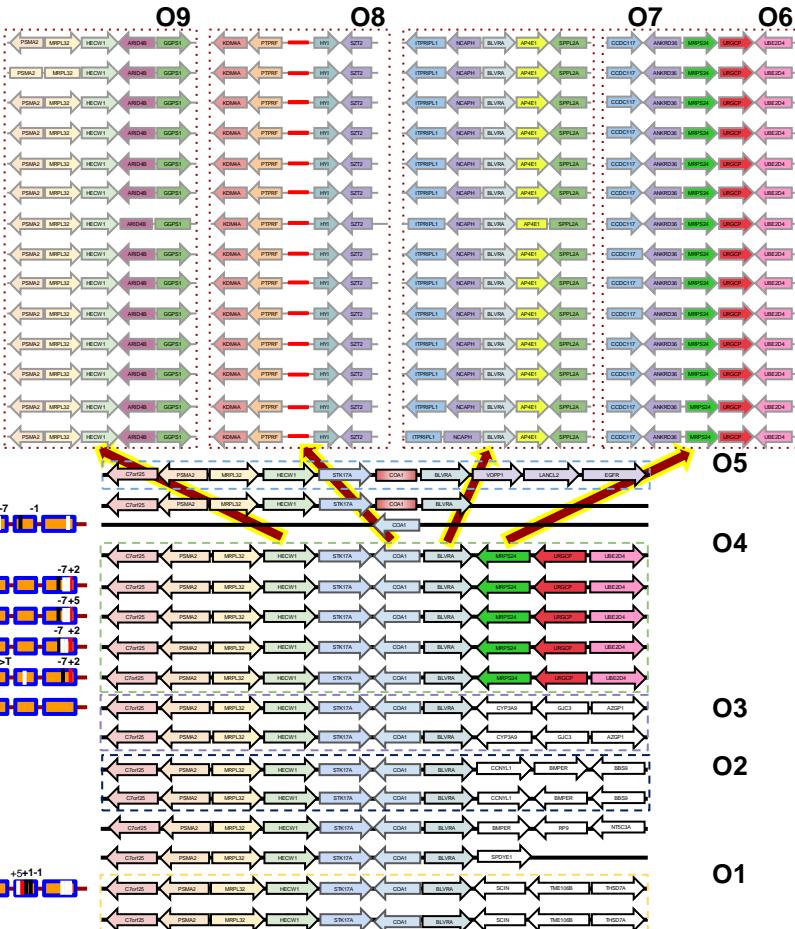
Figure 7

A Gene loss events

— *COA1* Present — *COA1* in EBR

B Gene Synteny

⇐⇒ Direction of the transcript   ▭ Unclear direction of the transcript

Time in MY